Deel I

Voortgezette Analyse

Aanbevolen achtergrondliteratuur met veel opgaven (en oplossingen):

- Robert C. Wrede, Murray Spiegel: (Schaum's Outline of Theory and Problems of) Advanced Calculus. McGraw-Hill Companies, 2002, 433 p., ISBN: 0071375678.

Meer wiskundige (maar goed leesbare) boeken over functies van meerdere veranderlijken:

- Jerrold Marsden, Alan Weinstein: Calculus III. Springer, 1998, 341 p., ISBN: 3540909850
- Jerrold E. Marsden, Anthony J. Tromba: Vector Calculus. Freeman, 2004, 676 p., ISBN: 0716749920

Les 1 Functies van meerdere variabelen

In het Calculus gedeelte van Wiskunde 1 hebben we ons bijna altijd beperkt tot functies van één variabel, dus functies van de vorm $y = f(x)$ die we makkelijk door hun grafiek in het $x-y$-vlak konden representeren. Helaas is de wereld niet zo eenvoudig dat zich alles door dit soort functies makkelijk laat beschrijven, denk bijvoorbeeld aan het volgende:

(1) Een steentje dat je in een meer gooit zal een cirkelvormige golf veroorzaken, waarvan de hoogte van de afstand r van het centrum van de cirkel en van het tijdstip t waarop je kijkt afhangt. De hoogte h is dus een functie van r en t, bijvoorbeeld $h(r,t) = \sin(r+t)\, e^{-t}$.

(2) Voor een ideaal gas geldt (volgens het algemene gaswet van Gay-Lussac en Boyle) de relatie $V = \frac{nRT}{p}$ tussen het volume V, de temperatuur T, de hoeveelheid n van de stof (in mol) en de druk p, waarbij R de universele gasconstante is. Het volume is dus een functie $V = f(n,T,p) = \frac{nRT}{p}$ van de variabelen n, T en p, maar net zo goed is de druk een functie $p = g(n,T,V) = \frac{nRT}{V}$ van n, T en V.

Er zijn verschillende voor de hand liggende gevallen van functies van meerdere veranderlijken die we moeten bekijken:

(i) $f : \mathbb{R}^n \to \mathbb{R}$: Dit zijn functies die van meerdere (n) parameters afhangen, maar slechts één waarde als resultaat opleveren. Een voorbeeld is de functie die de afstand van een punt (x,y,z) in de 3-dimensionale ruimte van de oorsprong aangeeft, namelijk $f(x,y,z) := \sqrt{x^2+y^2+z^2}$.

(ii) $f : \mathbb{R} \to \mathbb{R}^m$: Dit zijn functies die slechts van één variabel afhangen, maar meerdere waarden opleveren. Een voorbeeld is de functie $f(t) := (\cos(t), \sin(t))$ die het interval $[0, 2\pi]$ op de eenheidscirkel in het 2-dimensionale vlak afbeeldt.

(iii) $f : \mathbb{R}^n \to \mathbb{R}^m$: Dit is het algemene geval, waarbij de functie van meerdere parameters afhangt en ook meerdere waarden oplevert. Een voorbeeld van dit soort functies zijn de lineaire functies van een n-dimensionale vectorruimte naar een m-dimensionale vectorruimte, maar ook de functie $f(x,y,z) := (\frac{x}{\sqrt{x^2+y^2+z^2}}, \frac{y}{\sqrt{x^2+y^2+z^2}}, \frac{z}{\sqrt{x^2+y^2+z^2}})$ die een kubus rond de oorsprong op de eenheidskogel afbeeldt.

Als we naar de algemene functies van type (iii) kijken, zien we dat het resultaat uit m componenten opgebouwd is, die zelf functies van de eenvoudigere type (i) zijn. We kunnen namelijk elke functie $f : \mathbb{R}^n \to \mathbb{R}^m$ beschrijven door

$$f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), f_2(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)).$$

De functies $f_i(x_1, \ldots, x_m)$ noemen we de *componenten van* f en deze componenten zijn functies $f_i : \mathbb{R}^n \to \mathbb{R}$.

Voorbeeld: De functie $f(x,y,z) : \mathbb{R}^3 \to \mathbb{R}^3$ gegeven door

$$f(x,y,z) := (x\sin(y)\sin(z), x\sin(y)\cos(z), x\cos(y))$$

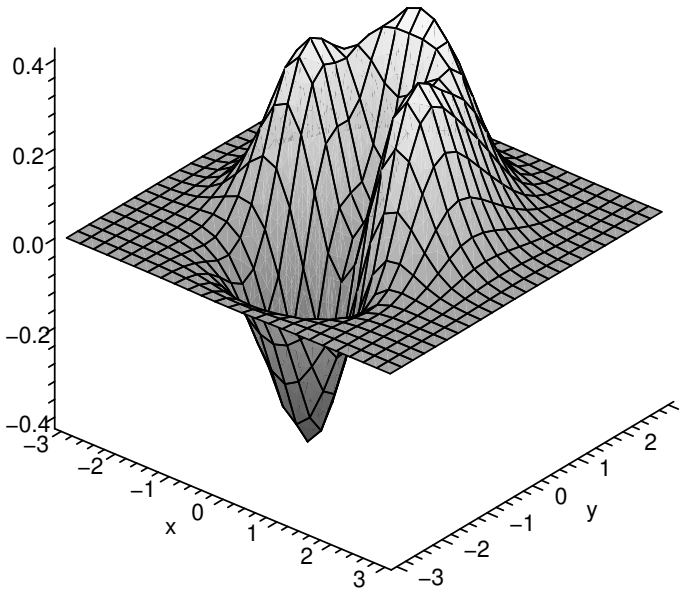
heeft de componenten

$$f_1(x,y,z) = x\sin(y)\sin(z),\ f_2(x,y,z) = x\sin(y)\cos(z),\ f_3(x,y,z) = x\cos(y).$$

Door naar de m componenten $f_i(x_1,\ldots,x_n)$ te kijken kunnen we ons dus meestal beperken tot het geval (i) van functies die van meerdere variabelen afhangen, maar slechts één waarde opleveren.

Om het schrijfwerk te beperken zullen we vaak voorbeelden van functies van twee variabelen behandelen, dus $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$. Hieraan worden de meeste ideeën wel duidelijk, de veralgemening op grotere aantallen van variabelen vergt meestal weinig nieuwe inzicht.

Een bijzonder voordeel van functies $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$ is dat we hiervan nog net grafieken kunnen tekenen, namelijk door de punten $(x,y,z) = (x,y,f(x,y))$ in de 3-dimensionale ruimte te bekijken. De functiewaarden vormen een soort *gebergte* boven het $x-y$-vlak waarin we het domein van de functie vinden. In Figuur I.1 is bijvoorbeeld de grafiek van de functie $f(x,y) := (x^2+y^3)\,e^{-x^2-y^2}$ te zien.



Figuur I.1: Grafiek van een functie van twee variabelen.

1.1 Continuïteit

We zullen zien dat de behandeling van functies van meerdere veranderlijken vaak analoog met gewone functies van één veranderlijke loopt, maar er zijn ook belangrijke verschillen waarvan we ons bewust moeten zijn.

Een eerste vraag die we ons kunnen stellen, is, wanneer we een functie continu noemen. Bij een gewone functie hadden we dit intuïtief zo gedefinieerd, dat een continue functie geen sprongen mag hebben. Bij functies van meerdere variabelen zouden we nu eerst moeten zeggen, wat een sprong eigenlijk is. Maar we hadden ook een meer formele definitie gegeven en deze kunnen we heel makkelijk naar functies van meerdere variabelen vertalen, in feite hoeven we alleen maar de absolute waarde op $\mathbb{R}$ te vervangen door de Euclidische afstand in de n-dimensionale ruimte.

Voor een gewone functie f van één variabele hadden we de volgende definitie gehanteerd:

Definitie: De functie $f : \mathbb{R} \to \mathbb{R}$ heet *continu in het punt* x als er voor iedere $\varepsilon > 0$ een $\delta > 0$ bestaat, zo dat $|f(x) - f(y)| < \varepsilon$ wanneer $|x - y| < \delta$.

In woorden betekent dit dat we voor een gekozen (klein) interval $[f(x) - \varepsilon, f(x) + \varepsilon]$ rond $f(x)$ een interval $[x - \delta, x + \delta]$ rond x kunnen vinden, waarvoor de functiewaarden alle in het interval rond $f(x)$ vallen.

Precies hetzelfde idee passen we nu ook bij een functie $f : \mathbb{R}^n \to R$ van n variabelen toe. Het enige verschil is, dat het domein waarop f gedefinieerd is nu vectoren $\mathbf{x} \in \mathbb{R}^n$ in plaats van reële getallen bevat.

> We zullen in deze cursus vectoren altijd met vet gedrukt letters aanduiden, zoals $\mathbf{x}$, $\mathbf{y}$ of $\mathbf{v}$, in tegenstellingen tot gewone reële variabelen zo als x en y.

Van het Lineaire Algebra gedeelte van Wiskunde 1 weten we nog, hoe we de afstand van twee vectoren $\mathbf{x}$ en $\mathbf{y}$ in de n-dimensionale vectorruimte $\mathbb{R}^n$ bepalen, namelijk door de *Euclidische lengte* $\|\mathbf{x} - \mathbf{y}\|$ van de verschilvector $\mathbf{x} - \mathbf{y}$. We zeggen dus, dat de vector $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ dicht bij de vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ ligt, als de afstand

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$$

klein is. We zullen de functie f dus continu in het punt $\mathbf{x}$ noemen als voor alle vectoren $\mathbf{y}$ die dicht bij $\mathbf{x}$ liggen ook de functiewaarden $f(\mathbf{y})$ dicht bij $f(\mathbf{x})$ liggen. De precieze definitie luidt:

Definitie: Een functie $f : \mathbb{R}^n \to \mathbb{R}$ heet *continu in het punt* $\mathbf{x}$ als er voor iedere $\varepsilon > 0$ een $\delta > 0$ bestaat, zo dat $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$ wanneer $\|\mathbf{x} - \mathbf{y}\| < \delta$.

Als we nu echt willen aantonen dat een zekere functie continu is, zien we dat er wel verschillen zijn tussen functies van één en van meerdere variabelen.

Voor een gewone functie van één variabel testen we continuïteit in principe zo: We bepalen de limiet $\lim_{y \to x+} f(y)$ van $f(y)$ als we met y van rechts tegen x aanlopen, en de limiet $\lim_{y \to x-} f(y)$ van $f(y)$ als we met y van links tegen x aanlopen. Als deze twee limieten bestaan en dezelfde waarde hebben is de functie continu.

Bij een functie van meerdere variabelen is dit niet meer zo makkelijk. We kunnen namelijk uit elke willekeurige richting tegen $\mathbf{x}$ aanlopen. En dat hoeft niet eens op een rechte lijn te gebeuren, we kunnen ook in een spiraal rond $\mathbf{x}$ lopen, in een zigzag of langs elke willekeurige kromme die uiteindelijk steeds dichter bij $\mathbf{x}$ komt. Omdat er zo veel mogelijkheden zijn, komt het er op neer dat men de continuïteit rechtstreeks met de definitie bewijst, die onafhankelijk van een gekozen pad naar $\mathbf{x}$ toe is.

Maar meestal is er een slimmere methode om de continuïteit van een functie te bewijzen: Men heeft ergens een lijst van heel eenvoudige functies waarvan de continuïteit bekend is, bijvoorbeeld veeltermfuncties zo als $f(x,y) = x^2 - 3xy + 2y^3$ en standaardfuncties zo als de *exponentiële functie*, de *sinus-* en de *cosinusfunctie*. De grap is nu dat sommen producten en samenstellingen van continue functies ook weer continu zijn. Voor de meeste functies volgt dus de continuïteit heel makkelijk, omdat ze met deze operaties uit eenvoudige continue functies opgebouwd kunnen worden. De volgende stelling geeft de precieze voorwaarden aan:

Stelling: Laten $f(\mathbf{x})$, $g(\mathbf{x})$ functies zijn die op een gemeenschappelijk domein $D \subseteq \mathbb{R}^n$ continu zijn.

(i) De som $f(\mathbf{x}) + g(\mathbf{x})$, het product $f(\mathbf{x}) \cdot g(\mathbf{x})$ en de schaling $c \cdot f(\mathbf{x})$ (met $c \in \mathbb{R}$) zijn continu op D.

(ii) Als $f(\mathbf{x}) \neq 0$ op D, dan is ook $\frac{1}{f(\mathbf{x})}$ continu op D.

(iii) Zij $h(x) : \mathbb{R} \to \mathbb{R}$ een functie die continu op $I \subseteq \mathbb{R}$ is en stel dat $f(\mathbf{x}) \in I$ voor alle $\mathbf{x} \in D$. Dan is de samenstelling $h(f(\mathbf{x}))$ continu op D.

Uit punt (iii) volgt in het bijzonder dat functies zo als $f(x,y) = e^{x^2+y^2}$ of $f(x,y,z) = \sin(xy - z)$ continu zijn. Problemen leveren meestal alleen maar breuken op, waar de noemer niet 0 mag worden, en de *logaritme* en wortels, die een positief argument moeten hebben. Daarom is bijvoorbeeld de functie $f(x,y) = \log(1-x^2-y^2)$ alleen maar continu op het gebied waar $1-x^2-y^2 > 0$ is, d.w.z. voor $x^2 + y^2 < 1$, met andere woorden binnen een cirkel met straal 1 rond $(0,0)$.

Soms kan een functie met nulpunten in de noemer wel door een geschikte definitie van functiewaarden continu naar de nulpunten van de noemer voortgezet worden. Voor gewone functies kennen we dit al: De functie $f(x) := \frac{x^2-1}{x-1}$ is voor $x = 1$ niet gedefinieerd, maar omdat voor $x \neq 1$ geldt dat $f(x) = \frac{(x-1)(x+1)}{x-1} = x + 1$, laat zich $f(x)$ door $f(1) := 2$ tot een continue functie voortzetten.

Om een functie continu naar een nulpunt van de noemer voort te zetten, is het noodzakelijk dat de nulpunten van de noemer ook nulpunten van de teller zijn (want anders gaat de functie naar oneindig). In zo'n geval moet onderzocht worden, hoe de functie zich in de buurt van het kritieke punt precies gedraagt. De volgende voorbeelden geven een illustratie hoe dit in speciale gevallen kan gebeuren.

Voorbeelden:

(1) De functie $f(x,y) := \frac{x^2}{\sqrt{x^2+y^2}}$ laat zich door $f(0,0) := 0$ tot een continue functie voortzetten. Er geldt namelijk voor $(x,y) \neq (0,0)$ dat $\frac{x^2}{\sqrt{x^2+y^2}} \leq \frac{x^2+y^2}{\sqrt{x^2+y^2}} = \sqrt{x^2+y^2}$ en voor $(x,y) \to (0,0)$ gaat $\sqrt{x^2+y^2} \to 0$.

(2) De functie $f(x,y) := \frac{x^2-y^2}{x-y}$ is op de lijn met $x = y$ niet gedefinieerd. Maar voor $x \neq y$ geldt dat $f(x,y) = \frac{(x-y)(x+y)}{x-y} = x+y$. Daarom laat zich $f(x,y)$ door $f(x,x) := 2x$ op de lijn met $x = y$ tot een continue functie voortzetten.

(3) De functie $f(x,y) := \frac{\sin(xy)}{x^2+y^2}$ laat zich niet continu naar $(x,y) = (0,0)$ voortzetten. In Wiskunde 1 hadden we namelijk gezien dat uit $\sin(x)' = \cos(x)$ volgt dat $\lim_{x\to 0} \frac{\sin(x)-\sin(0)}{x} = \cos(0) = 1$, en hieruit volgt dat op de lijn $x = y$ geldt dat $f(x,x) = \frac{\sin(x^2)}{2x^2} \to \frac{1}{2}$ voor $x \to 0$. Maar als we op de x-as tegen $(0,0)$ aan lopen, is $f(x,0) = \frac{\sin(0)}{x^2} = 0$. We zouden de functie dus tegelijkertijd met $f(0,0) = \frac{1}{2}$ en met $f(0,0) = 0$ moeten voortzetten en dit is natuurlijk onmogelijk.

Opdracht 1

(i) Laat zien dat de functie $f(x,y) := \frac{x^3+2x^2+xy^2+2y^2}{x^2+y^2}$ een continue voortzetting naar het punt $(0,0)$ heeft. (Hint: Probeer de noemer als factor van de teller te vinden.)

(ii) Hoe moet de functie $f(x,y) := \frac{x^3-y^3}{x-y}$ op de lijn $x = y$ gedefinieerd worden zo dat de functie op het hele $x-y$-vlak continu is?

(iii) Ga na dat de functie $f(x,y,z) := \frac{\sin(xyz)}{xyz}$ voor $(x,y,z) \to (0,0,0)$ tegen de waarde 1 gaat. (Hint: Denk aan de limiet $\frac{\sin(x)}{x}$ voor $x \to 0$.)

Een afschrikkend voorbeeld: De grotere vrijheid van paden die men in $\mathbb{R}^n$ tegenover $\mathbb{R}$ heeft, heeft wel soms verrassende effecten. We zullen hier een voorbeeld van bekijken, namelijk de functie

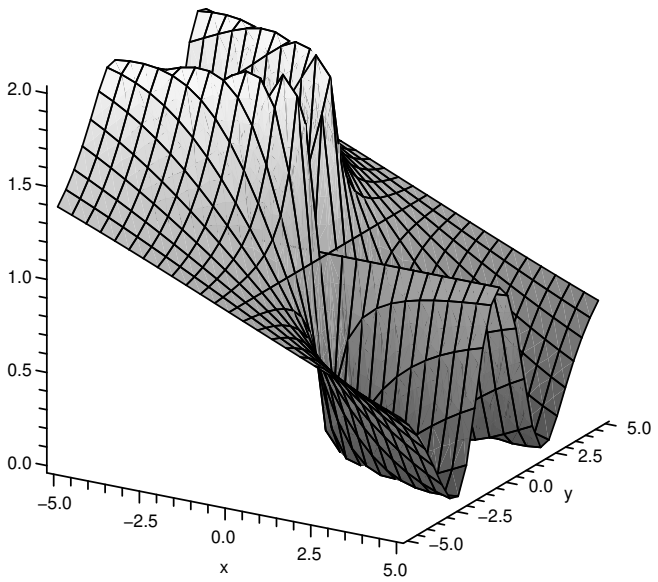
$$f(x,y) := \frac{(y^2-x)^2}{y^4+x^2}$$

waarvan de grafiek in Figuur I.2 afgebeeld is.

Voor $(x,y) = (0,0)$ is $f(x,y)$ natuurlijk niet gedefinieerd, maar het zou kunnen dat zich $f(x,y)$ door een geschikte definitie van $f(0,0)$ tot een continue functie laat voortzetten.

Als we langs de x-as naar $(0,0)$ lopen is $y = 0$, dus $f(x,y) = \frac{x^2}{x^2} = 1$. Lopen we langs de y-as, is $x = 0$ en $f(x,y) = \frac{y^4}{y^4} = 1$. Als de functie überhaupt een continue voortzetting heeft, dan moet deze dus noodzakelijk de waarde $f(0,0) = 1$ hebben.

Als we nu langs een willekeurige lijn lopen (behalve de x- en y-as) kunnen we dit door de punten $(x,y) = (t,ct)$ beschrijven, die op de lijn met $y = cx$



Figuur I.2: Grafiek van de functie $f(x,y) := \frac{(y^2-x)^2}{y^4+x^2}$.

liggen en voor $t \to 0$ lopen we op zo'n lijn tegen $(0,0)$ aan. Als we (t, ct) voor (x, y) invullen, krijgen we

$$f(x,y) = \frac{(c^2t^2 - t)^2}{c^4t^4 + t^2} = \frac{c^4t^4 + t^2 - 2c^2t^3}{c^4t^4 + t^2} = 1 - \frac{2c^2t^3}{c^4t^4 + t^2}$$
$$= 1 - \frac{2c^2t}{c^4t^2 + 1} \to 1 - \frac{0}{0+1} = 1 \text{ voor } t \to 0.$$

We hebben dus aangetoond dat $f(x, y)$ op elke lijn naar $(0, 0)$ de limiet 1 heeft.

Ook al zouden we nu misschien denken, dat de functie zich inderdaad met $f(0, 0) = 1$ tot een continue functie laat voortzetten, laten we nu zien dat $f(x, y)$ met deze voortzetting *niet continu* in $(0, 0)$ is.

Als we namelijk nog eens goed naar de teller $(y^2 - x)^2$ van $f(x, y)$ kijken, zien we dat die voor $x = y^2$ gelijk aan 0 is. Dit betekent dat voor alle punten op de parabool $(x, y) = (t^2, t)$ de functiewaarde 0 is. Maar voor $t \to 0$ lopen we op deze parabool ook tegen het punt $(0, 0)$ aan, dus zijn er punten die willekeurig dicht bij $(0, 0)$ liggen waarvoor $f(x, y)$ de waarde 0 heeft en die ligt helaas niet dicht bij 1 (neem bijvoorbeeld $\varepsilon = \frac{1}{2}$).

> Nu dat we het weten, kunnen we ook in Figuur I.2 zien, dat er onderaan een soort rand is, waar de functie op de functiewaarde 0 blijft, terwijl alle lijnen door $(x, y) = (0, 0)$ inderdaad door de functiewaarde 1 gaan.

Het voorbeeld laat zien dat de vraag of een functie continu is bij functies

van meerdere variabelen soms enigszins subtiel kan zijn. Maar omdat we in de praktijk nauwelijks functies tegen komen die niet continu zijn, zullen we ons hier niet verder in verdiepen.

OPDRACHT 2 *Laat zien dat de functie* $f(x,y) := \frac{x^2-y^2}{x^2+y^2}$ *geen continue voortzetting naar* $(x,y) = (0,0)$ *heeft. (Hint: Bekijk verschillende lijnen door het nulpunt.)*

1.2 Partiële afgeleide en richtingsafgeleide

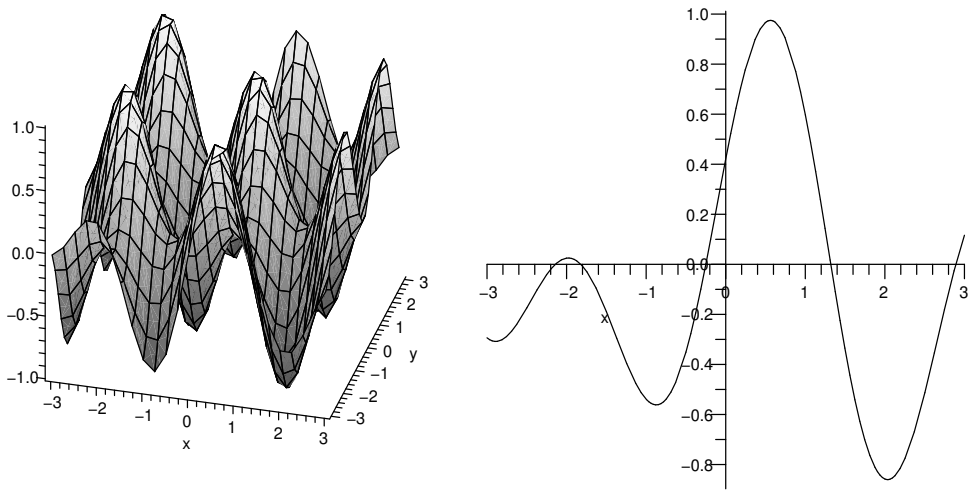
Bij functies van één veranderlijke hebben we gezien dat de afgeleide van een functie de snelheid aangeeft waarmee de functie in een punt verandert. Dit was heel nuttig om minima en maxima van functies te vinden, maar ook om te zien of de functie stijgt of daalt en waar de punten zijn waar ze het snelste stijgt. Een van de interpretaties van de afgeleide was, dat hij de richtingscoëfficiënt van de raaklijn aan de grafiek van de functie aangeeft. Dit kunnen we helaas niet rechtstreeks op functies van meerdere veranderlijken veralgemenen. Zo is bijvoorbeeld de grafiek van een functie van twee veranderlijken (met één waarde) een soort gebergte boven het $x-y$-vlak waar men in een punt voor elke richting van het $x-y$-vlak een raaklijn kan aanleggen, en voor verschillende richtingen zullen deze raaklijnen zeker verschillende stijgingen hebben. We moeten daarom iets beter kijken, wat we als afgeleide willen definiëren.

Een eerste idee (die we in Wiskunde 1 al kort hebben bekeken) is, dat we de richtingen van de raaklijnen op richtingen langs de coördinaatassen beperken. Als we bij een functie $f(x,y)$ de raaklijn in het punt (x_0,y_0) in de richting van de x-as willen bepalen, leggen we een lijn door de punten (x_0,y_0) en (x_0+h,y_0) en laten h tegen 0 lopen. Bij dit proces blijft y_0 altijd ongedeerd, want omdat we de raaklijn in de richting van de x-as bepalen, moet de y-waarde constant blijven. Maar als de y-waarde de vaste waarde y_0 heeft, is de functie $f(x,y)$ eigenlijk een functie $g(x) = f(x,y_0)$ van slechts één variabel.

Voor een functie van twee variabelen kunnen we dit idee ook grafisch illustreren: De grafiek van zo'n functie kunnen we zien als de verzameling van punten $(x,y,z) = (x,y,f(x,y))$ in de 3-dimensionale ruimte, net zo als we de grafiek van een gewone functie als de verzameling van punten $(x,y) = (x,f(x))$ in het 2-dimensionale vlak bekijken. Als we nu y tot een constante y_0 verklaren, dan kijken we naar de doorsnede van de grafiek $(x,y,f(x,y))$ met het vlak dat bepaald is door de vergelijking $y = y_0$, dus we bekijken de punten $(x,y_0,f(x,y_0))$. Maar de tweede coördinaat hierbij is natuurlijk volstrekt overbodig, en als we deze schrappen, houden we de punten $(x,f(x,y_0))$ over. Dit zijn gewoon de punten van de grafiek van de functie $g(x) := f(x,y_0)$ van één variabel (want y_0 is een constante).

De plaatjes in Figuur I.3 laten links de functie $f(x,y) := \sin(2x+y)\cos(\frac{x}{2}-y)$ en rechts de doorsnede door deze grafiek voor $y = \frac{1}{2}$ zien, dus de functie $g(x) := f(x,\frac{1}{2}) = \sin(2x+\frac{1}{2})\cos(\frac{x}{2}-\frac{1}{2})$.

Het is nu voor de hand liggend hoe we de afgeleide in het punt (x_0,y_0) in de richting van de x-as definiëren, namelijk gewoon als de richtingscoëfficiënt van



Figuur I.3: Functie $f(x, y)$ en doorsnede door deze functie voor $y = 0.5$

de raaklijn aan de gewone functie $g(x) = f(x, y_0)$ in het punt x_0. De afgeleide in de richting van de x-as noemen we de *partiële afgeleide* van f naar x.

Als we ons nu aan de definitie

$$g'(x) = \lim_{h \to 0} \frac{g(x+h) - g(x)}{h}$$

van de gewone afgeleide van een functie van één variabel herinneren, kunnen we rechtstreeks de definitie voor de partiële afgeleide in dit geval aangeven, dit is namelijk de limiet

$$\lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}.$$

Natuurlijk is er geen reden, waarom de x-as beter dan de y-as zou zijn, we kunnen net zo goed ook de raaklijn in de richting van de y-as bekijken en krijgen dan de partiële afgeleide van f naar y. Maar nu dat we het idee hebben gezien dat tot op één na alle variabelen tot constanten verklaard worden, kunnen we ook meteen de algemene definitie van de partiële afgeleide geven.

Definitie: Voor een functie $f(\mathbf{x}) = f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ definiëren we de *partiële afgeleide* van f naar de variabel x_i als de limiet

$$\lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}$$

als deze limiet bestaat. De partiële afgeleide wordt meestal met $\frac{\partial f}{\partial x_i}$ genoteerd, maar vaak ook kort als f_{x_i} geschreven.

Merk op: Om een partiële afgeleide uit te rekenen, gebruiken we nooit deze limiet-definitie. Omdat we alleen maar de variabel x_i gaan veranderen, kunnen we de andere variabelen $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ als constanten (net zo als de constanten 5 of π in de functie $g(x) = e^{5x} \sin(\pi x)$) behandelen. Op deze manier

interpreteren we de functie $f(x_1, \ldots, x_n)$ als een nieuwe functie $g(x_i)$ van één variabel (waarin de andere variabelen x_j als constanten voorkomen) en deze functie leiden we nu met de bekende rekenregels naar zijn variabel x_i af.

Voorbeelden:

(1) Zij $f(x, y)$ gegeven door $f(x, y) := x^3 y + e^{xy^2}$. Dan geldt (let op de kettingregel):

$$\frac{\partial f}{\partial x} = 3x^2 y + y^2\, e^{xy^2} \quad \text{en} \quad \frac{\partial f}{\partial y} = x^3 + 2xy\, e^{xy^2}.$$

(2) Zij $f(x, y, z)$ de functie van drie variabelen, gegeven door $f(x, y, z) := x \log(yz)$, dan geldt:

$$\frac{\partial f}{\partial x} = \log(yz), \quad \frac{\partial f}{\partial y} = x(yz)^{-1} z = \frac{x}{y}, \quad \frac{\partial f}{\partial z} = x(yz)^{-1} y = \frac{x}{z}.$$

OPDRACHT 3 *Bepaal de partiële afgeleiden* $\frac{\partial f}{\partial x}$ *en* $\frac{\partial f}{\partial y}$ *van de functies* $f(x, y) := 2x^2 - xy + y^2$ *en* $g(x, y) := \cos(xy) + x \cos(y)$.

Merk op: De partiële afgeleiden $\frac{\partial f}{\partial x_i}$ zijn zelfs ook weer functies van $\mathbb{R}^n$ naar $\mathbb{R}$, want ook al behandelen we $x_2, x_3, \ldots$ bij het afleiden naar x_1 als constanten, heeft de partiële afgeleide afhankelijk van deze constanten toch verschillende waarden.

> Als we voor een functie $f(x, y)$ van twee variabelen de waarde van $\frac{\partial f}{\partial x}$ in een bepaald punt (x_0, y_0) willen aanduiden, noteren we dit met $\frac{\partial f}{\partial x}(x_0, y_0)$ of $\frac{\partial f}{\partial x}|_{x_0, y_0}$. Net zo schrijven we bij een algemene functie $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ van n variabelen $\frac{\partial f}{\partial x_i}(\mathbf{x}_0)$ of $\frac{\partial f}{\partial x_i}|_{\mathbf{x}_0}$ voor de waarde van de i-de partiële afgeleide in het punt $\mathbf{x}_0$.

Voorbeeld: Voor $f(x, y) := x^2 y + y^3$ geldt $\frac{\partial f}{\partial x} = 2xy$ en $\frac{\partial f}{\partial y} = x^2 + 3y^2$. De waarden van de twee partiële afgeleiden in het punt $(x, y) = (1, 2)$ zijn $\frac{\partial f}{\partial x}(1, 2) = 2 \cdot 1 \cdot 2 = 4$ en $\frac{\partial f}{\partial y}(1, 2) = 1^2 + 3 \cdot 2^2 = 13$.

Hogere partiële afgeleiden

Net zo als we bij gewone functies de afgeleide van de afgeleide kunnen bepalen en zo tot de hogere afgeleiden $f''(x), f'''(x), f^{(i)}(x)$ komen, kunnen we ook partiële afgeleide itereren. Hierbij hebben we echter veel meer keuze want we kunnen iedere keer een andere variabele kiezen waar we naar afleiden.

In het voorbeeld met $f(x, y, z) = x \log(yz)$ hadden we de partiële afgeleide $\frac{\partial f}{\partial x} = \log(yz)$ gevonden. Als we dit bijvoorbeeld partieel naar y afleiden, schrijven we dit als

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y\, \partial x} = \frac{1}{yz} z = \frac{1}{y}.$$

De alternatieve notatie f_x in plaats van $\frac{\partial f}{\partial x}$ voor de partiële afgeleide geeft aanleiding tot de notatie $f_{xy} = (f_x)_y = \frac{\partial^2 f}{\partial y\,\partial x}$ voor de geïtereerde partiële afgeleide. Het lijkt erg onhandig dat de volgorde van de variabelen in de twee notaties verruild is, maar we zullen straks zien dat dit geen probleem voorstelt.

Als we meerdere keer naar dezelfde variabel afleiden, is er een verdere notatie gebruikelijk: $\frac{\partial^2 f}{\partial x\,\partial x} = f_{xx}$ wordt kort geschreven als $\frac{\partial^2 f}{\partial x^2}$. Merk op dat hier niet naar een nieuwe variabel x^2 wordt afgeleid, maar twee keer naar de variabel x (de ∂^2 in de teller geeft aan dat het om een tweede afgeleide gaat).

Als we voor een functie van twee variabelen de tweede partiële afgeleiden bepalen, hebben we hiervoor $2^2 = 4$ mogelijkheden, want voor de eerste en voor de tweede afgeleide kunnen we telkens een van de twee variabelen kiezen. Bij een functie van drie variabelen krijgen we op deze manier zelfs $3^2 = 9$ tweede afgeleiden en bij een functie van n variabelen zijn het er n^2.

Voorbeeld:

(1) Voor de functie $f(x,y) := x^3y + e^{xy^2}$ (waarvan we boven al de eerste partiële afgeleiden hebben bepaald) krijgen we de volgende tweede partiële afgeleiden:

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}(3x^2y + y^2\,e^{xy^2}) = 6xy + y^2\;y^2\,e^{xy^2} = 6xy + y^4\,e^{xy^2}\\
\frac{\partial^2 f}{\partial y\,\partial x} &= \frac{\partial}{\partial y}(3x^2y + y^2\,e^{xy^2}) = 3x^2 + 2y\,e^{xy^2} + y^2\;2xy\,e^{xy^2}\\
&= 3x^2 + (2y + 2xy^3)\,e^{xy^2}\\
\frac{\partial^2 f}{\partial x\,\partial y} &= \frac{\partial}{\partial x}(x^3 + 2xy\,e^{xy^2}) = 3x^2 + 2y\,e^{xy^2} + 2xy\;y^2\,e^{xy^2}\\
&= 3x^2 + (2y + 2xy^3)\,e^{xy^2}\\
\frac{\partial^2 f}{\partial y^2} &= \frac{\partial}{\partial y}(x^3 + 2xy\,e^{xy^2}) = 2x\,e^{xy^2} + 2xy\;2xy\,e^{xy^2} = (2x + 4x^2y^2)\,e^{xy^2}
\end{aligned}$$

(2) Voor de functie $f(x,y,z) := x\log(yz)$ schrijven we de tweede partiële afgeleiden in een 3×3 schema, waarbij de rijen met de variabel voor de eerste afgeleide en de kolommen met de variabel voor de tweede afgeleide corresponderen. Het schema ziet er als volgt uit:

$$\begin{array}{ccc}
\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\log(yz) = 0 & \frac{\partial^2 f}{\partial y\,\partial x} = \frac{\partial}{\partial y}\log(yz) = \frac{1}{y} & \frac{\partial^2 f}{\partial z\,\partial x} = \frac{\partial}{\partial z}\log(yz) = \frac{1}{z}\\
\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial}{\partial x}\frac{x}{y} = \frac{1}{y} & \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\frac{x}{y} = -\frac{x}{y^2} & \frac{\partial^2 f}{\partial z\,\partial y} = \frac{\partial}{\partial z}\frac{x}{y} = 0\\
\frac{\partial^2 f}{\partial x\,\partial z} = \frac{\partial}{\partial x}\frac{x}{z} = \frac{1}{z} & \frac{\partial^2 f}{\partial y\,\partial z} = \frac{\partial}{\partial y}\frac{x}{z} = 0 & \frac{\partial^2 f}{\partial z^2} = \frac{\partial}{\partial z}\frac{x}{z} = -\frac{x}{z^2}
\end{array}$$

Merk op: In beide voorbeelden kunnen we constateren, dat het geen verschil maakt of we eerst partieel naar x en dan naar y afleiden, of andersom, en dit geldt ook voor alle andere paren van variabelen. Na twee voorbeelden te

controleren geloven we natuurlijk niet meer aan een toeval, maar het is ook niet vanzelfsprekend dat dit daadwerkelijk altijd zou gelden. Hier zit inderdaad een serieuze stelling achter, de *Stelling van Schwarz*, die een niet helemaal triviaal bewijs heeft (die door Leonhard Euler werd gevonden). Gelukkig is de uitspraak van de stelling wel zo eenvoudig als we maar zouden kunnen hopen, namelijk dat we (onder zwakke voorwaarden) de volgorde van partiële afgeleiden mogen verruilen.

Stelling van Schwarz: Als voor een functie $f(\mathbf{x}) = f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ de tweede partiële afgeleiden $\frac{\partial^2 f}{\partial x_i \, \partial x_j}$ en $\frac{\partial^2 f}{\partial x_j \, \partial x_i}$ bestaan en continu zijn, dan zijn ze gelijk. De volgorde van de partiële afgeleiden speelt in dit geval dus geen rol.

Als we eens ervan uit gaan dat we het altijd met voldoende goedaardige functies te maken hebben (met continue hogere partiële afgeleiden, dus), hoeven we niet op de volgorde van de partiële afgeleiden te letten. We hoeven ons dus ook van de verwarring over de notaties $\frac{\partial^2 f}{\partial y \, \partial x} = f_{xy}$ niets aan te trekken.

We merken nog op dat we door herhaalde toepassing van de Stelling van Schwarz ook voor derde, vierde en hogere partiële afgeleiden kunnen laten zien dat de volgorde van de afgeleiden geen rol speelt. Er geldt dus bijvoorbeeld dat

$$\frac{\partial^3 f}{\partial x \, \partial y \, \partial z} = \frac{\partial^2}{\partial x \, \partial y}(\frac{\partial f}{\partial z}) = \frac{\partial^2}{\partial y \, \partial x}(\frac{\partial f}{\partial z}) = \frac{\partial}{\partial y}(\frac{\partial^2 f}{\partial x \, \partial z}) = \frac{\partial}{\partial y}(\frac{\partial^2 f}{\partial z \, \partial x}) = \frac{\partial^3 f}{\partial y \, \partial z \, \partial x}.$$

Op eenzelfde manier laat zich aantonen dat

$$\frac{\partial^4 g}{\partial y \, \partial x^3} = \frac{\partial^4 g}{\partial x \, \partial y \, \partial x^2} = \frac{\partial^4 g}{\partial x^2 \, \partial y \, \partial x} = \frac{\partial^4 g}{\partial x^3 \, \partial y}.$$

OPDRACHT 4 *Ga voor* $f(x, y, z) := z\, e^{xy} + yz^3x^2$ *door expliciet partieel af te leiden na dat* $\frac{\partial^3 f}{\partial x \, \partial y \, \partial z} = \frac{\partial^3 f}{\partial z \, \partial y \, \partial x}$.

De richtingsafgeleide

Bij de partiële afgeleide hebben we de raaklijnen in de richting van de coördinaatassen bekeken. Het is natuurlijk net zo goed toegestaan de raaklijn in een andere richting te bekijken, bijvoorbeeld als men de verandering van de functie in de richting van de lijn met $x = y$ wil weten. Algemeen kunnen we een richting steeds met een *richtingsvector* $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ in het domein $D \subseteq \mathbb{R}^n$ van de functie $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ aangeven. Hierbij nemen we altijd aan dat $\mathbf{v}$ een vector van lengte 1 is, dus dat $\|\mathbf{v}\| = \sqrt{v_1^2 + \ldots + v_n^2} = 1$.

Als we nu een raaklijn in de richting van zo'n richtingsvector $\mathbf{v}$ willen bepalen, bekijken we net als bij de partiële afgeleide de lijn door de punten $\mathbf{x}$ en $\mathbf{x} + h \cdot \mathbf{v}$ en laten h tegen 0 gaan. Hierbij is het belangrijk dat de richtingsvector $\mathbf{v}$ op lengte 1 genormeerd is. We noemen dan de limiet

$$\lim_{h \to 0} \frac{f(\mathbf{x} + h \cdot \mathbf{v}) - f(\mathbf{x})}{h}$$

de *richtingsafgeleide* van f in de richting $\mathbf{v}$ en noteren dit met $\frac{\partial f}{\partial \mathbf{v}}$.

De gewone partiële afgeleiden vinden we nu als speciale richtingsafgeleiden terug, namelijk als richtingsafgeleiden met betrekking tot de basisvectoren $\mathbf{e}_i$ van de standaardbasis $B = \left\{ \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}$.

Als we de partiële afgeleiden van een functie al kennen, kunnen we hieruit de richtingsafgeleide voor een willekeurige richting makkelijk berekenen, er geldt namelijk voor $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$:

$$\frac{\partial f}{\partial \mathbf{v}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot v_i = \frac{\partial f}{\partial x_1} \cdot v_1 + \ldots + \frac{\partial f}{\partial x_n} \cdot v_n.$$

Helaas zullen we de reden voor deze samenhang pas in de volgende les nader toelichten.

Voorbeeld: We bepalen de richtingsafgeleide van de functie $f(x, y, z) := xyz$ in de richting van de vector $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. Omdat deze vector lengte $\sqrt{2}$ heeft, hebben we de richtingsvector $\mathbf{v} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ nodig. Voor de partiële afgeleiden geldt $\frac{\partial f}{\partial x} = yz$, $\frac{\partial f}{\partial y} = xz$ en $\frac{\partial f}{\partial z} = xy$, dus krijgen we de richtingsafgeleide $\frac{\partial f}{\partial \mathbf{v}} = yz \cdot \frac{1}{\sqrt{2}} + yz \cdot 0 + xy \cdot \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}(yz + xy)$.

OPDRACHT 5 *Bepaal de richtingsafgeleide van $f(x, y, z) := e^x + yz$ in de richting van de vector $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ (let op dat de vector niet lengte 1 heeft).*

1.3 De gradiënt

We zijn nu klaar om een van de meest belangrijke begrippen voor functies van meerdere variabelen te behandelen, namelijk de *gradiënt* van een functie. Dit is eigenlijk helemaal niets nieuws, we schrijven gewoon de verschillende partiële afgeleiden $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ van een functie $f(\mathbf{x}) = f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ in een vector.

Definitie: De vector

$$\nabla f(\mathbf{x}) := \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

heet de *gradiënt van f in* $\mathbf{x}$, waarbij het symbool ∇ als *nabla* te lezen is. Soms wordt ook de notatie grad f in plaats van ∇f gehanteerd.

De gradiënt van een functie geeft voor ieder punt $\mathbf{x}$ in het domein van de functie een vector aan, dus is ∇f zelf een functie van $\mathbb{R}^n$ naar $\mathbb{R}^n$.

Voorbeeld: Zij $f(x,y) := e^{xy} + \sin(xy)$, dan is $\frac{\partial f}{\partial x} = y\,e^{xy} + y\cos(xy)$ en $\frac{\partial f}{\partial y} = x\,e^{xy} + x\cos(xy)$, dus

$$\nabla f(x,y) = \begin{pmatrix} y\,e^{xy} + y\cos(xy) \\ x\,e^{xy} + x\cos(xy) \end{pmatrix} = (e^{xy} + \cos(xy)) \begin{pmatrix} y \\ x \end{pmatrix}.$$

OPDRACHT 6 *Bepaal de gradiënten van de functies $f(x,y) := x\,e^{x^2+y^2}$ en $g(x,y,z) := xy^2 + yz^2 + zx^2$.*

Als we nu nog eens naar de richtingsafgeleide in de richting van de vector $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ kijken, kunnen we dit met behulp van de gradiënt herschrijven als

$$\frac{\partial f}{\partial \mathbf{v}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot v_i = \frac{\partial f}{\partial x_1} \cdot v_1 + \ldots + \frac{\partial f}{\partial x_n} \cdot v_n = \nabla f \cdot \mathbf{v}$$

waarbij we met het product $\mathbf{a} \cdot \mathbf{b}$ van twee vectoren $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ en $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ het *(standaard)inproduct*

$$\sum_{i=1}^{n} a_i b_i = a_1 b_1 + \ldots + a_n b_n$$

van deze twee vectoren bedoelen. We hebben dus een alternatieve definitie voor de richtingsafgeleide gevonden:

Definitie: De richtingsafgeleide $\frac{\partial f}{\partial \mathbf{v}}$ van een functie $f(\mathbf{x})$ in de richting van de vector $\mathbf{v}$ (van lengte 1) is het inproduct $\nabla f \cdot \mathbf{v}$ van $\mathbf{v}$ met de gradiënt ∇f.

Als we deze samenhang tussen richtingsafgeleide en gradiënt combineren met het feit dat de richtingsafgeleide de verandering van de functie in de richting van $\mathbf{v}$ aangeeft, komen we al een heel stuk verder met de interpretatie van de gradiënt. Als we namelijk het inproduct van de gradiënt ∇f met alle vectoren $\mathbf{v}$ met lengte 1 bekijken, kunnen we meteen zeggen voor welke $\mathbf{v}$ het inproduct de maximale waarde aanneemt. In Wiskunde 1 hadden we namelijk gezien dat voor het inproduct $\nabla f \cdot \mathbf{v}$ geldt dat

$$\nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \|\mathbf{v}\| \cdot \cos(\varphi)$$

waarbij φ de hoek tussen de vectoren ∇f en $\mathbf{v}$ is. Maar omdat de richtingsvectoren $\mathbf{v}$ alle dezelfde lengte $\|\mathbf{v}\| = 1$ hebben, geldt $\nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \cos(\varphi)$ en dit

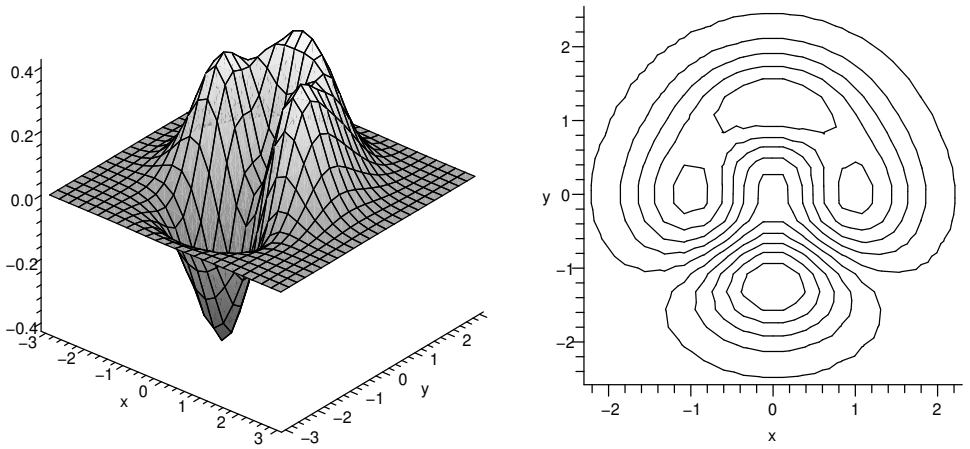
is maximaal als $\cos(\varphi) = 1$, dus als $\varphi = 0$. Het inproduct is dus maximaal als $\mathbf{v}$ precies in de richting van ∇f wijst. Omgekeerd is het inproduct minimaal als $\cos(\varphi) = -1$, dus als $\varphi = 180^\circ$ en dit betekent dat $\mathbf{v}$ in de tegengestelde richting van ∇f wijst. We hebben dus de volgende stelling ingezien:

Stelling: De gradiënt $\nabla f(\mathbf{x})$ wijst in de richting van de maximale toename van de functie f in het punt $\mathbf{x}$. De tegengestelde richting $-\nabla f(\mathbf{x})$ is de richting van de snelste afname van de functie.

> Deze stelling laat zich goed onthouden door te zeggen dat een knikker in de richting $-\nabla f(\mathbf{x})$ loopt als hij in het punt $\mathbf{x}$ op de grafiek van de functie $f(\mathbf{x})$ neergezet wordt.

In plaats van de maximale waarde van het inproduct kunnen we ook eens naar het geval kijken, dat het inproduct $\nabla f \cdot \mathbf{v} = 0$ is. Dit betekent aan de ene kant dat de gradiënt loodrecht op $\mathbf{v}$ staat want $\cos(\varphi) = 0$ voor een rechte hoek φ. Maar aan de andere kant geeft $\nabla f \cdot \mathbf{v}$ de verandering van f in de richting van $\mathbf{v}$ aan en $\nabla f \cdot \mathbf{v} = 0$ betekent dus dat de functie in de richting van $\mathbf{v}$ niet van waarde verandert.

We komen zo naar het begrip van *niveaukrommen*: Dit zijn de krommen in het $x - y$-vlak waarop de functie $f(x, y)$ dezelfde waarde heeft. Iedereen heeft wel eens niveaukrommen gezien, dit zijn namelijk juist de hoogtelijnen op een topografische kaart. In Figuur I.4 zijn naast de grafiek ook niveaukrommen voor de functie $f(x, y) = (x^2 + y^3)\, e^{-x^2-y^2}$ te zien.



Figuur I.4: Grafiek en niveaukrommen voor de functie $f(x, y) = (x^2 + y^3)\, e^{-x^2-y^2}$.

De richting in die een functie in een punt $\mathbf{x}$ niet verandert is juist de raaklijn aan de niveaukromme door $\mathbf{x}$. En deze richting staat loodrecht op de gradiënt. Natuurlijk zijn er twee mogelijke vectoren, maar de raaklijn heeft ook twee richtingen. Onze tweede nieuwe inzicht over de gradiënt luidt dus:

Stelling: De gradiënt $\nabla f(\mathbf{x})$ staat loodrecht op de raaklijn aan de niveaukromme door het punt $\mathbf{x}$.

Voorbeeld: Voor de functie $f(x,y) := (x^2+y^3)\,e^{-x^2-y^2}$ geldt:

$$\frac{\partial f}{\partial x} = 2x\,e^{-x^2-y^2} - 2x(x^2+y^3)\,e^{-x^2-y^2},\ \frac{\partial f}{\partial y} = 3y^2\,e^{-x^2-y^2} - 2y(x^2+y^3)\,e^{-x^2-y^2},$$

en dus

$$\nabla f(x,y) = e^{-x^2-y^2}\begin{pmatrix} 2x - 2x^3 - 2xy^3 \\ 3y^2 - 2x^2y - 2y^4 \end{pmatrix}.$$

Voor $(x,y) = (1,2)$ hebben we bijvoorbeeld $\nabla f(1,2) = e^{-5}\begin{pmatrix} -16 \\ -24 \end{pmatrix}$ en we kunnen in het rechterplaatje van Figuur I.4 controleren dat deze vector inderdaad loodrecht op de niveaukromme staat.

Voor de punten met $x = 0$, dus de punten op de y-as, hebben we $\nabla f(0,y) = e^{-y^2}\begin{pmatrix} 0 \\ 3y^2 - 2y^4 \end{pmatrix}$ en dit betekent dat de niveaukrommen loodrecht op de y-as staan. Op een soortgelijke manier volgt uit $\nabla f(x,0) = e^{-x^2}\begin{pmatrix} 2x - 2x^3 \\ 0 \end{pmatrix}$ dat de niveaukrommen ook loodrecht op de x-as staan. Ook dit kunnen we in het plaatje terugvinden.

In de discussie hierboven hebben we functies van twee veranderlijken bekeken. Als we nu naar het algemene geval van functies van n variabelen kijken, verandert niet zo vreselijk veel. De gradiënt geeft nog steeds de richting in $\mathbb{R}^n$ aan waar de functie het snelste toeneemt.

Om te zien wat er met de niveaukrommen gebeurt, kijken we eerst naar het geval van een functie $f(x,y,z) : \mathbb{R}^3 \to \mathbb{R}$. Als voorbeeld nemen we de functie $f(x,y,z) := \sqrt{x^2+y^2+z^2}$ die de afstand van het nulpunt aangeeft. Voor een vaste waarde r zijn de punten met $f(x,y,z) = r$ juist het oppervlak van een kogel met straal r. Maar het woord opper*vlak* geeft het al aan: De punten met een vaste waarde voor $f(x,y,z)$ liggen op een krom vlak in de 3-dimensionale ruimte. Net zo als een kromme in het 2-dimensionale vlak bij uitvergroten steeds meer op een rechte lijn lijkt, wordt ook een oppervlak in de 3-dimensionale ruimte bij uitvergroten bij benadering een plat vlak.

Als we ons tot kleine omgevingen van een punt beperken, kunnen we ook in het algemeen zeggen dat de punten in $\mathbb{R}^3$ waarop een functie $f(x,y,z)$ dezelfde waarde heeft op een oppervlak in de 3-dimensionale ruimte liggen en dit noemen we een *niveauvlak*. Aan het niveauvlak door een gegeven punt $\mathbf{x} \in \mathbb{R}^3$ kunnen we ook weer een raakvlak aanleggen, en het feit dat de functie $f(x,y,z)$ in de richtingen van dit raakvlak niet verandert laat zien, dat de gradiënt $\nabla f(\mathbf{x})$ loodrecht op dit raakvlak staat.

Voor een algemene functie $f(x_1,\ldots,x_n)$ van n variabelen vormen de punten met constante functiewaarden veralgemeende oppervlakken, die bij uitvergroten op $n-1$-dimensionale hypervlakken lijken (het woord *hypervlak* wordt algemeen voor een $n-1$-dimensionale deelruimte van een n-dimensionale vectorruimte

gebruikt). We noemen de oppervlakken waarop de functie dezelfde waarde heeft *niveauhypervlakken* en we kunnen ook hier aan het niveauhypervlak door een punt $\mathbf{x} \in \mathbb{R}^n$ een raakhypervlak leggen. Dit is een $n-1$-dimensionale deelruimte van $\mathbb{R}^n$ met de richtingen waarin de functie niet van waarde verandert. Dit betekent dat het juist het hypervlak is dat loodrecht op de gradiënt $\nabla f(\mathbf{x})$ staat.

De gradiëntenmethode

Een belangrijke toepassing van de gradiënt zijn benaderingsmethoden voor het vinden van maxima (of minima) van functies, te weten de *gradiëntenmethoden*. Vaak is het namelijk (zelfs bij gewone functies van één veranderlijke) zo, dat maxima niet expliciet bepaald kunnen worden en daarom numeriek benaderd worden. Voor gewone functies kan men dit heel naïf doen:

Kies een stapgrootte Δx en loop zo lang in stappen van Δx naar rechts tot dat de functie niet meer toeneemt. (Als de functie bij de eerste stap al kleiner wordt loop je naar links in plaats van rechts.) Op deze manier kom je (als de functie enigszins glad en Δx voldoende klein is) in de buurt van een lokaal maximum van de functie. Als de benadering nog niet goed genoeg is, kan je natuurlijk vanaf het gevonden punt met een kleinere Δx nog verder door gaan.

Als we hetzelfde idee voor functies van meerdere variabelen willen toepassen, hebben we het probleem een richting te kiezen, want er zijn nu oneindig veel richtingen in plaats van slechts twee (rechts en links). Maar omdat men zo snel mogelijk een maximum wil bereiken, is het voor de hand liggend de richting van de snelste toename van de functie te kiezen, en dat is juist de richting van de gradiënt.

Men kiest daarom ook hier een stapgrootte $\Delta\mathbf{x}$ en loopt bij iedere stap om $\Delta\mathbf{x}$ in de richting van de gradiënt in het laatste punt. Omdat het bepalen van de gradiënt een dure operatie is, wordt in de praktijk meestal pas een nieuwe richting gekozen, als op de lijn in de richting van de gradiënt een maximum van de functie is gevonden (zo als bij gewone functies van één veranderlijke).

Natuurlijk zijn bij ingewikkelde functies vaak ook de partiële afgeleiden niet analytisch te berekenen, maar deze laten zich benaderen door

$$\frac{\partial f}{\partial x_i} \approx \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}$$

voor een voldoende kleine waarde van h.

1.4 De algemene afgeleide

Met de begrippen die we tot nu toe hebben behandeld, kunnen we nu ook het begrip van de algemene afgeleide van een functie van n veranderlijken formuleren.

Een van de ideeën achter de afgeleide van een gewone functies is, dat de functie in een kleine omgeving van een punt door een lijn benaderd kan worden. Als we namelijk de grafiek van een functie $f(x)$ rond een gekozen punt x_0 steeds

meer uitvergroten, lijkt de grafiek (behalve voor exotische functies, die we hier buiten beschouwing laten) steeds meer op de raaklijn in het punt $(x_0, f(x_0))$ met stijging $f'(x_0)$.

Als we nu over een functie van twee veranderlijken nadenken, hebben we ons de grafiek van de functie nu al een paar keer als een soort gebergte voorgesteld. Als we dit nu uitvergroten, wordt het gebergte gewoon een vlakte in de ruimte die de grafiek in het gegeven punt raakt, dus het raakvlak.

Een belangrijke vraag is nu hoe we de functiewaarden op het raakvlak kunnen bepalen, maar in principe hebben we al gezien dat dit juist door de richtingsafgeleide aangegeven wordt:

We hadden gezien dat de functie in de richting van een richtingsvector $\mathbf{v}$ van lengte 1 om het inproduct $\nabla f \cdot \mathbf{v}$ toeneemt (of afneemt). Maar als we nu op een vlak lopen, is de toename langs een veelvoud $c\mathbf{v}$ van $\mathbf{v}$ juist c keer de toename langs $\mathbf{v}$ en omdat het inproduct lineair is, geldt $c(\nabla f \cdot \mathbf{v}) = \nabla f \cdot (c\mathbf{v})$. Maar dit betekent dat de toename van f langs een willekeurige vector $\mathbf{v}$ gegeven is door het inproduct $\nabla f \cdot \mathbf{v}$. Er geldt dus:

Stelling: Op het raakvlak aan de grafiek van $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ is de toename van de functie $f(\mathbf{x})$ langs een willekeurige vector $\mathbf{v}$ gegeven door het inproduct $\nabla f(\mathbf{x}_0) \cdot \mathbf{v}$.

Analoog met de benadering $f(x) - f(x_0) \approx f'(x_0)(x - x_0)$ voor een gewone functie van één veranderlijke krijgen we zo de uitspraak dat

$$f(\mathbf{x}) - f(\mathbf{x}_0) \approx \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$$

voor vectoren $\mathbf{x}$ die dicht bij $\mathbf{x}_0$ liggen.

De limiet definitie voor de afgeleide van een gewone functie was

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \text{ of te wel } \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) = 0$$

als deze limiet bestaat. We kunnen dit makkelijk herschrijven tot de volgende definitie: De functie $f(x) : \mathbb{R} \to \mathbb{R}$ heeft in het punt x_0 de afgeleide $f'(x_0)$ als geldt dat

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0) \cdot (x - x_0)}{x - x_0} = 0.$$

Merk op dat vermenigvuldiging met $f'(x_0)$ een lineaire afbeelding $h \mapsto f'(x_0) \cdot h$ op de 1-dimensionale vectorruimte $\mathbb{R} = \mathbb{R}^1$ geeft. Deze definitie laat zich nu als volgt op functies $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ van n veranderlijken en met m componenten veralgemenen:

Definitie: De functie $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$ heeft in het punt $\mathbf{x}_0$ de *afgeleide* $T := T_{\mathbf{x}_0}$ als $T : \mathbb{R}^n \to \mathbb{R}^m$ een lineaire afbeelding is waarvoor geldt dat

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{f(\mathbf{x}) - f(\mathbf{x}_0) - T(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

Hierbij is de toepassing $T(\mathbf{x} - \mathbf{x}_0)$ van T op de vector $\mathbf{x} - \mathbf{x}_0$ gegeven door het matrix product van de matrix van T met $\mathbf{x} - \mathbf{x}_0$.

Uit de discussie van boven weten we dat voor functies $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ met een enkele component geldt dat $f(\mathbf{x}) - f(\mathbf{x}_0) \approx \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$. Hieruit volgt dat er helemaal geen keuze voor de lineaire afbeelding $T_{\mathbf{x}_0}$ is, er geldt namelijk in dit geval noodzakelijk dat

$$T_{\mathbf{x}_0} = (\nabla f(\mathbf{x}_0))^{tr} = \left(\tfrac{\partial f}{\partial x_1}(\mathbf{x}_0), \ldots, \tfrac{\partial f}{\partial x_n}(\mathbf{x}_0) \right),$$

de afgeleide is dus juist de getransponeerde (gespiegelde) van de gradiënt. Merk op dat voor een vector $\mathbf{v} \in \mathbb{R}^n$ het matrix product $\left(\tfrac{\partial f}{\partial x_1}, \ldots, \tfrac{\partial f}{\partial x_n} \right)(\mathbf{v}) = (\nabla f)^{tr}(\mathbf{v})$ hetzelfde is als het inproduct $\nabla f \cdot \mathbf{v}$.

De Jacobi matrix

Ten slotte komen we terug op een algemene functie $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^m$, die door de componenten $f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})$ beschreven is. Voor iedere van de componenten $f_i(\mathbf{x})$ krijgen we de afgeleide $T_i = (\nabla f_i)^{tr}$ als een rijvector. Als we deze rijvectoren nu als rijen in een matrix schrijven, krijgen we

$$J := \begin{pmatrix} T_1 \\ \vdots \\ T_m \end{pmatrix} = \begin{pmatrix} (\nabla f_1)^{tr} \\ \vdots \\ (\nabla f_m)^{tr} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Deze $m \times n$-matrix $J = (\frac{\partial f_i}{\partial x_j})$ met de partiële afgeleiden van de componenten van $f(\mathbf{x})$ heet de *Jacobi matrix* of kort *Jacobiaan* van $f(\mathbf{x})$. (Let op, soms wordt de term *Jacobiaan* ook voor de determinant van deze matrix gebruikt.)

We weten nu dat voor iedere component f_i van f geldt dat

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - T_i(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

Maar als we nu de componenten tot een m-dimensionale vector samenvoegen, krijgen we hieruit meteen dat

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{f(\mathbf{x}) - f(\mathbf{x}_0) - J(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

We hebben dus de volgende belangrijke stelling ingezien:

Stelling: De afgeleide van een functie $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) : \mathbb{R}^n \to \mathbb{R}^m$ is gegeven door de Jacobi matrix

$$J := \begin{pmatrix} (\nabla f_1)^{tr} \\ \vdots \\ (\nabla f_m)^{tr} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

Voorbeeld: De functie $f(x, y, z) := (z\, e^x, -y\, e^z)$ heeft de Jacobi matrix

$$\begin{pmatrix} z\, e^x & 0 & e^x \\ 0 & -e^z & -y\, e^z \end{pmatrix}.$$

OPDRACHT 7 *Bepaal de Jacobi matrices van de volgende functies:*

(i) $f : \mathbb{R}^3 \to \mathbb{R}^2$ *gegeven door* $f(x, y, z) := (x - y, y + z)$*;*

(ii) $f : \mathbb{R}^2 \to \mathbb{R}^3$ *gegeven door* $f(x, y) := (x + y, x - y, xy)$.

BELANGRIJKE BEGRIPPEN IN DEZE LES

- functie van meerdere variabelen
- continuïteit
- partiële afgeleide
- verruilen van de volgorde van partiële afgeleiden
- richtingsafgeleide
- gradiënt
- niveaukrommen
- Jacobi matrix

OPGAVEN

1. Geef voor de volgende functies de maximale domeinen aan, waarop ze continu gedefinieerd kunnen worden en maak een schets van deze domeinen:
   (i) $f(x, y) := \log((16 - x^2 - y^2)(x^2 + y^2 - 4))$,
   (ii) $f(x, y) := \sqrt{6 - (2x + 3y)}$.
2. Ga na of de volgende functies continu naar de aangegeven punten voortgezet kunnen worden.
   (i) $f(x, y) = \frac{x}{x^2+y^2}$ naar $(x, y) = (0, 0)$,
   (ii) $f(x, y) = \frac{x^2+y^2}{y}$ naar $(x, y) = (0, 0)$,
   (iii) $f(x, y) = \frac{y^3}{x^2+y^2}$ naar $(x, y) = (0, 0)$,
   (iv) $f(x, y) = \frac{\sin(x-y)}{\cos(x+y)}$ naar $(x, y) = (0, 0)$,
   (v) $f(x, y) = \frac{\cos(xy)}{1-x-\cos(y)}$ naar $(x, y) = (1, \pi)$,
   (vi) $f(x, y) = \frac{\sin(xy)}{x^2+y^2}$ naar $(x, y) = (0, 0)$.
3. Bepaal voor de volgende functies de partiële afgeleiden $\frac{\partial f}{\partial x}$ en $\frac{\partial f}{\partial x}$:
   (i) $f(x, y) := xy$;
   (ii) $f(x, y) := e^{xy}$;
   (iii) $f(x, y) := x \cos(x) \cos(y)$;
   (iv) $f(x, y) := (x^2 + y^2) \log(x^2 + y^2)$.

4. Bepaal voor de volgende functies de partiële afgeleiden $\frac{\partial f}{\partial x}$ en $\frac{\partial f}{\partial x}$:

   (i) $f(x,y) := x\, e^{x^2+y^2}$;

   (ii) $f(x,y) := \frac{x^2+y^2}{x^2-y^2}$;

   (iii) $f(x,y) := e^{xy}\log(x^2+y^2)$;

   (iv) $f(x,y) := \frac{x}{y}$;

   (v) $f(x,y) := \cos(y\, e^{xy})\sin(x)$.

5. Bepaal voor de volgende functies de tweede partiële afgeleiden $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial x\,\partial y}$, $\frac{\partial^2 f}{\partial y\,\partial x}$ en $\frac{\partial^2 f}{\partial y^2}$ en laat hiermee in het bijzonder zien dat $\frac{\partial^2 f}{\partial x\,\partial y} = \frac{\partial^2 f}{\partial y\,\partial x}$:

   (i) $f(x,y) := \cos(xy^2)$;

   (ii) $f(x,y) := e^{-xy^2} + y^3x^4$;

   (iii) $f(x,y) := \log(x-y)$;

   (iv) $f(x,y) := \sin(x^2-3xy)$.

6. Laat voor $f(x,y,z,w) := e^{xyz}\sin(xw)$ expliciet zien dat $\frac{\partial^3 f}{\partial w\,\partial z\,\partial x} = \frac{\partial^3 f}{\partial x\,\partial w\,\partial z}$.

7. Bepaal voor de volgende functies de richtingsafgeleiden in de richtingen van de aangegeven vectoren (die niet op lengte 1 genormeerd zijn):

   (i) $f(x,y) := x + 2xy - 3y^2$ in de richting van $\begin{pmatrix}3\\4\end{pmatrix}$;

   (ii) $f(x,y) := \log(x^2+y^2)$ in de richting van $\begin{pmatrix}2\\1\end{pmatrix}$;

   (iii) $f(x,y) := e^x\cos(\pi y)$ in de richting van $\begin{pmatrix}-1\\2\end{pmatrix}$.

8. Bepaal de gradiënten van de functies:

   (i) $f(x,y,z) := x\, e^{-x^2-y^2-z^2}$;

   (ii) $f(x,y,z) := \frac{xyz}{x^2+y^2+z^2}$;

   (iii) $f(x,y,z) := z^2\, e^x\cos(y)$.

9. Bepaal de gradiënten $\nabla f(\mathbf{x})$ van de volgende functies in de aangegeven punten $\mathbf{x}$:

   (i) $f(x,y) := x^2 + 2y^3$ in $(x,y) = (1,1)$;

   (ii) $f(x,y,z) := (x+z)\, e^{x-y}$ in $(x,y,z) = (1,1,1)$;

   (iii) $f(x,y,z) := x^2 + y^2 - z^2$ in $(x,y,z) = (0,0,1)$;

   (iv) $f(x,y,z) := \log(x^2+y^2+z^2)$ in $(x,y,z) = (1,0,1)$.

10. Bepaal voor de volgende functies de richting waarin de functie in het punt $(x,y) = (1,1)$ het snelste toeneemt:

    (i) $f(x,y) := x^2 + 2y^2$;

    (ii) $f(x,y) := x^2 - 2y^2$;

    (iii) $f(x,y) := e^x\sin(y)$;

    (iv) $f(x,y) := e^x\sin(x) - e^{-x}\cos(y)$.

11. Bepaal de Jacobi matrices van de volgende functies:

(i) $f : \mathbb{R}^2 \to \mathbb{R}^2$ gegeven door $f(x, y) := (e^x, \sin(xy))$;

(ii) $f : \mathbb{R}^3 \to \mathbb{R}^3$ gegeven door $f(x, y, z) := (x + z, y - 5z, x - y)$;

(iii) $f : \mathbb{R}^2 \to \mathbb{R}^3$ gegeven door $f(x, y) := (x\, e^y + \cos(y), x, x + e^y)$;

(iv) $f : \mathbb{R}^3 \to \mathbb{R}^2$ gegeven door $f(x, y, z) := (x + e^z + y, yx^2)$;

(v) $f : \mathbb{R}^2 \to \mathbb{R}^3$ gegeven door $f(x, y) := (xy\, e^{xy}, x \sin(y), 5xy^2)$.

Les 2 Taylor reeksen

We hebben in Wiskunde 1 een aantal belangrijke reële functies gezien, bijvoorbeeld de exponentiële functie $\exp(x)$ of de trigonometrische functies $\sin(x)$ en $\cos(x)$. Toen hebben we wel eigenschappen van deze functies aangegeven, bijvoorbeeld dat $\exp(x)$ gekarakteriseerd is door $\exp(x)' = \exp(x)$ en $\exp(0) = 1$. Maar hoe we de waarde van zo'n functie echt kunnen berekenen, of hoe een zakrekenmachine, GRM of computer de waarden van zo'n functie berekent, hebben we toen niet gezien.

De methode die hiervoor (met zekere variaties) wordt gebruikt, is een functie te benaderen door een veelterm. Dit lijkt op het eerste gezicht te eenvoudig om efficiënt te kunnen werken, maar men kan bewijzen dat op een begrensd gebied een continue functie door veeltermen willekeurig goed benaderd kan worden. Naarmate de gewenste nauwkeurigheid toeneemt zijn hiervoor echter veeltermen van hogere graad nodig.

2.1 Interpolatie

Eén mogelijkheid om een benaderende veelterm te vinden, gebruikt de methode van *interpolatie*. We weten dat door 2 punten (x_1, y_1) en (x_2, y_2) met verschillende x-coördinaten $x_1 \neq x_2$ een eenduidige lineaire functie vastgelegd is, met als grafiek de lijn door deze twee punten, te weten

$$l(x) = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1.$$

Net zo leggen 3 punten (weer met verschillende x-waarden) in het $x - y$-vlak eenduidig een parabool, dus een kwadratische functie vast. Op een analoge manier laat zich aantonen dat door $n + 1$ punten met verschillende x-waarden een eenduidige veelterm van graad n vastgelegd is. De punten x_i waarop de waarden gegeven zijn noemt men ook *basispunten* of *roosterpunten*. De veelterm door de punten laat zich met behulp van de *Lagrange interpolatie* zelfs expliciet aangeven:

Stelling: De eenduidige veelterm $p(x)$ van graad n die door de $n + 1$ punten $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$ met verschillende x-coördinaten x_i loopt, is gegeven door

$$p(x) := y_0 \cdot L_0(x) + y_1 \cdot L_1(x) + \ldots + y_n \cdot L_n(x), \text{ waarbij}$$

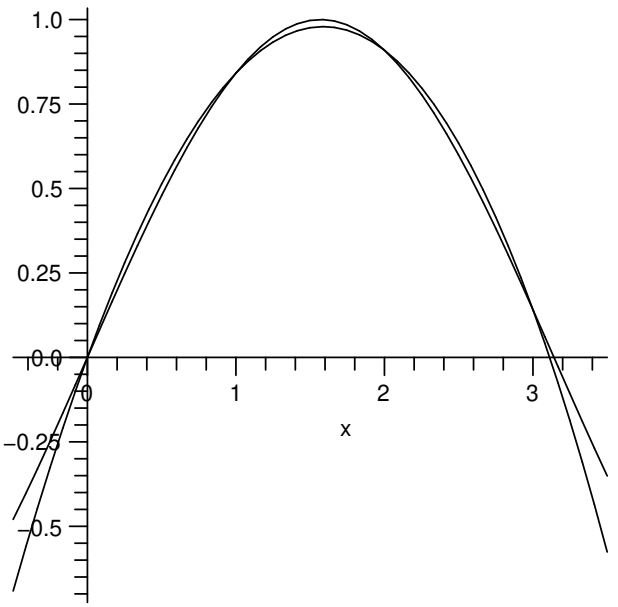
$$L_k(x) := \frac{x - x_0}{x_k - x_0} \cdot \frac{x - x_1}{x_k - x_1} \ldots \frac{x - x_{k-1}}{x_k - x_{k-1}} \cdot \frac{x - x_{k+1}}{x_k - x_{k+1}} \ldots \frac{x - x_n}{x_k - x_n}.$$

De grap bij deze stelling ligt in het feit dat de hulpfuncties $L_k(x)$ juist zo gemaakt zijn dat $L_k(x_k) = 1$ en $L_k(x_i) = 0$ voor $i \neq k$.

In Figuur I.5 zien we dat de sinus functie goed door een derdegraads interpolatie $p(x)$ benaderd wordt, die door de vier basispunten 0, 1, 2 en 3 loopt en

gegeven is door:

$$\begin{aligned} p(x) &= \frac{\sin(1)}{2}x(x-2)(x-3) - \frac{\sin(2)}{2}x(x-1)(x-3) + \frac{\sin(3)}{6}x(x-1)(x-2) \\ &= (\frac{\sin(1)}{2} - \frac{\sin(2)}{2} + \frac{\sin(3)}{6})\,x^3 + (-\frac{5\sin(1)}{2} + 2\sin(2) - \frac{\sin(3)}{2})\,x^2 \\ &+ (3\sin(1) - \frac{3\sin(2)}{2} + \frac{\sin(3)}{3})\,x \\ &\approx -0.0104\,x^3 - 0.3556\,x^2 + 1.2075\,x. \end{aligned}$$



Figuur I.5: Benadering van $\sin(x)$ met behulp van de Lagrange interpolatie door de punten met x-waarden 0, 1, 2 en 3.

> De interpolatie laat zich op een enigszins voor de hand liggende manier ook op functies van meerdere variabelen veralgemenen. Dit heeft in het bijzonder voor punten in de 3-dimensionale ruimte veel toepassingen. Zo worden bijvoorbeeld voor oppervlakken (zo als gezichten) alleen maar een aantal roosterpunten opgeslagen en de punten ertussen door interpolatie berekend om een glad oppervlak te krijgen. Op deze manier laat zich de beweging van een *computer animated* figuur in de *virtual reality* reconstrueren uit gemeten bewegingen van een echte figuur, waarbij alleen maar op enkele punten sensoren zitten.

Er zijn echter ook een paar problemen als we de interpolatie willen gebruiken om een functie te benaderen:

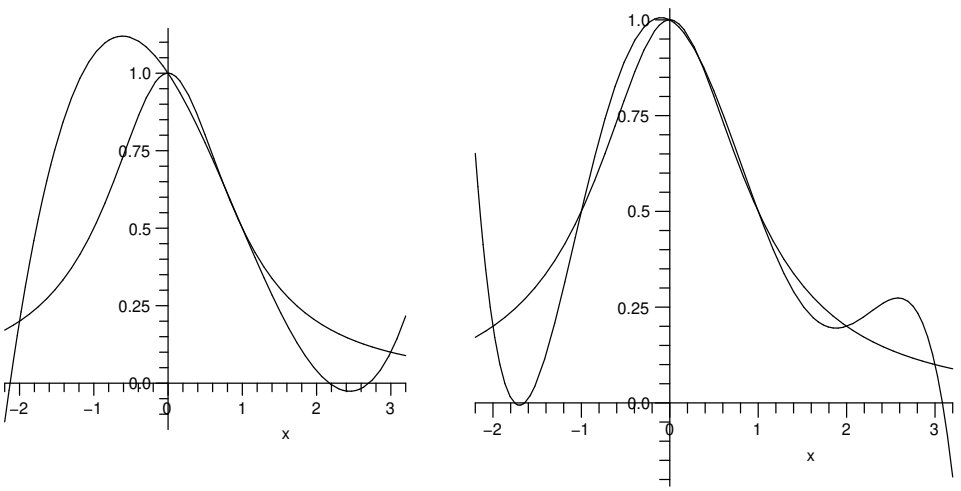
(i) Omdat we een functie $f(x)$ door een veelterm willen benaderen, moeten we de y_k in principe al als functiewaarden $y_k = f(x_k)$ kennen, terwijl we eigenlijk de veelterm juist bepalen om de functiewaarden te kunnen benaderen.

(ii) Afhankelijk van hoe we de x_k kiezen, kan de veelterm $p(x)$ sterk van de functie $f(x)$ afwijkingen, bijvoorbeeld als de x_k te grote afstanden van elkaar hebben.

Het laatste punt is in Figuur I.6 geïllustreerd: We benaderen de functie

$$f(x) := \frac{1}{1+x^2}$$

op het interval $[-2, 3]$ door interpolaties van graad 3 en van graad 5. In het linkerplaatje is de interpolatie met de vier basispunten $-2, 0, 1, 3$ te zien, in het rechterplaatje de interpolatie met de zes basispunten $-2, -1, 0, 1, 2, 3$. In beide gevallen wijkt $f(x)$ sterk van de interpolatie af, alleen maar op het interval $[0, 1.5]$ is de benadering enigszins redelijk.



Figuur I.6: Interpolatie van $\frac{1}{1+x^2}$ met 4 en 6 basispunten.

In principe zouden we op intervallen waar de functie $f(x)$ heel sterk verandert veel basispunten x_k willen hebben, terwijl op stukken waar de functie bijna lineair is niet zo veel basispunten nodig zijn. Om hierover te kunnen beslissen moeten we naar de hogere afgeleiden van $f(x)$ kijken en dit idee geeft aanleiding tot een andere manier om een benaderende veelterm te bepalen. Deze gebruikt inderdaad de hogere afgeleiden van $f(x)$, maar slechts in een enkele punt.

2.2 Taylor veeltermen

Het idee bij de *Taylor veelterm* is, een functie $f(x)$ in de omgeving van één punt x_0 te benaderen. Hiervoor wordt een veelterm geconstrueerd die in het punt x_0

niet alleen maar dezelfde functiewaarde als $f(x)$ heeft, maar ook dezelfde eerste afgeleide, dezelfde tweede afgeleide, enzovoorts.

Definitie: De *Taylor veelterm* $p(x) := p_{f,n,x_0}(x)$ van graad n in het punt x_0 is voor een in x_0 (minstens) n keer differentieerbare functie $f(x)$ vastgelegd door de eigenschappen:

$$p(x_0) = f(x_0), \quad p'(x_0) = f'(x_0), \quad p''(x_0) = f''(x_0), \ \ldots, \ p^{(n)}(x_0) = f^{(n)}(x_0),$$

de Taylor veelterm $p(x)$ heeft in het punt x_0 dus precies dezelfde afleidingen (tot orde n) als $f(x)$.

Men gaat na dat deze eigenschappen inderdaad een eenduidige veelterm van graad n bepalen, namelijk de veelterm

$$p(x) = c_n(x - x_0)^n + c_{n-1}(x - x_0)^{n-1} + \ldots + c_1(x - x_0) + c_0$$

met coëfficiënten

$$c_0 = f(x_0), \quad c_1 = f'(x_0), \quad c_2 = \frac{f''(x_0)}{2!}, \quad \ldots \quad c_n = \frac{f^{(n)}(x_0)}{n!}.$$

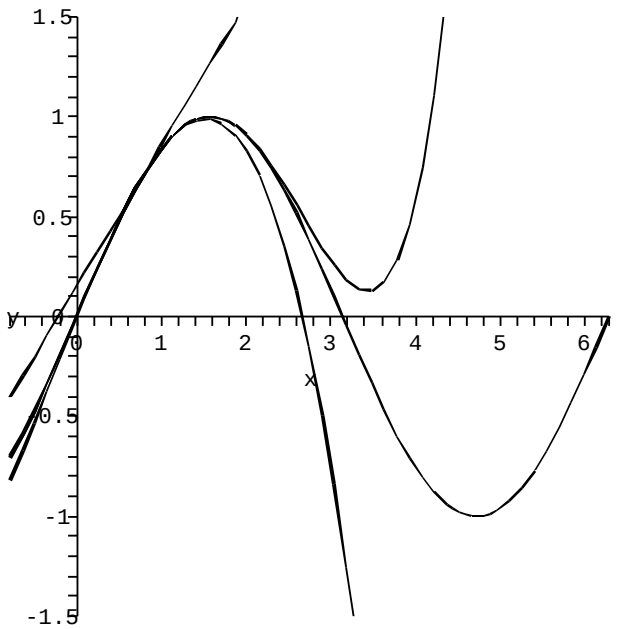
Dat dit inderdaad de coëfficiënten zijn, ziet men door het uitschrijven van de afgeleiden van $p(x)$, invullen van x_0 en vergelijken met $f^{(n)}(x_0)$:

$$\begin{aligned}
p(x) &= c_n(x - x_0)^n + \ldots + c_3(x - x_0)^3 + c_2(x - x_0)^2 + c_1(x - x_0) + c_0 \\
&\Rightarrow p(x_0) = c_0, \text{ uit } f(x_0) = p(x_0) = c_0 \text{ volgt dan } c_0 = f(x_0) \\
p'(x) &= n \cdot c_n(x - x_0)^{n-1} + \ldots + 3 \cdot c_3(x - x_0)^2 + 2 \cdot c_2(x - x_0) + c_1 \\
&\Rightarrow p'(x_0) = c_1, \text{ uit } f'(x_0) = p'(x_0) = c_1 \text{ volgt dan } c_1 = f'(x_0) \\
p''(x) &= n(n-1) \cdot c_n(x - x_0)^{n-2} + \ldots + 3 \cdot 2 \cdot c_3(x - x_0) + 2 \cdot c_2 \\
&\Rightarrow p''(x_0) = 2 \cdot c_2, \text{ uit } f''(x_0) = p''(x_0) = 2 \cdot c_2 \text{ volgt dan } c_2 = \frac{f''(x_0)}{2} \\
p'''(x) &= n(n-1)(n-2) \cdot c_n(x - x_0)^{n-3} + \ldots + 3 \cdot 2 \cdot c_3 \\
&\Rightarrow p'''(x_0) = 3 \cdot 2 \cdot c_3 \Rightarrow c_3 = \frac{f'''(x_0)}{3!} \quad \text{enzovoorts.}
\end{aligned}$$

Het idee achter de geëiste eigenschappen is, dat het overeenstemmen van de hogere afgeleiden ervoor zorgt dat de grafiek van de Taylor veelterm zich rond het punt x_0 steeds beter aan de grafiek van $f(x)$ vlijt.

In Figuur I.7 zien we het effect van Taylor veeltermen van verschillende graad. Terwijl de veelterm van graad 1 de grafiek van $\sin(x)$ alleen maar raakt (omdat $\sin(x)$ geen rechte stukken heeft), is de veelterm van graad 3 al een redelijke benadering tussen ongeveer 0.5 en 1.5, voordat hij naar beneden wegduikt. Met de veelterm van graad 5 gaat het iets langer goed, maar dan loopt deze naar boven weg.

Meestal wordt de Taylor veelterm niet in de boven aangegeven volgorde van de termen geschreven (dus beginnend met de hoogste), maar andersom. Verder wordt het feit dat de Taylor veelterm de functie slechts in een kleine omgeving



Figuur I.7: Benadering van $\sin(x)$ door de Taylor veeltermen van graad 1, 3 en 5 in het punt $x_0 = \frac{\pi}{4}$

van x_0 goed benadert weergegeven door x in de vorm $x = x_0 + h$ te schrijven (waarbij men stiekem verondersteld dat h klein is). Dit geeft de volgende vorm van de Taylor veelterm:

$$p(x_0 + h) = f(x_0) + f'(x_0) \cdot h + \frac{f''(x_0)}{2} \cdot h^2 + \ldots + \frac{f^{(n)}(x_0)}{n!} \cdot h^n.$$

OPDRACHT 8 *Bepaal voor* $f(x) := \sqrt{x}$ *de Taylor veelterm van graad* 4 *in het punt* $x_0 = 1$.

Over het verband tussen de benaderde functie $f(x)$ en de Taylor veelterm $p(x)$ valt echter nog veel meer te zeggen, in het bijzonder laat zich expliciet een afschatting voor de fout $|f(x_0 + h) - p(x_0 + h)|$ aangeven (die natuurlijk van h afhangt). We definiëren hiervoor de n-de *foutterm* $R_n(h)$ door

$$f(x_0 + h) = p(x_0 + h) + R_n(h) \quad \text{of} \quad R_n(h) = f(x_0 + h) - p(x_0 + h).$$

Er zijn verschillende manieren om de foutterm $R_n(h)$ aan te geven, de drie

meest belangrijke zijn de volgende:

$$R_n(h) = \frac{1}{(n+1)!} \cdot f^{(n+1)}(t) \cdot h^{n+1} \text{ voor een } t \in [0,h] \quad \text{(Lagrange vorm)}$$

$$R_n(h) = \frac{1}{n!} \cdot f^{(n+1)}(t) \cdot (h-t)^n h \text{ voor een } t \in [0,h] \quad \text{(Cauchy vorm)}$$

$$R_n(h) = \frac{1}{n!} \int_0^h f^{(n+1)}(t) \cdot (h-t)^n \, dt \quad \text{(integraal vorm)}$$

> Het bewijs dat de foutterm op deze manieren geschreven kan worden is niet eens zo moeilijk. Bijvoorbeeld berust de Lagrange vorm op de *middelwaardestelling*, die zegt, dat er voor een (differentieerbare) functie $f(x)$ op een interval $[a,b]$ een punt $c \in [a,b]$ in het interval ligt waar de raaklijn aan $f(x)$ dezelfde stijging heeft als de gemiddelde stijging $\frac{f(b)-f(a)}{b-a}$ van $f(x)$ op het interval $[a,b]$, d.w.z. waar geldt dat $f(b)-f(a) = f'(c)(b-a)$. Dit klopt omdat de raaklijn niet overal een grotere of overal een kleinere stijging dan de gemiddelde stijging kan hebben.

Vooral de Lagrange vorm van de foutterm is vaak geschikt om de fout bij de benadering door de Taylor veelterm af te schatten. Dit lukt in het bijzonder als zich makkelijk een grens voor de $(n+1)$-de afgeleide van $f(x)$ op het interval $[0,h]$ laat aangeven.

Bijvoorbeeld geldt voor $f(x) = \sin(x)$ dat $|f^{(n+1)}(t)| \leq 1$ voor alle t, omdat de afgeleiden $\pm\sin(t)$ of $\pm\cos(t)$ zijn en deze functies alleen maar waarden tussen -1 en 1 hebben.

Net zo goed ziet men in dat voor $f(x) = \exp(x)$ voor $t \in [0,h]$ (met $h > 0$) geldt dat $|f^{(n+1)}(t)| \leq e^h$ omdat de afgeleiden de exponentiële functie reproduceren en de exponentiële functie monotoon stijgend is.

Voorbeelden:

(1) Voor $f(x) := \sin(x)$ en $x_0 = 0$ hebben we $f(x_0) = \sin(0) = 0$, $f'(x_0) = \cos(0) = 1$, $f''(x_0) = -\sin(0) = 0$, $f'''(x_0) = -\cos(0) = -1$ en $f^{(4)}(x_0) = \sin(0) = 0$. De Taylor veelterm van graad 4 van $\sin(x)$ rond $x_0 = 0$ is dus

$$p(x) = 0 + x - 0 - \frac{1}{6}x^3 + 0 = x - \frac{1}{6}x^3.$$

Als we nu bijvoorbeeld de waarde van $\sin(\frac{\pi}{10})$ willen bepalen, krijgen we de benadering $\sin(\frac{\pi}{10}) \approx \frac{\pi}{10} - \frac{1}{6}(\frac{\pi}{10})^3 = \frac{\pi}{10} - \frac{\pi^3}{6000} \approx 0.30899$.

De fout kunnen we afschatten met $|R_4(\frac{\pi}{10})| < \frac{1}{5!} \cdot 1 \cdot (\frac{\pi}{10})^5 < 2.6 \cdot 10^{-5} = 0.000026$, dus hebben we aangetoond dat $0.30896 < \sin(\frac{\pi}{10}) < 0.30902$.

(2) We kunnen het Euler getal e benaderen met behulp van de Taylor ontwikkeling van $f(x) := \exp(x)$ in $x_0 = 0$. Omdat $f^{(n)}(x) = \exp(x)$, geldt $f^{(n)}(0) = 1$ voor alle n. We krijgen daarom de n-de Taylor veelterm

$$p(h) = 1 + h + \frac{h^2}{2} + \frac{h^3}{3!} + \ldots + \frac{h^n}{n!}.$$

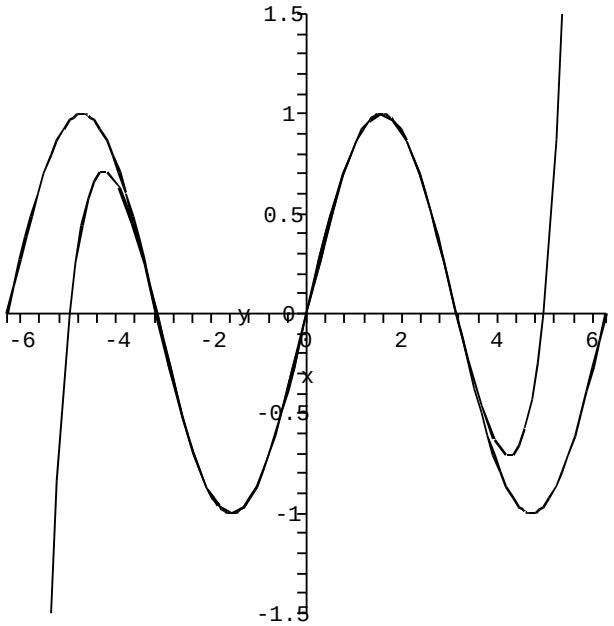
Als benadering van $e = f(1)$ krijgen we met $n = 6$ bijvoorbeeld $p(1) = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \frac{1}{120} + \frac{1}{720} \approx 2.718$.

De fout kunnen we afschatten met $\frac{1}{(n+1)!}\, e^t \cdot 1^{n+1}$ voor een $t \in [0, 1]$, maar omdat we e juist willen berekenen, moeten we hier een grove afschatting voor e nemen, bijvoorbeeld $e^t \leq e < 3$. Hieruit volgt dat $|R_n(1)| < \frac{3}{7!} = \frac{3}{5040} \approx 0.0006$, dus hebben we aangetoond dat $2.7174 < e < 2.7186$.

OPDRACHT 9 *Bepaal met behulp van een Taylor veelterm de waarde van* $\sin(1)$ *op* 8 *decimalen, d.w.z. met een fout van hoogstens* 10^{-8}.

2.3 Taylor reeksen

Door de graad van de Taylor veelterm te laten groeien, krijgen we (in goedaardige gevallen) een steeds betere benadering van een functie, en meestal ook op een groter interval. In Figuur I.8 zien we bijvoorbeeld, dat de Taylor veelterm van graad 10 (deze keer in het punt $x_0 = 0$ berekend) tussen $-\pi$ en π bijna niet van de functie $\sin(x)$ te onderscheiden is.



Figuur I.8: Benadering van $\sin(x)$ door de Taylor veelterm van graad 10 in het punt $x_0 = 0$

Men krijgt dus het idee dat de benadering steeds beter wordt als we de graad verhogen, en het beste zou zijn, helemaal niet te stoppen maar oneindig door te gaan. Dit doen we dus!

Hier wordt nu duidelijk waarom het nuttig is de Taylor veelterm niet met de hoogste term $\frac{f^{(n)}(x_0)}{n!}h^n$ te beginnen, maar met de laagste term $f(x_0)$, want

op deze manier kunnen we op een geschikte plek *enzovoorts* zeggen om aan te duiden, dat de veelterm met hogere termen door gaat.

Machtreeksen

Als men een algemene veeltermen met opstijgende termen schrijft, krijgt men iets als

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n = \sum_{i=0}^{n} a_ix^i.$$

Als we nu niet met een term a_nx^n stoppen, maar ook voor een willekeurig hoge graad i steeds nog een coëfficiënt a_i van x^i definiëren (die weliswaar 0 mag zijn), noemen we dit geen veelterm meer, maar een *reeks*, soms voor alle duidelijkheid zelfs een *oneindige reeks*. Een reeks is dus een uitdrukking van de vorm

$$r(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots = \sum_{n=0}^{\infty} a_nx^n.$$

Veeltermen zijn nu een speciale vorm van reeksen, een veelterm van graad n is namelijk een reeks waarbij vanaf $i = n + 1$ geldt dat $a_i = 0$ is.

Er zijn twee verschillende manieren hoe men tegen een oneindige reeks aan kan kijken, van een meer algebraïsch standpunt of van een meer analytisch standpunt.

- Algebraïsch kan men een reeks zien als een abstracte uitdrukking die gerepresenteerd is door de oneindige rij $(a_0, a_1, a_2, \ldots)$ van coëfficiënten. Twee reeksen laten zich net zo optellen als men veeltermen optelt, namelijk door de coëfficiënten van iedere term x^n bij elkaar op te tellen.

  Desnoods laten zich twee reeksen ook met elkaar vermenigvuldigen, dit geeft het zogeheten *convolutieproduct* of *Cauchy product* waarbij men in het product van de twee reeksen weer de coëfficiënten van de termen x^n bepaald. Dit gebeurt in principe ook hetzelfde als bij veeltermen, omdat alleen maar de termen tot graad n in de twee reeksen een bijdrage aan de term x^n in het product kunnen leveren. Er geldt

  $$(\sum_{n=0}^{\infty} a_nx^n) \cdot (\sum_{n=0}^{\infty} b_nx^n) = \sum_{n=0}^{\infty} c_nx^n \text{ met } c_n = \sum_{i=0}^{n} a_ib_{n-i},$$

  bijvoorbeeld is $c_3 = a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0$.
- Analytisch zien we $f(x) = \sum_{n=0}^{\infty} a_nx^n$ als functie, waarbij we voor de functiewaarde in het punt x een limiet moeten bekijken, namelijk de limiet

  $$f(x) = \lim_{n\to\infty} \sum_{i=0}^{n} a_ix^i.$$

  Deze limiet hoeft niet voor iedere x te bestaan, dus hebben we het alleen maar voor degene waarden x echt met een functie te maken, waar deze limiet inderdaad bestaat.

Voorbeeld: Als we een voorbeeld van een reeks willen bekijken die geen veelterm is, schiet misschien als eerste de reeks door het hoofd waarbij alle coëfficiënten $a_n = 1$ zijn, dus de reeks

$$r(x) = 1 + x + x^2 + \ldots = \sum_{n=0}^{\infty} x^n.$$

Deze reeks heet de *meetkundige reeks* en heeft de mooie eigenschap dat voor $|x| < 1$ geldt dat

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

> Dit gaat men na, door het product $(1 + x + \ldots + x^n)(1 - x)$ uit te schrijven als $(1 + x + \ldots + x^n)(1 - x) = (1 + x + \ldots + x^n) - (x + x^2 + \ldots + x^{n+1}) = 1 - x^{n+1}$. Voor $|x| < 1$ is $\lim_{n\to\infty} x^{n+1} = 0$, dus gaat $(1 + x + \ldots + x^n)(1 - x) \to 1$ en we moeten alleen nog door $(1 - x)$ delen om de boven aangegeven formule te krijgen.

De Taylor reeks van een functie

We gaan nu van de Taylor veeltermen een reeks maken door de graad van de veeltermen naar oneindig te laten gaan. De reeks die we zo krijgen noemen we natuurlijk Taylor reeks.

Definitie: De oneindige reeks $T(x) := T_{f,x_0}(x)$ gedefinieerd door

$$\begin{aligned} T(x) &:= \lim_{n\to\infty} p_{f,n,x_0}(x) \\ &= f(x_0) + f'(x_0)(x - x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \ldots \\ &= \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n. \end{aligned}$$

heet de *Taylor reeks* van $f(x)$ in het punt x_0.

Over de Taylor reeks als limiet van de Taylor veeltermen moeten we een paar belangrijke opmerkingen kwijt:

(1) Natuurlijk geldt $T(x_0) = f(x_0)$, omdat alle termen in de som voor $n > 0$ wegvallen. Maar het kan zijn dat de reeks $T(x)$ voor geen enkele $x \neq x_0$ convergeert, dus een limiet heeft. Voor goedaardige functies (zo als we die in de praktijk kunnen verwachten) geldt dit wel, tenminste in een (voldoende klein) interval rond x_0.

De vraag of en waar de Taylor reeks van een functie convergeert, geeft aanleiding tot belangrijke en diepe stellingen in de wiskunde, maar is in deze cursus een minder belangrijke vraagstelling, omdat we veronderstellen dat we het met voldoende gladde functies te maken hebben.

(2) Zelfs als de Taylor reeks $T(x)$ voor een zekere waarde x convergeert, hoeft de limiet niet de juiste functiewaarde $f(x)$ te zijn. Een afschrikkend voorbeeld hiervoor is de functie

$$f(x) := e^{-\frac{1}{x^2}}$$

die door $f(0) := 0$ continu naar 0 voortgezet wordt. Deze functie is zelfs willekeurig vaak differentieerbaar, en er geldt $f^{(n)}(0) = 0$ voor alle n. Dit betekent, dat de coëfficiënten van de Taylor reeks van $f(x)$ in het punt $x_0 = 0$ alle 0 zijn en dus de Taylor reeks $T(x) = 0$ is. Aan de andere kant is $f(x) \neq 0$ voor $x \neq 0$, dus geeft de Taylor reeks de functie alleen maar in het nulpunt $x_0 = 0$ weer.

(3) De grootste afstand r zo dat $T(x)$ voor alle x met $|x - x_0| < r$ naar de goede waarde $f(x)$ convergeert, noemt men de *convergentiestraal* van $T(x)$.

Als we bij een functie $f(x)$ ervan uit gaan dat de Taylor reeks $T(x)$ in de punten waar hij convergeert ook naar de goede waarde $f(x)$ convergeert, schrijven we gewoon $f(x)$ in plaats van $T(x)$. Net als bij de Taylor veeltermen is het ook hier gebruikelijk, duidelijk te maken dat x dicht bij het punt x_0 ligt door $x = x_0 + h$ te schrijven. Op deze manier krijgt men de volgende vorm van de Taylor reeks:

$$f(x_0 + h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}\, h^n.$$

Voorbeelden:

(1) $\exp(x)$:
We bekijken de Taylor reeks in het punt $x_0 = 0$, daar geldt $\exp^{(n)}(x_0) = \exp(0) = 1$ voor alle $n \in \mathbb{N}$. De Taylor veelterm van graad n voor $\exp(x)$ in het punt 0 is dus

$$1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \ldots + \frac{1}{n!}x^n = \sum_{i=0}^{n} \frac{x^i}{i!}$$

en de Taylor reeks in het punt 0 is de limiet $n \to \infty$ van deze veeltermen, dus

$$T(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots.$$

Er laat zich aantonen dat deze Taylor reeks convergentiestraal ∞ heeft, d.w.z. dat $T(x)$ voor elke $x \in \mathbb{R}$ naar $\exp(x)$ convergeert. Omdat de noemers met $n!$ heel snel groeien, is de convergentie erg goed en kunnen we de exponentiële functie al met weinig termen goed benaderen.

(2) $\sin(x)$:
Ook voor de sinus functie bepalen we de Taylor reeks in $x_0 = 0$. Merk op dat $\sin''(x) = -\sin(x)$, $\sin'''(x) = -\sin'(x)$ en $\sin^{(4)}(x) = \sin(x)$, hieruit

volgt dat de afgeleiden (beginnend met de 0-de afgeleide $\sin(0) = 0$) van $\sin(x)$ in het punt $x = 0$ periodiek $0, 1, 0, -1, 0, 1, 0, -1, 0$ enz. zijn. Dit betekent voor algemene n dat

$$\sin^{(2n)}(0) = 0 \quad \text{en} \quad \sin^{(2n+1)}(0) = (-1)^n.$$

De Taylor reeks van $\sin(x)$ in het punt 0 is dus

$$T(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{5040} + \dots.$$

(3) $\cos(x)$:
Analoog met de sinus geldt voor de cosinus dat $\cos''(x) = -\cos(x)$, $\cos'''(x) = -\cos'(x)$ en $\cos^{(4)}(x) = \cos(x)$, dus zijn hier de afgeleiden in het punt $x_0 = 0$ periodiek $1, 0, -1, 0, 1, 0, -1, 0, 1$ enz. We hebben dus

$$\cos^{(2n)}(0) = (-1)^n \quad \text{en} \quad \cos^{(2n+1)}(0) = 0.$$

De Taylor reeks van $\cos(x)$ in het punt 0 is dus

$$T(x) = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \dots.$$

Net als bij de exponentiële functie is de convergentiestraal ook bij de Taylor reeks van de sinus en cosinus functies oneindig. Verder is ook hier de convergentie van de Taylor reeks erg goed, zo dat we snel een goede benadering van $\sin(x)$ of $\cos(x)$ vinden.

(4) $\log(x)$:
Omdat de logaritme voor $x = 0$ niet gedefinieerd is, moeten we hier de Taylor reeks in een andere punt bepalen, en we kiezen hiervoor $x_0 = 1$. Maar omdat het uiteindelijk toch prettiger is om een reeks met termen x^n en niet $(x-1)^n$ te hebben, gebruiken we een klein trucje: In plaats van $\log(x)$ kijken we naar de functie $f(x) := \log(x+1)$ en bepalen hiervoor de Taylor reeks in het punt $x_0 = 0$.

Er geldt $f'(x) = \frac{1}{x+1} = (x+1)^{-1}$, $f''(x) = (-1) \cdot (x+1)^{-2}$, $f'''(x) = (-1)(-2) \cdot (x+1)^{-3}$, $f^{(4)}(x) = (-1)(-2)(-3) \cdot (x+1)^{-4}$ en algemeen

$$f^{(n)}(x) = (-1)(-2)\dots(-(n-1)) \cdot (x+1)^{-n} = (-1)^{n-1}(n-1)! \frac{1}{(x+1)^n}.$$

In het bijzonder is $f^{(n)}(0) = (-1)^{n-1}(n-1)!$ en dus $\frac{f^{(n)}(0)}{n!} = \frac{(-1)^{n-1}}{n}$. De Taylor reeks van $\log(x+1)$ in het punt 0 is dus

$$T(x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots.$$

Merk op dat deze Taylor reeks hoogstens convergentiestraal 1 kan hebben, omdat de logaritme in 0 niet gedefinieerd is. Dit is echter ook de convergentiestraal, we kunnen met deze reeks dus alleen maar waarden in het

interval $(0, 2)$ bepalen. Maar dit stelt geen probleem voor het berekenen van $\log(x)$ voor, want voor $x \geq 2$ is $\frac{1}{x} \leq \frac{1}{2}$ en dit ligt wel in het interval $[0, 2]$ en we berekenen $\log(x)$ met behulp van de relatie $\log(x) = -\log(\frac{1}{x})$.

Er valt wel nog op te merken, dat de convergentie van de reeks voor de logaritme veel slechter is dan die voor $\exp(x)$, $\sin(x)$ of $\cos(x)$, omdat de noemers met n en niet met $n!$ groeien.

OPDRACHT 10 *Bepaal de Taylor reeks van* $f(x) := \sqrt{x+1}$ *in het punt* $x_0 = 0$.

2.4 Taylor reeksen voor functies van meerdere variabelen

Tot nu toe hebben we in deze les erna gekeken hoe we een gewone functie van één variabel door een oneindige reeks kunnen beschrijven of door een veelterm kunnen benaderen die we door afbreken van de Taylor reeks krijgen. De benadering van een functie $f(x)$ door de Taylor veelterm van graad n wordt vaak ook kort de n-de benadering van $f(x)$ genoemd. De meest belangrijke van deze benaderingen zijn de 0de (ook al lijkt het flauw), de 1ste of *lineaire* en de 2de of *kwadratische* benadering.

We zullen nu kijken, hoe we Taylor veeltermen op functies van meerdere veranderlijken kunnen veralgemenen. Het idee dat we hierbij hanteren is hetzelfde: We benaderen een functie $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ door een veelterm $p(\mathbf{x})$ in de n variabelen die in een vaste punt $\mathbf{x}_0$ dezelfde functiewaarde heeft als $f(\mathbf{x})$ en ook dezelfde eerste, tweede enz. partiële afgeleiden.

Om de notatie overzichtelijk te houden, zullen we weer vooral naar functies van twee variabelen kijken, de resultaten laten zich dan makkelijk ook voor n variabelen formuleren.

Om te beginnen moeten we afspreken wat een veelterm van meerdere variabelen is, en wat de graad daarvan is.

- Een *veelterm* in de twee variabelen x en y is een (eindige) som van termen van de vorm $c\,x^i y^j$, waarbij we c de *coëfficiënt* van $x^i y^j$ noemen. Een zuiver product $x^i y^j$ van machten van de variabelen noemt men ook en *monoom*.

- De *graad* van een term $c\,x^i y^j$ is de som $i + j$ van de machten van de variabelen. De graad van een veelterm is het maximum van de graden van de termen.

- Een veelterm van graad 0 heet een *constante functie*, een veelterm van graad 1 een *lineaire veelterm* of *lineaire functie* en een veelterm van graad 2 een *kwadratische veelterm* of *kwadratische functie*.

Voorbeeld: De algemene kwadratische veelterm van twee variabelen is van de vorm

$$p(x, y) = a + b_1 x + b_2 y + c_1 x^2 + c_2 xy + c_3 y^2.$$

We zullen nu de algemene definitie van de begrippen *veelterm* en *graad* voor n variabelen geven. Dit vergt enigszins veel indices en ingewikkelde notaties, maar met het geval van twee variabelen voor ogen zal het geen probleem zijn.

Definitie: Een *veelterm* in de n variabelen $x_1, \ldots, x_n$ is een eindige som van termen van de vorm

$$c\, x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}.$$

De *graad* van een term $c\, x_1^{i_1} \cdot \ldots \cdot x_n^{i_n}$ is de som $i_1 + \ldots + i_n$ van de machten van de variabelen.

De *graad van een veelterm* is het maximum van de graden van de termen in de veelterm.

Voorbeelden:

(1) De algemene lineaire veelterm in $\mathbf{x} = (x_1, \ldots, x_n)$ is

$$p(\mathbf{x}) = a + b_1 x_1 + \ldots + b_n x_n = a + \sum_{i=1}^{n} b_i x_i.$$

(2) De algemene kwadratische veelterm is van de vorm

$$p(\mathbf{x}) = a + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} c_{ij} x_i x_j.$$

Lineaire benadering

Als we een functie $f(x, y)$ door een lineaire veelterm $p(x, y)$ willen benaderen, kunnen we eisen dat $p(x_0, y_0) = f(x_0, y_0)$ voor een vast gekozen punt (x_0, y_0) en dat de eerste partiële afgeleiden van $p(x, y)$ overeenkomen met de eerste partiële afgeleiden van $f(x, y)$ in het punt (x_0, y_0).

Voor het gemak nemen we nu nog aan, dat het vast gekozen punt (x_0, y_0) het nulpunt $(0, 0)$ is, we komen naar een algemeen punt terug door x door $h_1 := x - x_0$ en y door $h_2 := y - y_0$ te vervangen.

Voor de algemene lineaire veelterm

$$p(x, y) = a + b_1 x + b_2 y$$

geldt

$$\frac{\partial p}{\partial x} = b_1 \quad \text{en} \quad \frac{\partial p}{\partial y} = b_2.$$

Hieruit volgt

$$p(0,0) = a, \quad \frac{\partial p}{\partial x}(0,0) = b_1, \quad \frac{\partial p}{\partial y}(0,0) = b_2$$

en om te bereiken dat

$$p(0,0) = f(0,0), \quad \frac{\partial p}{\partial x}(0,0) = \frac{\partial f}{\partial x}(0,0), \quad \frac{\partial p}{\partial y}(0,0) = \frac{\partial f}{\partial y}(0,0)$$

moet dus gelden dat

$$a = f(0,0), \quad b_1 = \frac{\partial f}{\partial x}(0,0), \quad b_2 = \frac{\partial f}{\partial y}(0,0).$$

Voor de *lineaire benadering* van $f(x,y)$ in het punt $(x,y) = (0,0)$ krijgen we dus de lineaire veelterm

$$p(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)\,x + \frac{\partial f}{\partial y}(0,0)\,y.$$

> Omdat $p(x,y)$ zo is gekozen dat de functiewaarde en de eerste partiële afgeleiden met $f(x,y)$ overeenkomen, geeft $p(x,y)$ juist het raakvlak aan de grafiek van $f(x,y)$ in het punt (x_0,y_0) aan.

Met behulp van de gradiënt kunnen we de lineaire benadering nog iets eenvoudiger schrijven: Met de notaties $\mathbf{x}_0 = \begin{pmatrix}0\\0\end{pmatrix}$ en $\mathbf{h} = \begin{pmatrix}x\\y\end{pmatrix}$ geldt:

$$p(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h}.$$

Met deze notatie hebben we inderdaad de algemene vorm van de Taylor veelterm van graad 1 gevonden:

Stelling: Voor een functie $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ van n variabelen is de Taylor veelterm van graad 1 in het punt $\mathbf{x}_0$ gegeven door

$$p(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h} = f(\mathbf{x}_0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\,h_i$$

waarbij h_i de i-de component van de vector $\mathbf{h}$ is, dus $\mathbf{h} = \begin{pmatrix}h_1\\ \vdots \\ h_n\end{pmatrix}$.

Kwadratische benadering

Voor de algemene kwadratische veelterm

$$p(x,y) = a + b_1x + b_2y + c_1x^2 + c_2xy + c_3y^2$$

geldt

$$\frac{\partial p}{\partial x} = b_1 + 2c_1x + c_2y \quad \text{en} \quad \frac{\partial p}{\partial y} = b_2 + c_2x + 2c_3y.$$

Om naar de Taylor veelterm van graad 2 te komen, moeten we ervoor zorgen dat naast de eerste partiële afgeleiden ook de tweede partiële afgeleiden van de algemene kwadratische veelterm

$$p(x,y) = a + b_1x + b_2y + c_1x^2 + c_2xy + c_3y^2$$

in het punt $(0,0)$ met de tweede partiële afgeleiden van $f(x,y)$ in dit punt overeenkomen.

Voor de eerste partiële afgeleiden geldt

$$\frac{\partial p}{\partial x} = b_1 + 2c_1x + c_2y \quad \text{en} \quad \frac{\partial p}{\partial y} = b_2 + c_2x + 2c_3y.$$

en net als boven volgt uit $\frac{\partial p}{\partial x}(0,0) = \frac{\partial f}{\partial x}(0,0)$ en $\frac{\partial p}{\partial y}(0,0) = \frac{\partial f}{\partial y}(0,0)$ dat

$$a = f(0,0), \quad b_1 = \frac{\partial f}{\partial x}(0,0), \quad b_2 = \frac{\partial f}{\partial y}(0,0).$$

Voor de tweede partiële van $p(x,y)$ krijgen we:

$$\frac{\partial^2 p}{\partial x^2} = 2c_1, \quad \frac{\partial^2 p}{\partial x\,\partial y} = \frac{\partial^2 p}{\partial y\,\partial x} = c_2, \quad \frac{\partial^2 p}{\partial y^2} = 2c_3$$

en gelijk zetten met de tweede partiële afgeleiden van $f(x,y)$ in het punt $(0,0)$ geeft:

$$c_1 = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0), \quad c_2 = \frac{\partial^2 f}{\partial x\,\partial y}(0,0), \quad c_3 = \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(0,0).$$

Voor de *kwadratische benadering* van $f(x,y)$ in het punt $(x,y) = (0,0)$ krijgen we dus de kwadratische veelterm

$$\begin{aligned} p(x,y) &= f(0,0) + \frac{\partial f}{\partial x}(0,0)\,x + \frac{\partial f}{\partial y}(0,0)\,y \\ &\quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0)\,x^2 + \frac{\partial^2 f}{\partial x\,\partial y}(0,0)\,xy + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(0,0)\,y^2 \\ &= f(0,0) + \frac{\partial f}{\partial x}(0,0)\,x + \frac{\partial f}{\partial y}(0,0)\,y \\ &\quad + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x^2}(0,0)\,x^2 + \frac{\partial^2 f}{\partial x\,\partial y}(0,0)\,xy + \frac{\partial^2 f}{\partial y\,\partial x}(0,0)\,yx + \frac{\partial^2 f}{\partial y^2}(0,0)\,y^2\right). \end{aligned}$$

> Het opsplitsen van de term $\frac{\partial^2 f}{\partial x\,\partial y}(0,0)\,xy$ in de som $\frac{1}{2}(\frac{\partial^2 f}{\partial x\,\partial y}(0,0)\,xy + \frac{\partial^2 f}{\partial y\,\partial x}(0,0)\,yx)$ lijkt niet echt een vereenvoudiging, maar we zullen nu zien dat we hier wel iets aan hebben.

Net zo als we de eerste partiële afgeleiden in een *vector*, die we de gradiënt noemen, samengevat hebben, kunnen we de tweede partiële afgeleiden in een *matrix* samenvatten. Voor een functie $f(x,y)$ van twee variabelen geeft dit de 2×2-matrix

$$H := H_f := \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y\,\partial x} \\ \frac{\partial^2 f}{\partial x\,\partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$$

Deze matrix heet de *Hesse matrix* van f. Voor een algemene functie $f(\mathbf{x})$ van n variabelen is de Hesse matrix een $n \times n$-matrix met in de i-de rij de partiële afgeleiden naar de i-de variabel x_i en in de j-de kolom de partiële afgeleiden naar x_j, dus op positie (i,j) staat de tweede partiële afgeleide naar x_i en x_j.

Definitie: Voor een functie $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ heet de matrix

$$H := H_f := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n\,\partial x_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1\,\partial x_n} & \frac{\partial^2 f}{\partial x_2\,\partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

met $H_{ij} = \frac{\partial^2 f}{\partial x_j \, \partial x_i}$ de *Hesse matrix* van $f(\mathbf{x})$.

Omdat volgens de Stelling van Schwarz geldt dat $\frac{\partial^2 f}{\partial x_j \, \partial x_i} = \frac{\partial^2 f}{\partial x_i \, \partial x_j}$, is de Hesse matrix een symmetrische matrix, d.w.z. er geldt $H = H^{tr}$.

We noteren met $H(\mathbf{x}_0)$ (of $H(x, y)$ in het geval van twee variabelen) de Hesse matrix waarvoor de partiële afgeleiden in het punt $\mathbf{x}_0$ (of het punt (x, y)) geëvalueerd zijn.

Voor een vector $\mathbf{h} = \begin{pmatrix} x \\ y \end{pmatrix}$ geldt dan:

$$\begin{aligned} \mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h} &= \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_0)\, x^2 + \frac{\partial^2 f}{\partial y \, \partial x}(\mathbf{x}_0)\, xy + \frac{\partial^2 f}{\partial x \, \partial y}(\mathbf{x}_0)\, yx + \frac{\partial^2 f}{\partial y^2}(\mathbf{x}_0)\, y^2 \\ &= \frac{\partial^2 f}{\partial x^2}(\mathbf{x}_0)\, x^2 + 2\frac{\partial^2 f}{\partial y \, \partial x}(\mathbf{x}_0)\, xy + \frac{\partial^2 f}{\partial y^2}(\mathbf{x}_0)\, y^2. \end{aligned}$$

Met behulp van de Hesse matrix kunnen we Taylor veelterm van graad 2 daarom als volgt herschrijven:

$$p(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h} + \frac{1}{2}\, \mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h}.$$

Ook hier hebben we met deze notatie de algemene vorm van de Taylor veelterm van graad 2 gevonden, er geldt:

Stelling: Voor een functie $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ van n variabelen is de Taylor veelterm van graad 2 in het punt $\mathbf{x}_0$ gegeven door

$$p(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h} + \frac{1}{2}\, \mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h}.$$

waarbij h_i de i-de component van de vector $\mathbf{h}$ is.

Voorbeeld 1: We bekijken de functie

$$f(x, y) := e^x \cos(y)$$

in het punt $(x_0, y_0) = (0, 0)$. Er geldt

$$\frac{\partial f}{\partial x} = e^x \cos(y), \quad \frac{\partial f}{\partial y} = -e^x \sin(y),$$

en dus geldt voor de gradiënt

$$\nabla f = \begin{pmatrix} e^x \cos(y) \\ -e^x \sin(y) \end{pmatrix} \quad \text{en} \quad \nabla f(0, 0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Voor de tweede partiële afgeleiden geldt

$$\frac{\partial^2 f}{\partial x^2} = e^x \cos(y), \quad \frac{\partial^2 f}{\partial y \, \partial x} = -e^x \sin(y), \quad \frac{\partial^2 f}{\partial y^2} = -e^x \cos(y),$$

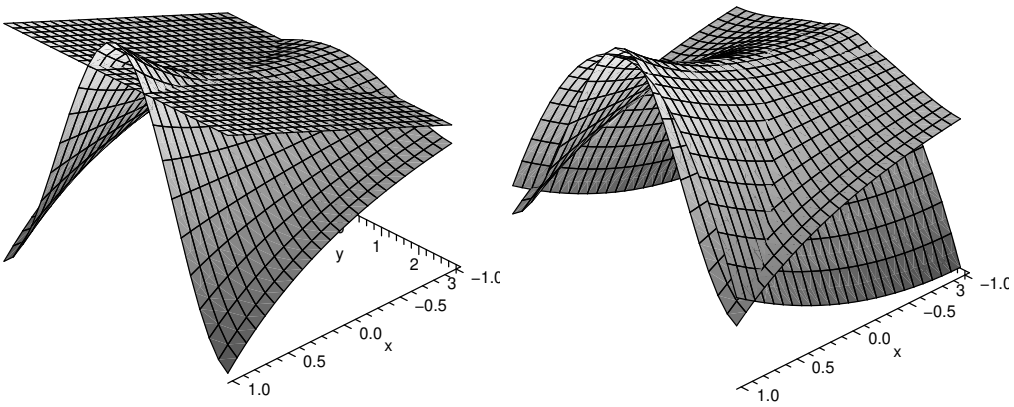
en dus krijgen we de Hesse matrix

$$H = \begin{pmatrix} e^x \cos(y) & -e^x \sin(y) \\ -e^x \sin(y) & -e^x \cos(y) \end{pmatrix} \quad \text{met} \quad H(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Omdat $f(0,0) = 1$, zijn de Taylor veeltermen van graad 1 en 2 van $f(x,y)$ gegeven door

$$p_1(x,y) = 1 + x \quad \text{en} \quad p_2(x,y) = 1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2.$$

In Figuur I.9 zijn de grafiek van de functie en de grafieken van de benadering door de Taylor veeltermen van graad 1 en 2 te zien. Het is duidelijk dat de lineaire benadering het raakvlak aan de grafiek geeft en men ziet goed dat de kwadratische benadering in een omgeving van $(0,0)$ al redelijk goed is.



Figuur I.9: Benadering van $e^x \cos(y)$ door Taylor veeltermen van graad 1 en 2.

Voorbeeld 2: We bekijken de functie

$$f(x,y) := \sin(xy)$$

in het punt $(x_0, y_0) = (1, \frac{\pi}{2})$. Er geldt

$$\frac{\partial f}{\partial x} = y\,\cos(xy), \quad \frac{\partial f}{\partial y} = x\,\cos(xy),$$

en dus geldt voor de gradiënt

$$\nabla f = \begin{pmatrix} y\,\cos(xy) \\ x\,\cos(xy) \end{pmatrix} \quad \text{en} \quad \nabla f(1, \frac{\pi}{2}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

De lineaire benadering van $f(x,y)$ is dus de constante $f(0,0) = 1$.

Voor de tweede partiële afgeleiden geldt

$$\frac{\partial^2 f}{\partial x^2} = -y^2 \sin(xy), \ \frac{\partial^2 f}{\partial y\,\partial x} = \cos(xy) - xy\,\sin(xy), \ \frac{\partial^2 f}{\partial y^2} = -x^2 \sin(xy),$$

en dus krijgen we de Hesse matrix

$$H = \begin{pmatrix} -y^2 \sin(xy) & \cos(xy) - xy\,\sin(xy) \\ \cos(xy) - xy\,\sin(xy) & -x^2 \sin(xy) \end{pmatrix}$$

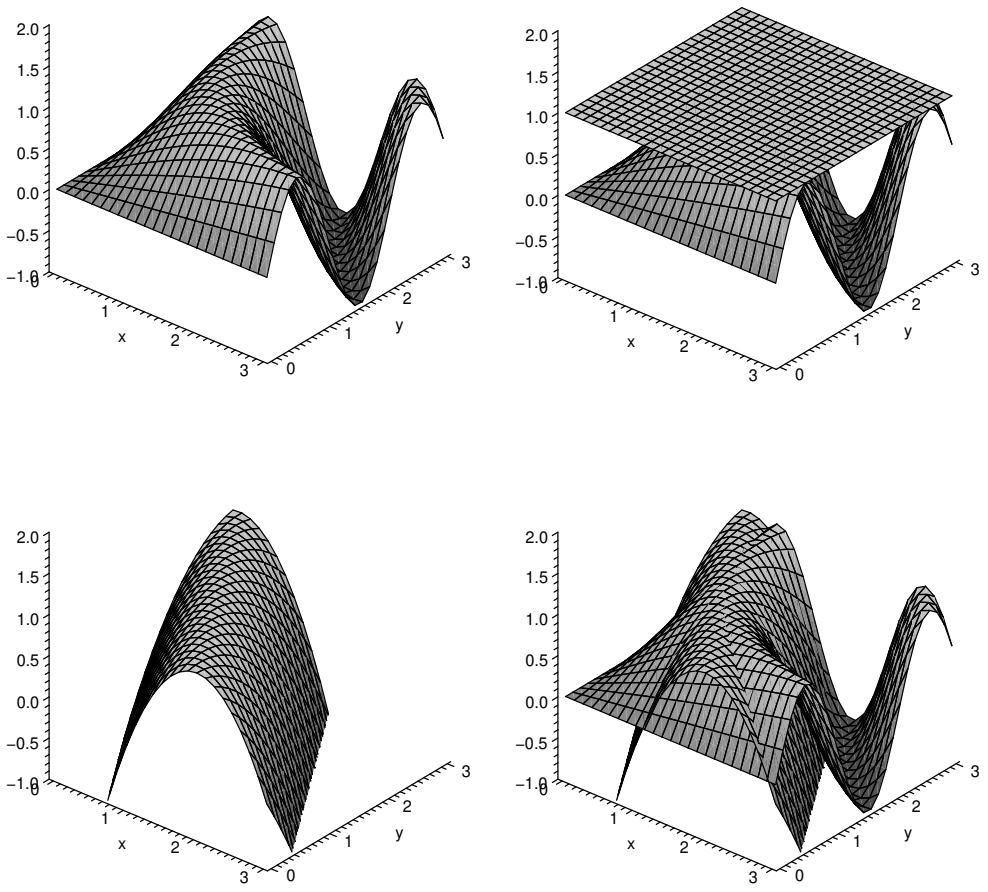
en dus

$$H(1,\frac{\pi}{2}) = \begin{pmatrix} -\frac{\pi^2}{4} & -\frac{\pi}{2} \\ -\frac{\pi}{2} & -1 \end{pmatrix}.$$

De Taylor veeltermen van graad 2 van $f(x,y)$ is dus gegeven door

$$\begin{aligned} p(x,y) &= 1 + \frac{1}{2}\left(-\frac{\pi^2}{4}(x-1)^2 - 2\frac{\pi}{2}(x-1)(y-\frac{\pi}{2}) - (y-\frac{\pi}{2})^2\right) \\ &= 1 - \frac{\pi^2}{8}(x-1)^2 - \frac{\pi}{2}(x-1)(y-\frac{\pi}{2}) - \frac{1}{2}(y-\frac{\pi}{2})^2). \end{aligned}$$

In Figuur I.10 zijn boven de grafiek van de functie en de (constante) lineaire benadering te zien, beneden de grafiek van de kwadratische benadering apart en de kwadratische benadering samen met de grafiek van de functie.



Figuur I.10: Benadering van $\sin(xy)$ door Taylor veeltermen van graad 1 (boven) en 2 (beneden) in het punt $(x_0, y_0) = (1, \frac{\pi}{2})$.

OPDRACHT 11 *Bepaal voor* $f(x,y) := e^{x^2+y^2}$ *de Taylor veelterm van graad* 2 *in het punt* $(0,0)$.

Algemene vorm van de Taylor reeks

Als we tot Taylor veeltermen van hogere graden dan 2 en misschien zelfs tot Taylor reeksen voor functies van meerdere variabelen willen komen, wordt het vergelijken van de partiële afgeleiden van een veelterm met algemene coëfficiënten met de partiële afgeleiden van $f(\mathbf{x})$ snel erg onhandig. Maar gelukkig kunnen we de Taylor veeltermen en Taylor reeks voor functies van meerdere variabelen met een klein trucje ook afleiden uit de Taylor reeks voor een functie van één variabel.

Als we een functie $f(\mathbf{x})$ in een kleine omgeving van een vast gekozen punt $\mathbf{x}_0$ bekijken, kunnen we dit schrijven als $f(\mathbf{x}_0 + \mathbf{h})$ voor een (korte) vector $\mathbf{h}$. Hieruit maken we nu (kunstmatig) een nieuwe functie $g(t)$ van een enkele variabel t, namelijk

$$g(t) := f(\mathbf{x}_0 + t \cdot \mathbf{h}).$$

We krijgen de Taylor reeks van $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ door $g(t)$ rond $t = 0$ in een Taylor reeks te ontwikkelen en vinden $f(\mathbf{x}_0+\mathbf{h})$ door $t = 1$ in deze Taylor reeks in te vullen, want $f(\mathbf{x}_0 + \mathbf{h}) = g(1)$. Op deze manier berekenen we $f(\mathbf{x}_0 + \mathbf{h})$ door:

$$\begin{aligned} f(\mathbf{x}_0 + h) = g(1) &= g(0) + g'(0) \cdot 1 + \frac{1}{2} g''(0) \cdot 1^2 + \ldots + \frac{1}{n!} g^{(n)}(0) \cdot 1^n + \ldots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} g^{(n)}(0). \end{aligned}$$

Voor de Taylor reeks van $f(\mathbf{x})$ moeten we dus alleen maar de afgeleiden van $g(t)$ in het punt $t = 0$ bepalen. De cruciale waarneming is nu als volgt:

Merk op: De gewone afgeleide $g'(t)$ komt precies overeen met de definitie van de richtingsafgeleide van $f(\mathbf{x})$ in de richting van $\mathbf{h}$.

Voor een vector $\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ met componenten h_i geldt daarom

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i \quad \text{en dus} \quad g'(0) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_0)\, h_i.$$

Voor de tweede afgeleide $g''(t)$ moeten we weer de richtingsafgeleide in de richting van $\mathbf{h}$ nemen, dit geeft

$$g''(t) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j\, \partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j \quad \text{en} \quad g''(0) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_j\, \partial x_i}(\mathbf{x}_0)\, h_i h_j.$$

Op een soortgelijke manier krijgen we voor de derde afgeleide

$$g'''(t) = \sum_{k=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^3 f}{\partial x_k\, \partial x_j\, \partial x_i}(\mathbf{x}_0 + t\mathbf{h})\, h_i h_j h_k,$$

$$g'''(0) = \sum_{k=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^3 f}{\partial x_k\, \partial x_j\, \partial x_i}(\mathbf{x}_0)\, h_i h_j h_k$$

en we kunnen in principe zo doorgaan tot hogere afgeleiden, de n-de term is dan een n-voudige som van de n-de partiële afgeleiden van $f(\mathbf{x})$. Maar ook deze sommen zijn nog niet echt prettig om op te schrijven, daarom zullen we nu nog naar een veel gebruikte notatie kijken, waarmee de Taylor reeks voor meerdere variabelen bijna net zo eenvoudig wordt als de Taylor reeks voor functies van één variabel.

Het idee is, de partiële afgeleide als een soort bewerkingsvoorschrift te beschouwen, die we op een functie toepassen en die we een *operator* noemen. In principe is dit niets nieuws, want ook de gewone afgeleide kunnen we zien als een operator $'$ die uit een functie $f(x)$ een andere functie $f'(x)$ maakt. Net zo interpreteren we nu de partiële afgeleide als een operator $\frac{\partial}{\partial x}$ die uit de functie $f(\mathbf{x})$ de nieuwe functie $\frac{\partial}{\partial x}f = \frac{\partial f}{\partial x}$ maakt.

Met deze operatoren kunnen we nu op een voor de hand liggende manier rekenen, voor de som $\frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ van twee operatoren moeten we hiervoor aangeven, welke nieuwe functie de toepassing van $\frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ op een functie $f(\mathbf{x})$ geeft. Als verdere bewerkingen definiëren we ook de vermenigvuldiging van twee operatoren en het vermenigvuldigen van een operator met een constante:

$$\begin{aligned}(c \cdot \frac{\partial}{\partial x_i})f &:= c \cdot \frac{\partial f}{\partial x_i} \\ (\frac{\partial}{\partial x_i} + \frac{\partial}{\partial x_j})f &:= \frac{\partial f}{\partial x_i} + \frac{\partial f}{\partial x_j} \\ (\frac{\partial}{\partial x_i} \cdot \frac{\partial}{\partial x_j})f &:= \frac{\partial^2 f}{\partial x_i \, \partial x_j}.\end{aligned}$$

Hiermee krijgen we bijvoorbeeld

$$(\frac{\partial}{\partial x} + \frac{\partial}{\partial y})^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y \, \partial x} + \frac{\partial^2}{\partial x \, \partial y} + \frac{\partial^2}{\partial y^2} = \frac{\partial^2}{\partial x^2} + 2\frac{\partial^2}{\partial y \, \partial x} + \frac{\partial^2}{\partial y^2}.$$

Merk op: Dit is in principe dezelfde formule als de binomische formule

$$(a+b)^2 = a^2 + 2ab + b^2$$

en dit betekent dat we met de partiële afgeleiden als operatoren juist zo rekenen als met gewone variabelen.

Met behulp van de operatoren kunnen we het inproduct van een vector $\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ met de gradiënt ∇f herschrijven als

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{h} = (h_1 \cdot \frac{\partial}{\partial x_1} + \ldots + h_n \cdot \frac{\partial}{\partial x_n})f(\mathbf{x}_0)$$

waarbij het invullen van $\mathbf{x}_0$ aan de rechterkant natuurlijk *na* het toepassen van de operatoren gebeurt.

Net zo krijgen we voor het product met de Hesse matrix

$$\mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h} = (h_1 \cdot \frac{\partial}{\partial x_1} + \ldots + h_n \cdot \frac{\partial}{\partial x_n})^2 f(\mathbf{x}_0).$$

De Taylor reeks voor een functie van n variabelen laat zich nu schrijven als:

$$f(\mathbf{x}_0 + \mathbf{h}) = \sum_{n=0}^{\infty} \frac{1}{n!}(h_1 \cdot \frac{\partial}{\partial x_1} + \ldots + h_n \cdot \frac{\partial}{\partial x_n})^n f(\mathbf{x}_0).$$

Voorbeeld: We berekenen de derdegraads term van de Taylor reeks voor een functie $f(x, y)$ van twee variabelen. Er geldt $(a+b)^3 = a^3+3a^2b+3ab^2+b^3$, hieruit volgt

$$(h_1\frac{\partial}{\partial x} + h_2\frac{\partial}{\partial y})^3 = h_1^3\frac{\partial^3}{\partial x^3} + 3h_1^2h_2\frac{\partial^3}{\partial y\,\partial x^2} + 3h_1h_2^2\frac{\partial^3}{\partial y^2\,\partial x} + h_2^3\frac{\partial^3}{\partial y^3}$$

dus is de derdegraads term van de Taylor reeks

$$\frac{1}{3!}\left(\frac{\partial^3 f}{\partial x^3}(\mathbf{x}_0)\,h_1^3 + 3\frac{\partial^3 f}{\partial y\,\partial x^2}(\mathbf{x}_0)\,h_1^2h_2 + 3\frac{\partial^3 f}{\partial y^2\,\partial x}(\mathbf{x}_0)\,h_1h_2^2 + \frac{\partial^3 f}{\partial y^3}(\mathbf{x}_0)\,h_2^3\right).$$

Belangrijke begrippen in deze les

- Lagrange interpolatie
- Taylor veelterm
- foutterm
- oneindige reeksen
- Taylor reeks
- lineaire/kwadratische benadering
- Hesse matrix
- partiële afgeleide als operator

Opgaven

12. Bepaal voor $f(x) := (1 + x)^a$ met $a \in \mathbb{R}$ de Taylor veelterm van graad 3 in $x_0 = 0$.
13. Bepaal voor de volgende functies de Taylor veelterm van graad 5 in $x_0 = 0$:
    (i) $f(x) := \sqrt[3]{1 + x} = (1 + x)^{\frac{1}{3}}$;
    (ii) $f(x) := \frac{1}{1+x^2}$;
    (iii) $f(x) := \frac{1}{\sqrt{1-x^2}}$;

(iv) $f(x) := \frac{1}{3+x}$;
(v) $f(x) := \sin(2x^2)$;
(vi) $f(x) := e^{-3x}$.

14. Zij $f(x) := \sqrt[3]{8+x} = (8+x)^{\frac{1}{3}}$.
    (i) Bepaal voor $f(x)$ de Taylor veelterm van graad 4 in $x_0 = 0$.
    (ii) Vind met behulp van de Taylor veelterm in (i) een benadering voor $\sqrt[3]{9}$.
    (iii) Geef met behulp van de Lagrange vorm van de foutterm een afschatting voor de mogelijke fout van de benadering in (ii).

15. Bepaal voor de volgende functies de Taylor reeksen in de aangeven punten:
    (i) $f(x) := \frac{1}{x}$ in $x_0 = 1$;
    (ii) $f(x) := e^x$ in $x_0 = 2$;
    (iii) $f(x) := \sin(x+1)$ in $x_0 = -1$;
    (iv) $f(x) := \log(x)$ in $x_0 = 2$.

16. Vind de Taylor reeks voor $f(x) := \arctan(x)$ in $x_0 = 0$. (Hint: Er geldt $\arctan'(x) = \frac{1}{1+x^2}$ en dit laat zich voor $x^2 < 1$ schrijven als meetkundige reeks $\frac{1}{1+x^2} = 1 - x^2 + x^4 - x^6 + \ldots = \sum_{n=0}^{\infty}(-1)^n x^{2n}$.)

17. Bepaal voor de functie $f(x,y) := \sqrt{x^2+y^3}$ de Taylor veelterm van graad 2 in het punt $(1,2)$. Gebruik de Taylor veelterm om de waarde van $\sqrt{1.02^2 + 1.97^3}$ te benaderen en vergelijk het resultaat met de *juiste* waarde (volgens een rekenmachine).

18. Bepaal voor de volgende functies de Taylor veelterm van graad 2 in het punt $(x_0, y_0) = (0,0)$:
    (i) $f(x,y) := \frac{1}{1+x^2+y^2}$;
    (ii) $f(x,y) := e^{x+y}$;
    (iii) $f(x,y) := e^{-x^2-y^2}\cos(xy)$;
    (iv) $f(x,y) := \sin(xy) + \cos(xy)$.

19. Bepaal voor de volgende functies de Taylor veelterm van graad 2 in de aangegeven punten (x_0, y_0):
    (i) $f(x,y) := \frac{1}{2+xy^2}$ in $(0,0)$;
    (ii) $f(x,y) := \log(1+x+y+xy)$ in $(0,0)$;
    (iii) $f(x,y) := \arctan(x+xy)$ in $(0,-1)$;
    (iv) $f(x,y) := x^2 + xy + y^3$ in $(1,-1)$;
    (v) $f(x,y) := \sin(2x+3y)$ in $(0,0)$;
    (vi) $f(x,y) := \frac{\sin(x)}{y}$ in $(\frac{\pi}{2}, 1)$;
    (vii) $f(x,y) := \frac{1+x}{1+x^2+y^4}$ in $(0,0)$;
    (viii) $f(x,y) := e^{(x-1)^2}\cos(y)$ in $(1,0)$.

20. Bepaal voor de volgende functies de Taylor veelterm van graad 3 in de aangegeven punten (x_0, y_0):
    (i) $f(x,y) := \frac{1}{2+x-2y}$ in $(2,1)$;
    (ii) $f(x,y) := \log(x^2+y^2)$ in $(1,0)$;
    (iii) $f(x,y) := \cos(x+\sin(y))$ in $(0,0)$.

Les 3 Extrema van functies van meerdere variabelen

Bij gewone functies van één variabel hebben we in Wiskunde 1 de vraag behandeld hoe we minima en maxima van een functie kunnen vinden. Het belangrijkste criterium was dat een differentieerbare functie in een lokaal extremum (minimum of maximum) een horizontale raaklijn heeft, de afgeleide van de functie in een extremum is dus noodzakelijk nul. Punten met deze eigenschap hebben we *kritieke punten* genoemd, naast de speciale punten waar de functie niet differentieerbaar is en de randpunten van het interval waarop we de functie bekijken.

3.1 Classificatie van kritieke punten

Het kan zijn dat een functie in een punt een horizontale raaklijn heeft, zonder in dit punt een extremum te hebben. Dit is bijvoorbeeld het geval voor de functie $f(x) = x^3$ in het punt $x = 0$. Zo'n punt noemt men ook een *zadelpunt*. Het verschil tussen een zadelpunt en een echt extremum laat zich aan de hand van de tweede afgeleide $f''(x)$ beschrijven:

Bij een minimum in het punt x_0 is de functie links van x_0 (dus voor $x < x_0$) dalend en rechts van x_0 stijgend, dus is de eerste afgeleide $f'(x)$ links van x_0 negatief en rechts van x_0 positief. Dit betekent dat $f'(x)$ rond x_0 stijgend is en dus geldt $f''(x_0) > 0$. Net zo volgt uit $f'(x_0) = 0$ en $f''(x_0) < 0$, dat de functie $f(x)$ in het punt x_0 een maximum heeft.

Met onze kennis van Taylor reeksen kunnen we dit nu ook van een andere kant bekijken. We ontwikkelen $f(x)$ rond een kritiek punt x_0 met $f'(x_0) = 0$ in een Taylor reeks, dit geeft:

$$f(x_0 + h) = f(x_0) + \frac{1}{2} f''(x_0)\, h^2 + \dots$$

In een kleine omgeving van x_0 (dus voor kleine waarden van h) kunnen we de hogere termen h^3, h^4, enz. tegenover h^2 verwaarlozen en we kunnen $f(x)$ door de kwadratische functie

$$g(h) := f(x_0) + \frac{1}{2} f''(x_0)\, h^2$$

benaderen. Omdat het bestaan van een lokaal extremum in x_0 alleen maar van een kleine omgeving van x_0 afhangt, heeft $f(x)$ in x_0 een maximum/minimum dan en slechts dan als $g(h)$ in 0 een maximum/minimum heeft. Maar de vergelijking van $g(h)$ beschrijft juist een parabool met toppunt $(0, f(x_0))$ en deze parabool is naar boven geopend als $f''(x_0) > 0$ en naar beneden geopend als $f''(x_0) < 0$. De volgende twee gevallen zijn dus duidelijk:

(1) Voor $f''(x_0) > 0$ heeft $g(h)$ in $h = 0$ een minimum, dus heeft ook $f(x)$ een minimum in x_0.

(2) Voor $f''(x_0) < 0$ heeft $g(h)$ in $h = 0$ een maximum, dus heeft ook $f(x)$ een maximum in x_0.

Als in een kritiek punt x_0 ook de tweede afgeleide $f''(x_0) = 0$ is, kunnen we nog steeds niet beslissen of de functie een minimum, maximum of een zadelpunt heeft. In dit geval moeten we de hogere afgeleiden bepalen tot dat we naar een n komen met $f^{(n)}(x_0) \neq 0$. Dan kunnen we weer de Taylor reeks van $f(x)$ in x_0 bepalen, deze is

$$f(x_0 + h) = f(x_0) + \frac{1}{n!} f^{(n)}(x_0)\, h^n + \dots$$

en in een kleine omgeving van x_0 kunnen we in plaats van $f(x)$ naar de functie

$$g(h) := f(x_0) + \frac{1}{n!} f^{(n)}(x_0)\, h^n$$

kijken. Nu zijn er drie mogelijke gevallen:

(1) Als n oneven is, heeft $f(x)$ in x_0 een zadelpunt, want $g(h)$ is (tot op een verschuiving en een schaling na) een functie van de vorm h^3, h^5, enz.

(2) Als n even is en $f^{(n)}(x_0) > 0$, heeft $f(x)$ in x_0 een minimum, want in dit geval is $g(h)$ naar boven geopend (net zo als de functies $2x^4$ of πx^6).

(3) Als n even is en $f^{(n)}(x_0) < 0$, heeft $f(x)$ in x_0 een maximum, want in dit geval is $g(h)$ naar beneden geopend (net zo als $-3x^4$ of $-\sqrt{2}x^6$).

We krijgen dus de volgende classificatie voor kritieke punten van differentieerbare functies:

Stelling: Zij $f(x)$ een in het punt x_0 differentieerbare functie met $f'(x_0) = 0$ en zij $n \geq 2$ de kleinste n met $f^{(n)}(x_0) \neq 0$. Dan geldt:

(i) $f(x)$ heeft in x_0 een minimum als n even is en $f^{(n)}(x_0) > 0$;

(ii) $f(x)$ heeft in x_0 een maximum als n even is en $f^{(n)}(x_0) < 0$;

(iii) $f(x)$ heeft in x_0 een zadelpunt als n oneven is.

3.2 Kritieke punten van functies van meerdere variabelen

We kijken nu naar de vraag hoe we lokale extrema van functies van meerdere veranderlijken kunnen vinden. We zullen zien dat de ideeën hiervoor in principe hetzelfde blijven als bij de gewone functies, het komt erop neer dat we de gewone afgeleide door de partiële afgeleiden moeten vervangen.

We starten weer met de beschrijving van de consequenties van een extremum. Als een functie $f(\mathbf{x})$ in een punt $\mathbf{x}_0$ een lokaal extremum heeft, kunnen we naar de partiële afgeleiden in dit punt kijken. Maar bij de partiële afgeleide $\frac{\partial f}{\partial x_i}$ bekijken we de verandering van een functie $g(x_i)$ die we uit $f(\mathbf{x})$ krijgen, door de andere variabelen $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ als constanten te beschouwen. Hieruit volgt dat de functie $f(\mathbf{x})$ alleen maar een extremum kan hebben, als de functie $g(x_i)$ van één variabel een extremum heeft, en dit betekent dat $\frac{\partial f}{\partial x_i}(\mathbf{x}_0) = 0$ moet zijn. Omdat dit argument voor iedere variabel x_i geldt, krijgen we als noodzakelijke voorwaarde voor een extremum in $\mathbf{x}_0$, dat de gradiënt in $\mathbf{x}_0$ nul moet zijn, dus:

Stelling: Als een functie $f(\mathbf{x})$ in een punt $\mathbf{x}_0$ een lokaal minimum of maximum heeft, dan geldt

$$\nabla f(\mathbf{x}_0) = 0.$$

Deze stelling kunnen we ook uit de interpretatie van de gradiënt afleiden, want de gradiënt $\nabla f(\mathbf{x}_0)$ wijst in de richting van de snelste toename van $f(\mathbf{x})$. Maar in een maximum mag de functie in geen enkele richting toenemen, dus moet de gradiënt nul zijn. Net zo wijst $-\nabla f(\mathbf{x}_0)$ in de richting van de snelste afname van de functie, en in een minimum mag de functie in geen richting afnemen, dus moet ook hier de gradiënt nul zijn.

We kunnen ook vanuit het perspectief van de Taylor reeks argumenteren. De lineaire benadering

$$f(\mathbf{x}_0 + \mathbf{h}) \approx T(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h}$$

geeft het raakvlak aan de grafiek van $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ aan. Maar in een lokaal extremum moet het raakvlak horizontaal zijn, dus moet de lineaire benadering in een minimum of maximum een constante zijn, en dit betekent ook weer dat $\nabla f(\mathbf{x}_0) = 0$ moet zijn.

Net zo als bij gewone functies noemen we de punten $\mathbf{x}_0$ die aan de noodzakelijke voorwaarde $\nabla f(\mathbf{x}_0) = 0$ voldoen, de *kritieke punten* van $f(\mathbf{x})$.

Definitie: De punten $\mathbf{x}_0$ waar voor een differentieerbare functie $f(\mathbf{x})$ de gradiënt $\nabla f(\mathbf{x}_0) = 0$ is, heten *kritieke punten* van $f(\mathbf{x})$. De kritieke punten zijn juist de kandidaten voor lokale minima of maxima van $f(\mathbf{x})$.

Voorbeelden:

(1) Een open doos met rechthoekig grondvlak moet een bepaald volume V bevatten. Wat zijn de optimale afmetingen van de doos zo dat we zo weinig materiaal als mogelijk nodig hebben?

Als de zijden van het grondvlak afmetingen x en y hebben, moet de hoogte z van de doos $z = \frac{V}{xy}$ zijn. De oppervlakte van de doos is dus een functie $A(x, y)$ van de afmetingen van het grondvlak en er geldt

$$A(x, y) = xy + 2xz + 2yz = xy + 2x\frac{V}{xy} + 2y\frac{V}{xy} = xy + 2\frac{V}{y} + 2\frac{V}{x}.$$

Voor de partiële afgeleiden geldt

$$\frac{\partial A}{\partial x} = y - \frac{2V}{x^2} \quad \text{en} \quad \frac{\partial A}{\partial y} = x - \frac{2V}{y^2}.$$

Uit $\frac{\partial A}{\partial x} = 0$ volgt $y = \frac{2V}{x^2}$. Dit ingevuld in $\frac{\partial A}{\partial y} = 0$ geeft $x = \frac{2V}{y^2} = \frac{x^4}{2V}$. Hieruit volgt $x = 0$ of $x^3 = 2V$, waarbij de eerste oplossing wegvalt, omdat de doos niet lengte 0 kan hebben.

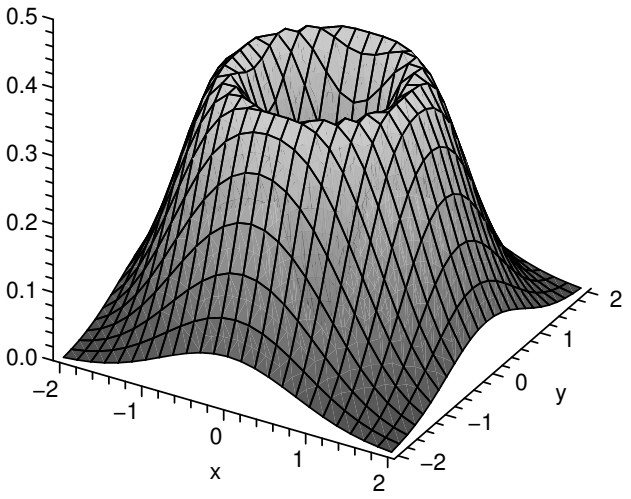
Uit $y = \frac{2V}{x^2}$ volgt nu $y^3 = \frac{8V^3}{x^6} = \frac{8V^3}{4V^2} = 2V$, dus is $x = y = \sqrt[3]{2V}$ en we krijgen het niet erg verrassende resultaat dat het grondvlak een vierkant is.

Voor de hoogte z van de doos geldt dat $z = \frac{V}{xy} = \sqrt[3]{\frac{V^3}{4V^2}} = \sqrt[3]{\frac{V}{4}} = \frac{1}{2}\sqrt[3]{2V} = \frac{1}{2}x$, dus is de doos half zo hoog als lang en breed.

(2) We bepalen de kritieke punten van de functie

$$f(x,y) := (x^2+y^2)\,e^{-x^2-y^2}$$

waarvan de grafiek (in de vorm van een vulkaan) in Figuur I.11 te zien is.



Figuur I.11: Grafiek van de functie $f(x,y) = (x^2+y^2)e^{-x^2-y^2}$.

Voor de partiële afgeleiden geldt

$$\frac{\partial f}{\partial x} = 2x\,e^{-x^2-y^2} - 2x(x^2+y^2)\,e^{-x^2-y^2} = 2x\,e^{-x^2-y^2}(1-x^2-y^2)$$
$$\frac{\partial f}{\partial y} = 2y\,e^{-x^2-y^2} - 2y(x^2+y^2)\,e^{-x^2-y^2} = 2y\,e^{-x^2-y^2}(1-x^2-y^2)$$

en hieruit volgt dat $\nabla f(x,y) = 0$ voor $(x,y) = (0,0)$ en voor (x,y) met $x^2+y^2 = 1$, dus voor punten op een cirkel met straal 1 rond $(0,0)$. Het eerste geval geeft het minimum in het centrum van de vulkaan, het tweede geval geeft de lokale maxima op de rand van de vulkaan.

Merk op dat we de kritieke punten alleen maar met behulp van de grafiek van de functie als minima of maxima hebben geïdentificeerd. Hoe we dit zonder grafiek kunnen herkennen, gaan we straks behandelen.

(3) We bekijken de functie $f(x,y) := x^2y + xy^2$. Er geldt

$$\frac{\partial f}{\partial x} = 2xy + y^2 = y(2x+y) \quad \text{en} \quad \frac{\partial f}{\partial y} = x^2 + 2xy = x(x+2y).$$

Uit $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$ volgt $x = y = 0$, want voor $x \neq 0$ volgt $y = -2x$ en $y = -\frac{1}{2}x$ en dit is onmogelijk. Dus is $(x,y) = (0,0)$ het enige kritieke

punt. Maar op de lijn $x = y$ is de functie $f(x, y)$ gelijk aan $2x^3$, en is dus negatief voor $x < 0$ en positief voor $x > 0$, dus heeft de functie in $(0, 0)$ geen maximum of minimum.

OPDRACHT 12 *Vind de kritieke punten voor* $f(x, y) := x^3 + y^3 - 3x - 12y + 20$.

3.3 Criterium voor lokale extrema

Het voorbeeld (3) van de functie $f(x, y) = x^2y + xy^2$ laat zien dat (net als bij gewone functies van één variabel) een functie van meerdere veranderlijken in een kritiek punt niet noodzakelijk een maximum of minimum hoeft te hebben.

Definitie: Een kritiek punt $\mathbf{x}_0$ met $\nabla f(\mathbf{x}_0) = 0$ die geen extremum van de functie $f(\mathbf{x})$ is noemt men een *zadelpunt* van $f(\mathbf{x})$. In een zadelpunt vindt men in iedere (willekeurig kleine) omgeving van $\mathbf{x}_0$ punten $\mathbf{x}$ met $f(\mathbf{x}) > f(\mathbf{x}_0)$ en punten met $f(\mathbf{x}) < f(\mathbf{x}_0)$.

De vraag is nu, hoe we erover kunnen beslissen of een kritiek punt een minimum, maximum of een zadelpunt is. Hiervoor zullen we analoog met het geval van gewone functies de Taylor reeks in het kritieke punt $\mathbf{x}_0$ gebruiken, beter gezegd bekijken we de kwadratische benadering van $f(\mathbf{x})$ door de Taylor veelterm van graad 2.

We veronderstellen vanaf nu dat $\mathbf{x}_0$ een kritiek punt van de functie $f(\mathbf{x})$ is, dus dat $\nabla f(\mathbf{x}_0) = 0$ en we noteren met $H(\mathbf{x}_0)$ de Hesse matrix geëvalueerd in het punt $\mathbf{x}_0$. Dan is de kwadratische benadering van $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ gegeven door

$$f(\mathbf{x}_0 + \mathbf{h}) \approx T(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \frac{1}{2}\,\mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h}.$$

Als we de functie $f(\mathbf{x})$ alleen maar in een kleine omgeving van $\mathbf{x}_0$ bekijken, kunnen we de hogere termen van de Taylor reeks verwaarlozen, het gedrag van de functie wordt dan door de kwadratische benadering weergegeven. We krijgen nu rechtstreeks het volgende criterium voor minima en maxima:

Criterium:

(1) De functie $T(\mathbf{x})$ en dus ook de functie $f(\mathbf{x})$ heeft een *minimum* in $\mathbf{x}_0$ als $T(\mathbf{x})$ vanuit $\mathbf{x}_0$ in alle richtingen $\mathbf{h}$ toeneemt, d.w.z. als $\mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h} > 0$ voor alle richtingen $\mathbf{h} \neq 0$.

(2) De functie $T(\mathbf{x})$ en dus ook de functie $f(\mathbf{x})$ heeft een *maximum* in $\mathbf{x}_0$ als $T(\mathbf{x})$ vanuit $\mathbf{x}_0$ in alle richtingen $\mathbf{h}$ afneemt, d.w.z. als $\mathbf{h}^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h} < 0$ voor alle richtingen $\mathbf{h} \neq 0$.

(3) Als er een richting $\mathbf{h}_1$ bestaat met $\mathbf{h}_1^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h}_1 > 0$ en een andere richting $\mathbf{h}_2$ met $\mathbf{h}_2^{tr} \cdot H(\mathbf{x}_0) \cdot \mathbf{h}_2 < 0$, dan heeft $T(\mathbf{x})$ en dus ook $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ een zadelpunt.

Om dit criterium toe te passen, moeten we dus voor de symmetrische matrix $H := H(\mathbf{x}_0)$ beslissen of de producten $\mathbf{h}^{tr} \cdot H \cdot \mathbf{h}$ altijd positief, altijd negatief of geen van de twee zijn. Dit is eigenlijk een vraagstelling uit de Lineaire Algebra, die in het verband met inproducten ter sprake komt.

Positief definiete matrices

Definitie: Zij A een symmetrische $n \times n$-matrix, d.w.z. $A_{ij} = A_{ji}$ voor alle i, j (kort: $A^{tr} = A$).

(i) A heet *positief definiet* als $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} > 0$ voor alle $\mathbf{v} \neq 0$.

(ii) A heet *positief semidefiniet* als $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} \geq 0$ voor alle $\mathbf{v}$.

(iii) A heet *negatief definiet* als $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} < 0$ voor alle $\mathbf{v} \neq 0$.

(iv) A heet *negatief semidefiniet* als $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} \leq 0$ voor alle $\mathbf{v}$.

(v) Als er vectoren $\mathbf{v}_1$ en $\mathbf{v}_2$ bestaan met $\mathbf{v}_1^{tr} \cdot A \cdot \mathbf{v}_1 > 0$ en $\mathbf{v}_2^{tr} \cdot A \cdot \mathbf{v}_2 < 0$, heet A *indefiniet*.

> Het idee achter deze definitie is, dat men algemeen met behulp van een symmetrische matrix A een bilineaire afbeelding op paren van vectoren kan definiëren door $(v, w) \mapsto v^{tr} \cdot A \cdot w$. Deze afbeelding is lineair in beide argumenten en is symmetrisch, d.w.z. verruilen van de argumenten verandert de waarde niet. Als A positief definiet is, laat zich met $\|v\| := \sqrt{v^{tr} \cdot A \cdot v}$ een lengte voor de vectoren definiëren.

Merk op: Uit de definite en uit $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} > 0 \Leftrightarrow \mathbf{v}^{tr} \cdot (-A) \cdot \mathbf{v} < 0$ volgt rechtstreeks, dat een matrix A positief definiet is dan en slechts dan als de tegengestelde matrix $-A$ negatief definiet is. Evenzo volgt dat A negatief definiet is dan en slechts dan als $-A$ positief definiet is.

Voorbeelden:

(1) De matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is positief definiet, want voor $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ geldt $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} = x^2 + y^2 > 0$ voor $\mathbf{v} \neq 0$.

(2) Voor de matrix $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ en $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ geldt $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} = ax^2 + by^2$. Voor $a, b > 0$ is A positief definiet, voor $a, b < 0$ is A negatief definiet. In het geval $a \cdot b < 0$ is A indefiniet, met $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ en $\mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ volgt bijvoorbeeld voor $a > 0$ en $b < 0$ dat $\mathbf{v}_1^{tr} \cdot A \cdot \mathbf{v}_1 = a > 0$ en $\mathbf{v}_2^{tr} \cdot A \cdot \mathbf{v}_2 = b < 0$.

(3) De matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is positief semidefiniet, want voor $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ geldt $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} = x^2 + 2xy + y^2 = (x + y)^2 \geq 0$. De matrix is niet positief definiet, want voor $y = -x$ is $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} = 0$ zonder dat $\mathbf{v} = 0$ is.

(4) De matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is indefiniet, want voor $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$ geldt $\mathbf{v}^{tr} \cdot A \cdot \mathbf{v} = 2xy$, voor $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is dus $\mathbf{v}_1^{tr} \cdot A \cdot \mathbf{v}_1 = 2 > 0$ en voor $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is $\mathbf{v}_2^{tr} \cdot A \cdot \mathbf{v}_2 = -2 < 0$.

Algemeen voorbeeld: We bekijken de $n \times n$-diagonaalmatrix

$$A := \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \text{ met } A_{ii} = d_i \text{ en } A_{ij} = 0 \text{ voor } i \neq j.$$

Er geldt:

(a) A is positief definiet als $d_i > 0$ voor alle i;

(b) A is positief semidefiniet als $d_i \geq 0$ voor alle i;

(c) A is negatief definiet als $d_i < 0$ voor alle i;

(d) A is negatief semidefiniet als $d_i \leq 0$ voor alle i;

(e) A is indefiniet als er i en j bestaan met $d_i > 0$ en $d_j < 0$.

In het bijzonder geldt voor een diagonaalmatrix A die positief (of negatief) semidefiniet maar niet positief (of negatief) definiet is, dat $\det(A) = 0$, omdat in dit geval minstens een $d_i = 0$ is en $\det(A) = d_1 \cdot d_2 \cdot \ldots \cdot d_n$ geldt.

Het algemene voorbeeld geeft het cruciale idee, hoe we kunnen testen of een matrix positief of negatief definiet is. De volgende stelling geeft hiervoor een criterium aan. Hierbij bedoelen met de *linksboven $k \times k$-deelmatrix* van een matrix A de deelmatrix van A waarvoor de indices slechts tussen 1 en k lopen (in plaats van tussen 1 en n):

$$\left(\begin{array}{ccc|cc} A_{11} & \ldots & A_{1k} & \ldots & A_{1n} \\ \vdots & & \vdots & & \vdots \\ A_{k1} & \ldots & A_{kk} & \ldots & A_{kn} \\ \hline \vdots & & \vdots & & \vdots \\ A_{n1} & \ldots & A_{nk} & \ldots & A_{nn} \end{array}\right).$$

Stelling: Zij A een symmetrische $n \times n$-matrix, dan geldt:

(i) A is positief definiet dan en slechts dan als alle linksboven $k \times k$-deelmatrices van A positieve determinant hebben.

(ii) A is negatief definiet dan en slechts dan als de linksboven $k \times k$-deelmatrices van A alternerend negatieve en positieve determinant hebben, dus als de 1×1-deelmatrix negatieve determinant heeft, de 2×2-deelmatrix positieve determinant, de 3×3-deelmatrix negatieve determinant enz.

Equivalent (en eenvoudiger) geldt: A is negatief definiet dan en slechts dan als de matrix $-A$ positief definiet is, dus als alle linksboven $k \times k$-deelmatrices van $-A$ positieve determinant hebben.

(iii) Als $\det(A) \neq 0$ is, is A indefiniet dan en slechts dan als nog A nog $-A$ positief definiet zijn.

We zien rechtstreeks in dat deze stelling voor diagonaalmatrices geldt. De grap is nu, dat we door een basistransformatie iedere symmetrische matrix op diagonaalvorm kunnen brengen, en een basistransformatie bewaart de eigenschap van een matrix positief of negatief definiet te zijn.

> De attente lezer is natuurlijk gewaargeworden dat de stelling in het geval $\det(A) = 0$ geen uitspraak erover maakt of de matrix positief of negatief semidefiniet is of indefiniet. Hiervoor zou men de matrix A inderdaad door een basistransformatie op diagonaalvorm moeten brengen, dan laat het zich weer makkelijk aan de diagonaalelementen aflezen.
>
> We gaan dit probleem hier echter niet verdiepen, omdat het geval dat de Hesse matrix determinant 0 heeft in de praktijk nauwelijks een rol speelt. Om in zo'n geval te beslissen of een kritiek punt een minimum, maximum of zadelpunt is, zou men net zo als in het geval $f''(x_0) = 0$ voor gewone functies naar hogere partiële afgeleiden dan de tweede moeten kijken.

Voorbeeld: Een 2×2-matrix $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is positief definiet als $a > 0$ en $\det(A) = ac - b^2 > 0$. De matrix A is negatief definiet als $a < 0$ en $\det(A) = ac - b^2 > 0$. Als $\det(A) < 0$, is A indefiniet.

Toepassing op functies van twee variabelen

Als we de uitspraak van het vorige voorbeeld voor de Hesse matrix

$$H := H(x_0, y_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x_0, y_0) & \frac{\partial^2 f}{\partial x\, \partial y}(x_0, y_0) \\ \frac{\partial^2 f}{\partial x\, \partial y}(x_0, y_0) & \frac{\partial^2 f}{\partial y^2}(x_0, y_0) \end{pmatrix}$$

van een functie van twee veranderlijken herformuleren, krijgen we een handige stelling over de kritieke punten van een functie van twee veranderlijken.

Stelling: Zij $f(x, y)$ een functie van twee variabelen en zij (x_0, y_0) een kritiek punt, d.w.z. een punt met $\nabla f(x_0, y_0) = (0, 0)$.

Verder zij $H = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix}$ de Hesse matrix van $f(x, y)$ in (x_0, y_0), d.w.z.

$$H_{11} = \frac{\partial^2 f}{\partial x^2}(x_0, y_0), \quad H_{12} = \frac{\partial^2 f}{\partial x\, \partial y}(x_0, y_0), \quad H_{22} = \frac{\partial^2 f}{\partial y^2}(x_0, y_0).$$

(1) $f(x, y)$ heeft in het punt (x_0, y_0) een *lokaal extremum* als

$$\det(H) = H_{11}H_{22} - H_{12}^2 > 0, \text{ dus als}$$

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) \cdot \frac{\partial^2 f}{\partial y^2}(x_0, y_0) - (\frac{\partial^2 f}{\partial x\, \partial y}(x_0, y_0))^2 > 0.$$

(i) Het lokale extremum van $f(x,y)$ in (x_0,y_0) is een *lokaal minimum* als
$$H_{11} > 0, \text{ dus als } \frac{\partial^2 f}{\partial x^2}(x_0,y_0) > 0.$$

(ii) Het lokale extremum van $f(x,y)$ in (x_0,y_0) is een *lokaal maximum* als
$$H_{11} < 0, \text{ dus als } \frac{\partial^2 f}{\partial x^2}(x_0,y_0) < 0.$$

(2) $f(x,y)$ heeft in het punt (x_0,y_0) een *zadelpunt* als
$$\det(H) = H_{11}H_{22} - H_{12}^2 < 0, \text{ dus als}$$
$$\frac{\partial^2 f}{\partial x^2}(x_0,y_0)\cdot\frac{\partial^2 f}{\partial y^2}(x_0,y_0) - (\frac{\partial^2 f}{\partial x\,\partial y}(x_0,y_0))^2 < 0.$$

Zo als eerder opgemerkt is deze stelling niet van toepassing als $\det(H) = 0$. In dit geval is de Hesse matrix positief of negatief semidefiniet en is er een vector $\mathbf{h} \neq 0$ met $\mathbf{h}^{tr}\cdot H\cdot\mathbf{h} = 0$. Deze situatie is analoog met het geval $f''(x_0) = 0$ voor gewone functies, waar pas de hogere termen van de Taylor reeks aangeven of het punt een extremum of een zadelpunt is. Dit geldt ook voor de functies van meerdere variabelen, men moet de hogere termen van de Taylor reeks raadplegen om te beslissen hoe zich de functie in een richting $\mathbf{h}$ met $\mathbf{h}^{tr}\cdot H\cdot\mathbf{h} = 0$ gedraagt. In de praktijk speelt dit probleem echter een minder belangrijke rol, dus zullen we genoegen nemen met het geval $\det(H) \neq 0$.

Voorbeeld 1: We bekijken de functie
$$f(x,y) := x^3 + 6xy^2 - 2y^3 - 12x.$$
Er geldt
$$\frac{\partial f}{\partial x} = 3x^2 + 6y^2 - 12 \quad \text{en} \quad \frac{\partial f}{\partial y} = 12xy - 6y^2 = 6y(2x - y).$$
Uit $\frac{\partial f}{\partial y} = 0$ volgt $y = 0$ of $y = 2x$. In het eerste geval volgt uit $\frac{\partial f}{\partial x} = 0$ dat $3x^2 = 12$, dus $x = \pm 2$. In het geval $y = 2x$ moet gelden dat $3x^2 + 24x^2 = 12$, dus $x^2 = \frac{12}{27} = \frac{4}{9}$, dus $x = \pm\frac{2}{3}$. Er zijn dus vier kritieke punten:
$$(2,0), \quad (-2,0), \quad (\frac{2}{3},\frac{4}{3}), \quad (-\frac{2}{3},-\frac{4}{3}).$$
Voor de analyse van de kritieke punten hebben we de tweede partiële afgeleiden nodig, er geldt
$$\frac{\partial^2 f}{\partial x^2} = 6x, \quad \frac{\partial^2 f}{\partial x\,\partial y} = 12y, \quad \frac{\partial^2 f}{\partial y^2} = 12x - 12y.$$
De Hesse matrix is dus
$$H = \begin{pmatrix} 6x & 12y \\ 12y & 12x - 12y \end{pmatrix}$$
en $\det(H) = 6x(12x - 12y) - (12y)^2 = 72x^2 - 72xy - 144y^2$. Voor de kritieke punten geeft dit de volgende tabel:

kritiek punt	H_{11}	$\det(H)$	type
$(2,0)$	12	288	minimum
$(-2,0)$	-12	288	maximum
$(\frac{2}{3},\frac{4}{3})$	4	-288	zadelpunt
$(-\frac{2}{3},-\frac{4}{3})$	-4	-288	zadelpunt

Voorbeeld 2: We onderzoeken de kritieke punten van de functie

$$f(x,y) := (x^2-y^2)\, e^{\frac{-x^2-y^2}{2}}.$$

Er geldt

$$\frac{\partial f}{\partial x} = (2x - x(x^2-y^2))\, e^{\frac{-x^2-y^2}{2}} \quad \text{en} \quad \frac{\partial f}{\partial y} = (-2y - y(x^2-y^2))\, e^{\frac{-x^2-y^2}{2}}.$$

De exponentiële functie wordt nooit 0, dus vinden we de kritieke punten als oplossingen van de vergelijkingen

$$x(2-(x^2-y^2)) = 0 \quad \text{en} \quad y(-2-(x^2-y^2)) = 0.$$

Omdat x^2-y^2 niet tegelijkertijd de waarden 2 en -2 kan hebben, moet $x=0$ of $y=0$ zijn, dit geeft de kritieke punten $(0,0)$, $(\pm\sqrt{2},0)$ en $(0,\pm\sqrt{2})$.

Voor de tweede partiële afgeleiden, dus de elementen H_{ij} van de Hesse matrix, krijgen we:

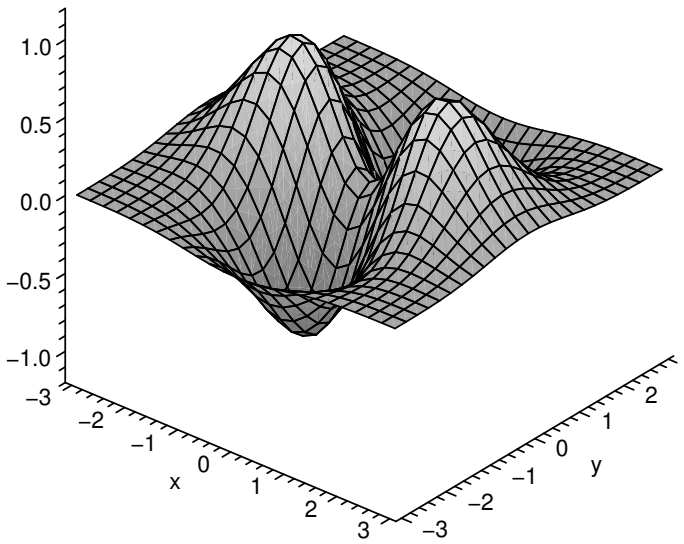
$$\begin{aligned} H_{11} = \frac{\partial^2 f}{\partial x^2} &= (2-5x^2+x^2(x^2-y^2)+y^2)\, e^{\frac{-x^2-y^2}{2}} \\ H_{12} = H_{21} = \frac{\partial^2 f}{\partial x\, \partial y} &= xy(x^2-y^2)\, e^{\frac{-x^2-y^2}{2}} \\ H_{22} = \frac{\partial^2 f}{\partial y^2} &= (5y^2-2+y^2(x^2-y^2)-x^2)\, e^{\frac{-x^2-y^2}{2}}. \end{aligned}$$

Hiermee vinden we de volgende tabel voor de kritieke punten:

kritiek punt	H_{11}	H_{12}	H_{22}	$\det(H) = H_{11}H_{22} - H_{12}^2$	type
$(0,0)$	2	0	-2	-4	zadelpunt
$(\sqrt{2},0)$	$-4e^{-1}$	0	$-4e^{-1}$	$16e^{-2}$	maximum
$(-\sqrt{2},0)$	$-4e^{-1}$	0	$-4e^{-1}$	$16e^{-2}$	maximum
$(0,\sqrt{2})$	$4e^{-1}$	0	$4e^{-1}$	$16e^{-2}$	minimum
$(0,-\sqrt{2})$	$4e^{-1}$	0	$4e^{-1}$	$16e^{-2}$	minimum

In de grafiek van de functie in Figuur I.12 kunnen we controleren dat onze analyse van de kritieke punten inderdaad klopt.

OPDRACHT 13 *Bepaal de kritieke punten van de functie* $f(x,y) := 4x^3 - 3x^2y + y^3 - 9y$ *en ga na of de punten minima, maxima of zadelpunten zijn.*



Figuur I.12: Grafiek van de functie $f(x,y) = (x^2 - y^2)e^{\frac{-x^2-y^2}{2}}$.

Functies van meer dan twee variabelen

Als we bij een functie $f(\mathbf{x})$ van meer dan twee variabelen willen testen of een kritiek punt een minimum, maximum of een zadelpunt is, is het meestal het eenvoudigste het criterium van de linksboven deelmatrices op het concrete voorbeeld toe te passen. Algemene formules zijn al voor 3 variabelen behoorlijk afschrikkend, we beperken ons daarom tot de algemene stelling en een voorbeeld.

Stelling: Zij $f(\mathbf{x})$ een functie van n variabelen, zij $\mathbf{x}_0$ een kritiek punt van $f(\mathbf{x})$, d.w.z. $\nabla f(\mathbf{x}_0) = 0$ en zij $H := H(\mathbf{x}_0)$ de Hesse matrix van $f(\mathbf{x})$ in het punt $\mathbf{x}_0$, d.w.z. $H_{ij} = \frac{\partial^2 f}{\partial x_i \, \partial x_j}(\mathbf{x}_0)$. Dan geldt:

(i) $f(\mathbf{x})$ heeft in het punt $\mathbf{x}_0$ een lokaal minimum als H positief definiet is;

(ii) $f(\mathbf{x})$ heeft in het punt $\mathbf{x}_0$ een lokaal maximum als H negatief definiet is;

(iii) $f(\mathbf{x})$ heeft in het punt $\mathbf{x}_0$ een zadelpunt als H indefiniet is.

Voorbeeld: Zij de functie $f(x,y,z)$ gegeven door

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

Om de kritieke punten te vinden moeten we de eerste partiële afgeleiden bepalen. Er geldt

$$\frac{\partial f}{\partial x} = 2xy - 2, \quad \frac{\partial f}{\partial y} = x^2 + 2yz, \quad \frac{\partial f}{\partial z} = y^2 + 2z.$$

Uit $\frac{\partial f}{\partial z} = 0$ volgt $z = -\frac{y^2}{2}$. Dit ingevuld in $\frac{\partial f}{\partial y} = 0$ geeft $x^2 = y^3$ en hiermee volgt uit $\frac{\partial f}{\partial x} = 0$ dat $y^{\frac{5}{2}} = 1$. Dit geeft $y = 1$, $x = 1$ en $z = -\frac{1}{2}$, het enige kritieke punt is dus $\mathbf{x}_0 = (1, 1, -\frac{1}{2})$.

Om na te gaan of dit een minimum, maximum of een zadelpunt is, moeten we nu de tweede partiële afgeleiden bepalen. We krijgen

$$\frac{\partial^2 f}{\partial x^2} = 2y, \quad \frac{\partial^2 f}{\partial x\,\partial y} = 2x, \quad \frac{\partial^2 f}{\partial x\,\partial z} = 0, \quad \frac{\partial^2 f}{\partial y^2} = 2z, \quad \frac{\partial^2 f}{\partial y\,\partial z} = 2y, \quad \frac{\partial^2 f}{\partial z^2} = 2$$

dit geeft in het punt $\mathbf{x}_0 = (1, 1, -\frac{1}{2})$ de Hesse matrix

$$H(\mathbf{x}_0) = \begin{pmatrix} 2 & 2 & 0 \\ 2 & -1 & 2 \\ 0 & 2 & 2 \end{pmatrix}$$

Voor de linksboven deelmatrices van $H(\mathbf{x}_0)$ geldt

$$\det(\begin{pmatrix} 2 \end{pmatrix}) = 2 > 0, \quad \det(\begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}) = -6 < 0, \quad \det(H) = -20.$$

De matrix is dus indefiniet en het kritieke punt $\mathbf{x}_0$ is een zadelpunt.

Toepassing: Regressielijn

Als belangrijke toepassing voor extrema van functies van meerdere veranderlijken bekijken we het bepalen van de regressielijn

$$L(x) := ax + b$$

door een verzameling $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ van punten. Het idee is hierbij, dat er een lineaire samenhang $y = Ax + B$ tussen de x- en y-coördinaten van de punten verondersteld wordt, maar dat de parameters A en B onbekend zijn. Hiervoor wordt er uit een steekproef van punten een schatting a voor A en b voor B gemaakt door een lijn zo door de punten te leggen, dat de som van de kwadratische afwijkingen $(y_i - (ax_i + b))^2$ tussen de *gemeten* y-waarden y_i en de *berekende* waarden $\hat{y}_i := ax_i + b$ minimaal wordt. We kijken dus naar de functie

$$f(a, b) := \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

en moeten a en b zo bepalen dat $f(a, b)$ minimaal wordt.

Voor de partiële afgeleiden geldt

$$\frac{\partial f}{\partial a} = -2 \sum_{i=1}^{n} x_i (y_i - ax_i - b) \quad \text{en} \quad \frac{\partial f}{\partial b} = -2 \sum_{i=1}^{n} (y_i - ax_i - b)$$

en uit $\nabla f = 0$ volgt dat de gezochte parameters a en b oplossingen zijn van het lineaire stelsel vergelijkingen

$$\begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix}.$$

Uit de tweede vergelijking volgt in het bijzonder dat de regressielijn $ax+b$ door het zwaartepunt $(\overline{x},\overline{y})$ met $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ en $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$ loopt, met deze notatie luidt de tweede vergelijking namelijk $n\overline{x}a + nb = n\overline{y}$, dus geldt

$$b = \overline{y} - a\overline{x}.$$

Ook de eerste vergelijking kunnen we met een geschikte notatie voor gemiddelden iets eenvoudiger schrijven, met $\overline{x^2} = \frac{1}{n}\sum_{i=1}^n x_i^2$ en $\overline{xy} = \frac{1}{n}\sum_{i=1}^n x_i y_i$ krijgen we de vergelijking $\overline{x^2}a + \overline{x}b = \overline{xy}$. Als we in deze vergelijking b door $\overline{y} - a\overline{x}$ vervangen, krijgen we $\overline{x^2}a + \overline{x}(\overline{y} - a\overline{x}) = \overline{xy}$ en opgelost naar a geeft dit de bekende vergelijking

$$a = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}$$

voor de richtingscoëfficiënt van de regressielijn.

We moeten nu nog nagaan dat we inderdaad een minimum van de functie hebben gevonden. Hiervoor bepalen we de tweede partiële afgeleiden, er geldt

$$\frac{\partial^2 f}{\partial a^2} = 2\sum_{i=1}^n x_i^2, \quad \frac{\partial^2 f}{\partial a\,\partial b} = 2\sum_{i=1}^n x_i, \quad \frac{\partial^2 f}{\partial b^2} = 2n,$$

we krijgen dus de Hesse matrix

$$H = 2\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix}.$$

Het is duidelijk dat $H_{11} = 2\sum_{i=1}^n x_i^2 > 0$, omdat dit een som van kwadraten is. Met de notatie $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ krijgen we verder

$$\begin{aligned} n(\sum_{i=1}^n (x_i - \overline{x})^2) &= n\sum_{i=1}^n x_i^2 - n\sum_{i=1}^n 2x_i\overline{x} + n\cdot n\overline{x} = n\sum_{i=1}^n x_i^2 - n^2\overline{x} \\ &= n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 = \det(\frac{1}{2}H). \end{aligned}$$

Aan de linkerkant staat een som van kwadraten, dus is $\det(\frac{1}{2}H) > 0$ en dus ook $\det(H) > 0$. De Hesse matrix is dus altijd positief definiet, in het bijzonder ook in het kritieke punt (a,b) en dus hebben we een minimum gevonden.

3.4 Extrema onder randvoorwaarden

Vaak zijn problemen waarbij we minima of maxima van functies moeten bepalen zo geformuleerd, dat de oplossingen aan zekere randvoorwaarden moeten voldoen. In principe hebben we zo iets al gezien. In het voorbeeld van de open doos hadden we gezegd, dat de doos een bepaald volume V moet hebben. In principe moeten we dus voor een doos met de afmetingen x, y en z de oppervlakte $f(x,y,z) = xy + 2xz + 2yz$ onder de randvoorwaarde minimaliseren dat $xyz = V$ geldt. Deze randvoorwaarde hadden we gebruikt om de variabel z te verwijderen, want we konden z door $\frac{V}{xy}$ vervangen.

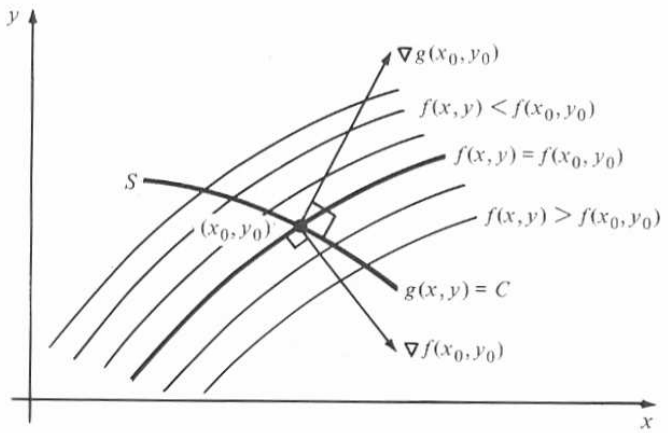
Meestal is het helaas zo, dat we een randvoorwaarde niet zo makkelijk naar één van de variabelen kunnen oplossen, of dat de functie die we dan krijgen erg ingewikkeld wordt. We zullen daarom nu naar een methode kijken, hoe we bij gegeven randvoorwaarden een minimum of maximum van een functie kunnen bepalen.

Het probleem voor functies van twee variabelen luidt als volgt: Voor een functie $f(x, y)$ is een extremum (maximum of minimum) gezocht onder de randvoorwaarde dat $g(x, y) = 0$.

> Een typische randvoorwaarde is, dat we een extremum op de rand van een begrensd oppervlak zo als een cirkelschijf willen bepalen. Als we bijvoorbeeld alleen maar de punten op een cirkel van straal 3 rond het punt $(1, 1)$ willen bekijken, hebben we de punten nodig waarvoor geldt dat $(x-1)^2 + (y-1)^2 = 3^2$, de randvoorwaarde is dan
>
> $$g(x, y) = (x-1)^2 + (y-1)^2 - 3^2 = 0.$$

We gaan nu na dat in een extremum (x_0, y_0) noodzakelijk geldt dat de gradiënten van $f(x, y)$ en $g(x, y)$ in dit punt parallel zijn, dus dat $\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$ voor een geschikte λ.

> Het fundamentele idee waarom in een extremum onder een randvoorwaarde de gradiënten van $f(x, y)$ en $g(x, y)$ parallel moeten zijn, kunnen we aan de hand van de schets in Figuur I.13 inzien.



Figuur I.13: $\nabla f(x_0, y_0)$ en $\nabla g(x_0, y_0)$ niet parallel.

> De punten (x, y) met $g(x, y) = 0$ vormen een niveaukromme van $g(x, y)$, die in het plaatje met S genoteerd is. Omdat S een niveaukromme is, staat ∇g in alle punten van S loodrecht op de raaklijn aan $g(x, y)$. Stel nu dat in een punt (x_0, y_0) de gradiënt ∇f niet loodrecht op de raaklijn aan $g(x, y)$ staat, dan is de projectie $\nabla f_{||}$ van ∇f op de raaklijn aan $g(x, y)$ niet 0.
>
> Aan de ene kant kunnen we nu op de niveaukromme $g(x, y) = 0$ in de richting van $\nabla f_{||}$ lopen, omdat dit de richting van de raaklijn aan

$g(x,y)$ is. Aan de andere kant neemt $f(x,y)$ in de richting van $\nabla f_{||}$ toe, omdat $\nabla f_{||}$ (als projectie van ∇f) niet loodrecht of ∇f staat.

Als we vanuit het punt (x_0, y_0) in de richting van $\nabla f_{||}$ lopen, neemt $f(x,y)$ dus toe, als we in de tegengestelde richting $-\nabla f_{||}$ lopen, neemt $f(x,y)$ af. Omdat we in ieder geval op de niveaukromme $g(x,y) = 0$ lopen en dus aan de randvoorwaarde blijven voldoen, is (x_0, y_0) nog een maximum nog een minimum.

Met behulp van Taylor reeksen kunnen we de redenering van boven zo aanpassen, dat we dezelfde uitspraak voor functies $f(\mathbf{x})$ van n variabelen krijgen. In dit geval is de randvoorwaarde gegeven door $g(\mathbf{x}) = 0$, waarbij ook $g(\mathbf{x})$ een functie van n variabelen is.

Omdat $g(\mathbf{x}) = 0$ een niveauoppervlak is, staat de gradiënt $\nabla g(\mathbf{x}_0)$ loodrecht op het raakvlak aan $g(\mathbf{x})$ in het punt $\mathbf{x}_0$. De richtingen $\mathbf{h}$ met $\nabla g(\mathbf{x}_0) \cdot h = 0$ zijn dus juist de richtingen in die we vanuit $\mathbf{x}_0$ mogen lopen, om verder aan de randvoorwaarde $g(\mathbf{x}) = 0$ te voldoen.

Dit kunnen we ook uit de Taylor reeks voor $g(\mathbf{x})$ afleiden: Voor kleine $\mathbf{h}$ geldt $g(\mathbf{x}_0 + \mathbf{h}) = g(\mathbf{x}_0) + \nabla g(\mathbf{x}_0) \cdot \mathbf{h}$ en als ook $\mathbf{x}_0 + \mathbf{h}$ aan de randvoorwaarde voldoet, is $g(\mathbf{x}_0 + \mathbf{h}) = g(\mathbf{x}_0)$, dus moet $\nabla g(\mathbf{x}_0) \cdot \mathbf{h} = 0$ gelden.

Nu bekijken we de Taylor reeks van $f(\mathbf{x})$ in het punt $\mathbf{x}_0$. De lineaire benadering is

$$f(\mathbf{x}_0 + \mathbf{h}) \approx f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) \cdot \mathbf{h}$$

en in een extremum onder de randvoorwaarde $g(\mathbf{x}) = 0$ moet gelden, dat $\nabla f(\mathbf{x}_0) \cdot \mathbf{h} = 0$ voor alle richtingen $\mathbf{h}$ in die we onder de randvoorwaarde $g(\mathbf{x}) = 0$ mogen lopen. Maar we hebben net ingezien dat de mogelijke richtingen $\mathbf{h}$ juist de richtingen zijn die loodrecht op $\nabla g(\mathbf{x}_0)$ staan, dus moet in een extremum gelden dat

$$\nabla f(\mathbf{x}_0) \cdot \mathbf{h} = 0 \text{ voor alle } \mathbf{h} \text{ met } \nabla g(\mathbf{x}_0) \cdot \mathbf{h} = 0.$$

Dit betekent dat $\nabla f(\mathbf{x}_0)$ loodrecht op alle $\mathbf{h}$ staat, die zelf loodrecht op $\nabla g(\mathbf{x}_0)$ staan, dus moet $\nabla f(\mathbf{x}_0)$ loodrecht op het raakvlak aan $g(\mathbf{x})$ in het punt $\mathbf{x}_0$ staan. Maar de enige vectoren die loodrecht op dit raakvlak staan, zijn de (positieve en negatieve) veelvouden van $\nabla g(\mathbf{x}_0)$ en hieruit volgt dat

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0).$$

We hebben dus de volgende stelling ingezien:

Stelling: In een lokaal extremum $\mathbf{x}_0$ onder de randvoorwaarde $g(\mathbf{x}) = 0$ geldt dat $\nabla f(\mathbf{x}_0)$ en $\nabla g(\mathbf{x}_0)$ lineair afhankelijke vectoren zijn, dus dat er een λ bestaat met

$$\nabla f(\mathbf{x}_0) + \lambda \nabla g(\mathbf{x}_0) = 0.$$

(Merk op dat we hier tegenover de formulering $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$ van boven λ door $-\lambda$ hebben vervangen.)

3.5 De methode van Lagrange multiplicatoren

De stelling hierboven geeft aanleiding tot een belangrijke methode om extrema onder randvoorwaarden te bepalen. De functie

$$L(\mathbf{x}, \lambda) := f(\mathbf{x}) + \lambda g(\mathbf{x})$$

van de $n+1$ variabelen $x_1, \ldots, x_n$ en λ heet de *Lagrange functie* van $f(\mathbf{x})$ onder de randvoorwaarde $g(\mathbf{x}) = 0$. De variabel λ heet hierbij de *Lagrange multiplicator*. Voor de Lagrange functie geldt:

Stelling: Als $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ een extremum onder de randvoorwaarde $g(\mathbf{x}) = 0$ heeft, dan heeft de Lagrange functie $L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$ in $\mathbf{x}_0$ een kritiek punt.

Voor een extremum onder een randvoorwaarde geldt namelijk volgens de stelling van boven, dat $\nabla f(\mathbf{x}_0) = -\lambda \nabla g(\mathbf{x}_0)$, en daarom is $\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \lambda \frac{\partial g}{\partial x_i} = 0$ voor alle $1 \leq i \leq n$.

Omdat $\frac{\partial L}{\partial \lambda} = g(\mathbf{x})$, is $\frac{\partial L}{\partial \lambda} = 0$ equivalent met de uitspraak dat het punt $\mathbf{x}_0$ aan de randvoorwaarde voldoet.

Voorbeeld 1: We bepalen de extrema van de functie

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

op de rand van de cirkel met straal 1 rond $(0, 0)$. De punten op de cirkel kunnen we beschrijven door de randvoorwaarde $x^2 + y^2 = 1$, dus is de functie $g(x, y)$ gegeven door $g(x, y) := x^2 + y^2 - 1$. De Lagrange functie is dus

$$L(x, y, \lambda) = x^2 + 2y^2 + \lambda(x^2 + y^2 - 1).$$

Er geldt

$$\frac{\partial L}{\partial x} = 2x + 2\lambda x, \quad \frac{\partial L}{\partial y} = 4y + 2\lambda y, \quad \frac{\partial L}{\partial \lambda} = x^2 + y^2 - 1.$$

Uit $\frac{\partial L}{\partial x} = 0$ volgt $2x(1 + \lambda) = 0$, dus $x = 0$ of $\lambda = -1$. Uit $\frac{\partial L}{\partial y} = 0$ volgt $2y(2 + \lambda) = 0$, dus $y = 0$ of $\lambda = -2$. Omdat λ niet tegelijkertijd -1 en -2 kan zijn, is noodzakelijk $x = 0$ of $y = 0$.

We krijgen dus de kritieke punten $(1, 0)$, $(-1, 0)$, $(0, 1)$ en $(0, -1)$ van de Lagrange functie.

Het is natuurlijk in dit voorbeeld niet moeilijk om in te zien dat onder de randvoorwaarde $x^2 + y^2 = 1$ geldt, dat $f(x, y) = (x^2 + y^2) + y^2 = 1 + y^2$, dus vinden we minima in $(\pm 1, 0)$ en maxima in $(0, \pm 1)$.

Voorbeeld 2: We bepalen de minima en maxima van de functie

$$f(x, y) := xy$$

op de cirkel met $x^2 + y^2 = 1$. De Lagrange functie is

$$L(x, y, \lambda) = xy + \lambda(x^2 + y^2 - 1)$$

en voor de partiële afgeleiden krijgen we

$$\frac{\partial L}{\partial x} = y + 2\lambda x, \quad \frac{\partial L}{\partial y} = x + 2\lambda y, \quad \frac{\partial L}{\partial \lambda} = x^2 + y^2 - 1.$$

Voor $x = 0$ volgt rechtstreeks $y = 0$ en andersom, dus moet volgens de randvoorwaarde noodzakelijk gelden dat $x \neq 0$ en $y \neq 0$. Uit de eerste twee vergelijkingen volgt dat $\frac{x}{y} = \frac{y}{x} = -2\lambda$, dit geeft $x^2 = y^2$ en hieruit volgt met de derde vergelijking dat $2x^2 = 1$, dus $x = \pm\frac{1}{\sqrt{2}}$ en evenzo $y = \pm\frac{1}{\sqrt{2}}$.

Men gaat na dat de vier punten $(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}})$ inderdaad voldoen aan $\nabla L = 0$, dus kritieke punten van de Lagrange functie zijn, en er geldt $f(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = f(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}) = \frac{1}{2}$ en $f(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}) = f(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = -\frac{1}{2}$, dus zijn de eerste twee punten maxima en de laatste twee punten minima van de functie.

Voorbeeld 3: We willen het punt (x, y, z) op het oppervlak gegeven door de vergelijking $z = x^2 + y^2$ bepalen, dat het dichtst bij het punt $P := (1, 1, \frac{1}{2})$ ligt. De functie $f(x, y, z)$ die we moeten bekijken is in dit geval de afstand van (x, y, z) van het punt P. Maar omdat wortels vaak onhandig zijn, kijken we liever naar het kwadraat van de afstand, dus naar de functie

$$f(x, y, z) := (x-1)^2 + (y-1)^2 + (z - \frac{1}{2})^2 = x^2 - 2x + y^2 - 2y + z^2 - z + \frac{9}{4}.$$

De randvoorwaarde is in dit geval gegeven door $g(x, y, z) = 0$ met $g(x, y, z) = x^2 + y^2 - z$ en we krijgen de Lagrange functie

$$L(x, y, z, \lambda) = x^2 - 2x + y^2 - 2y + z^2 - z + \frac{9}{4} + \lambda(x^2 + y^2 - z).$$

Voor de partiële afgeleiden geldt

$$\frac{\partial L}{\partial x} = 2x - 2 + 2\lambda x, \ \frac{\partial L}{\partial y} = 2y - 2 + 2\lambda y, \ \frac{\partial L}{\partial z} = 2z - 1 - \lambda, \ \frac{\partial L}{\partial \lambda} = x^2 + y^2 - z$$

en op 0 zetten van de partiële afgeleiden naar x, y en z geeft

$$x = \frac{1}{1+\lambda}, \quad y = \frac{1}{1+\lambda}, \quad z = \frac{1+\lambda}{2}.$$

Dit ingevuld in $x^2+y^2 = z$ geeft $\frac{2}{(1+\lambda)^2} = \frac{1+\lambda}{2}$ en dus $(1+\lambda)^3 = 4$ of $\lambda = \sqrt[3]{4}-1$. Hieruit krijgen we voor x, y en z:

$$x = \frac{1}{\sqrt[3]{4}}, \quad y = \frac{1}{\sqrt[3]{4}}, \quad z = \frac{\sqrt[3]{4}}{2} = \frac{1}{\sqrt[3]{2}}.$$

Het gezochte punt op het oppervlak $z = x^2 + y^2$ met de kleinste afstand van P is dus $(\frac{1}{\sqrt[3]{4}}, \frac{1}{\sqrt[3]{4}}, \frac{1}{\sqrt[3]{2}})$.

Natuurlijk hadden we de vergelijking $x^2 + y^2 - z = 0$ ook naar z kunnen oplossen en $x^2 + y^2$ in plaats van z in de functie $f(x, y, z)$ invullen. Op deze manier krijgen we de nieuwe functie

$$h(x, y) = x^2-2x+y^2-2y+(x^2+y^2)^2-(x^2+y^2)+\frac{9}{4} = x^4+2x^2y^2+y^4-2x-2y+\frac{9}{4}$$

waarvan we het minimum zonder randvoorwaarden mogen bepalen. Er geldt

$$\frac{\partial h}{\partial x} = 4x^3+4xy^2-2 = 4x(x^2+y^2)-2 \quad \text{en} \quad \frac{\partial h}{\partial y} = 4y^3+4x^2y-2 = 4y(x^2+y^2)-2$$

en uit $\frac{\partial h}{\partial x} = 0$ volgt $x^2 + y^2 = \frac{1}{2x}$. Dit ingevuld in $\frac{\partial h}{\partial y} = 0$ geeft $\frac{4y}{2x} = 2$, dus moet $x = y$ gelden. Als we dit weer in $\frac{\partial h}{\partial x} = 0$ invullen, krijgen we $8x^3 = 2$ en dus $x = y = \frac{1}{\sqrt[3]{4}}$. Uit $z = x^2 + y^2$ volgt nu weer dat $z = \frac{1}{\sqrt[3]{2}}$.

> Gelukkig geven beide methodes dezelfde oplossing. In het algemeen is het echter niet mogelijk, een variabel (zo als hier z) te elimineren, maar zelfs als dit lukt zijn de vergelijkingen $\nabla f = 0$ vaak moeilijk op te lossen. Zo als in dit voorbeeld geeft de methode met Lagrange multiplicatoren meestal eenvoudigere vergelijkingen, maar het zijn er natuurlijk meer vergelijkingen en meer onbekenden. Alles heeft zijn prijs!

OPDRACHT 14 *Bepaal het maximum van de functie $f(x, y, z) := x + z$ onder de randvoorwaarde $x^2 + y^2 + z^2 = 1$.*

Meerdere randvoorwaarden

Tot nu toe zijn we er altijd van uitgegaan dat de randvoorwaarde door één functie $g(\mathbf{x}) = 0$ gegeven is. Maar soms moet men ook naar meerdere randvoorwaarden kijken, bijvoorbeeld als het maximum van een functie op een kromme gezocht is, die gegeven is als doorsnede van twee oppervlakken in de ruimte.

In feite verandert bij meerdere randvoorwaarden niet zo erg veel: Als r randvoorwaarden gegeven zijn door $g_k(\mathbf{x}) = 0$ voor $1 \leq k \leq r$ moet in een extremum $\mathbf{x}_0$ gelden dat de gradiënt $\nabla f(\mathbf{x}_0)$ een lineaire combinatie van de gradiënten $\nabla g_k(\mathbf{x}_0)$ van de randvoorwaarden is. Er moeten dus coëfficiënten $\lambda_1, \ldots, \lambda_r$ bestaan met

$$\nabla f(\mathbf{x}_0) = -\lambda_1 \nabla g_1(\mathbf{x}_0) - \ldots - \lambda_r \nabla g_r(\mathbf{x}_0).$$

> Ook dit is niet erg moeilijk in te zien: De toegelaten richtingen $\mathbf{h}$ in die we onder de randvoorwaarden vanuit een punt $\mathbf{x}_0$ mogen lopen, liggen in de doorsnede van de raakvlakken aan de $g_k(\mathbf{x})$ in het punt $\mathbf{x}_0$. Maar de vectoren die loodrecht op deze doorsnede van raakvlakken staan, zijn juist de lineaire combinaties van de gradiënten $\nabla g_k(\mathbf{x}_0)$.

Men definieert nu de Lagrange functie

$$L(\mathbf{x}, \lambda_1, \ldots, \lambda_r) := f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \ldots + \lambda_r g_r(\mathbf{x})$$

met Lagrange multiplicatoren $\lambda_1, \ldots, \lambda_r$ en vindt de extrema van $f(\mathbf{x})$ onder de randvoorwaarden $g_k(\mathbf{x})$ met behulp van de volgende stelling:

Stelling: Als $f(\mathbf{x})$ in het punt $\mathbf{x}_0$ een extremum onder de randvoorwaarden $g_1(\mathbf{x}) = 0, \ldots, g_r(\mathbf{x}) = 0$ heeft, dan is $\mathbf{x}_0$ een kritiek punt van de Lagrange functie $L(\mathbf{x}, \lambda_1, \ldots, \lambda_r) = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \ldots + \lambda_r g_r(\mathbf{x})$.

Voorbeeld: We bepalen de extrema van de functie $f(x,y,z) := 5x+y-3z$ op de doorsnede van het vlak met de vergelijking $x+y+z=0$ en de kogel van straal 1 rond $(0,0,0)$, dus onder de randvoorwaarden $g_1(x,y,z) := x+y+z=0$ en $g_2(x,y,z) := x^2+y^2+z^2-1=0$. De Lagrange functie is

$$L(x,y,z,\lambda_1,\lambda_2) := 5x+y-3z+\lambda_1(x+y+z)+\lambda_2(x^2+y^2+z^2-1)$$

en we krijgen de partiële afgeleiden

$$\frac{\partial L}{\partial x} = 5+\lambda_1+2\lambda_2 x, \quad \frac{\partial L}{\partial y} = 1+\lambda_1+2\lambda_2 y, \quad \frac{\partial L}{\partial z} = -3+\lambda_1+2\lambda_2 z$$

$$\frac{\partial L}{\partial \lambda_1} = x+y+z, \quad \frac{\partial L}{\partial \lambda_2} = x^2+y^2+z^2-1.$$

Optellen van de drie eerste vergelijkingen geeft in verband met de vierde, dat $3+3\lambda_1 = 0$, dus $\lambda_1 = -1$. Hieruit volgt met de tweede dat $y=0$ en dus $z=-x$. De laatste vergelijking geeft nu $2x^2=1$, dus $x = \pm\frac{1}{\sqrt{2}}$ en we krijgen als kritieke punten

$$\mathbf{x}_1 = (\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}) \quad \text{en} \quad \mathbf{x}_2 = (-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}).$$

Men gaat na dat $f(\mathbf{x}_1) = 4\sqrt{2}$ en $f(\mathbf{x}_2) = -4\sqrt{2}$, het eerste punt is dus een maximum, het tweede een minimum.

> De kritieke punten van de Lagrange functie geven net als de kritieke punten van functies zonder randvoorwaarden alleen maar kandidaten voor minima of maxima. Om erover te beslissen of een punt inderdaad een minimum of maximum is, moet men op een iets slimmere manier dan zonder randvoorwaarden naar de tweede partiële afgeleiden kijken.
>
> Maar bij dit soort vraagstukken is het bepalen van de kritieke punten meestal het grotere probleem, vaak volgt uit de samenhang dat een kritiek punt alleen maar een minimum of maximum kan zijn. We zullen deze vraag dus buiten beschouwing laten.

Toepassing: Entropie

Voor een discrete kansverdeling met kansen $p_1,\ldots,p_n$ voor n mogelijke uitkomsten definieert men de *entropie* H door

$$H := -\sum_{i=1}^{n} p_i \; {}^2\log(p_i),$$

waarbij we met ${}^2\log(x)$ de logaritme met basis 2 noteren (dus ${}^2\log(2^x) = x$). De entropie is een maat voor de onzekerheid die we over de uitkomsten volgens de gegeven kansverdeling hebben. Bijvoorbeeld zijn we bij een experiment met 2 mogelijke uitkomsten met $p_1 = 0.9$ en $p_2 = 0.1$ veel minder onzeker over de uitkomst dan bij een experiment met $p_1 = p_2 = 0.5$.

We zullen nu aantonen, dat de uniforme verdeling met $p_i = \frac{1}{n}$ voor alle kansverdelingen voor n uitkomsten de hoogste entropie heeft. De randvoorwaarde die we hanteren is natuurlijk $p_1 + \ldots + p_n = 1$, omdat we het over kansverdelingen hebben. Als Lagrange functie krijgen we

$$L(p_1, \ldots, p_n, \lambda) = -\sum_{i=1}^{n} p_i \ {}^2\log(p_i) + \lambda(\sum_{i=1}^{n} p_i) - \lambda.$$

Merk op dat ${}^2\log(x) = \frac{\log(x)}{\log(2)}$, dus is ${}^2\log'(x) = \frac{1}{\log(2)x}$. Er geldt

$$\frac{\partial L}{\partial p_i} = {}^2\log(p_i) + p_i \frac{1}{\log(2)p_i} + \lambda = {}^2\log(p_i) + \frac{1}{\log(2)} + \lambda$$

en uit $\frac{\partial L}{\partial p_i} = 0$ volgt dus ${}^2\log(p_i) = -\frac{1}{\log(2)} - \lambda$. In het bijzonder moeten dus alle p_i gelijk zijn en uit de randvoorwaarde volgt dan natuurlijk $p_i = \frac{1}{n}$.

Voor de entropie van de uniforme verdeling geldt dat

$$H = -\sum_{i=1}^{n} \frac{1}{n} \ {}^2\log(\frac{1}{n}) = -\ {}^2\log(\frac{1}{n}) = \ {}^2\log(n),$$

dus heeft een uniforme verdeling met 2^m mogelijke uitkomsten de entropie m. Omgekeerd is de onzekerheid bij een kansverdeling met entropie H even groot als bij een uniforme verdeling met 2^H uitkomsten.

De entropie H van een kansexperiment geeft ook aan hoeveel bits gemiddeld minstens nodig zijn om de uitkomsten van het experiment te coderen.

Belangrijke begrippen in deze les

- kritieke punten
- lokale maxima/minima
- positief/negatief definiet
- extrema onder randvoorwaarden
- Lagrange functie, Lagrange multiplicatoren

Opgaven

21. De uitwerking van een hoeveelheid van x μg (microgram) van een medicijn is op een tijdstip t na de inneming gegeven door

$$f(x,t) = x^2(a-x)t^2e^{-t},$$

waarbij a de maximaal mogelijke hoeveelheid van de medicijn is. Wat is de maximale uitwerking van de medicijn die bereikt kan worden, en voor welke hoeveelheid wordt deze op welk tijdstip bereikt?

22. Vind de kritieke punten, lokale maxima, minima en zadelpunten van de volgende functies:

    (i) $f(x,y) := x^3 + 6xy^2 - 2y^3 - 12x$;

    (ii) $f(x,y) := xy\, e^{-(x^2+y^2)}$;

    (iii) $f(x,y) := \frac{x}{1+x^2+y^2}$;

    (iv) $f(x,y,z) := x^2y + y^2z + z^2 - 2x$.

23. Vind de kritieke punten van de volgende functies en beslis waar minima, maxima of zadelpunten liggen:

    (i) $f(x,y) := x^2 - y^2 + xy$;

    (ii) $f(x,y) := x^2 + y^2 + 3xy$;

    (iii) $f(x,y) := e^{1+x^2-y^2}$;

    (iv) $f(x,y) := x^2 - 3xy + 5x - 2y + 6y^2 + 8$;

    (v) $f(x,y) := \sin(x^2+y^2)$;

    (vi) $f(x,y) := \cos(x^2+y^2)$;

    (vii) $f(x,y) := y + x\sin(y)$;

    (viii) $f(x,y) := e^x \cos(y)$.

24. Vind de kritieke punten van de volgende functies en beslis waar minima, maxima of zadelpunten liggen:

    (i) $f(x,y) := xy + \frac{1}{x} + \frac{1}{y}$;

    (ii) $f(x,y) := \log(2 + \sin(xy))$;

    (iii) $f(x,y) := x\sin(y)$;

    (iv) $f(x,y) := (x+y)(xy+1)$.

25. Zij $f(x,y) := x^2 + y^2 + kxy$ waarbij k een constante is.

    (i) Bepaal de kritieke punten van $f(x,y)$.

    (ii) Voor welke waarden van k heeft $f(x,y)$ een extremum, voor welke een zadelpunt?

    (iii) Voor welke waarde van k heeft de Hesse matrix van $f(x,y)$ determinant 0 en hoe ziet in dit geval de grafiek van $f(x,y)$ er uit?

26. Zij $f(x,y) := \frac{1}{xy}$. Vind het punt op de grafiek van $f(x,y)$ dat het dicht bij de oorsprong $(0,0,0)$ in $\mathbb{R}^3$ ligt, d.w.z. vind het punt $(x, y, \frac{1}{xy})$ met de kleinste afstand van $(0,0,0)$.

27. Vind de extrema van de volgende functies onder de aangegeven randvoorwaarden:

    (i) $f(x,y,z) := x - y + z$ onder de randvoorwaarde $x^2 + y^2 + z^2 = 2$;

    (ii) $f(x,y) := x - y$ onder de randvoorwaarde $x^2 - y^2 = 2$;

    (iii) $f(x,y) := x$ onder de randvoorwaarde $x^2 + 2y^2 = 3$;

    (iv) $f(x,y) := 3x + 2y$ onder de randvoorwaarde $2x^2 + 3y^2 = 3$;

    (v) $f(x,y,z) := x + y + z$ onder de randvoorwaarden $x^2 - y^2 = 1$ en $2x + z = 1$.

28. De temperatuur $T = T(x,y,z)$ op het oppervlak van een kogel van straal 1 rond de oorsprong $(0,0,0)$ is gegeven door $T(x,y,z) = xz + yz$. Vind de *hot spots* op de kogel, dus de punten met maxima van de temperatuur.

29. Een open doos moet volume V hebben. De onderkant van de doos moet stabiel zijn en de voorkant van de doos moet er mooi uitzien, daarom is het materiaal voor deze twee zijden van de doos vijf keer zo duur als het materiaal voor de andere drie zijden. Wat zijn de afmetingen van de goedkoopste doos met deze eigenschappen?

30. Bepaal het punt op de kromme gegeven door $17x^2+12xy+8y^2 = 100$ die het dichtst bij de oorsprong $(0, 0)$ ligt. (De kromme is in feite een ellips die schuin in het vlak ligt.)

31. Bepaal (bij benadering) het punt op de grafiek van $y = \log(x)$ die het dichtst bij het punt $(1, 1)$ ligt.

    (Je moet hierbij uiteindelijk een nulpunt van een gewone functie van x bepalen, die alleen maar numeriek maar niet analytisch te vinden is.)

32. Bepaal het maximum van de functie $f(x_1, \ldots, x_n) = x_1 \cdot \ldots \cdot x_n$ onder de randvoorwaarde $x_1 + \ldots + x_n = a$ voor een zekere $a > 0$. Veronderstel hierbij dat $x_i \geq 0$ voor alle i.

    Concludeer uit het resultaat dat het meetkundig gemiddelde $\sqrt[n]{x_1 \cdot \ldots \cdot x_n}$ uit positieve getallen x_i steeds kleiner of gelijk aan het rekenkundig gemiddelde $\frac{1}{n}(x_1 + \ldots + x_n)$ is.

Les 4 Integratie voor meerdere variabelen

In deze les bekijken we het omgekeerde van de afgeleide, de integratie, en gaan na hoe we een integraal voor functies van meerdere variabelen kunnen uitrekenen. De functies waar we het hierbij over hebben zijn weer functies van n variabelen die waarden in $\mathbb{R}$ hebben, dus functies $f(\mathbf{x}) = f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$.

Net zo als we met de integraal voor een gewone functie van één variabel de oppervlakte onder een grafiek berekenen, geeft de integraal voor een functie van twee variabelen het volume onder de grafiek van de functie aan. Analoog geeft voor een algemene functie van n variabelen de integraal een (veralgemeend) volume in de $n+1$-dimensionale ruimte aan.

> Meerdimensionale integralen hebben veel toepassingen in de patroonverwerking, bijvoorbeeld is de intensiteit van een plaatje een functie van de $x-y$-coördinaten en is de totale intensiteit op een gebied de integraal van de intensiteit over dit gebied. Maar ook voor kansverdelingen van gecombineerde stochasten die door een dichtheidsfunctie gegeven zijn, moeten we meerdimensionale integralen berekenen om de kans op uitkomsten in een zeker interval te vinden of de verwachtingswaarde te bepalen.

We zullen in deze les vooral functies van twee of drie variabelen behandelen, omdat deze belangrijke toepassingen hebben en het schrijfwerk hierbij nog beperkt is. Het algemene geval werkt echter op een analoge manier en bevat geen verdere complicaties.

4.1 Integratie op (veralgemeende) rechthoeken

Voor een functie $f(x)$ van één variabel hebben we de integraal $\int_a^b f(x)\,dx$ gedefinieerd als limiet

$$\int_a^b f(x)\,dx := \lim_{N\to\infty} \sum_{i=0}^{N-1} f(a+i\Delta x)\cdot \Delta x,$$

waarbij het interval $[a,b]$ in N even grote deelintervallen van lengte $\Delta x = \frac{b-a}{N}$ onderverdeeld wordt. Het idee achter deze definitie is, de oppervlakte onder de grafiek van $f(x)$ te benaderen door een rij van rechthoeken van breedte Δx en hoogte $f(a+i\Delta x)$. Naarmate N groter (en dus Δx kleiner) wordt, geven de rechthoeken een steeds betere benadering van de echte oppervlakte.

> In de wiskunde is een iets algemenere definitie gebruikelijk, waarbij men punten $a = x_0 < x_1 < \ldots < x_{N-1} < x_N = b$ kiest en de som $\sum_{i=0}^{N-1} f(x_i)(x_{i+1}-x_i)$ bekijkt. Dit betekent gewoon dat de rechthoeken niet alle even breed hoeven te zijn. Voor de limiet is het dan wel noodzakelijk dat het maximum van de intervallen $(x_{i+1}-x_i)$ tegen 0 gaat.

> Voor de functies waar we het hier over hebben is onze eenvoudigere definitie echter voldoende, de gevallen waar de definities tot verschillende resultaten leiden, zijn erg pathologisch.

We kunnen het idee achter de definitie van de gewone integraal nu als volgt op functies van meerdere veranderlijke veralgemenen: Uit een interval $[a, b]$ voor de variabel x wordt bij twee variabelen x en y een rechthoek in het $x - y$-vlak, dit is namelijk de combinatie van twee intervallen, één voor x en één voor y. In het geval van functies van twee variabelen worden de gewone intervallen $[a, b]$ dus vervangen door rechthoeken $[a, b] \times [c, d]$ van de vorm

$$[a, b] \times [c, d] := \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b,\, c \leq y \leq d\}.$$

Analoog krijgen we voor een functie van drie variabelen door de combinatie van de drie intervallen $x \in [a, b]$, $y \in [c, d]$, $z \in [e, f]$ een blok

$$[a, b] \times [c, d] \times [e, f] := \{(x, y, z) \in \mathbb{R}^3 \mid a \leq x \leq b,\, c \leq y \leq d,\, e \leq z \leq f\}$$

in de 3-dimensionale ruimte.

Algemeen geeft bij n variabelen $x_1, \ldots, x_n$ de combinatie van n intervallen $[a_1, b_1], \ldots, [a_n, b_n]$ de n-dimensionale rechthoek

$$[a_1, b_1] \times \ldots \times [a_n, b_n] := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i \text{ voor } i = 1, \ldots, n\}.$$

Integratie voor twee variabelen

De integratie over een rechthoek wordt analoog met het geval van één variabel gedefinieerd als limiet van de som over pilaren met als grondvlak een rechthoek met zijden $\Delta x, \Delta y$ en hoogte $f(a + i\Delta x, c + j\Delta y)$. Het volume van zo'n pilaar is natuurlijk gelijk aan de hoogte maal het grondvlak, dus $f(a+i\Delta x, c+j\Delta y) \cdot \Delta x \Delta y$. Dit leidt tot de volgende definitie van de intergraal voor functies van twee variabelen over een rechthoek:

Definitie: Voor een functie $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ is de integraal van $f(x, y)$ over de rechthoek $R := [a, b] \times [c, d]$ gedefinieerd door

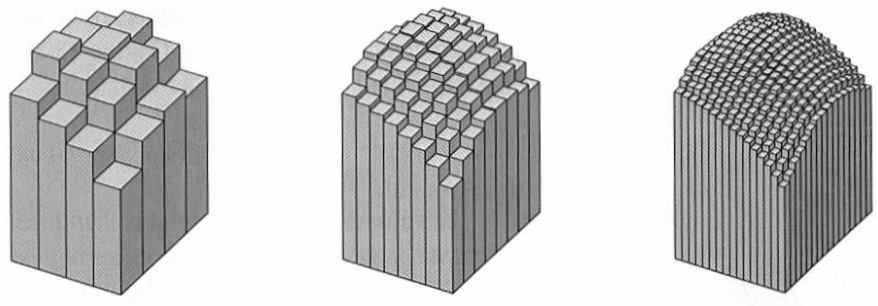
$$\int_R f(x, y)\, dA = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} f(a + i\Delta x, c + j\Delta y) \cdot \Delta x \Delta y,$$

waarbij $N = \frac{b-a}{\Delta x}$ en $M = \frac{d-c}{\Delta y}$. Hierbij schrijven we het symbool dA voor het *differentiaal* van een oppervlakte element, dus voor de limiet $\Delta x, \Delta y \to 0$ van de rechthoeken met zijden $\Delta x, \Delta y$.

> In een algemenere definitie wordt de rechthoek $[a, b] \times [c, d]$ in kleine stukken ΔA_i gesplitst, die niet noodzakelijk rechthoekig hoeven te zijn. Men kiest nu in elk stuk A_i een punt (x_i, y_i) en benadert de integraal door de som $\sum_{i=1}^{N} f(x_i, y_i) \cdot \Delta A_i$. Voor de limiet moet de diameter van de ΔA_i tegen 0 gaan.

> Ook hier geldt, dat dit voor redelijke functies geen verschil met onze eenvoudigere definitie geeft. Als redelijk beschouwen we hierbij functies, die stuksgewijs continu zijn.

Het principe van het berekenen van het volume onder de grafiek van een functie van twee variabelen middels de benadering door steeds fijner wordende rechthoekige pilaren is in Figuur I.14 geïllustreerd. In het eerste plaatje zijn beide intervallen $[a, b]$ en $[c, d]$ in 4 deelintervallen onderverdeeld, in het tweede plaatje in 8 deelintervallen en in het derde plaatje in 16 deelintervallen.



Figuur I.14: Benadering van een volume door rechthoekige pilaren.

In de definitie van de integraal voor functies van twee variabelen hebben we het met twee limieten tegelijkertijd te maken, met de limiet $\Delta x \to 0$ en de limiet $\Delta y \to 0$. Deze kunnen we op verschillende manieren berekenen, we kunnen of eerst de limiet over Δx en dan die over Δy uitvoeren, of andersom, of we kunnen de twee tegelijkertijd tegen 0 laten gaan. Het is niet vanzelfsprekend dat de verschillende manieren in elk geval hetzelfde resultaat geven, en bij zekere functies is dit helaas ook niet het geval. Maar we mogen hier weer ervan uitgaan, dat het voor de functies die we in de praktijk tegenkomen wel goed gaat en dat we altijd in de aangename situatie zijn die door de stelling van Fubini weergegeven wordt:

Stelling van Fubini: Als de integraal $\int_{[a,b]\times[c,d]} f(x,y)\ dA$ bestaat, dan bestaan ook de functies $g(y) := \int_a^b f(x,y)\ dx$ en $h(x) := \int_c^d f(x,y)\ dy$ en er geldt

$$\begin{aligned}\int_{[a,b]\times[c,d]} f(x,y)\ dA &= \int_c^d g(y)\ dy = \int_c^d \left(\int_a^b f(x,y)\ dx\right) dy \\ &= \int_a^b h(x)\ dx = \int_a^b \left(\int_c^d f(x,y)\ dy\right) dx.\end{aligned}$$

Merk op dat we in $g(y) := \int_a^b f(x,y)\ dx$ bij de integratie de variabel y als constante beschouwen, dit is dus een gewone integratie van één veranderlijke. Hetzelfde geldt voor de functie $h(x)$.

Merk op: In de praktijk kunnen we een integraal over een (voldoende goedaardige) functie van twee variabelen uitwerken door eerst over een van de

variabelen te integreren, en vervolgens over de andere, dus door *geïtereerde integraties* van één veranderlijke.

Voorbeeld 1: Zij $f(x,y) := 2x + 3y$ en $R := [0,2] \times [3,4]$. Dan is

$$\begin{aligned}\int_R f(x,y)\ dA &= \int_0^2 \left(\int_3^4 (2x+3y)\ dy \right) dx = \int_0^2 \left((2xy + \frac{3}{2}y^2)\big|_3^4 \right) dx \\ &= \int_0^2 (8x + 24 - 6x - \frac{27}{2})\ dx = \int_0^2 (2x + \frac{21}{2})\ dx \\ &= (x^2 + \frac{21}{2}x)\big|_0^2 = 4 + 21 = 25.\end{aligned}$$

We kunnen ook eerst over x en dan over y integreren:

$$\begin{aligned}\int_R f(x,y)\ dA &= \int_3^4 \left(\int_0^2 (2x+3y)\ dx \right) dy = \int_3^4 \left((x^2 + 3xy)\big|_0^2 \right) dy \\ &= \int_3^4 (4 + 6y)\ dy = (4y + 3y^2)\big|_3^4 = 16 + 48 - 12 - 27 = 25.\end{aligned}$$

We zien dat in dit voorbeeld de tweede manier iets makkelijker is dan de eerste, maar de resultaten zijn natuurlijk hetzelfde.

Voorbeeld 2: Zij $f(x,y) := e^{x+y}$ en $R := [1,2] \times [1,2]$. Dan is

$$\begin{aligned}\int_R f(x,y)\ dA &= \int_1^2 \left(\int_1^2 e^{x+y}\ dy \right) dx = \int_1^2 \left(\int_1^2 e^x \cdot e^y\ dy \right) dx \\ &= \int_1^2 \left(\int_1^2 e^y\ dy \right) e^x\ dx = \int_1^2 \left(e^y\big|_1^2 \right) e^x\ dx \\ &= \int_1^2 (e^2 - e) e^x\ dx = (e^2 - e) \int_1^2 e^x\ dx = (e^2 - e) \cdot e^x\big|_1^2 \\ &= (e^2 - e)(e^2 - e) = (e^2 - e)^2.\end{aligned}$$

Integratie voor drie (of meer) variabelen

Voor functies van drie variabelen geldt iets soortgelijks als voor functies van twee variabelen, we moeten nu over kleine volume elementen (blokken) $\Delta x \Delta y \Delta z$ integreren, die in de limiet tot een differentiaal dV van een volume element wordt. Ook de integratie over de kleine volume elementen kunnen we weer opsplitsen in drie gewone integraties, er geldt:

$$\int_{[a,b]\times[c,d]\times[e,f]} f(x,y,z)\ dV = \int_a^b \left(\int_c^d \left(\int_e^f f(x,y,z)\ dz \right) dy \right) dx.$$

Ook hier kunnen we een andere volgorde voor de integraties kiezen, het maakt niets uit of we eerst over x, y of z integreren. Soms scheelt een geschikte keuze van de volgorde zelfs een hoop rekenwerk.

Voorbeeld: Zij $f(x,y,z) := \frac{x^2z^3}{1+y^2}$ en $R := [0,1]\times[0,1]\times[0,1]$. Dan is

$$\begin{aligned}\int_R f(x,y,z)\,dV &= \int_0^1\left(\int_0^1\left(\int_0^1 \frac{x^2z^3}{1+y^2}\,dx\right)dz\right)dy\\ &= \int_0^1\left(\int_0^1\left(\frac{x^3z^3}{3(1+y^2)}\Big|_0^1\right)dz\right)dy = \int_0^1\left(\int_0^1 \frac{z^3}{3(1+y^2)}\,dz\right)dy\\ &= \int_0^1\left(\frac{z^4}{12(1+y^2)}\Big|_0^1\right)dy = \int_0^1 \frac{1}{12(1+y^2)}\,dy\\ &= \frac{1}{12}\arctan(y)\Big|_0^1 = \frac{1}{12}(\frac{\pi}{4}-0) = \frac{\pi}{48}.\end{aligned}$$

Merk op: In principe is het natuurlijk logisch, dat de eerste integraal $\int_a^b$ met grenzen a en b bij de laatste differentiaal dx hoort, de tweede integraal $\int_c^d$ bij de voorlaatste differentiaal dy enzovoorts. Maar men is vaak iets slordig met de haakjes en ook met de volgorde, en bij ingewikkelde functies wordt de notatie alsnog onoverzichtelijk. Daarom is er een vaak gebruikte conventie, de differentiaal meteen achter de bijhorende integraal te plaatsen om zo duidelijk te maken voor welke integratie variabel de grenzen van dit integraalteken gelden. In plaats van de schrijfwijze hierboven vind je dus ook vaak:

$$\int_{[a,b]\times[c,d]\times[e,f]} f(x,y,z)\,dV = \int_a^b dx\int_c^d dy\int_e^f dz\; f(x,y,z).$$

OPDRACHT 15 *Bepaal voor $f(x,y) := 2xy+3y^2$ de integraal $\int_R f(x,y)\,dA$ voor de rechthoek $R = [a,b]\times[c,d] = \{(x,y) \mid x\in[a,b], y\in[c,d]\}$ door geïtereerde integratie over x en y. Laat zien dat het resultaat niet van de volgorde van de integraties afhangt.*

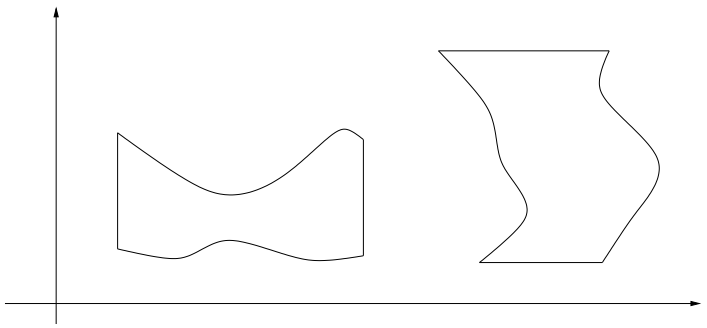
4.2 Integratie over normaalgebieden

Het lijkt natuurlijk erg beperkend als we alleen maar over rechthoek gebieden kunnen integreren. In feite is de beperking niet zo groot, want we kunnen een willekeurig gebied benaderen door een combinatie van kleine rechthoeken en als de onderverdeling voldoende fijn is, kunnen we ervan uit gaan dat de fout die we hierbij maken klein (verwaarloosbaar) is. Maar voor gebieden die alleen maar door krommen begrensd zijn (zo als een cirkel), moeten we hiervoor vaak een redelijk groot aantal rechthoeken bekijken om een redelijke benadering te krijgen, en dit is ook weer een beetje vervelend.

Er is echter een algemenere klasse van gebieden dan rechthoek gebieden, waarvoor we de integraal rechtstreeks kunnen uitrekenen, dit zijn de *normaalgebieden*. In het 2-dimensionale geval zijn dit gebieden die door twee evenredige rechte lijnen evenredig met een van de coördinaatassen en twee elkaar niet snijdende krommen begrensd zijn, zo als de voorbeelden in Figuur I.15.

Een normaalgebied met rechte lijnen evenredig met de y-as is van de vorm

$$B = \{(x,y)\in\mathbb{R}^2 \mid a\le x\le b, \varphi_1(x)\le y\le\varphi_2(x)\},$$



Figuur I.15: Voorbeelden van normaalgebieden.

d.w.z. het gebied is begrensd door de rechte lijnen $x = a$ en $x = b$ (evenredig met de y-as) en de twee krommen $\varphi_1(x)$ en $\varphi_2(x)$.

Net als de integraal over een rechthoek kan de integraal over zo'n normaalgebied door twee in elkaar geschakelde gewone integralen berekend worden:

$$\int_B f(x,y)\ dA = \int_a^b \left(\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\ dy \right) dx.$$

Analoog is een normaalgebied met rechte lijnen evenredig met de x-as begrensd door lijnen $y = c$ en $y = d$ en door twee krommen $\psi_1(y)$ en $\psi_2(y)$. Op deze manier krijgt men het normaalgebied

$$B = \{(x,y) \in \mathbb{R}^2 \mid c \le y \le d, \psi_1(y) \le x \le \psi_2(y)\}$$

en de integraal over dit gebied wordt berekend door

$$\int_B f(x,y)\ dA = \int_c^d \left(\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\ dx \right) dy.$$

Voorbeeld 1: We berekenen de integraal van de functie $f(x,y) := x^2 y$ over de halfcirkel B van straal 1 rond $(0,0)$ die boven de x-as ligt. De halfcirkel B is begrensd door de lijnen $x = -1$ en $x = 1$ en de krommen $\varphi_1(x) = 0$ en $\varphi_2(x) = \sqrt{1-x^2}$. We hebben dus

$$\begin{aligned}
\int_B f(x,y)\ dA &= \int_{-1}^{1} \left(\int_0^{\sqrt{1-x^2}} x^2 y\ dy \right) dx = \int_{-1}^{1} \left(\frac{1}{2} x^2 y^2 \Big|_0^{\sqrt{1-x^2}} \right) dx \\
&= \int_{-1}^{1} \frac{1}{2} x^2 (1-x^2)\ dx = \frac{1}{2} \int_{-1}^{1} (x^2 - x^4)\ dx = \frac{1}{2} \left(\frac{x^3}{3} - \frac{x^5}{5} \right) \Big|_{-1}^{1} \\
&= \frac{1}{2} (\frac{1}{3} - \frac{1}{5} - \frac{-1}{3} + \frac{-1}{5}) = \frac{1}{3} - \frac{1}{5} = \frac{2}{15}.
\end{aligned}$$

Voorbeeld 2: We bepalen de integraal $\int_D x^3 y + \cos(x)\ dA$ op de driehoek D met hoekpunten $(0,0)$, $(\frac{\pi}{2}, 0)$, $(\frac{\pi}{2}, \frac{\pi}{2})$. De driehoek is begrensd door de lijnen

$x = 0$ en $x = \frac{\pi}{2}$ en door de functies $\varphi_1(x) = 0$ en $\varphi_2(x) = x$. Hiermee krijgen we:

$$\begin{aligned}\int_D x^3y + \cos(x)\ dA &= \int_0^{\frac{\pi}{2}} \left(\int_0^x x^3y + \cos(x)\ dy \right) dx \\ &= \int_0^{\frac{\pi}{2}} (\frac{1}{2}x^3y^2 + \cos(x)y)\big|_0^x\ dx = \int_0^{\frac{\pi}{2}} (\frac{1}{2}x^5 + x\cos(x))\ dx \\ &= (\frac{1}{12}x^6 + x\sin(x) + \cos(x))\big|_0^{\frac{\pi}{2}} = \frac{\pi^6}{768} + \frac{\pi}{2} - 1\end{aligned}$$

(merk op dat met partiële integratie geldt dat $\int x\cos(x)\ dx = x\sin(x) - \int \sin(x)\ dx = x\sin(x) + \cos(x)$).

OPDRACHT 16 *Bepaal de integraal $\int_G f(x,y)\ dA$ van de functie $f(x,y) := x + y$ op het gebied G gegeven door $G := \{(x,y) \mid 0 \le x \le 1, 1 \le y \le e^x\}$.*

In drie dimensies zijn normaalgebieden begrensd door een gebied B in het $x-y$-vlak (bijvoorbeeld) en twee functies $\varphi_1(x,y)$ en $\varphi_2(x,y)$, die de variabel z inschakelen. Dan geldt

$$\int_V f(x,y,z)\ dV = \int_B \left(\int_{\varphi_1(x,y)}^{\varphi_2(x,y)} f(x,y,z)\ dz \right) dA.$$

Na het uitwerken van de binnenste integraal over z is dit terug gebracht tot een integratie met twee variabelen op het 2-dimensionale gebied B, en het zou dus handig zijn als B ook weer een normaalgebied is.

OPDRACHT 17 *Laat zien dat het gebied B dat tussen de grafieken van $y = x^2$ en $y = x$ ligt een normaalgebied is en bepaal de oppervlakte van het gebied B. Bereken verder de integraal $\int_B 1 + 2xy\ dA$.*

4.3 Substitutie

Een belangrijke methode in de integratie van gewone functies van één variabel is de substitutie. Het idee hierbij is, de integratievariabel x door een geschikte nieuwe variabel u te vervangen zo dat de integratie makkelijker wordt. Als we in de integraal $\int f(x)\ dx$ de variabel x door een nieuwe variabel u willen vervangen, moeten we de samenhang van x en u kennen, en dit drukken we uit door x te schrijven als een functie $x = x(u)$ van u. Als we nu in de functie $f(x)$ de variabele x door u vervangen, creëren we een nieuwe functie $g(u)$ die gegeven is door $g(u) := f(x(u))$.

De integraal over $f(x)$ kan nu ook met behulp van de nieuwe functie $g(u)$ bepaald worden, en de *substitutieregel* geeft aan dat hiervoor geldt:

$$\int f(x)\ dx = \int g(u)x'(u)\ du = \int f(x(u))x'(u)\ du.$$

Als we ons nu nog eens herinneren dat de integraal gedefinieerd is als de limiet van de som $\sum f(x_i)(x_i - x_{i-1}) = \sum f(x_i)\Delta x$, kunnen we precies de reden

zien, waarom de differentiaal dx door de nieuwe differentiaal $x'(u)du$ vervangen moet worden. In een kleine omgeving van u vervangen we de functie $x(u)$ door de lineaire benadering van de Taylor reeks, dus door de lineaire functie $x(u + \Delta u) \approx x(u) + x'(u)\Delta u$. Maar hieruit volgt dat

$$\Delta x = x(u + \Delta u) - x(u) = x'(u)\Delta u,$$

de afgeleide $x'(u)$ geeft dus juist aan hoe groot de stappen Δx worden waarin we x veranderen als we u in stappen van Δu veranderen.

Als we nu weer naar de limiet $\Delta x \to 0$ kijken, krijgen we de relatie

$$dx = x'(u)\, du$$

tussen de differentialen voor de oorspronkelijke variabel x en de nieuwe variabel u.

De Jacobi matrix

Het idee van de substitutie voor gewone functies gaan we nu veralgemenen op functies van meerdere variabelen. Zij $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ een functie van de n variabelen $x_1, \ldots, x_n$.

Stel we willen nu nieuwe variabelen $u_1, \ldots, u_n$ hanteren, dan hangen de x_i van de nieuwe variabelen u_j af, en we schrijven x_i als functie

$$x_i(\mathbf{u}) = x_i(u_1, \ldots, u_n).$$

Net zo als boven kunnen we nu in een kleine omgeving van $\mathbf{u}$ de functie $x_i(\mathbf{u})$ door de lineaire benadering vervangen, dit geeft

$$x_i(\mathbf{u} + \Delta\mathbf{u}) = x_i(\mathbf{u}) + \nabla x_i(\mathbf{u}) \cdot \Delta\mathbf{u}.$$

Als we de componenten $x_i(\mathbf{u})$ nu in een vector schrijven, krijgen we een functie $\mathbf{x}(\mathbf{u}) : \mathbb{R}^n \to \mathbb{R}^n$ gegeven door

$$\mathbf{x}(\mathbf{u}) := \begin{pmatrix} x_1(\mathbf{u}) \\ \vdots \\ x_n(\mathbf{u}) \end{pmatrix}$$

en als lineaire benadering hiervan krijgen we:

$$\mathbf{x}(\mathbf{u} + \Delta\mathbf{u}) = \begin{pmatrix} x_1(\mathbf{u} + \Delta\mathbf{u}) \\ \vdots \\ x_n(\mathbf{u} + \Delta\mathbf{u}) \end{pmatrix} = \begin{pmatrix} x_1(\mathbf{u}) \\ \vdots \\ x_n(\mathbf{u}) \end{pmatrix} + \begin{pmatrix} \nabla x_1(\mathbf{u})^{tr} \\ \vdots \\ \nabla x_n(\mathbf{u})^{tr} \end{pmatrix} \Delta\mathbf{u}.$$

Maar de matrix $J := \begin{pmatrix} \nabla x_1(\mathbf{u})^{tr} \\ \vdots \\ \nabla x_n(\mathbf{u})^{tr} \end{pmatrix}$ is een oude bekende, in de i-de rij staat namelijk in de j-de kolom de afgeleide van $x_i(\mathbf{u})$ naar de variabel u_j, dus de partiële afgeleide $\frac{\partial x_i}{\partial u_j}$, dus is J juist de *Jacobi matrix* van $\mathbf{x}(\mathbf{u})$.

Merk op: Als de oude variabelen $\mathbf{x} = x_1, \ldots, x_n$ van de nieuwe variabelen $\mathbf{u} = u_1, \ldots, u_n$ afhangen volgens de functies $x_i = x_i(\mathbf{u})$, dan geeft de Jacobi matrix J van de functies x_i aan, hoe (in de lineaire benadering) de stap $\Delta\mathbf{x}$ in de oude variabelen er uit ziet, als we een stap $\Delta\mathbf{u}$ in de nieuwe variabelen zetten. Er geldt namelijk

$$\mathbf{x}(\mathbf{u} + \Delta\mathbf{u}) = \mathbf{x}(\mathbf{u}) + J \cdot \Delta\mathbf{u} \quad \text{en dus} \quad \Delta\mathbf{x} := \mathbf{x}(\mathbf{u} + \Delta\mathbf{u}) - \mathbf{x}(\mathbf{u}) = J \cdot \Delta\mathbf{u}.$$

Dit is volledig analoog met de formule $\Delta x = x'(u)\Delta u$ voor gewone functies, de Jacobi matrix J is dus de veralgemening van de afgeleide $x'(u)$.

Betekenis van de Jacobiaan voor de substitutie

We zullen nu de rol van de Jacobi matrix voor de substitutie van functies van meerdere variabelen toelichten. Hiervoor kijken we eerst naar een functie van twee variabelen, x en y. We kiezen twee nieuwe variabelen u en v en schrijven x en y als functies van u en v, dus $x = x(u, v)$, $y = y(u, v)$. Met behulp van de Jacobi matrix J kunnen we nu de functie

$$(x(u,v), y(u,v)) : \begin{pmatrix} u \\ v \end{pmatrix} \to \begin{pmatrix} x(u,v) \\ y(u,v) \end{pmatrix}$$

in een omgeving van (u, v) door de Taylor veelterm van graad 1 benaderen, voor het verschil van de functiewaarden geldt dan:

$$\begin{aligned} \begin{pmatrix} \Delta x(u,v) \\ \Delta y(u,v) \end{pmatrix} := \begin{pmatrix} x(u+\Delta u, v+\Delta v) - x(u,v) \\ y(u+\Delta u, v+\Delta v) - y(u,v) \end{pmatrix} &= J \cdot \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial x(u,v)}{\partial u} & \frac{\partial x(u,v)}{\partial v} \\ \frac{\partial y(u,v)}{\partial u} & \frac{\partial y(u,v)}{\partial v} \end{pmatrix} \cdot \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}. \end{aligned}$$

Bij de substitutie van functies met één variabel hebben we gezien dat we de differentiaal dx door $x'(u)\,du$ moeten vervangen. De vraag is nu, hoe in het geval van twee variabelen de differentiaal $dA = dx\,dy$ met de nieuwe differentiaal $du\,dv$ samenhangt.

Om hier uit te komen, gaan we even een stap terug en interpreteren de integraal weer als som van pilaren over kleine rechthoeken met zijden $\Delta x, \Delta y$. Zo'n rechthoek moeten we nu door de nieuwe variabelen u en v beschrijven, en bij benadering lukt dit in een punt (x, y) met behulp van de Jacobi matrix door de vergelijkingen

$$\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = J \cdot \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} \quad \text{en} \quad \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix} = J^{-1} \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}.$$

Omdat J en dus ook J^{-1} een lineaire afbeelding is, is het beeld van de rechthoek met zijden $\Delta x, \Delta y$ onder J^{-1} een parallellogram. De vraag is nu wat de oppervlakte van dit parallellogram is. Het antwoord is verrassend eenvoudig, we hebben namelijk alleen maar de determinant van J nodig.

Stelling: De absolute waarde van de determinant $\det(A)$ van een $n \times n$-matrix A geeft het *volume* van het parallellepipedum aan, dat door de kolommen van de matrix A opgespannen wordt.

Met andere woorden is $|\det(A)|$ het volume van het beeld onder A van de eenheidsvierkant (eenheidskubus, eenheidshyperkubus, enz.) die door de standaardbasis opgespannen wordt.

De geïnteresseerde lezer kan deze stelling als volgt inzien:

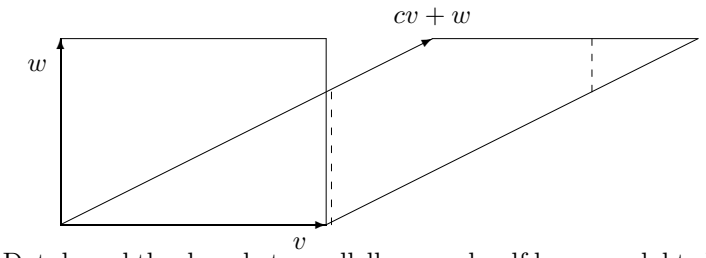
Voor een diagonaalmatrix A is het opspansel van de kolommen van A een rechthoek, blok, enzovoorts, en het volume hiervan is het product van de absolute waarden van de elementen op de diagonaal. Maar dit is ook de absolute waarde van de determinant van de matrix A.

Verder weten we dat we elke matrix door elementaire transformaties op diagonaal vorm kunnen brengen, we moeten dus alleen maar kijken, wat er met het volume gebeurt als we een elementaire transformatie toepassen:

(i) Als we twee kolommen verwisselen, verandert het parallellepipedum niet, het volume blijft dus hetzelfde. De determinant wordt hierbij met -1 vermenigvuldigd, maar de absolute waarde blijft gelijk.

(ii) Als we een kolom met een factor $c \neq 0$ vermenigvuldigen, wordt ook het volume van het parallellepipedum $|c|$ keer zo groot. Maar in dit geval wordt ook de determinant met c vermenigvuldigd.

(iii) Als we een veelvoud van een vector bij een andere optellen, verandert de determinant niet, dus mag ook het volume bij deze transformatie niet veranderen. Omdat hierbij alleen maar twee vectoren een rol spelen, is het voldoende dit in het 2-dimensionale geval te bekijken. De schets hieronder licht dit toe.
De rechthoek opgespannen door de vectoren v en w en het parallellogram opgespannen door v en $cv + w$ hebben dezelfde oppervlakte, omdat de oppervlakte van een parallellogram gelijk is aan het product van de grondzijde en de hoogte.



Dat de rechthoek en het parallellogram dezelfde oppervlakte hebben, laat zich ook door knippen en plakken aantonen, als we het parallellogram langs de twee stippellijnen in stukken snijden, zien we makkelijk in dat de delen de rechthoek precies overdekken.

De stelling hierboven toegepast op de Jacobi matrix betekent dat het parallellogram met zijden $\Delta u, \Delta v$ oppervlakte $|\det(J^{-1})| \cdot \Delta x \Delta y$ heeft en hieruit volgt omgekeerd dat

$$\Delta x \Delta y = |\det(J)| \cdot \Delta u \Delta v.$$

Door nu weer de limieten $\Delta x \to 0$ en $\Delta y \to 0$ te nemen, krijgen we dat voor de differentialen geldt dat

$$dx\,dy = |\det(J)|\,du\,dv.$$

Omdat de determinant van de Jacobi matrix zo'n belangrijke rol speelt, heeft deze ook een eigen naam, ze heet *Jacobiaan*.

Het argument dat we net op twee variabelen hebben toegepast, geldt natuurlijk volledig analoog voor functies van meerdere veranderlijken. We transformeren de variabelen $x_1, \ldots, x_n$ op nieuwe variabelen $u_1, \ldots, u_n$ met $x_i = x_i(u_1, \ldots, u_n)$ en bepalen de Jacobi matrix J met $J_{ij} = \frac{\partial x_i}{\partial u_j}$, dan geldt:

$$\begin{pmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} = J \cdot \begin{pmatrix} \Delta u_1 \\ \vdots \\ \Delta u_n \end{pmatrix}$$

en tussen de n-dimensionale volumes van de blok $\Delta x_1 \ldots \Delta x_n$ en het parallellepipedum $\Delta u_1 \ldots \Delta u_n$ bestaat de relatie

$$\Delta x_1 \ldots \Delta x_n = |\det(J)| \cdot \Delta u_1 \ldots \Delta u_n.$$

Voor de differentialen van de volume elementen geldt dus:

$$dx_1\,dx_2 \ldots dx_n = |\det(J)|\,du_1\,du_2 \ldots du_n.$$

De Jacobiaan speelt dus bij functies van meerdere veranderlijken precies de rol van de afgeleide in het geval van functies van één veranderlijke.

Substitutieregel voor functies van meerdere variabelen

We kunnen nu de substitutieregel voor functies van meerdere veranderlijken formuleren. Voor het gemak doen we dit eerst voor functies van twee veranderlijken en geven dan de algemene regel aan.

Substitutieregel voor twee variabelen: Voor een coördinatentransformatie naar nieuwe variabelen u en v met $x = x(u,v)$ en $y = y(u,v)$ en met Jacobi matrix $J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$ wordt een functie $f(x,y)$ met betrekking tot de nieuwe coördinaten geschreven als $g(u,v)$ met $g(u,v) := f(x(u,v), y(u,v))$.

Voor de integraal van $f(x,y)$ over een gebied $B \subseteq \mathbb{R}^2$ geldt dan in de nieuwe coördinaten:

$$\int_B f(x,y)\,dx\,dy = \int_{B'} g(u,v)\,|\det(J)|\,du\,dv.$$

Hierbij moet het gebied B' in het $u-v$-vlak zo gekozen worden, dat (x,y) over B loopt als (u,v) over B' loopt, waarbij elke punt in B precies een keer voorkomt. Dit betekent dat de afbeelding $\begin{pmatrix} u \\ v \end{pmatrix} \to \begin{pmatrix} x(u,v) \\ y(u,v) \end{pmatrix}$ een bijectieve (omkeerbare) afbeelding van B' naar B is.

In de praktijk spelen vooral speciale coördinatentransformaties een rol die we hieronder gaan bespreken. Bij deze transformaties laat zich de vraag of de afbeelding omkeerbaar is eenvoudig beantwoorden.

Algemeen is de omkeerbaarheid een lastige vraag. Als de Jacobi matrix in een punt een inverteerbare matrix is, is de functie in een kleine omgeving van dit punt omkeerbaar (men noemt de functie dan lokaal inverteerbaar in dit punt). Maar hieruit volgt helaas niet dat de functie globaal omkeerbaar op een gebied B is, er bestaan zelfs functies die in ieder punt van een gebied B lokaal inverteerbaar zijn, maar niet omkeerbaar op B.

Uitgebreid voorbeeld: We bepalen de integraal van $f(x,y) := e^{\frac{y}{x+y}}$ op de driehoek gegeven door $B = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\, 0 \le y \le 1-x\}$. De driehoek is een normaalgebied en in principe kunnen we de integratie opsplitsen in twee in elkaar geschakelde gewone integraties, namelijk

$$\int_B f(x,y)\, dA = \int_0^1 \int_0^{1-x} e^{\frac{y}{x+y}}\, dy\, dx.$$

Het probleem is, dat deze integraal niet zo eenvoudig op te lossen is.

Een slimme transformatie van de variabelen is

$$x + y = u, \quad y = uv, \quad \text{dus } x = u - uv, \quad y = uv.$$

Merk op dat de transformatie juist zo gekozen is dat $e^{\frac{y}{x+y}} = e^{\frac{uv}{u}} = e^v$ wordt.

We gaan na dat (x,y) over B loopt als (u,v) over de eenheidsvierkant $B' = \{(u,v) \in \mathbb{R}^2 \mid 0 \le u \le 1,\, 0 \le v \le 1\}$ loopt: Ten eerste is duidelijk dat $x \ge 0$ en $y \ge 0$ voor $(u,v) \in B'$. Verder is $x = u(1-v) \le 1$, omdat $u \le 1$ en $1 - v \le 1$ zijn. Net zo is $y = uv \le 1$. Ten slotte is $y \le 1 - x \Leftrightarrow uv \le 1 - u + uv \Leftrightarrow u \le 1$, dus geldt ook $y \le 1 - x$.

Omgekeerd moeten we nagaan dat alle punten van B echt doorlopen worden. Maar we kunnen de transformatie expliciet inverteren, er geldt

$$u = x + y \quad \text{en} \quad v = \frac{y}{x+y}$$

en omdat $x + y \le 1$ en $\frac{y}{x+y} \le 1$ voor $x, y \ge 0$ kunnen we voor iedere punt (x,y) een punt (u,v) aangeven, die door de transformatie op (x,y) wordt afgebeeld.

De Jacobi matrix en de Jacobiaan van de transformatie zijn

$$J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} 1-v & -u \\ v & u \end{pmatrix} \quad \text{en} \quad \det(J) = (1-v)u - (-uv) = u.$$

Met de transformatie op de nieuwe variabelen u en v krijgen we dus:

$$\int_B f(x,y)\, dA = \int_0^1 \int_0^1 e^v \cdot u\, du\, dv = \int_0^1 e^v \int_0^1 u\, du\, dv = \int_0^1 e^v \frac{1}{2}\, dv = \frac{1}{2}(e-1).$$

Algemene substitutieregel voor meerdere veranderlijken: We vervangen de coördinaten $x_1, \ldots, x_n$ door nieuwe coördinaten $u_1, \ldots, u_n$ zo dat

$x_i = x_i(u_1, \ldots, u_n)$ een functie van de nieuwe coördinaten wordt en noteren met J de Jacobi matrix van de coördinatentransformatie, d.w.z. $J_{ij} = \frac{\partial x_i}{\partial u_j}$.

Herschrijven van een functie $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ in de nieuwe coördinaten geeft een nieuwe functie

$$g(\mathbf{u}) = g(u_1, \ldots, u_n) := f(x_1(u_1, \ldots, u_n), \ldots, x_n(u_1, \ldots, u_n)).$$

Voor de integraal van $f(\mathbf{x})$ over een gebied $B \subseteq \mathbb{R}^n$ geldt dan met betrekking tot de nieuwe coördinaten:

$$\int_B f(\mathbf{x})\, dx_1 \ldots dx_n = \int_{B'} g(\mathbf{u})\, |\det(J)|\, du_1 \ldots du_n.$$

4.4 Poolcoördinaten, cilindercoördinaten, sferische coördinaten

De belangrijkste toepassingen van substitutie bij functies van meerdere variabelen zijn transformaties tussen verschillende standaard stelsels van coördinaten.

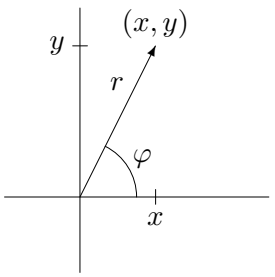
Als functies in het 2-dimensionale vlak alleen maar van de afstand van de oorsprong afhangen, is het vaak handig het probleem op *poolcoördinaten* te transformeren. Hierbij wordt een punt (x, y) door zijn afstand van de oorsprong en door een hoek beschreven.

Ook in de 3-dimensionale ruimte zijn er naast de gewone cartesische coördinaten nog twee andere stelsels coördinaten, die geschikt zijn voor zekere situaties, namelijk de *cilindercoördinaten* en de *sferische coördinaten* (ook *kogelcoördinaten* genoemd).

Poolcoördinaten

Bij functies van twee variabelen zijn vaak *poolcoördinaten* handig, in het bijzonder als het over integratie van functies op ronde gebieden gaat.

Het idee bij de poolcoördinaten is, een punt (x, y) door zijn afstand r van de oorsprong en door de hoek tussen de lijn door de oorsprong en (x, y) en de positieve x-as te beschrijven, zo als in de schets hieronder te zien:



Figuur I.16: Poolcoördinaten

Tussen de gewone coördinaten x, y en de poolcoördinaten r, φ bestaat het volgende verband:

$$x = r\cos(\varphi), \qquad y = r\sin(\varphi)$$
$$r = x^2 + y^2, \qquad \tan(\varphi) = \frac{y}{x}.$$

In het bijzonder wordt een cirkelschijf $B(0,R) := \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq R^2\}$ in poolcoördinaten een rechthoek, namelijk $[0,R] \times [0,2\pi]$.

Voor de partiële afgeleiden van de transformatie $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos(\varphi) \\ r\sin(\varphi) \end{pmatrix}$ geldt:

$$\frac{\partial x}{\partial r} = \cos(\varphi), \quad \frac{\partial x}{\partial \varphi} = -r\sin(\varphi), \quad \frac{\partial y}{\partial r} = \sin(\varphi), \quad \frac{\partial y}{\partial \varphi} = r\cos(\varphi).$$

Hieruit volgt dat de Jacobi matrix J gelijk is aan

$$J = \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) \\ \sin(\varphi) & r\cos(\varphi) \end{pmatrix}$$

en de Jacobiaan $\det(J)$ is dus

$$\det(J) = r\cos^2(\varphi) + r\sin^2(\varphi) = r.$$

Voor een functie $f(x,y)$ en $g(r,\varphi) := f(r\cos(\varphi), r\sin(\varphi))$ geldt dus de substitutieregel

$$\int f(x,y)\, dx\, dy = \int g(r,\varphi)\, r\, dr\, d\varphi.$$

Een eerste toepassing van de poolcoördinaten is natuurlijk het berekenen van de oppervlakte van een cirkel met straal R. Dit kunnen we berekenen door de constante functie $f(x,y) = 1$ over het gebied $B := B(0,R) := \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq R^2\}$ te integreren. Maar in poolcoördinaten wordt B de rechthoek $B' = [0,R] \times [0,2\pi]$, want als (r,φ) over B' loopt, loopt (x,y) precies een keer over B. We hebben dus

$$\begin{aligned} \int_B 1\, dx\, dy &= \int_{B'} r\, dr\, d\varphi = \int_0^{2\pi} \left(\int_0^R r\, dr \right) d\varphi = \int_0^{2\pi} \left(\frac{1}{2} r^2 \Big|_0^R \right) d\varphi \\ &= \int_0^{2\pi} \frac{1}{2} R^2\, d\varphi = \frac{1}{2} R^2 \varphi \Big|_0^{2\pi} = \pi R^2. \end{aligned}$$

OPDRACHT 18 *Bepaal de integraal $\int_B (x^2+2xy)\, dA$ op de halfcirkel $B = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq 1, y \geq 0$. Transformeer de functie (en het gebied) hiervoor op poolcoördinaten. (Herinnering: Met partiële integratie volgt $\int \cos^2(x) = \sin(x)\cos(x) + \int \sin^2(x) = \sin(x)\cos(x) + \int (1-\cos^2(x)) = \frac{1}{2}(\sin(x)\cos(x)+x)$.)*

Toepassing: Normale verdeling

Een iets verrassendere toepassing van poolcoördinaten is dat we nu de integraal $\int_{-\infty}^{\infty} e^{-x^2}\, dx$ over de Gauss-functie kunnen berekenen die we in de normale verdeling tegenkomen en waarvan we tot nu toe de integraal niet analytisch konden bepalen.

Hiervoor bekijken we de analoge functie in twee variabelen, namelijk de functie $f(x,y) := e^{-(x^2+y^2)}$ en integreren deze functie één keer over een cirkel van straal R en één keer over een vierkant met lengte $2a$.

Zij eerst $B(0,R) := \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 \leq R^2\}$ de cirkel met straal R rond 0, dan is

$$\begin{aligned}\int_{B(0,R)} e^{-(x^2+y^2)}\,dx\,dy &= \int_0^R \left(\int_0^{2\pi} d\varphi\right) e^{-r^2} r\,dr = 2\pi \int_0^R e^{-r^2} r\,dr \\ &= 2\pi(-\frac{1}{2}e^{-r^2})\Big|_0^R = -\pi(e^{-R^2}-1) = \pi(1-e^{-R^2})\end{aligned}$$

Zij nu $V(-a,a) := \{(x,y) \in \mathbb{R}^2 \mid |x| \leq a, |y| \leq a\}$ het vierkant met lengte $2a$ rond 0, dan is

$$\begin{aligned}\int_{V(-a,a)} e^{-(x^2+y^2)}\,dx\,dy &= \int_{-a}^a \left(\int_{-a}^a e^{-(x^2+y^2)}dx\right) dy = \int_{-a}^a \left(\int_{-a}^a e^{-x^2}dx\right) e^{-y^2}\,dy \\ &= (\int_{-a}^a e^{-x^2}\,dx)(\int_{-a}^a e^{-y^2}\,dy) = (\int_{-a}^a e^{-x^2}\,dx)^2\end{aligned}$$

Maar de cirkel $B(0,a)$ van straal a ligt volledig in het vierkant $V(-a,a)$ en dit ligt wederom volledig in de cirkel $B(0,\sqrt{2}\,a)$ met straal $\sqrt{2}\,a$. Omdat de functie $e^{-(x^2+y^2)} > 0$ is, volgt hieruit

$$\pi(1-e^{-a^2}) \leq (\int_{-a}^a e^{-x^2}\,dx)^2 \leq \pi(1-e^{-2a^2}).$$

Als we nu de limiet $a \to \infty$ laten lopen, gaat $e^{-a^2} \to 0$ en $e^{-2a^2} \to 0$. Maar dan wordt de rechter- en de linkerzijde van deze ongelijkheden gelijk aan π, en dus hebben we bewezen dat

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

Cilindercoördinaten

In de 3-dimensionale ruimte komt het vaak voor dat een probleem symmetrisch ten opzichte van een rotatie as is. Dit is bijvoorbeeld het geval voor het elektrische veld rond een rechte geleider. Bij dit soort problemen zijn *cilindercoördinaten* heel praktisch, die veronderstellen dat de rotatie-as de z-as is. Het idee van de cilindercoördinaten is, een punt (x,y,z) te beschrijven door poolcoördinaten voor het $x-y$-vlak en de gewone z-coördinaat.

Dit geeft:

$$x = r\cos(\varphi), \qquad y = r\sin(\varphi), \qquad z = z,$$

waarbij $r > 0$ en $\varphi \in [0, 2\pi)$. De Jacobi matrix J hiervan is

$$J = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial z} \end{pmatrix} = \begin{pmatrix} \cos(\varphi) & -r\sin(\varphi) & 0 \\ \sin(\varphi) & r\cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

en de Jacobiaan is

$$\det(J) = r\cos^2(\varphi) + r\sin^2(\varphi) = r.$$

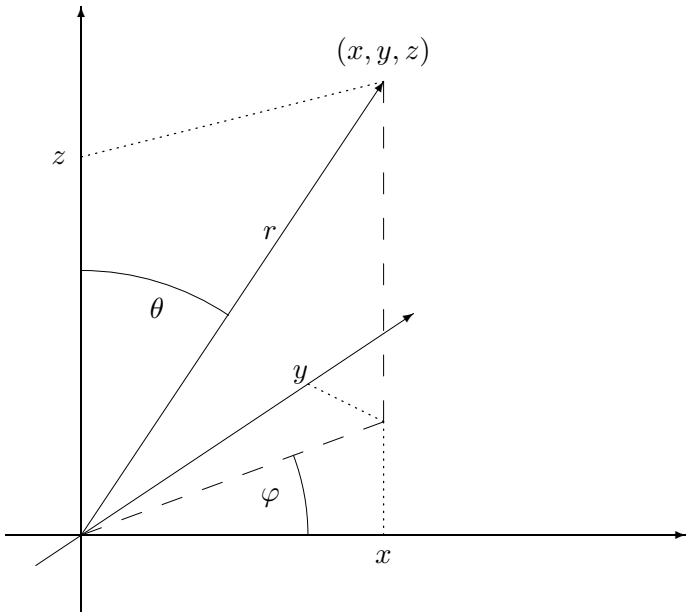
Hieruit volgt voor een functie $f(x,y,z)$ en $g(r,\varphi,z) := f(r\cos(\varphi), r\sin(\varphi), z)$:

$$\int f(x,y,z)\,dx\,dy\,dz = \int g(r,\varphi,z)\,r\,\,dr\,d\varphi\,dz.$$

Sferische coördinaten

Bij functies op de 3-dimensionale ruimte die eigenlijk alleen maar van de afstand van een punt afhangen (zo als de gravitatie kracht of de intensiteit van een geïdealiseerde bron van licht) worden vaak *sferische coördinaten* toegepast. Het idee is, een punt door zijn afstand en twee ruimtelijke hoeken aan te geven.

Men splitst de vector van de oorsprong naar het punt (x, y, z) in zijn projecties in het $x - y$-vlak en op de z-as. De projectie in het $x - y$-vlak wordt door poolcoördinaten r en φ aangegeven en de projectie op de z-as met behulp van de hoek θ tussen (x, y, z) en de z-as (zie de schets in Figuur I.17).



Figuur I.17: Sferische coördinaten

De coördinatentransformatie luidt:

$$x = r\cos(\varphi)\sin(\theta), \qquad y = r\sin(\varphi)\sin(\theta), \qquad z = r\cos(\theta)$$

waarbij $r > 0$, $\varphi \in [0, 2\pi)$, $\theta \in [0, \pi]$.

De Jacobi matrix J hiervan is

$$J = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial \theta}{\partial r} & \frac{\partial \theta}{\partial \varphi} & \frac{\partial \theta}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos(\varphi)\sin(\theta) & -r\sin(\varphi)\sin(\theta) & r\cos(\varphi)\cos(\theta) \\ \sin(\varphi)\sin(\theta) & r\cos(\varphi)\sin(\theta) & r\sin(\varphi)\cos(\theta) \\ \cos(\theta) & 0 & -r\sin(\theta) \end{pmatrix}$$

en voor de Jacobiaan $\det(J)$ krijgt men in dit geval

$$\begin{aligned}\det(J) &= -r^2\cos^2(\varphi)\sin^3(\theta) - r^2\sin^2(\varphi)\cos^2(\theta)\sin(\theta)\\ &\quad - r^2\cos^2(\varphi)\cos^2(\theta)\sin(\theta) - r^2\sin^2(\varphi)\sin^3(\theta)\\ &= -r^2\sin^3(\theta) - r^2\cos^2(\theta)\sin(\theta)\\ &= -r^2\sin(\theta).\end{aligned}$$

Omdat $\sin(\theta) > 0$ is $|\det(J)| = r^2\sin(\theta)$ en dus

$$\int f(x,y,z)\; dx\, dy\, dz = \int g(r,\varphi,\theta)\, r^2\sin(\theta)\; dr\, d\varphi\, d\theta,$$

waarbij $g(r,\varphi,\theta) := f(r\cos(\varphi)\sin(\theta), r\sin(\varphi)\sin(\theta), r\cos(\theta))$.

> Een alternatieve versie van de sferische coördinaten gebruikt voor de hoek θ in plaats van de hoek tussen (x,y,z) en de z-as de hoek tussen (x,y,z) en het $x-y$-vlak. Dit geeft
>
> $$x = r\cos(\varphi)\cos(\theta), \quad y = r\sin(\varphi)\cos(\theta), \quad z = r\sin(\theta),$$
>
> waarbij $r > 0$, $\varphi \in [0, 2\pi)$, $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.
> In dit geval wordt $dx\, dy\, dz = r^2\cos(\theta)\; dr\, d\varphi\, d\theta$.

De eenvoudigste toepassing van sferische coördinaten is het bepalen van het volume V van een kogel $B := B(0,R)$ van straal R. De functie $f(x,y,z)$ is in dit geval $f(x,y,z) = 1$, dus hebben we

$$\begin{aligned}V &= \int_B 1\; dx\, dy\, dz = \int_0^{\pi}\int_0^{2\pi}\int_0^{R} r^2\sin(\theta)\; dr\, d\varphi\, d\theta\\ &= \int_0^{\pi}\int_0^{2\pi}\frac{1}{3}R^3\sin(\theta)\; d\varphi\, d\theta = \int_0^{\pi}\frac{2\pi}{3}R^3\sin(\theta)\; d\theta = \frac{2\pi}{3}R^3(-\cos(\theta))\Big|_0^{\pi}\\ &= \frac{4\pi}{3}R^3.\end{aligned}$$

4.5 Toepassingen

Oppervlaktes, volumes

Een belangrijke toepassing voor integralen over functies van meerdere variabelen is het bepalen van oppervlaktes en volumes. Voorbeelden hiervan hebben we al gezien, namelijk de oppervlakte van een cirkel en het volume van een kogel. De manier van aanpak is steeds dezelfde: Men integreert de constante functie die overal de waarde 1 heeft over het gebied waarvan men de oppervlakte of het volume wil bepalen. De kunst ligt hierbij meestal niet zo zeer in de integratie, maar in het beschrijven van het gebied. Soms is het mogelijk een gebied in meerdere delen te splitsen die als normaalgebieden te beschrijven zijn en vaak helpt een geschikte keuze van nieuwe coördinaten.

Vaak is het ook handig één van de standaard coördinatentransformaties te combineren met een verdere substitutie. Een voorbeeld hiervoor is het berekenen van de oppervlakte van een ellips.

Zij E een ellips rond het nulpunt $(0,0)$ met hoofdassen van lengte a en b in de richtingen van de x-as en de y-as, dan wordt E beschreven door:

$$E = \{(x,y) \in \mathbb{R}^2 \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\}.$$

Door de transformatie op nieuwe coördinaten u, v met $x = au$ en $y = bv$ wordt E op de eenheidscirkel $B(0,1)$ getransformeerd, want $\frac{x^2}{a^2}+\frac{y^2}{b^2} \leq 1 \Leftrightarrow \frac{(au)^2}{a^2}+\frac{(bv)^2}{b^2} = u^2+v^2 \leq 1$. De Jacobi matrix voor deze substitutie is heel eenvoudig, we hebben $J = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ en dus $\det(J) = ab$. Hieruit volgt dat

$$\int_E 1 \; dx\, dy = \int_{B(0,1)} ab \; du\, dv = ab \int_{B(0,1)} 1 \; du\, dv = \pi ab.$$

Zwaartepunten

Een verdere toepassing van meerdimensionale integralen is het berekenen van zwaartepunten van objecten. Het zwaartepunt is een soort gemiddelde van het object en in drie dimensies kunnen we de coördinaten (x_s, y_s, z_s) van het zwaartepunt van een object B met volume V berekenen als $x_s = \frac{1}{V}\int_B x \; dx\, dy\, dz$ $y_s = \frac{1}{V}\int_B y \; dx\, dy\, dz$ $z_s = \frac{1}{V}\int_B z \; dx\, dy\, dz$ waarbij we veronderstellen dat de dichtheid van het object overal hetzelfde is.

Maar we kunnen het zwaartepunt ook berekenen als de dichtheid niet constant is, maar een functie $\rho(x,y,z)$. De massa van B berekenen we door $M = \int_B \rho(x,y,z) \; dx\, dy\, dz$, dus geeft de functie $\frac{1}{M}\int_B \rho(x,y,z) \; dx\, dy\, dz$ de verdeling van de massa over B aan. Deze verdelingsfunctie moeten we nu gewoon in de integralen invullen en krijgen zo $x_s = \frac{1}{M}\int_B x\rho(x,y,z) \; dx\, dy\, dz$ $y_s = \frac{1}{M}\int_B y\rho(x,y,z) \; dx\, dy\, dz$ $z_s = \frac{1}{M}\int_B z\rho(x,y,z) \; dx\, dy\, dz$ Het speciaal geval voor constante dichtheid $\rho(x,y,z) = \rho$ volgt hieruit met $V = \int_B dx\, dy\, dz$, omdat dat $M = \int_B \rho \; dx\, dy\, dz = \rho \cdot V$, dus $\frac{\rho}{M} = \frac{1}{V}$.

> In het kader van de kansrekening is het taalgebruik iets anders en het zwaartepunt pakt uit als een oude bekende. Als $f(x,y)$ of $f(x,y,z)$ de dichtheidsfunctie van een meerdimensionale kansverdeling is, heet het zwaartepunt namelijk de *verwachtingswaarde* van de kansverdeling. De dichtheidsfunctie speelt precies de rol van de functie $\frac{1}{M}\rho(x,y,z)$ voor de verdeling van de massa, want de integraal over het hele gebied is gelijk aan 1.

Als voorbeeld berekenen we het zwaartepunt van een halfkogel H met straal R rond het nulpunt $(0,0,0)$ die boven het $x-y$-vlak ligt. We gaan van constante dichtheid uit. Een halfkogel beschrijven we het makkelijkste met sferische coördinaten, we moeten alleen maar over de hoek θ tussen (x,y,z) en de z-as nadenken. Bij een volle kogel loopt die van 0 tot π en voor punten in het

$x-y$-vlak is $\theta = \frac{\pi}{2}$, dus loopt θ nu van 0 tot $\frac{\pi}{2}$. Uit symmetrie redenen is het duidelijk dat het zwaartepunt op de z-as moet liggen, daarom hoeven we alleen maar de z-coördinaat uit te rekenen. Omdat een volle kogel van straal R het volume $\frac{4}{3}\pi R^3$ heeft, heeft H het volume $V = \frac{2}{3}\pi R^3$. Er geldt:

$$\begin{aligned} z_s &= \frac{1}{V}\int_H z\ dx\,dy\,dz = \frac{1}{V}\int_0^R \int_0^{2\pi} \int_0^{\frac{\pi}{2}} r\cos(\theta)\, r^2 \sin(\theta)\ d\theta\, d\varphi\, dr \\ &= \frac{1}{V}\int_0^R \int_0^{2\pi} r^3 (\frac{1}{2}\sin^2(\theta)\big|_0^{\frac{\pi}{2}})\ d\varphi\, dr = \frac{1}{V}\int_0^R \int_0^{2\pi} \frac{1}{2} r^3\ d\varphi\, dr \\ &= \frac{1}{V}\int_0^R \pi r^3\ dr = \frac{1}{V}\frac{1}{4}\pi r^4\big|_0^R = \frac{\pi}{4V}R^4 = \frac{\pi}{4\frac{2}{3}\pi R^3}R^4 = \frac{3}{8}R. \end{aligned}$$

Hetzelfde voorbeeld kunnen we ook in cilindercoördinaten uitwerken. In het $x-y$-vlak loopt de straal r dan van 0 tot R, de hoek φ van 0 tot 2π en de z-variabel loopt van 0 tot $\sqrt{R^2-r^2}$. Hiermee krijgen we:

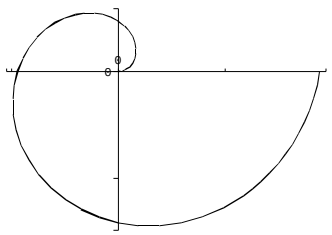
$$\begin{aligned} z_s &= \frac{1}{V}\int_H z\ dx\,dy\,dz = \frac{1}{V}\int_0^R \int_0^{2\pi} \int_0^{\sqrt{R^2-r^2}} z\, r\ dz\, d\varphi\, dr \\ &= \frac{1}{V}\int_0^R \int_0^{2\pi} (\frac{1}{2}z^2\big|_0^{\sqrt{R^2-r^2}})\, r\ d\varphi\, dr = \frac{1}{V}\int_0^R \int_0^{2\pi} \frac{1}{2}(R^2-r^2) r\ d\varphi\, dr \\ &= \frac{1}{V}\int_0^R \pi(R^2-r^2) r\ dr = \frac{\pi}{V}\int_0^R (rR^2 - r^3)\ dr = \frac{\pi}{V}(\frac{1}{2}r^2R^2 - \frac{1}{4}r^4)\big|_0^R \\ &= \frac{\pi}{V}\frac{1}{4}R^4 = \frac{3}{8}R. \end{aligned}$$

Belangrijke begrippen in deze les

- integratie van functies van meerdere variabelen
- geïtereerde integratie
- integratie over rechthoek gebieden
- integratie over normaalgebieden
- Jacobi matrix, Jacobiaan
- substitutie voor functies van meerdere variabelen
- coördinatentransformatie
- poolcoördinaten
- cilindercoördinaten, sferische coördinaten

Opgaven

33. Bereken de volgende 2-dimensionale integralen:
    (i) $\int_B (xy + y^2)\, dA$ met $B = [0,1] \times [0,1]$,
    (ii) $\int_B \sin(x+y)\, dA$ met $B = [0, \frac{\pi}{2}] \times [0, \frac{\pi}{2}]$,
    (iii) $\int_B (x + y^2)\, dA$, waarbij B de driehoek met hoekpunten $(0,0)$, $(1,0)$ en $(0,1)$ is.
    (iv) $\int_B (x^2 + y^2)\, dA$, waarbij B de driehoek met hoekpunten $(0,0)$, $(1,0)$ en $(\frac{1}{2}, \frac{1}{2})$ is.
34. Zij B het gebied tussen de grafieken van $\varphi_1(x) := x^3$ en $\varphi_2(x) := x^2$. Bereken de integralen $\int_B x\, dA$ en $\int_B y\, dA$.
35. Bereken de integraal $\int_B \frac{\sin(x)}{x}\, dA$ voor de driehoek B met hoekpunten $(0,0)$, $(1,0)$ en $(1,1)$. Let op dat hierbij de volgorde van de integraties een rol speelt, want de integraal $\int \frac{\sin(x)}{x}\, dx$ laat zich niet zonder integraal teken schrijven.
36. Beschrijf het gebied $B := \{(x,y) \in \mathbb{R}^2 \mid 3y \leq x \leq 3, 0 \leq y \leq 1\}$. en bereken de integraal $\int_B e^{x^2}\, dA$. Dit lukt helaas alleen maar voor een van de twee mogelijke volgordes van integratie.
37. Bereken de oppervlakte van het gebied tussen de Archimedische spiraal (zie Figuur I.18) gegeven door $r = a\varphi$, $0 \leq \varphi \leq 2\pi$ en de x-as. Merk op dat de spiraal in poolcoördinaten aangegeven is.



Figuur I.18: Archimedische spiraal

38. Bereken op het gebied $B := \{(x,y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1, x^2 \leq y \leq \sqrt{x}\}$ de integraal $\int_B (x^2 + y)\, dx\, dy$.
39. Bepaal de integraal $\int_0^\infty \int_0^\infty e^{-(x^2+y^2)} x^2\, dA$ met behulp van een transformatie op poolcoördinaten.
40. Bereken het volume van de (onregelmatige) tetraëder die begrensd is door de drie coördinaatvlakken $x = 0$, $y = 0$ en $z = 0$ en het vlak met $z = 2 - 2x - y$.
41. Een cirkelvormige boor van straal R snijdt uit een kogel van straal $2R$ een cilinder langs de z-as uit. Wat is het volume van de cilinder?
42. Een halfkogel H van straal R die op het $x-y$-vlak ligt heeft een niet constante dichtheidsfunctie, de dichtheid hangt namelijk af van de afstand van het grondvlak: $\rho(x,y,z) = az$ voor een $a > 0$. Bereken het zwaartepunt van de halfkogel.

Les 5 Complexe getallen

Iedereen weet, dat kwadraten van getallen positieve getallen zijn. Dat is vaak erg praktisch, we weten bijvoorbeeld dat de functie $f(x) := x^2+1$ steeds positief is en in het bijzonder geen nulpunten heeft. Daarom is bijvoorbeeld ook de functie $f(x) := \frac{1}{x^2+1}$ voor alle waarden van x gedefinieerd, omdat de noemer nooit 0 wordt.

Aan de andere kant is het feit, dat kwadraten positief zijn, ook een bron van frustratie, we kunnen namelijk vergelijkingen van de vorm $X^2 = a$ voor $a < 0$ niet oplossen.

Nu is het een typische eigenschap van wiskundigen, dat ze in een voor gewone mensen hopeloze situatie (een situatie zonder oplossing) toch vooruit gaan: ze definiëren gewoon iets, waarmee ze verder kunnen.

> Er zijn mensen die beweren dat wiskundigen mensen zijn die geen weg weten met de reële wereld en zich daarom hun eigen wereld definiëren waarin ze zich thuis voelen.

In sommige gevallen zijn de nieuw gedefinieerde objecten misschien niet zo erg nuttig, maar in het geval van de oplossingen van kwadratische vergelijkingen bleek de definitie van de *complexe getallen* een echt succesverhaal te zijn.

5.1 Constructie van de complexe getallen

Het idee achter de *complexe getallen* is dat we een oplossing van de vergelijking $X^2 = -1$ definiëren en kijken wat er gebeurt als we deze oplossing aan de reële getallen $\mathbb{R}$ toevoegen.

Definitie: We noteren de (symbolische) oplossing van de kwadratische vergelijking $X^2 = -1$ met de letter i en noemen i de *imaginaire eenheid*. Voor de imaginaire eenheid i geldt dus dat $i^2 = -1$.

> **Let op:** In de wiskunde staat steeds het symbool i voor de imaginaire eenheid. Maar omdat in de natuurkunde en elektrotechniek traditioneel de letter I (en vroeger ook i) voor de *stroomsterkte* gebruikt wordt, wordt in deze disciplines meestal j voor de imaginaire eenheid geschreven.

Wat betekent het nu dat we i aan de reële getallen *toevoegen*? We willen zeker dat we i met een willekeurig reëel getal y kunnen vermenigvuldigen, dit geeft getallen van de vorm $i \cdot y$.

Het aardige is dat we hiermee al uit ieder reëel getal de wortel kunnen trekken, want voor $a \geq 0$ konden we dit al eerder en voor $a < 0$ is $-a > 0$, dus bestaat er een $b \in \mathbb{R}$ met $b^2 = -a$ en we hebben $(i \cdot b)^2 = i^2 \cdot b^2 = (-1) \cdot (-a) = a$.

Maar we willen getallen natuurlijk ook optellen, daardoor krijgen we alle getallen van de vorm $x + i \cdot y$ met $x, y \in \mathbb{R}$. De positieve verrassing ligt in het

feit, dat dit ook voldoende is, d.w.z. dat we geen verdere getallen meer nodig hebben om goed met i te kunnen rekenen:

- Het *optellen* van getallen van de vorm $x + i \cdot y$ gebeurt componentsgewijs, waarbij we x en y als de componenten van het getal beschouwen, dus

$$(x_1 + i \cdot y_1) + (x_2 + i \cdot y_2) = (x_1 + x_2) + i \cdot (y_1 + y_2).$$

- Voor het *vermenigvuldigen* moeten we het product van twee getallen gewoon uitwerken, waarbij we natuurlijk gebruik ervan maken dat $i^2 = -1$ is:

$$\begin{aligned}(x_1 + i \cdot y_1)(x_2 + i \cdot y_2) &= x_1x_2 + i \cdot (x_1y_2 + y_1x_2) + i^2 \cdot y_1y_2 \\ &= (x_1x_2 - y_1y_2) + i \cdot (x_1y_2 + y_1x_2).\end{aligned}$$

Voorbeeld: We hebben $(1 + i \cdot 2) + (3 - i) = (1 + 3) + i \cdot (2 - 1) = 4 + i$ en $(1 + i \cdot 2)(3 - i) = (3 - (-2)) + i \cdot (-1 + 6) = 5 + i \cdot 5$.

Definitie: De verzameling

$$\mathbb{C} := \{x + i \cdot y \mid x, y \in \mathbb{R}\}$$

met de bewerkingen

$$\begin{aligned}(x_1 + i \cdot y_1) + (x_2 + i \cdot y_2) &= (x_1 + x_2) + i \cdot (y_1 + y_2) \\ (x_1 + i \cdot y_1)(x_2 + i \cdot y_2) &= (x_1x_2 - y_1y_2) + i \cdot (x_1y_2 + y_1x_2).\end{aligned}$$

heet het *lichaam* $\mathbb{C}$ *der complexe getallen.*

Notatie: Net zo goed als $1 + i \cdot 2$ kunnen we natuurlijk ook $1 + 2i$ schrijven en in feite is $x + yi$ in het algemeen de gebruikelijkere schrijfwijze dan $x + i \cdot y$.

We mogen de verzameling $\mathbb{C}$ een *lichaam* noemen, omdat optellen en vermenigvuldigen commutatief ($a + b = b + a$ en $ab = ba$) en associatief ($(a + b) + c = a + (b + c)$ en $(ab)c = a(bc)$) zijn en omdat de vermenigvuldiging distributief over het optellen is ($a(b + c) = ab + ac$ en $(a + b)c = ac + bc$). Deze eigenschappen erven de complexe getallen gewoon van de reële getallen.

In een lichaam moet verder nog gelden, dat we door ieder element $\neq 0$ kunnen delen. Ook aan deze eigenschap voldoen de complexe getallen, want er geldt

$$\frac{1}{x + yi} = \frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2}\, i.$$

Waar deze formule vandaan komt, zullen we pas iets later in deze les toelichten, maar we kunnen wel controleren dat het product van $x + yi$ en $\frac{1}{x+yi}$ inderdaad 1 is:

$$\begin{aligned}(x + yi) \cdot \left(\frac{x}{x^2 + y^2} - \frac{y}{x^2 + y^2}\, i\right) &= \frac{x^2}{x^2 + y^2} - \frac{xy}{x^2 + y^2}\, i + \frac{yx}{x^2 + y^2}\, i - \frac{y^2}{x^2 + y^2}\, i^2 \\ &= \frac{x^2 + y^2}{x^2 + y^2} = 1\end{aligned}$$

Omdat we de complexe getallen verkregen hebben door i aan de reële getallen toe te voegen, zijn de reële getallen in de complexe getallen bevat, namelijk als de getallen van de vorm $x + i \cdot 0$ met $x \in \mathbb{R}$. Aan de andere kant noemt men de getallen van de vorm $i \cdot y$ met $y \in \mathbb{R}$ *zuiver imaginair*.

We hebben al gezien dat een complex getal $z \in \mathbb{C}$ eenduidig door twee reële getallen beschreven wordt, namelijk door $x, y \in \mathbb{R}$ zo dat $z = x + i \cdot y$. Voor de twee componenten x en y van een complex getal zijn de volgende notaties gebruikelijk:

Definitie: De eerste component x van $z = x + i \cdot y$ heet het *reële deel* van z, genoteerd met $x = \Re(z)$ of $x = \mathrm{Re}(z)$ en de tweede component y heet het *imaginaire deel* van z, genoteerd met $y = \Im(z)$ of $y = \mathrm{Im}(z)$. Er geldt dus:

$$z = \Re(z) + i \cdot \Im(z).$$

> Het is gebruikelijk dat complexe getallen of complexe variabelen z heten, terwijl reële getallen x en y heten. Dit is natuurlijk geen garantie, maar als je de letter z in een formule tegen komt, is dit een sterk signaal dat je het misschien met complexe getallen te maken hebt.

Opdracht 19 *Schrijf de volgende complexe getallen in de standaard vorm $z = x + yi$:*

(i) $(1 + 2i) - 3(5 - 2i)$;

(ii) $(4 - 3i)(8 + i) + 5 - i$;

(iii) $(2 + i)^2$.

5.2 Oplossen van vergelijkingen

We hebben boven al gezien, dat we met behulp van de getallen $i \cdot y$ uit elk reëel getal de wortel kunnen trekken. Dit betekent, dat elke kwadratische veelterm $f(x) = ax^2 + bx + c$ een nulpunt heeft, want de *abc*-formule

$$x_{1,2} = -\frac{b}{2a} \pm \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}}$$

geeft de nulpunten expliciet aan en we hoeven alleen maar de wortel uit het reële getal $\frac{b^2}{4a^2} - \frac{c}{a}$ te trekken.

Maar de situatie is nog veel beter, we kunnen namelijk zelfs uit een willekeurig *complex getal* de wortel trekken: Voor een gegeven complex getal $a + i \cdot b$ is een complex getal $z = x + i \cdot y$ gezocht, waarvoor geldt dat $z^2 = a + i \cdot b$.

Uit $z^2 = (x^2 - y^2) + i \cdot (xy + yx) = (x^2 - y^2) + i \cdot 2xy$ volgt $x^2 - y^2 = a$ en $2xy = b$. Hieruit berekenen we $a^2 + b^2 = x^4 - 2x^2y^2 + y^4 + 4x^2y^2 = x^4 + 2x^2y^2 + y^4 = (x^2 + y^2)^2$, dus hebben we $x^2 + y^2 = \sqrt{a^2 + b^2}$ (merk op dat $a^2 + b^2$ positief is).

Door de vergelijkingen $x^2 + y^2 = \sqrt{a^2 + b^2}$ en $x^2 - y^2 = a$ bij elkaar op te tellen en van elkaar af te trekken krijgen we

$$x^2 = \frac{1}{2}\sqrt{a^2 + b^2} + \frac{1}{2}a \quad \text{en} \quad y^2 = \frac{1}{2}\sqrt{a^2 + b^2} - \frac{1}{2}a$$

en omdat $\sqrt{a^2+b^2} \geq |a|$ hebben deze vergelijkingen reële oplossingen x en y.

We moeten wel opletten of we voor x en y de positieve of de negatieve wortel kiezen, want er moet gelden dat $2xy = b$. Voor $b \geq 0$ moeten we bij x en y hetzelfde teken kiezen (beide positief of beide negatief), voor $b < 0$ moeten x en y verschillende tekens hebben.

Het zal geen verrassing zijn dat we in ieder geval twee oplossingen vinden, want met $z^2 = a + bi$ geldt ook $(-z)^2 = a + bi$.

Uit de *abc*-formule volgt nu rechtstreeks dat elke kwadratische veelterm met coëfficiënten in $\mathbb{C}$ ook een nulpunt in $\mathbb{C}$ heeft, of anders gezegd, dat elke kwadratische vergelijking een oplossing in $\mathbb{C}$ heeft. Maar er geldt een veel sterker resultaat, namelijk de

Hoofdstelling van de algebra (I): Elke veelterm met coëfficiënten in $\mathbb{C}$ heeft een nulpunt in $\mathbb{C}$.

Als een veelterm $f(z)$ een nulpunt a_1 heeft, dan kunnen we (met behulp van een staartdeling) $f(z)$ schrijven als $f(z) = (z - a_1)g(z)$ waarbij $g(z)$ een veelterm van lagere graad is. Maar ook $g(z)$ heeft volgens de hoofdstelling van de algebra een nulpunt a_2, en dus kunnen we doorgaan en $f(z)$ schrijven als $f(z) = (z - a_1)(z - a_2)h(z)$ waarbij de graad van $h(z)$ al om 2 kleiner is dan die van $f(z)$.

Uiteindelijk kunnen we een veelterm $f(z) = c_n z^n + c_{n-1}z^{n-1} + c_1 z + c_0$ op deze manier schrijven als $f(z) = c_n(z - a_1)(z - a_2)\dots(z - a_n)$, waarbij de a_i de (niet noodzakelijk verschillende) nulpunten van $f(z)$ zijn. Omdat $z - a$ een lineaire functie is hebben we zo de volgende variatie van de hoofdstelling van de algebra ingezien:

Hoofdstelling van de algebra (II): Elke veelterm met coëfficiënten in $\mathbb{C}$ laat zich (over $\mathbb{C}$) schrijven als een product van lineaire factoren.

Merk op: Over de reële getallen geldt slechts de zwakkere uitspraak: Elke veelterm met coëfficiënten in $\mathbb{R}$ laat zich (over $\mathbb{R}$) schrijven als een product van lineaire en kwadratische factoren.

Opdracht 20 *Bepaal de (complexe) nulpunten van de volgende veeltermen:*

(i) $z^2 - 2z + 5$;

(ii) $z^2 + \frac{1}{3}z + \frac{1}{2}$;

(iii) $z^3 + 2z^2 + 2z + 1$ *(een nulpunt is makkelijk te gokken).*

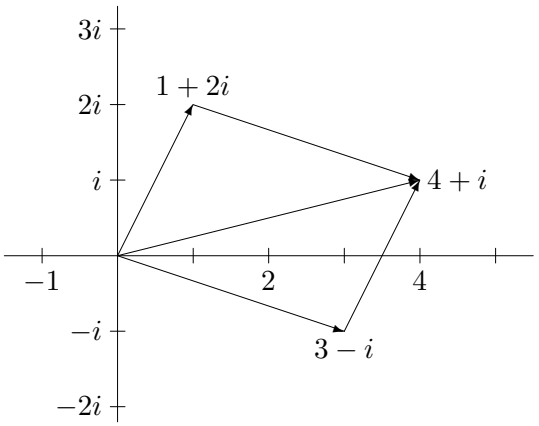
5.3 Meetkunde van de complexe getallen

We hebben ons tot nu toe tot algebraïsche eigenschappen van de complexe getallen beperkt, maar een belangrijke rol spelen ook de meetkundige eigenschappen. We hebben gezien, dat een reëel getal via het reële en imaginaire deel met een paar van reële getallen correspondeert. Dit geeft een identificatie van de complexe getallen met het gewone 2-dimensionale vlak $\mathbb{R}^2$, het getal $z = x + i \cdot y$ correspondeert hierbij met het punt (x, y) en op grond van deze correspondentie spreekt men ook vaak van het *complexe vlak* in plaats van de complexe getallen.

We hebben al gezien dat het optellen van complexe getallen componentsgewijs voor het reële en imaginaire deel gebeurt. Maar dat is precies de manier hoe we vectoren optellen en daarom is het redelijk voor de hand liggend het getal $z = x + i \cdot y$ met de 2-dimensionale vector $\begin{pmatrix} x \\ y \end{pmatrix}$ te identificeren. Het voordeel ervan, bij complexe getallen aan vectoren in plaats van punten te denken, is dat we van vectoren weten hoe we ze optellen, terwijl we hiervoor bij punten toch stiekem weer vectoren zouden gebruiken.

In de taal van de lineaire algebra zeggen we, dat de complexe getallen $\mathbb{C}$ een 2-dimensionale $\mathbb{R}$-vectorruimte vormen, en de boven aangegeven correspondentie met $\mathbb{R}^2$ vinden we door de standaardbasis $\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$ van $\mathbb{R}^2$ met de basis $(1, i)$ van $\mathbb{C}$ te identificeren.

Als voorbeeld geeft het plaatje in Figuur I.19 het optellen van de complexe getallen $1 + 2i$ en $3 - i$ in het complexe vlak weer.



Figuur I.19: Optellen in het complexe vlak

Een voor de hand liggende vraag is nu natuurlijk, of ook de vermenigvuldiging van complexe getallen een mooie meetkundige interpretatie heeft. Dit is inderdaad het geval, maar het verhaal is iets ingewikkelder dan voor het optellen.

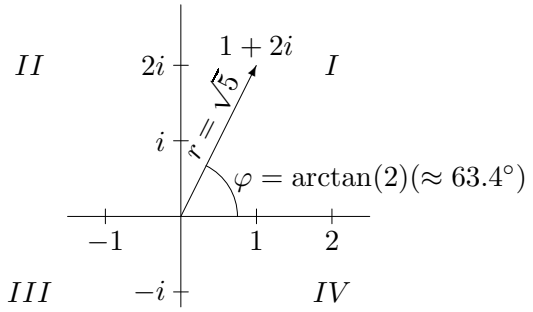
Poolcoördinaten

Om te beginnen, hebben we hiervoor en andere manier van beschrijving van punten in het 2-dimensionale vlak nodig, namelijk de *poolcoördinaten* die we ook bij de integratie van functies van meerdere veranderlijken al zijn tegengekomen.

Elk punt P in het vlak $\mathbb{R}^2$ kan behalve door zijn *cartesische* coördinaten (x, y) ook in de vorm (r, φ) aangegeven worden, waarbij r de afstand van het nulpunt is en φ de hoek tussen de x-as en de verbinding van het nulpunt met P (tegen de klok gemeten). Tussen de gewone (cartesische) coördinaten (x, y)

en de poolcoördinaten (r, φ) bestaat het volgende verband:

$$x = r\cos(\varphi), \quad y = r\sin(\varphi) \quad \longleftrightarrow \quad r = \sqrt{x^2 + y^2}, \quad \tan(\varphi) = \frac{y}{x}.$$



Figuur I.20: Poolcoördinaten

Merk op dat de relatie $\tan(\varphi) = \frac{y}{x}$ de hoek φ nog niet eenduidig vast legt, omdat $\tan(x)$ een periode van π en niet van 2π heeft. De omkeerfunctie $\arctan(x)$ heeft waarden tussen $-\frac{\pi}{2}$ en $\frac{\pi}{2}$ en is negatief voor $x < 0$ en positief voor $x > 0$. We moeten daarom voor de vier kwadranten tussen de assen van het complexe vlak aparte definities nemen:

I : $x > 0, y \geq 0$: $\varphi = \arctan(\frac{y}{x})$

II : $x < 0, y \geq 0$: $\varphi = \arctan(\frac{y}{x}) + \pi$

III : $x < 0, y < 0$: $\varphi = \arctan(\frac{y}{x}) + \pi$

IV : $x > 0, y < 0$: $\varphi = \arctan(\frac{y}{x}) + 2\pi$

Voor (x, y) met $x = 0$ hebben we $\varphi = \frac{\pi}{2}$ als $y > 0$ en $\varphi = \frac{3\pi}{2}$ als $y < 0$. Voor het nulpunt zelf is de hoek niet gedefinieerd.

Opdracht 21 *Bepaal de poolcoördinaten van de volgende complexe getallen:*

$$2i, \quad 1 - i, \quad \sqrt{3} + i, \quad -2 - \sqrt{12}\, i.$$

We hadden eerder al gezegd dat de gewone coördinaten van een punt z in het complexe vlak het reële en imaginaire deel van z heten. Maar ook voor de poolcoördinaten r en φ van een complex getal z bestaat een speciale notatie:

Definitie: Zij $z = x + yi$ een complex getal met poolcoördinaten (r, φ):

(i) Het getal $r = \sqrt{x^2 + y^2}$ heet de *absolute waarde* of *modulus* van z en wordt genoteerd met $|z|$. Dit is de gewone (euclidische) afstand van het nulpunt in $\mathbb{R}^2$ en komt voor $z \in \mathbb{R}$ overeen met de gewone absolute waarde van een reëel getal.

(ii) Het getal φ met $\tan(\varphi) = \frac{y}{x}$ heet het *argument* van z en wordt met $\arg(z)$ genoteerd.

Voor een complex getal z met $|z| = r$ en $\arg(z) = \varphi$ geldt dus

$$z = r \cdot (\cos(\varphi) + i \cdot \sin(\varphi)).$$

We hebben al gezien dat voor twee complexe getallen $z_1 = x_1 + i \cdot y_1$ en $z_2 = x_2 + i \cdot y_2$ het product $z_1 z_2$ gegeven is door $z_1 z_2 = (x_1 x_2 - y_1 y_2) + i \cdot (x_1 y_2 + y_1 x_2)$. Als we z_1 en z_2 in poolcoördinaten schrijven, dus $z_1 = (r_1, \varphi_1)$ en $z_2 = (r_2, \varphi_2)$, geeft dit volgens de boven aangegeven transformaties:

$$\begin{aligned} z_1 \cdot z_2 &= r_1(\cos(\varphi_1) + i \cdot \sin(\varphi_1)) \cdot r_2(\cos(\varphi_2) + i \cdot \sin(\varphi_2)) \\ &= r_1 \cos(\varphi_1) r_2 \cos(\varphi_2) - r_1 \sin(\varphi_1) r_2 \sin(\varphi_2) \\ &\quad + i \cdot (r_1 \cos(\varphi_1) r_2 \sin(\varphi_2) + r_1 \sin(\varphi_1) r_2 \cos(\varphi_2)) \\ &= r_1 r_2 (\cos(\varphi_1)\cos(\varphi_2) - \sin(\varphi_1)\sin(\varphi_2)) \\ &\quad + i \cdot r_1 r_2 (\cos(\varphi_1)\sin(\varphi_2) + \sin(\varphi_1)\cos(\varphi_2)) \\ &= r_1 r_2 \cos(\varphi_1 + \varphi_2) + i \cdot r_1 r_2 \sin(\varphi_1 + \varphi_2). \end{aligned}$$

In de laatste stap hebben we een optelstheorema toegepast dat we in Wiskunde 1 al een keer zijn tegengekomen. Hier is een korte herinnering: Een rotatie in het 2-dimensionale vlak om een hoek van φ beschrijven we met betrekking tot de standaardbasis $\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right)$ door de matrix $\begin{pmatrix}\cos(\varphi) & -\sin(\varphi)\\ \sin(\varphi) & \cos(\varphi)\end{pmatrix}$. Maar een rotatie om $\varphi_1 + \varphi_2$ kunnen we ook zien als de samenstelling van eerst een rotatie om φ_1 en vervolgens een rotatie om φ_2. De matrix van de samenstelling van twee rotaties is het product van de matrices van de enkele rotaties. Dit geeft de matrix vergelijking

$$\begin{pmatrix}\cos(\varphi_1+\varphi_2) & -\sin(\varphi_1+\varphi_2)\\ \sin(\varphi_1+\varphi_2) & \cos(\varphi_1+\varphi_2)\end{pmatrix} = \begin{pmatrix}\cos(\varphi_1) & -\sin(\varphi_1)\\ \sin(\varphi_1) & \cos(\varphi_1)\end{pmatrix} \cdot \begin{pmatrix}\cos(\varphi_2) & -\sin(\varphi_2)\\ \sin(\varphi_2) & \cos(\varphi_2)\end{pmatrix}$$

$$= \begin{pmatrix}\cos(\varphi_1)\cos(\varphi_2) - \sin(\varphi_1)\cos(\varphi_2) & -\cos(\varphi_1)\sin(\varphi_2) - \sin(\varphi_1)\cos(\varphi_2)\\ \sin(\varphi_1)\cos(\varphi_2) + \cos(\varphi_1)\sin(\varphi_2) & -\sin(\varphi_1)\cos(\varphi_2) + \cos(\varphi_1)\cos(\varphi_2)\end{pmatrix}$$

en een vergelijk van de matrixelementen geeft in het bijzonder:

$$\begin{aligned} \cos(\varphi_1 + \varphi_2) &= \cos(\varphi_1)\cos(\varphi_2) - \sin(\varphi_1)\sin(\varphi_2) \\ \sin(\varphi_1 + \varphi_2) &= \cos(\varphi_1)\sin(\varphi_2) + \sin(\varphi_1)\cos(\varphi_2). \end{aligned}$$

De gevonden relatie

$$z_1 z_2 = r_1 r_2 \cos(\varphi_1+\varphi_2) + i \cdot r_1 r_2 \sin(\varphi_1+\varphi_2) = r_1 r_2 (\cos(\varphi_1+\varphi_2) + i \cdot \sin(\varphi_1+\varphi_2))$$

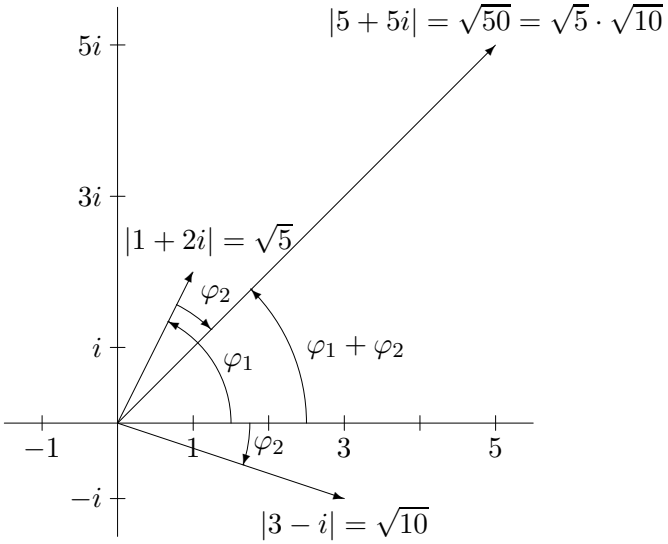
betekent nu het volgende:

Voor complexe getallen z_1 met poolcoördinaten (r_1, φ_1) en z_2 met poolcoördinaten (r_2, φ_2) heeft het product $z_1 z_2$ de poolcoördinaten $(r_1 r_2, \varphi_1 + \varphi_2)$, d.w.z. de absolute waarden r_1 en r_2 worden vermenigvuldigd en de argumenten φ_1 en φ_2 opgeteld.

Merk op: Twee complexe getallen worden met elkaar vermenigvuldigd door hun absolute waarden te vermenigvuldigen en hun argumenten bij elkaar op te tellen.

Meetkundig uitgedrukt vermenigvuldigen we een complex getal z_1 met een complex getal z_2 door z_1 met de lengte (absolute waarde) van z_2 te schalen en vervolgens om het argument van z_2 (tegen de klok) te draaien.

Voor het product $(1+2i)(3-i)$ geeft het plaatje in Figuur I.21 de meetkundige interpretatie van de vermenigvuldiging weer.



Figuur I.21: Vermenigvuldigen in het complexe vlak

Worteltrekken

We komen even terug op het worteltrekken voor complexe getallen. We hadden gezien hoe we voor een gegeven complex getal $x+i\cdot y$ een complex getal $z = a+i\cdot b$ kunnen vinden met $z^2 = x + i \cdot y$. Maar met de meetkundige interpretatie van de vermenigvuldiging is dit eigenlijk veel makkelijker.

Een complex getal w met poolcoördinaten (r, φ) heeft de wortel z met poolcoördinaten $(\sqrt{r}, \frac{\varphi}{2})$. Merk op dat ook het getal met poolcoördinaten $(\sqrt{r}, \frac{\varphi}{2} + \pi)$ een wortel is, want omdat het argument steeds tussen 0 en 2π ligt, is $2 \cdot (\varphi/2 + \pi) = \varphi + 2\pi = \varphi$. Dit is geen verrassing, want voor een complex getal z met argument $\arg(z) = \varphi$ is $\arg(-z) = \varphi + \pi$ en natuurlijk weten we dat $z^2 = (-z)^2$, dus met z is ook $-z$ een wortel uit w.

Op dezelfde manier kunnen we ook n-de machtswortels trekken. Een complex getal w met $|w| = r$ en $\arg(w) = \varphi$ heeft als n-de machtswortel het getal $z = \sqrt[n]{r}(\cos(\frac{\varphi}{n}) + i \cdot \sin(\frac{\varphi}{n}))$, dus moeten we uit de absolute waarde de n-de wortel trekken en het argument door n delen. Ook hier zijn behalve van het getal z met $\arg(z) = \frac{\varphi}{n}$ de getallen met absolute waarde $\sqrt[n]{r}$ en argumenten $\frac{\varphi}{n} + \frac{2\pi k}{n}$ voor $k = 1, \ldots, n-1$ n-de machtswortels uit w, want bij het vermenigvuldigen met n worden al deze hoeken gelijk aan φ.

Een belangrijke toepassing van de meetkundige interpretatie van het vermenigvuldigen van complexe getallen is de *Regel van de Moivre*. Een complex getal z met absolute waarde 1 kunnen we schrijven als $z = \cos(\varphi) + i \cdot \sin(\varphi)$, waarbij $\varphi = \arg(z)$. Maar de n-de macht z^n kunnen we nu makkelijk berekenen, de absolute waarde is nog steeds 1 en het argument is het n-voud van het argument van z, dus $\arg(z^n) = n \arg(z) = n\varphi$. Dit betekent dat $z^n = \cos(n\varphi) + i \cdot \sin(n\varphi)$ en dit geeft de

Regel van de Moivre:

$$(\cos(\varphi) + i \cdot \sin(\varphi))^n = \cos(n\varphi) + i \cdot \sin(n\varphi).$$

Als toepassing hiervan kunnen we eenvoudig formules voor de *sinus* of *cosinus* van het dubbele of drievoud van een hoek afleiden, bijvoorbeeld:

$$\begin{aligned}\cos(2x) &= \Re(\cos(2x) + i \cdot \sin(2x)) = \Re((\cos(x) + i \cdot \sin(x))^2) \\ &= \Re(\cos^2(x) - \sin^2(x) + 2i \cdot \cos(x)\sin(x)) \\ &= \cos^2(x) - \sin^2(x)\end{aligned}$$

$$\begin{aligned}\sin(3x) &= \Im(\cos(3x) + i \cdot \sin(3x)) = \Im((\cos(x) + i \cdot \sin(x))^3) \\ &= \Im(\cos^3(x) + 3i \cdot \cos^2(x)\sin(x) - 3\cos(x)\sin^2(x) - i \cdot \sin^3(x)) \\ &= 3\cos^2(x)\sin(x) - \sin^3(x).\end{aligned}$$

5.4 Complexe conjugatie

Een belangrijke operatie op de complexe getallen is de *complexe conjugatie* die van een getal $z = x + i \cdot y$ het nieuwe getal

$$\overline{z} := x - i \cdot y$$

maakt, dat de *complex geconjugeerde van* z heet. Soms wordt de complex geconjugeerde ook met z^* in plaats van $\overline{z}$ genoteerd.

Omdat bij de complexe conjugatie het reële deel van een getal hetzelfde blijft en het imaginaire deel op zijn negatieve waarde gaat, is de complexe conjugatie gewoon de spiegeling in de x-as van het complexe vlak.

Men ziet rechtstreeks in dat z en $\overline{z}$ dezelfde absolute waarde $\sqrt{x^2 + y^2}$ hebben en dat het argument van $\overline{z}$ het negatieve van het argument van z is, dus $\arg(\overline{z}) = -\arg(z)$.

Verder geldt $z \cdot \overline{z} = |z|^2 \in \mathbb{R}$, want $(x + i \cdot y)(x - i \cdot y) = x^2 + y^2$. Dit geeft nu de toelichting voor de formule voor het inverteren van een complex getal, die we in het begin van deze les hebben gezien:

$$(x + yi)^{-1} = z^{-1} = \frac{1}{z} = \frac{\overline{z}}{z \cdot \overline{z}} = \frac{1}{|z|^2}\,\overline{z} = \frac{x - yi}{x^2 + y^2},$$

de inverse van z is dus de complex geconjugeerde $\overline{z}$ van z gedeeld door het kwadraat $|z|^2$ van de absolute waarde van z.

Dat de inverse van z een veelvoud van $\overline{z}$ moet zijn, hadden we natuurlijk ook al uit de argumenten kunnen aflezen, want uit $\arg(1) = 0$ volgt $\arg(\frac{1}{z}) = -\arg(z) = \arg(\overline{z})$.

We kunnen nu ook een willekeurige breuk van complexe getallen op standaardvorm brengen, want

$$\frac{z_1}{z_2} = \frac{z_1\overline{z_2}}{|z_2|^2} = \frac{1}{|z_2|^2}\, z_1\overline{z_2}.$$

Voorbeeld: Er geldt $\frac{2i}{1-i} = \frac{2i(1+i)}{2} = \frac{1}{2}(-2+2i) = -1+i$.

Met behulp van de complex geconjugeerde kunnen we ook reëel en imaginair deel van een getal z makkelijk uitdrukken:

$$\Re(z) = \frac{1}{2}(z+\overline{z}) \quad \text{en} \quad \Im(z) = \frac{1}{2i}(z-\overline{z}).$$

In het bijzonder is een getal $z \in \mathbb{C}$ een reëel getal, dan en slechts dan als $\overline{z} = z$.

Met behulp van de complexe conjugatie vinden we ook een belangrijke eigenschap van de complexe nulpunten van veeltermen met reële coëfficiënten. Stel $f(z) = c_n z^n + \ldots + c_1 z + c_0$ is een veelterm met *reële* coëfficiënten $c_i \in \mathbb{R}$ en stel dat $a \in \mathbb{C}$ met $f(a) = 0$. Dan is natuurlijk ook $\overline{f(a)} = 0$, dus $\overline{c_n a^n} + \ldots + \overline{c_1 a} + \overline{c_0} = c_n \overline{a}^n + \ldots + c_1\overline{a} + c_0 = 0$ en dus is ook $\overline{a}$ een nulpunt van $f(z)$. De niet-reële nulpunten van $f(z)$ komen dus in paren van complex geconjugeerden voor.

5.5 Machtsverheffen

We hebben inmiddels alle bewerkingen en operaties op de reële getallen kunnen uitbreiden tot de complexe getallen, met uitzondering van het machtsverheffen met complexe exponenten.

Om te beginnen moeten we zeker iets kunnen zeggen over $e^{i\cdot y}$ waarbij $y \in \mathbb{R}$. Als dit lukt, kunnen we ook voor $z = x + i \cdot y$ de e-macht definiëren, namelijk door $e^z = e^{x+i\cdot y} = e^x \cdot e^{i\cdot y}$. Uiteindelijk zullen we dan (net als voor de reële getallen) a^z definiëren door $a^z = e^{\log(a)z}$, maar zo ver zijn we nog niet.

In de volgende les zullen we de complexe exponentiële functie en logaritme nader bekijken die een zuivere motivatie voor de volgende definitie van $e^{i\varphi}$ geven, die bekend staat als

Formule van Euler:

$$e^{i\varphi} = \cos(\varphi) + i \cdot \sin(\varphi).$$

Volgens deze formule is het getal $e^{i\varphi}$ juist het complexe getal met absolute waarde 1 en argument φ. Als φ van 0 tot 2π loopt, loopt $e^{i\varphi}$ één keer rond de eenheidscirkel.

Zonder de complexe exponentiële functie kunnen we alvast twee redenen aangeven, die de Formule van Euler plausibel maken. Als we i (net als $\sqrt{2}$ of

π) als een constante beschouwen, is de functie $f(\varphi) := e^{i\varphi}$ een functie van een reële veranderlijke die we kunnen afleiden, en dit geeft

$$\begin{aligned} f'(\varphi) &= (e^{i\varphi})' = (\cos(\varphi) + i \cdot \sin(\varphi))' \\ &= -\sin(\varphi) + i \cdot \cos(\varphi) = i \cdot (\cos(\varphi) + i \cdot \sin(\varphi)) = i \cdot e^{i\varphi} = i \cdot f(\varphi). \end{aligned}$$

Maar dit is precies wat we volgens de kettingregel van de exponentiële functie zouden verwachten.

Verder geldt volgens onze definitie dat

$$e^{i\varphi_1} \cdot e^{i\varphi_2} = e^{i(\varphi_1+\varphi_2)}$$

omdat we getallen op de eenheidscirkel vermenigvuldigen door hun argumenten op te tellen, dus lijkt onze definitie ook met deze eigenschappen van de reële exponentiële functie overeen te komen.

Met behulp van de relatie $e^{i\varphi} := \cos(\varphi) + i \cdot \sin(\varphi)$ en de symmetrieeigenschappen $\cos(-x) = \cos(x)$ en $\sin(-x) = -\sin(x)$ kunnen we nu $\sin(x)$ en $\cos(x)$ op de exponentiële functie terug brengen, want er geldt:

$$\begin{aligned} e^{i\varphi} + e^{-i\varphi} &= \cos(\varphi) + \cos(-\varphi) + i \cdot (\sin(\varphi) + \sin(-\varphi)) = 2\cos(\varphi), \\ e^{i\varphi} - e^{-i\varphi} &= \cos(\varphi) - \cos(-\varphi) + i \cdot (\sin(\varphi) - \sin(-\varphi)) = 2i\sin(\varphi) \end{aligned}$$

en hiermee volgt

$$\cos(\varphi) = \frac{e^{i\varphi} + e^{-i\varphi}}{2} \quad \text{en} \quad \sin(\varphi) = \frac{e^{i\varphi} - e^{-i\varphi}}{2i}.$$

We komen nu nog een keer op de regel van de Moivre terug, met onze nieuwe definitie ziet die er namelijk heel eenvoudig uit:

$$(e^{i\varphi})^n = e^{i(n\varphi)}.$$

Ook de opteltheorema's $\cos(\varphi_1 + \varphi_2) = \cos(\varphi_1)\cos(\varphi_2) - \sin(\varphi_1)\sin(\varphi_2)$ en $\sin(\varphi_1 + \varphi_2) = \cos(\varphi_1)\sin(\varphi_2) + \sin(\varphi_1)\cos(\varphi_2)$ kunnen we meteen uit de reële en imaginaire delen van $e^{i(\varphi_1+\varphi_2)} = e^{i\varphi_1} \cdot e^{i\varphi_2}$ aflezen.

> We hebben gezien dat we met complexe getallen net zo goed als met reële getallen kunnen reken (ook al is de vermenigvuldiging iets ingewikkelder) en dat we vergelijkingen veel beter kunnen oplossen dan in $\mathbb{R}$. Maar er is ook een belangrijk nadeel van de complexe getallen tegenover de reële getallen: We kunnen van twee reële getallen steeds zeggen dat één van de twee groter is dan de andere (als ze niet gelijk zijn). We zeggen namelijk dat $x > y$ als $x - y > 0$ en voor elk getal $x \in \mathbb{R}$ geldt $x > 0$, $x = 0$ of $-x > 0$. Verder is voor twee positieve getallen $x, y > 0$ ook de som $x + y$ en het product xy positief.
>
> Een ordening met deze eigenschappen kunnen we op $\mathbb{C}$ niet construeren, want als er een $z \in \mathbb{C}$ is met $z > 0$, dan moet gelden dat $z^2 > 0$. Maar voor $z \neq 0$ is of $z > 0$ of $-z > 0$ en dus is in elk geval $z^2 > 0$. Omdat we elk complex getal a in de vorm $a = z^2$ kunnen schrijven, zijn dus

alle getallen $z \in \mathbb{C}$ positief. In het bijzonder is $1 > 0$ en $-1 > 0$ en dus $0 = -1 + 1 > 0$ (omdat -1 en 1 positief zijn) en dit is onmogelijk. De enige manier om de complexe getallen zo te ordenen dat sommen en producten van positieve getallen weer positief zijn, is de triviale ordeningen, waar alle getallen even groot zijn als 0, maar daar hebben we natuurlijk niets aan.

Als toegift geven we een beroemde formule aan, die de meest belangrijke constanten 0, 1, i, e en π in een relatie brengt:

$$e^{i\pi} + 1 = 0.$$

Toepassingen van de complexe getallen

Op grond van de samenhang $e^{i\varphi} = \cos(\varphi) + i \cdot \sin(\varphi)$ zijn complexe getallen in alle toepassingen belangrijk die met golven te maken hebben. Voorbeelden hiervoor zijn:

- Het berekenen van het overlappen van twee of meer golven (bijvoorbeeld watergolven, maar ook elektromagnetische golven).
- Kwantummechanica: een deeltje wordt door een golf-functie beschreven, waarvan de absolute waarde de kans aangeeft, het deeltje in een zeker gebied te vinden.
- Spraakherkenning: een spraaksignaal wordt beschreven door een som van *sinus*-functies voor verschillende frequenties, waarbij het patroon van frequenties (formanten) karakteristiek voor de klinkers is. Het bepalen van dit patroon uit een signaal wordt met behulp van de Fourieranalyse bepaald, die we later gaan behandelen.
- Beeldherkenning: een plaatje wordt gezien als een bron van lichtgolven, waarbij verschillende kleuren met verschillende frequenties corresponderen en de intensiteit met de amplitude van de golven.

Belangrijke begrippen in deze les

- complexe getallen
- reëel deel, imaginair deel
- poolcoördinaten, absolute waarde, argument
- complexe conjugatie
- relatie $e^{i\varphi} = \cos(\varphi) + i \cdot \sin(\varphi)$

Opgaven

43. Schrijf de volgende complexe getallen in de vorm $a + i \cdot b$ en in poolcoördinaten:

$$\text{(i) } (1 - i\sqrt{3})^2 \qquad \text{(ii) } \frac{1+i}{i-1} \qquad \text{(iii) } \frac{3+4i}{2-i}$$

Hoe kan men absolute waarde en argument van deze getallen bepalen, zonder de getallen eerst in de vorm $a + i \cdot b$ te brengen?

44. Teken een punt $z \in \mathbb{C}$ op de eenheidscirkel (d.w.z. met $|z| = 1$). Construeer de punten z^2, z^3, z^{-1}, $-z$, $\overline{z}$, $i \cdot z$, $-i \cdot z$. Ga in de figuur na dat $z + z^{-1}$ reëel is.

45. Bereken de (complexe) oplossingen van de vergelijking $z^2 + 3z + 4 = 0$.

46. Vind de oplossingen $z \in \mathbb{C}$ voor de volgende vergelijkingen:

(i) $z^3 = i$, (ii) $z^2 - 2z + 2 = 0$, (iii) $z^4 = -1$, (iv) $(3+4i)z^2 + 5z + (2-4i) = 0$.

Teken de wortels in het complexe vlak.

47. Beschrijf de volgende verzamelingen van complexe getallen in het complexe vlak:

(i) $z \in \mathbb{C}$ met $\Re(z^2) > 0$;

(ii) $z \in \mathbb{C}$ met $\Re\left(\dfrac{z+i}{z-2i}\right) = 1$.

48. Druk met behulp van de regel van de Moivre:

(i) $\cos(4\varphi)$ uit in $\cos(\varphi)$ en $\sin(\varphi)$,

(ii) $\sin(3\varphi)$ uit in $\sin(\varphi)$ (zonder $\cos(\varphi)$).

49. Maak een schets van de verzamelingen van complexe getallen z die aan de aangegeven voorwaarden voldoen:

(i) $|z| \leq 2$;

(ii) $|z - 2i| \leq 3$;

(iii) $|z - 3 + 4i| \leq 5$;

(iv) $\arg(z) = \frac{\pi}{3}$;

(v) $\pi \leq \arg(z) \leq \frac{7\pi}{4}$.

50. Zij $L_1 \subset \mathbb{C}$ de lijn met $\Re(z) = \Im(z)$ en $L_2 \subset \mathbb{C}$ de lijn met $\Im(z) = 1$. Wat zijn de beelden van deze lijnen onder de afbeelding $z \mapsto z^{-1}$ (d.w.z. de verzamelingen $\{z^{-1} \mid z \in L_1(L_2)\}$?

51. Welke baan doorloopt $w := \dfrac{z^2 - z + 1}{2z}$ als z de eenheidscirkel doorloopt (d.w.z. $z = e^{ix}$ met $x \in [0, 2\pi]$)?

52. Welke baan doorloopt $z := \dfrac{x - i}{x + i}$ als x langs de reële as loopt (in positieve richting)?

53. Beschrijf de baan die

$$z := 149597887\, e^{2\pi i \frac{t}{365.257}} + 384403\, e^{2\pi i \frac{t}{27.321}}$$

doorloopt, als t langs de positieve reële as loopt. Welk hemelsfenomeen wordt hierdoor weergegeven?

Hoe zou de baan veranderen, als de constante 384403 door 38440300 vervangen wordt?

Les 6 Complexe functies

Nadat we de complexe getallen hebben leren kennen, is het een voor de hand liggende vraag of hiervoor net als voor de reële getallen ook functies bestaan. Met een functie bedoelen we hierbij een *voorschrift* die aan elk complex getal $z \in B$ uit een deelverzameling $B \subseteq \mathbb{C}$ een eenduidige waarde $f(z) \in \mathbb{C}$ toewijst. Het gebied $B \subseteq \mathbb{C}$ heet dan (net als bij gewone functies) het *domein* van de functie $f(z)$.

Bij reële functies hebben we veel over een functie $f(x)$ kunnen zeggen, door de grafiek $(x, f(x))$ te bekijken. Dit is bij complexe functies echter moeilijk, want voor het domein $\mathbb{C}$ (waar een functie op gedefinieerd is) hebben we al een 2-dimensionaal vlak nodig, en voor de functiewaarden ook nog eens een 2-dimensionaal vlak, zo dat we voor de grafiek een 4-dimensionaal plaatje nodig zouden hebben.

Maar we kunnen wel een redelijke indruk van een complexe functie krijgen door de volgende methoden:

(1) Bekijk de reële en imaginaire delen van de functie apart. Dit betekent dat we een complexe functie $f(z) : \mathbb{C} \to \mathbb{C}$ opsplitsen in twee functies met reële waarden, namelijk

$$u(z) : \mathbb{C} \to \mathbb{R},\ z \mapsto \Re(f(z)) \quad \text{en} \quad v(z) : \mathbb{C} \to \mathbb{R},\ z \mapsto \Im(f(z))$$

zo dat $f(z) = u(z) + i \cdot v(z)$. Als we nu z schrijven als $z = x + iy$ met $x, y \in \mathbb{R}$, kunnen we $u(z)$ en $v(z)$ opvatten als functies $u(x, y)$ en $v(x, y)$ van de twee reële variabelen x en y met waarden in $\mathbb{R}$. Maar voor dit soort functies hebben we al eerder gezien dat we als grafiek een 3-dimensionaal plaatje krijgen, door de punten $(x, y, u(x, y))$ te bekijken.

(2) We kunnen kijken hoe een functie $f(z)$ zekere lijnen afbeeldt, bijvoorbeeld de lijnen parallel met de x-as (dus de complexe getallen met hetzelfde imaginaire deel), de lijnen parallel met de y-as (de complexe getallen met hetzelfde reële deel), lijnen door de oorsprong (de complexe getallen met hetzelfde argument). We kunnen ook kijken wat met cirkels rond de oorsprong gebeurt, dus met complexe getallen met dezelfde absolute waarde.

Als voorbeeld van de eerste methode laat Figuur I.22 de reële en imaginaire delen van de derdegraads veelterm

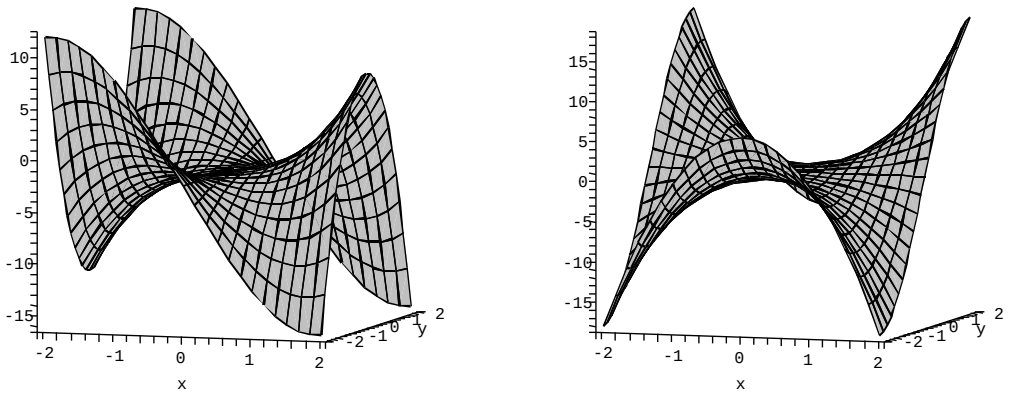
$$f(z) = z^3 + z - 2$$

zien. Invullen van $z = x + iy$ geeft

$$\begin{aligned} f(x + iy) &= (x + iy)^3 + (x + iy) - 2 = x^3 + 3ix^2y - 3xy^2 - iy^3 + x + iy - 2 \\ &= x^3 - 3xy^2 + x - 2 + i(3x^2y - y^3 + y) \end{aligned}$$

en dus geldt $f(z) = u(x, y) + i \cdot v(x, y)$ met

$$\begin{aligned} u(x, y) &= \Re(f(z)) = x^3 - 3xy^2 + x - 2, \\ v(x, y) &= \Im(f(z)) = 3x^2y - y^3 + y. \end{aligned}$$



Figuur I.22: Reëel en imaginair deel van $f(z) = z^3 + z - 2$

OPDRACHT 22 *Bepaal voor $z = x + iy$ de reële en imaginaire delen van $f(z) = \frac{1}{1-z} = (1-z)^{-1}$, d.w.z. bepaal reële functies $u(x,y)$ en $v(x,y)$ van twee variabelen zo dat $f(x+iy) = u(x,y) + i \cdot v(x,y)$.*

Omdat we weten hoe we complexe getallen optellen en vermenigvuldigen, hebben we met complexe functies die door een veelterm

$$f(z) = c_n z^n + c_{n-1} z^{n-1} + \ldots + c_1 z + c_0$$

gegeven zijn helemaal geen moeite. Hierbij mogen de coëfficiënten natuurlijk ook zelfs complexe getallen zijn, dit geeft complexe functies zo als de kwadratische veelterm $f(z) := z^2 + 2iz + \sqrt{-3}$.

Maar natuurlijk kunnen we niet verwachten dat alle complexe functies veeltermfuncties zijn, de vraag is echter, hoe we aan andere complexe functies zouden kunnen komen. De oplossing hiervoor is verrassend eenvoudig. We hadden gezien dat we een reële functie $f(x)$ door Taylor polynomen kunnen benaderen en dat (in een kleine omgeving van een punt x_0) de functie door de Taylor reeks beschreven kan worden. Dit brengt algemene functies op veeltermen en machtreeksen terug, en die kunnen we (met behulp van optellen en vermenigvuldigen) ook voor complexe getallen uitwerken.

Afspraak: We zullen ons in deze cursus beperken tot complexe functies $f(z)$ die door een machtreeks

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = a_0 + a_1 z + \ldots + a_n z^n + \ldots$$

gegeven zijn, waarbij we steeds veronderstellen dat de reeks voor de waarden z die we nodig hebben convergeert.

Een diepere analyse van complexe functies laat zien, dat de beperking tot complexe functies die door een machtreeks gegeven zijn helemaal geen sterke beperking is. Er laat zich namelijk aantonen dat een complexe functie die in een cirkel van straal R rond een punt z_0 *complex differentieerbaar* is (we komen hier in de Appendix voor deze les op terug) automatisch een convergente Taylor reeks rond z_0 heeft. Anders dan bij reële functies volgt namelijk uit het bestaan van de *eerste* afgeleide $f'(z)$ op een gebied B dat ook alle hogere afgeleiden $f^{(n)}(z)$ op B bestaan en continu zijn. De reden hiervoor is dat complexe differentieerbaarheid een veel sterkere eigenschap is dan reële differentieerbaarheid.

6.1 Complexe exponentiële functie

We hebben gezien dat we de (reële) exponentiële functie door de Taylor reeks $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ kunnen beschrijven, en omdat deze reeks voor alle x naar $\exp(x)$ convergeert mogen we zelfs zeggen, dat de twee gelijk zijn, dus dat

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Als we nu over een definitie voor de exponentiële functie op de complexe getallen nadenken, willen we natuurlijk dat die op de reële getallen met de reële exponentiële functie overeen komt. Het is nu een enigszins voor de hand liggende gedachte, de complexe exponentiële functie dusdanig te definiëren, dat we complexe getallen in de Taylor reeks van de reële exponentiële functie invullen. Dan weten we in ieder geval dat voor reële getallen inderdaad de functiewaarden hetzelfde blijven.

Algemeen noemt men het invullen van waarden $z \in B \subseteq \mathbb{C}$ in een machtreeks van een functie $f(z)$ die op een kleiner domein $B_1 \subseteq B$ gedefinieerd is, het *voortzetten* van $f(z)$ op B. In ons geval zetten we de reële exponentiële functie van de reële lijn $\mathbb{R}$ op het hele complexe vlak voort.

We hadden gezien dat er problemen met de Taylor reeks voor de functie $\exp(-\frac{1}{x^2})$ zijn, omdat de reeks de 0-functie is en (behalve voor $x = 0$) niet tegen de goede functiewaarden convergeert. Als we deze functie op de complexe getallen voortzetten, zien we dat we in het punt $z = 0$ helemaal geen continue voortzetting meer kunnen vinden (wat voor de reële getallen wel nog het geval was). Als we namelijk met $z = ix$ langs de imaginaire as lopen, hebben we $\exp(-\frac{1}{(ix)^2}) = \exp(\frac{1}{x^2})$ en dit gaat voor $x \to 0$ naar oneindig. Het feit dat we de functie in het complexe vlak niet continu in het punt 0 kunnen voortzetten hangt nauw samen met het feit dat de Taylor reeks op de reële getallen niet tegen de goede functie convergeert.

Definitie: De *complexe exponentiële functie* is gedefinieerd als voortzetting van de reële exponentiële functie op het volledige complexe vlak, dus door

$$\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \dots$$

Om te rechtvaardigen dat deze definitie zinvol is, merken we een aantal dingen op:

(1) Om te zien dat de reeks van $\exp(z)$ convergent is, is het voldoende dat de reeks over de absolute waarden van de termen convergent is, men zegt hiervoor dat de reeks *absoluut convergent* is. Maar

$$\sum_{n=0}^{\infty} \left| \frac{z^n}{n!} \right| = \sum_{n=0}^{\infty} \frac{|z|^n}{n!} = \exp(|z|),$$

dus volgt de absolute convergentie van de reeks voor $\exp(z)$ uit de convergentie van de Taylor reeks voor de reële exponentiële functie.

(2) We kunnen $\exp(z_1 + z_2)$ (in principe) uitrekenen door $z_1 + z_2$ in de reeks in te vullen, dus $\exp(z_1 + z_2) = \sum_{n=0}^{\infty} \frac{(z_1+z_2)^n}{n!}$. Aan de andere kant berekent men $\exp(z_1) \cdot \exp(z_2)$ door de reeksen $\sum_{n=0}^{\infty} \frac{z_1^n}{n!}$ en $\sum_{n=0}^{\infty} \frac{z_2^n}{n!}$ te vermenigvuldigen. Door de coëfficiënten van $z_1^j z_2^k$ in de uitdrukkingen voor $\exp(z_1 + z_2)$ en voor $\exp(z_1) \cdot \exp(z_2)$ te vergelijken, ziet men dat inderdaad

$$\exp(z_1 + z_2) = \exp(z_1) \cdot \exp(z_2),$$

net als we dat van de reële exponentiële functie gewend zijn. (In feite berust het bewijs dat $\exp(x + y) = \exp(x) \cdot \exp(y)$ voor $x, y \in \mathbb{R}$ precies op hetzelfde idee.)

(3) In Wiskunde 1 hadden we gezien dat de exponentiële functie gekarakteriseerd is door de eigenschappen dat

$$\exp(x)' = \exp(x) \quad \text{en} \quad \exp(0) = 1.$$

We zullen straks nader op het differentiëren van complexe functies ingaan, maar voor een functie die door een reeks gegeven is zou men hopen de afgeleide te vinden door de reeks termsgewijs af te leiden. Voor functies met een absoluut convergente reeks is dit inderdaad juist, dus hebben we voor de complexe exponentiële functie:

$$\exp(z)' = (\sum_{n=0}^{\infty} \frac{z^n}{n!})' = \sum_{n=1}^{\infty} n \cdot \frac{z^{n-1}}{n!} = \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} = \exp(z).$$

De complexe exponentiële functie heeft dus ook de eigenschappen die de reële exponentiële functie karakteriseren.

Met onze definitie van de complexe exponentiële functie kunnen we nu eindelijk ook de Formule van Euler

$$e^{i\varphi} = \cos(\varphi) + i \cdot \sin(\varphi)$$

uit de vorige les toelichten. Hiervoor vullen we $z = i\varphi$ in de reeks voor $\exp(z)$ in, waarbij we rekening ermee houden dat $i^2 = -1$, $i^3 = -i$ en $i^4 = 1$. We

krijgen:

$$\begin{aligned} e^{i\varphi} &= 1 + i\cdot\varphi - \frac{\varphi^2}{2!} - i\cdot\frac{\varphi^3}{3!} + \frac{\varphi^4}{4!} + i\cdot\frac{\varphi^5}{5!} - \frac{\varphi^6}{6!} - i\cdot\frac{\varphi^7}{7!} + \ldots \\ &= (1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \frac{\varphi^6}{6!} + \ldots) + i\cdot(\varphi - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} - \frac{\varphi^7}{7!} + \ldots) \\ &= \left(\sum_{n=0}^{\infty}(-1)^n\frac{\varphi^{2n}}{(2n)!}\right) + i\cdot\left(\sum_{n=0}^{\infty}(-1)^n\frac{\varphi^{2n+1}}{(2n+1)!}\right) \\ &= \cos(\varphi) + i\cdot\sin(\varphi). \end{aligned}$$

In de laatste stap hebben we hierbij gebruik gemaakt van de (allang bekende) Taylor reeksen voor $\cos(x)$ en $\sin(x)$.

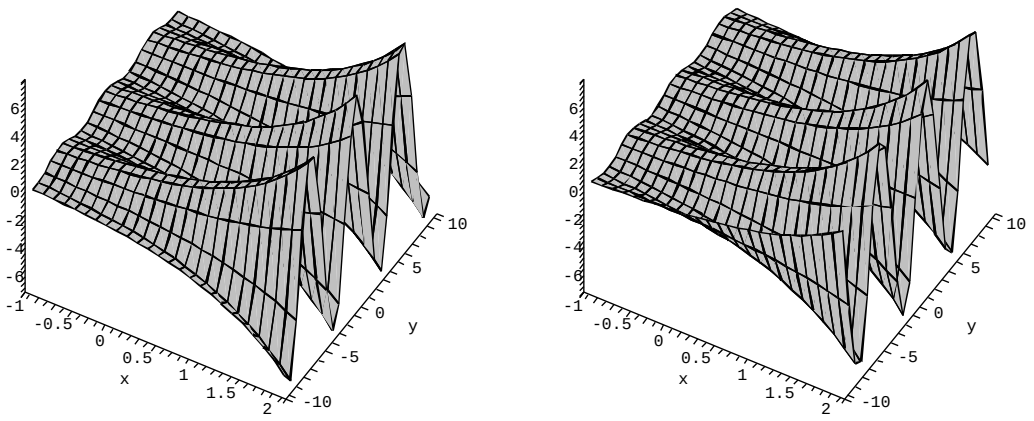
Om een beter idee van de complexe exponentiële functie te krijgen, is het verstandig naar de reële en imaginaire delen te kijken. Voor $z \in \mathbb{C}$ met $z = x + iy$, dus $x = \Re(z)$ en $y = \Im(z)$, geldt

$$\exp(x + iy) = \exp(x)\exp(iy) = \exp(x)(\cos(y) + i\sin(y))$$

waarbij we de formule van Euler $e^{iy} = \cos(y) + i\sin(y)$ hebben toegepast. Hieruit volgt:

$$\Re(\exp(x + iy)) = \exp(x)\cos(y) \quad \text{en} \quad \Im(\exp(x + iy)) = \exp(x)\sin(y).$$

In Figuur I.23 zijn de reële en imaginaire delen van de complexe exponentiële functie te zien.



Figuur I.23: Reëel en imaginair deel van $\exp(z)$

Zo zeer de complexe exponentiële functie in veel aspecten op de reële exponentiële functie lijkt, moeten we toch bij de overgang van de reële naar de

complexe exponentiële functie afscheid nemen van sommige vertrouwde eigenschappen van de exponentiële functie. Met name heeft de complexe exponentiële functie geen globale omkeerfunctie (zo als de logaritme voor de reële exponentiële functie), omdat er getallen z_1 en z_2 in het complexe vlak zijn, waarvoor de complexe exponentiële functie dezelfde functiewaarde $\exp(z_1) = \exp(z_2)$ oplevert. Dit kunnen we makkelijk inzien, als we even herhalen waarom de reële exponentiële functie voor verschillende getallen wel verschillende functiewaarden oplevert en wat er mis gaat als we proberen deze redenering op de complexe exponentiële functie uit te breiden.

Voor $x_1, x_2 \in \mathbb{R}$ geldt $e^{x_1} = e^{x_2} \Leftrightarrow e^{x_1}/e^{x_2} = e^{x_1 - x_2} = 1$. Maar omdat de reële exponentiële functie strikt stijgend is, is dit alleen maar het geval voor $x_1 - x_2 = 0$, dus voor $x_1 = x_2$.

Als we hetzelfde argument op de complexe exponentiële functie toepassen, beleven we een kleine verrassing. Er geldt weer dat $e^{z_1} = e^{z_2} \Leftrightarrow e^{z_1 - z_2} = 1$. Maar voor een getal $z = x + iy \in \mathbb{C}$ geldt $e^z = e^x \cdot (\cos(y) + i \sin(y))$ en $|e^z| = e^x$, dus geldt $e^z = 1 \Leftrightarrow e^x = 1, \cos(y) = 1, \sin(y) = 0$ en dit is precies het geval voor $x = 0$ en $y = 2\pi k$ met $k \in \mathbb{Z}$. Er geldt dus

$$e^z = 1 \Longleftrightarrow z = 2\pi i k \text{ met } k \in \mathbb{Z}.$$

Merk op: Hieruit volgt in het bijzonder dat voor alle $z \in \mathbb{C}$ geldt dat

$$e^z = e^{z+2\pi i} = e^{z+2\pi i \cdot k} \text{ voor alle } k \in \mathbb{Z}$$

en we zeggen daarom dat de complexe exponentiële functie $2\pi i$-*periodiek* is.

Toepassing: Gedempte trilling

De kracht F die een massa m ervaart, die aan een spiraalveer hangt, is proportioneel met de afwijking x van de massa tegenover de evenwichtspositie, dus $F = -kx$. Het minteken betekent dat de kracht altijd naar de evenwichtspositie terugtrekt.

Een kracht F die op m werkt leidt tot een versnelling $x''(t)$ van de massa met $F = mx''(t)$. Zonder verdere invloed van buiten zou de tijdelijke beweging $x(t)$ van de massa dus voldoen aan de differentiaalvergelijking $mx''(t) = -kx(t)$. We hadden in Wiskunde 1 al gezien, dat de oplossingen van deze vergelijking sinus- en cosinusfuncties zijn, namelijk $x(t) = \sin(\omega t)$ of $x(t) = \cos(\omega t)$ (of een lineaire combinatie hiervan) met $\omega = \sqrt{\frac{k}{m}}$. De massa zou dus in een sinustrilling bewegen.

Als we nu ook met wrijving rekening willen houden, moeten we hierover een aanname maken. Meestal wordt verondersteld dat de wrijving proportioneel met de snelheid van de massa is, dus geldt voor de wrijvingskracht F_w dat $F_w = -\gamma x'(t)$. Ook hier houdt het minteken rekening ermee dat de wrijvingskracht de massa remt. In totaal geldt nu $mx''(t) = -kx - \gamma x'(t)$ of te wel

$$x''(t) + \beta x'(t) + \omega^2 x(t) = 0 \text{ met } \beta = \frac{\gamma}{m},\ \omega^2 = \frac{k}{m}.$$

Om een oplossing voor deze differentiaalvergelijking te vinden, proberen we een functie $x(t)$ van de vorm $x(t) = Ce^{at}$. Als we dit invullen, krijgen we $Ce^{at}(a^2 + \beta a + \omega^2) = 0$, dus moet a een oplossing van de kwadratische vergelijking $X^2 + \beta X + \omega^2 = 0$ zijn. Met de *abc*-formule (of anders) vinden we de oplossingen

$$a_{1,2} = -\frac{\beta}{2} \pm \frac{1}{2}\sqrt{\beta^2 - 4\omega^2}.$$

Als $\beta^2 > 4\omega^2$, dus $\beta > 2\omega$, zijn er twee reële oplossingen en we krijgen voor $x(t)$ een functie van de vorm

$$x(t) = C_1 e^{a_1 t} + C_2 e^{a_2 t} \text{ met } a_{1,2} = -\frac{\beta}{2} \pm \frac{1}{2}\sqrt{\beta^2 - 4\omega^2}.$$

Deze beweging beschrijft een steeds langzamer wordend terugvallen in de evenwichtspositie, waarbij het mogelijk is dat de beweging een keer door de evenwichtspositie doorheen gaat.

Als $\beta^2 < 4\omega^2$, dus als de wrijving zwakker is, zijn de oplossingen complex. We definiëren

$$\overline{\omega} := \frac{1}{2}\sqrt{4\omega^2 - \beta^2} = \omega\sqrt{1 - \frac{\beta^2}{4\omega^2}} < \omega,$$

dan zijn de oplossingen $a_{1,2} = -\frac{\beta}{2} \pm i\,\overline{\omega}$. We krijgen dus $x(t)$ van de vorm

$$x(t) = C_1 e^{-\frac{\beta}{2}} e^{i\,\overline{\omega} t} + C_2 e^{-\frac{\beta}{2}} e^{-i\,\overline{\omega} t} = e^{-\frac{\beta}{2}}(c_1 \cos(\overline{\omega} t) + c_2 \sin(\overline{\omega} t)),$$

waarbij

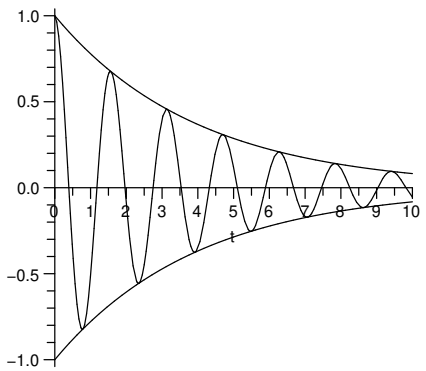
$$c_1 = (C_1 + C_2) \quad \text{en} \quad c_2 = i(C_1 - C_2).$$

De samenhang tussen de coëfficiënten C_1, C_2 en c_1, c_2 ziet men met behulp van het volgende trucje in:

$$\begin{aligned} & C_1 e^{i\omega t} + C_2 e^{-i\omega t} \\ &= \frac{C_1 + C_2}{2} e^{i\omega t} + \frac{C_1 - C_2}{2} e^{i\omega t} + \frac{C_1 + C_2}{2} e^{-i\omega t} - \frac{C_1 - C_2}{2} e^{-i\omega t} \\ &= (C_1 + C_2)\cos(\omega t) + i(C_1 - C_2)\sin(\omega t) \end{aligned}$$

De oplossing $x(t)$ geeft dus een sinus-vormige trilling met de frequentie $\overline{\omega} < \omega$ aan die gedempt is met de functie $e^{-\frac{\beta}{2}}$. De wrijving heeft dus naast het dempen van de trilling ook het effect dat de trilling om de factor $\sqrt{1 - \frac{\beta^2}{4\omega^2}}$ langzamer wordt dan in het vrije geval zonder wrijving.

In Figuur I.24 is de grafiek van een gedempte trilling $x(t) = e^{-\frac{\beta}{2}} \cos(\overline{\omega} t)$ te zien. Naast de functie $x(t)$ zijn ook de grensfuncties $\pm e^{-\frac{\beta}{2}}$ geschetst die de demping aangeven.



Figuur I.24: Gedempte trilling

6.2 Complexe sinus en cosinus functies

Nu dat we hebben gezien dat het voortzetten van de Taylor reeks van $\exp(x)$ op de complexe getallen een succes was, is het voor de hand liggend hetzelfde principe ook op de sinus en cosinus functies toe te passen. We definiëren dus de complexe sinus functie door

$$\sin(z) := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{1}{6}z^3 + \frac{1}{120}z^5 - \frac{1}{5040}z^7 + \dots$$

en de complexe cosinus functie door

$$\cos(z) := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{1}{2}z^2 + \frac{1}{24}z^4 - \frac{1}{720}z^6 + \dots$$

Als we nu nog een keer naar de berekening kijken waarmee we net hebben aangetoond dat $e^{i\varphi} = \cos(\varphi) + i \cdot \sin(\varphi)$ is, zien we dat we nergens iets speciaals over φ verondersteld hebben. Als we dus precies hetzelfde opschrijven met z in plaats van φ waarbij $z \in \mathbb{C}$ een willekeurig complex getal is, vinden we de relatie

$$e^{iz} = \cos(z) + i \sin(z) \text{ voor alle } z \in \mathbb{C}.$$

Maar voor de boven aangegeven definities van $\cos(z)$ en $\sin(z)$ geldt net als op de reële getallen, dat

$$\cos(-z) = \cos(z) \quad \text{en} \quad \sin(-z) = -\sin(z)$$

want $(-z)^{2n} = (-1)^{2n} \cdot z^{2n} = z^{2n}$ en $(-z)^{2n+1} = (-z) \cdot (-z)^{2n} = (-z) \cdot z^{2n} = -z^{2n+1}$. Hieruit volgt

$$e^{iz} + e^{-iz} = \cos(z) + \cos(-z) + i \cdot (\sin(z) + \sin(-z)) = 2\cos(z) \text{ en}$$
$$e^{iz} - e^{-iz} = \cos(z) - \cos(-z) + i \cdot (\sin(z) - \sin(-z)) = 2i \sin(z)$$

en we zien dus dat

$$\cos(z) = \frac{e^{iz} + e^{-iz}}{2} \quad \text{en} \quad \sin(z) = \frac{e^{iz} - e^{-iz}}{2i},$$

dus precies dezelfde relaties die we voor reële waarden van z al in de vorige les hadden verkregen.

> In de zuivere wiskunde wordt eigenlijk alleen maar de complexe exponentiële functie $\exp(z)$ door een reeks gedefinieerd, $\cos(z)$ en $\sin(z)$ worden vervolgens door de relaties $\cos(z) = \frac{e^{iz}+e^{-iz}}{2}$ en $\sin(z) = \frac{e^{iz}-e^{-iz}}{2i}$ *gedefinieerd*. Maar voor de toepassingen is uiteindelijk alleen maar de samenhang tussen deze functies belangrijk.

Voor de complexe cosinus en sinus is het uitwerken van de reële en imaginaire delen met iets meer rekenwerk verbonden. We zullen hierbij de hyperbolische functies $\sinh(x)$ en $\cosh(x)$ tegen komen, die we in Wiskunde 1 hebben leren kennen. Voor deze functies geldt:

$$\cosh(x) = \frac{e^{x} + e^{-x}}{2} \quad \text{en} \quad \sinh(x) = \frac{e^{x} - e^{-x}}{2}.$$

We hebben nu:

$$\begin{aligned}
\sin(x+iy) &= \frac{e^{i(x+iy)} - e^{-i(x+iy)}}{2i} = \frac{e^{ix}\cdot e^{-y} - e^{-ix}\cdot e^{y}}{2i} \\
&= \frac{1}{2i}(\cos(x) + i\sin(x))e^{-y} - \frac{1}{2i}(\cos(-x) + i\sin(-x))e^{y} \\
&= \frac{1}{2}(-i\cos(x) + \sin(x))e^{-y} - \frac{1}{2}(-i\cos(-x) + \sin(-x))e^{y} \\
&= \frac{1}{2}(-i\cos(x) + \sin(x))e^{-y} + \frac{1}{2}(i\cos(x) + \sin(x))e^{y} \\
&= \sin(x)\frac{1}{2}(e^{-y} + e^{y}) + i\cos(x)\frac{1}{2}(-e^{-y} + e^{y}) \\
&= \sin(x)\cosh(y) + i\cos(x)\sinh(y).
\end{aligned}$$

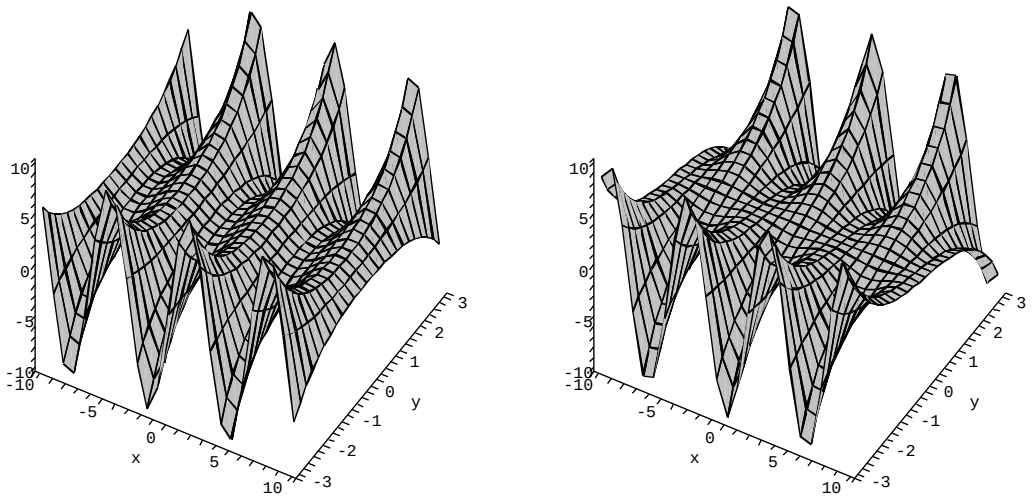
Er geldt dus:

$$\Re(\sin(x+iy)) = \sin(x)\cosh(y) \quad \text{en} \quad \Im(\sin(x+iy)) = \cos(x)\sinh(y).$$

In Figuur I.25 zijn de reële en imaginaire delen van de complexe sinus functie te zien.

Een soortgelijke berekening voor $\cos(z)$ levert het volgende op:

$$\begin{aligned}
\cos(x+iy) &= \frac{e^{i(x+iy)} + e^{-i(x+iy)}}{2} = \frac{e^{ix}\cdot e^{-y} + e^{-ix}\cdot e^{y}}{2} \\
&= \frac{1}{2}(\cos(x) + i\sin(x))e^{-y} + \frac{1}{2}(\cos(-x) + i\sin(-x))e^{y} \\
&= \frac{1}{2}(\cos(x) + i\sin(x))e^{-y} + \frac{1}{2}(\cos(x) - i\sin(x))e^{y} \\
&= \cos(x)\frac{1}{2}(e^{-y} + e^{y}) + i\sin(x)\frac{1}{2}(e^{-y} - e^{y}) \\
&= \cos(x)\cosh(y) - i\sin(x)\sinh(y).
\end{aligned}$$



Figuur I.25: Reëel en imaginair deel van $\sin(z)$

Er geldt dus:

$$\Re(\cos(x+iy)) = \cos(x)\cosh(y) \quad \text{en} \quad \Im(\cos(x+iy)) = \sin(x)\sinh(y).$$

Let op: We zijn gewend dat de reële sinus en cosinus *begrensde* functies zijn, de waarden liggen gewoon tussen -1 en 1. Voor de complexe versies van deze functies geldt dit echter niet meer: Als we in $\cos(z)$ met z langs de imaginaire as lopen, d.w.z. $z = ix$ invullen, hebben we $\cos(ix) = \frac{e^{-x}+e^{x}}{2} = \cosh(x)$ en dit is een onbegrensde functie.

De reden voor dit ongemaak is dat $\sin(z)$ en $\cos(z)$ zich langs de imaginaire as zo gedragen als de exponentiële functie langs de reële as en andersom.

Voor het gemak vatten we de formules voor reëel en imaginair deel van $\exp(z)$, $\cos(z)$ en $\sin(z)$ nog eens samen:

$$\begin{aligned}
\exp(x+iy) &= \exp(x)\cos(y) + i\exp(x)\sin(y);\\
\sin(x+iy) &= \sin(x)\cosh(y) + i\cos(x)\sinh(y);\\
\cos(x+iy) &= \cos(x)\cosh(y) - i\sin(x)\sinh(y).
\end{aligned}$$

Voor de volledigheid merken we nog op, dat ook de hyperbolische functies een voortzetting naar de complexe getallen hebben. Dit gebeurt heel makkelijk door de relaties $\cosh(x) = \frac{e^{x}+e^{-x}}{2}$ en $\sinh(x) = \frac{e^{x}-e^{-x}}{2}$ van reële x naar complexe z uit te breiden, we definiëren dus:

$$\cosh(z) := \frac{e^{z}+e^{-z}}{2} \quad \text{en} \quad \sinh(z) := \frac{e^{z}-e^{-z}}{2}.$$

Ook voor deze functies kunnen we uit de Taylor reeks voor de exponentiële functie machtreeksen afleiden die overeen komen met de reële

Taylor reeksen van $\cosh(x)$ en $\sinh(x)$, er geldt:

$$\cosh(z) = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!} = 1 + \frac{z^2}{2} + \frac{z^4}{24} + \frac{z^6}{720} + \dots$$

$$\sinh(z) = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!} = z + \frac{z^3}{6} + \frac{z^5}{120} + \frac{z^7}{5040} + \dots$$

OPDRACHT 23 *Laat zien dat* $\cosh(iz) = \cos(z)$ *en* $\sinh(iz) = i\sin(z)$.

6.3 Complexe logaritme

Als we een complexe logaritme willen definiëren hebben we (minstens) twee mogelijkheden om hieraan te beginnen. Aan de ene kant hebben we de Taylor reeks voor de reële logaritme en na onze goede ervaringen met deze aanpak zou het gek zijn als we deze reeks niet naar de complexe getallen zouden kunnen voortzetten. De Taylor reeks voor de reële logaritme is $\log(x+1) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n}$, dus kunnen we de complexe logaritme definiëren door

$$\log(z+1) := \sum_{n=1}^{\infty}(-1)^{n+1}\frac{z^n}{n} = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots$$

Het probleem is dat deze reeks niet voor alle waarden van z convergent is, maar alleen maar voor z met $|z| < 1$. We kunnen dus met deze machtreeks de waarden van de complexe logaritme alleen maar in een cirkel van straal 1 rond $z_0 = 1$ uitrekenen.

Aan de andere kant willen we natuurlijk dat de complexe logaritme de omkeerfunctie van de complexe exponentiële functie is, dus dat $\log(e^z) = z$ en $e^{\log(z)} = z$. Beide mogelijkheden leiden uiteindelijk tot hetzelfde resultaat dat we nu vanuit het perspectief van de logaritme als omkeerfunctie van $\exp(z)$ gaan bekijken.

Voor een complex getal $z = re^{i\varphi}$ (in poolcoördinaten) volgt uit de eis $e^{\log(z)} = z$ dat we $\log(z)$ noodzakelijk moeten definiëren door

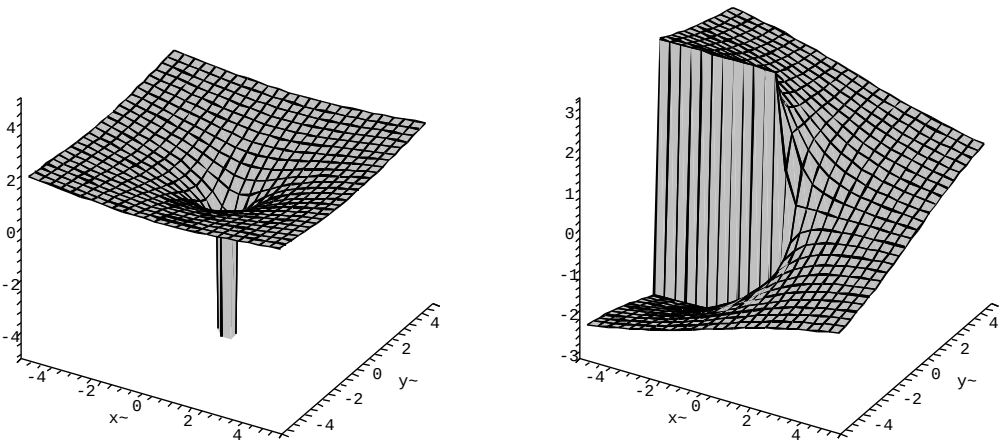
$$\log(re^{i\varphi}) := \log(r) + i\varphi$$

want $\log(z) = x + iy$ moet voldoen aan $re^{i\varphi} = z = e^{\log(z)} = e^{x+iy} = e^x \cdot e^{iy}$, dus hebben we nodig dat $e^x = r$ en $e^{iy} = e^{i\varphi}$ en dus $x = \log(r)$ en $y = \varphi$.

Er is wel een kleine complicatie bij deze definitie: Omdat de complexe exponentiële functie $2\pi i$-periodiek is, geldt ook voor $w = \log(z) + 2\pi i$ dat $e^w = z$, het imaginaire deel van $\log(z)$ is dus alleen maar tot op veelvouden van 2π na bepaald. De exponentiële functie beeld namelijk elke streep $S_a = \{z \in \mathbb{C} \mid \Im(z) \in (a, a+2\pi]\}$ op $\mathbb{C}\setminus\{0\}$ af en in principe is elke streep even goed. De conventie is echter, dat het imaginaire deel van $\log(z)$ in het interval $(-\pi, \pi]$ ligt. We hebben dus

$$\log(z) = \begin{cases} \log(|z|) + i\arg(z) & \text{als } \arg(z) \in [0, \pi] \\ \log(|z|) + i(\arg(z) - 2\pi) & \text{als } \arg(z) \in (\pi, 2\pi). \end{cases}$$

In Figuur I.26 zien we de reële en imaginaire delen van de complexe logaritme. Het is duidelijk dat het imaginaire deel op de negatieve reële as niet continu is, maar een sprong om 2π heeft.



Figuur I.26: Reëel en imaginair deel van $\log(z)$

Omdat we voor de complexe logaritme een keuze moeten maken in welke streep van breedte 2π het imaginaire deel van $\log(z)$ ligt, krijgen we een probleem dat we bij de reële logaritme niet kennen. Kijken we bijvoorbeeld naar $z_1 = z_2 = e^{i\frac{2}{3}\pi}$, dan is duidelijk $\log(z_1) = \log(z_2) = i\frac{2}{3}\pi$ en dus $\log(z_1) + \log(z_2) = i\frac{4}{3}\pi$. Maar $z_1 \cdot z_2 = e^{i(\frac{2}{3}+\frac{2}{3})\pi} = e^{i\frac{4}{3}\pi} = e^{i(-\frac{2}{3})\pi}$ en daarom is $\log(z_1 \cdot z_2) = -i\frac{2}{3}\pi$. De relatie $\log(z_1 \cdot z_2) = \log(z_1) + \log(z_2)$ geldt dus niet meer in elk geval, want de imaginaire delen aan de rechter en linker kant kunnen om veelvouden van 2π verschillen.

Merk op: De regel $\log(x_1 \cdot x_2) = \log(x_1) + \log(x_2)$ voor de reële logaritme verandert voor de complexe logaritme in de regel:

$$\log(z_1 \cdot z_2) = \log(z_1) + \log(z_2) \text{ modulo veelvouden van } 2\pi i.$$

Omdat n-de machten slechts een speciaal geval van producten zijn, geldt ook de regel $\log(z^n) = n \log(z)$ alleen maar modulo veelvouden van $2\pi i$.

Het voordeel van onze keuze van de streep $\Im(z) \in (-\pi, \pi)$ van breedte 2π is, dat de relatie

$$\log(\frac{1}{z}) = -\log(z)$$

nog wel geldt, want deze streep wordt onder de afbeelding $z \to -z$ op zich zelf afgebeeld.

OPDRACHT 24 *Schrijf* $\log(1+i)$, $\log(-i)$ *en* $\log(\frac{2+i}{2-i})$ *in de vorm* $x + iy$.

6.4 Differentiëren via Taylor reeksen

We zullen in de Appendix voor deze les de vraag nagaan, wanneer een functie complex differentieerbaar is. Voor het moment nemen we genoegen ermee dat we zeggen, dat een complexe functie $f(z)$ in het punt z_0 complex differentieerbaar is als de limiet

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

bestaat en onafhankelijk van het traject waarop z tegen z_0 aan loopt dezelfde waarde heeft. We zullen zien dat complexe differentieerbaarheid een veel sterkere eigenschap is dan de gewone differentieerbaarheid bij reële functies.

Definitie: Een functie $f(z)$ die in elk punt $z \in B$ van zijn domein complex differentieerbaar is, heet een (op B) *holomorfe* functie.

De *ontwikkelingsstelling van Cauchy-Taylor* zegt nu dat we een holomorfe functie altijd als een absoluut convergente Taylor reeks kunnen schrijven, dus dat

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \text{ waarbij } \sum_{n=0}^{\infty} |a_n z^n| \text{ convergent is.}$$

Omgekeerd geeft een absoluut convergente Taylor reeks steeds een holomorfe functie aan.

In de wereld van complex differentieerbare functies gaat eigenlijk alles goed, wat we zo maar zouden kunnen hopen, daarom is er ook een stelling die zegt dat we de afgeleide van een holomorfe functie krijgen door de (absoluut convergente) Taylor reeks van de functie termsgewijs af te leiden.

Stelling: Voor de holomorfe functie $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is de afgeleide $f'(z)$ gegeven door

$$f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

We weten dus dat een complexe functie $f(z)$ die door een absoluut convergente Taylor reeks gegeven is, complex differentieerbaar is en dat we de afgeleide $f'(z)$ vinden door de Taylor reeks termsgewijs af te leiden. Maar om een term in een Taylor reeks af te leiden, hebben we alleen maar de afgeleide van z^n nodig, en we zullen in de Appendix laten zien dat net als bij reële functies geldt dat

$$(z^n)' = n z^{n-1}.$$

Met behulp hiervan kunnen we nu de complexe functies afleiden die we tot nu toe hebben gezien:

$$(1) \qquad \exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \dots$$

$$\Rightarrow \exp'(z) = 1 + 2 \cdot \frac{z}{2!} + 3 \cdot \frac{z^2}{3!} + 4 \cdot \frac{z^3}{4!} + \dots = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots$$

$$= \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} = \sum_{n=0}^{\infty} \frac{z^n}{n!} = \exp(z).$$

$$(2)\qquad \sin(z) = \sum_{n=0}^{\infty}(-1)^n \frac{z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + \dots$$

$$\Rightarrow \sin'(z) = 1 - 3\cdot\frac{z^2}{3!} + 5\cdot\frac{z^4}{5!} - 7\cdot\frac{z^6}{7!} + \dots = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots$$

$$= \sum_{n=0}^{\infty}(-1)^n \frac{z^{2n}}{(2n)!} = \cos(z).$$

$$(3)\qquad \cos(z) = \sum_{n=0}^{\infty}(-1)^n \frac{z^{2n}}{(2n)!} = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + \dots$$

$$\Rightarrow \cos'(z) = -2\cdot\frac{z}{2!} + 4\cdot\frac{z^3}{4!} - 6\cdot\frac{z^5}{6!} + \dots = -z + \frac{z^3}{3!} - \frac{z^5}{5!} + \dots$$

$$= -\sum_{n=0}^{\infty}(-1)^n \frac{z^{2n+1}}{(2n+1)!} = -\sin(z).$$

$$(4)\qquad \log(z+1) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{z^n}{n} = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + \dots$$

$$\Rightarrow \log'(z+1) = 1 - 2\cdot\frac{z}{2} + 3\cdot\frac{z^2}{3} - 4\cdot\frac{z^3}{4} + \dots = 1 - z + z^2 - z^3 + \dots$$

$$= \sum_{n=0}^{\infty}(-1)^n z^n = \frac{1}{1+z} \text{ (meetkundige reeks).}$$

De meetkundige reeks $\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \dots$ is convergent als $|x| < 1$ en heeft in dit geval de waarde $\frac{1}{1-x}$. Dit ziet men in door uit te werken dat $(1 + x + x^2 + x^3 + \dots + x^n)(1 - x) = 1 - x^{n+1}$. Voor $|x| < 1$ gaat x^{n+1} voor $n \to \infty$ naar 0, dus is $(1 + x + x^2 + x^3 + \dots)(1 - x) = 1$.

We zien dus dat de afgeleiden precies zo zijn als we dat volgens de eigenschappen van de reële functies zouden verwachten, we hoeven alleen maar de reële variabel x door de complexe variabel z te vervangen. De resultaten van deze sectie laten zich dus in een enkele regel samenvatten die eigenlijk alleen maar bekende relaties bevat:

$$\exp(z)' = \exp(z), \quad \sin(z)' = \cos(z), \quad \cos(z)' = -\sin(z), \quad \log(z+1)' = \frac{1}{z+1}.$$

6.5 Appendix: Complexe differentieerbaarheid

Bij reële functies hadden we de afgeleide $f'(x_0)$ in een punt x_0 gedefinieerd als de stijging van de raaklijn in het punt x_0 aan de grafiek van $f(x)$. Voor een complexe functie hebben we al gezien, dat we de reële en imaginaire delen van de functie apart als driedimensionale landschappen (grafieken) kunnen representeren. In een punt van zo'n landschap kunnen we wel een raak*vlak* definiëren, maar het is onduidelijk hoe we uit de raakvlakken voor reëel en imaginair deel van de functie een complex getal zullen maken die we als afgeleide van de functie in dit punt definiëren.

Maar de eigenschap dat de afgeleide de stijging van de raaklijn aangeeft kunnen we ook nog iets anders formuleren: De functie $f(x)$ wordt in een kleine omgeving van een punt goed door de raaklijn benaderd, we noemen daarom de afgeleide ook de *linearisering* van de functie. Dit volgt uit de definitie dat

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \text{ als deze limiet bestaat.}$$

Als we namelijk de definitie van de afgeleide voor kleine waarden van $\Delta x = h$ (en zonder limiet) bekijken, hebben we

$$f'(x) \approx \frac{f(x+\Delta x)-f(x)}{\Delta x} \text{ en dus } f(x+\Delta x) \approx f(x) + f'(x)\Delta x.$$

In een kleine omgeving van x wordt de functie dus goed beschreven door een vermenigvuldiging met $f'(x)$. Preciezer gezegd beeld de functie het punt x naar $f(x)$ af en een afwijking Δx van x wordt door de functie met $f'(x)$ vermenigvuldigd en bij $f(x)$ opgeteld.

Deze interpretatie nemen we nu over als definitie van de complexe afgeleide: De afgeleide $f'(z)$ geeft aan, dat we in een (kleine) omgeving van z de functiewaarden van $f(z)$ kunnen benaderen door een afwijking Δz van z met $f'(z)$ te vermenigvuldigen en bij $f(z)$ op te tellen:

$$f(z+\Delta z) \approx f(z) + f'(z)\Delta z.$$

Dit kunnen we ook zuiver meetkundig interpreteren, want we weten wat vermenigvuldiging met een complex getal $f'(z) = re^{i\varphi}$ betekent, namelijk een schaling met een factor r en een draaiing om φ. In een omgeving van z wordt een complex differentieerbare functie $f(z)$ dus beschreven door een schaling gecombineerd met een draaiing.

De interpretatie van de complexe afgeleide als linearisering van $f(z)$ in een kleine omgeving maakt het noodzakelijk dat we de definitie van de reële afgeleide via de limiet $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ letterlijk overnemen voor complexe functies, waarbij we (volgens de conventies) de reële variabel x door een complexe variabel z vervangen. We krijgen dus de volgende definitie:

Definitie: Een complexe functie $f(z)$ heet in het punt z_0 (complex) *differentieerbaar*, als de limiet

$$\lim_{h\to 0} \frac{f(z_0+h)-f(z_0)}{h}$$

bestaat. In dit geval noteren we de afgeleide in het punt z_0 door $f'(z_0)$. Een functie $f(z)$ die in elk punt van zijn domein differentieerbaar is, heet ook een *holomorfe functie* of een *analytische functie*.

Het cruciale punt bij deze definitie is het *bestaan* van de limiet. Voor reële functies kan h alleen maar van links of van rechts naar 0 toe lopen. Dan is het voldoende als de limiet van links en van rechts bestaat en deze twee limieten hetzelfde zijn. Zo zien we bijvoorbeeld dat de functie $f(x) = |x|$ in het punt 0

niet differentieerbaar is omdat de limiet van $\frac{f(x+h)-f(x)}{h}$ voor $h > 0$ (dus van rechts) gelijk is aan 1 terwijl de limiet voor $h < 0$ (dus van links) gelijk is aan -1. Maar we hoeven inderdaad niet meer te doen dan van links en van rechts te kijken.

Voor complexe functies is dit een heel ander verhaal, want h kan van rechts of links op de reële as naar 0 lopen, maar ook van boven of beneden op de imaginaire as of langs een willekeurige lijn met $\Im(z) = a \cdot \Re(z)$. En h mag zelfs langs een heel kromme lijn lopen, bijvoorbeeld langs een spiraal die zich om het nulpunt wikkelt. En voor elk van de mogelijke trajecten van h moet de limiet bestaan en steeds dezelfde waarde hebben. Het feit dat h op een willekeurig traject naar 0 toe mag lopen maakt van de complexe differentieerbaarheid een heel sterke eigenschap die vergaande consequenties heeft.

> De complexe differentieerbaarheid heeft een aantal indrukwekkende consequenties. Bijvoorbeeld volgt uit de samenhang tussen holomorfe functies en hun Taylor reeksen de *Stelling van Liouville* die zegt dat een op $\mathbb{C}$ differentieerbare functie alleen maar begrensd kan zijn als hij constant is. We hadden al gezien dat de complexe sinus en cosinus functies langs de imaginaire as tegen oneindig gaan. De stelling van Liouville zegt nu dat globaal begrensde functies zo als de reële sinus of cosinus functies op het complexe vlak niet kunnen bestaan.
>
> Er is echter nog een veel sterker resultaat: De complexe exponentiële functie heeft alle complexe getallen als waarden behalve van 0. Dit is inderdaad voor alle holomorfe functies zo, want één van de stellingen van Picard (Charles Emile Picard, niet Jean-Luc) zegt dat een holomorfe functie die twee complexe getallen niet als waarde heeft noodzakelijk een constante functie $f(z) = c$ is.

Het voordeel ervan, de afgeleide van een complexe functie net zo te definiëren als voor reële functies, is dat de *rekenregels* voor de afgeleide hetzelfde blijven. Als $f(z)$ en $g(z)$ complex differentieerbare functies zijn, geldt dus:

$$
\begin{aligned}
&(f+g)'(z) = f'(z) + g'(z) \\
&(f \cdot g)'(z) = f'(z)g(z) + f(z)g'(z) \text{ (productregel)} \\
&\left(\frac{f}{g}\right)'(z) = \frac{f'(z)g(z) - f(z)g'(z)}{g(z)^2} \text{ (quotiëntregel)} \\
&(f \circ g)'(z) = f(g(z))' = f'(g(z))g'(z) \text{ (kettingregel)}
\end{aligned}
$$

Tot nu toe hebben we nog geen enkele complex differentieerbare functie gezien. Maar we hebben wel al een gok op de afgeleide van $f(z) = z^n$ gedaan, namelijk dat hiervoor $f'(z) = nz^{n-1}$ is, net als we dat van de reële functies gewend zijn. Voor de reële functies hebben we dit in Wiskunde 1 met behulp van de productregel per volledige inductie bewezen. Dit is wel elegant, maar een rechtstreekse berekening doet ook geen kwaad. Hierbij hebben we de binomische formule voor $(z+h)^n$ nodig, te weten

$$(z+h)^n = \sum_{k=0}^{n} \binom{n}{k} z^{n-k}h^k = z^n + nz^{n-1}h + \frac{n(n-1)}{2}z^{n-2}h^2 + \ldots + nzh^{n-1} + h^n.$$

Voor $f(z) = z^n$ geldt dus

$$\begin{aligned}\frac{f(z+h)-f(z)}{h} &= \frac{(z+h)^n - z^n}{h} \\ &= \frac{1}{h}(z^n + nz^{n-1}h + \frac{n(n-1)}{2}z^{n-2}h^2 + \ldots + nzh^{n-1} + h^n - z^n) \\ &= nz^{n-1} + \frac{n(n-1)}{2}z^{n-2}h + \ldots + nzh^{n-2} + h^{n-1}.\end{aligned}$$

In de laatste som bevat elke term vanaf de tweede een macht van h, en als we de limiet $h \to 0$ bekijken gaat dus ieder van deze termen naar 0. Merk op dat dit onafhankelijk van het traject is waarop h naar 0 loopt. Daarom bestaat de limiet $\lim_{h\to 0} \frac{f(z+h)-f(z)}{h}$ en we hebben bewezen:

Stelling: Voor $f(z) = z^n$ is $f'(z) = \lim\limits_{h\to 0} \dfrac{f(z+h)-f(z)}{h} = nz^{n-1}$.

Contrastvoorbeeld

Dat er bij complexe functies snel iets mis kan gaan zien we bij een heel eenvoudige, onschuldige functie, de complexe conjugatie

$$f(z) := \overline{z}.$$

We laten h eerst langs de reële as lopen, het maakt niet uit of van rechts of links. Voor $h \in \mathbb{R}$ geldt $\frac{\overline{z+h}-\overline{z}}{h} = \frac{\overline{z}+h-\overline{z}}{h} = \frac{h}{h} = 1$, dus is ook de limiet $h \to 0$ gelijk aan 1 als we langs de reële as lopen. Dit is natuurlijk geen verrassing, want op de reële as doet complexe conjugatie niets, en moet dus dezelfde afgeleide hebben als de reële functie $f(x) = x$.

Nu laten we h langs de imaginaire as lopen, hiervoor nemen we $h = i\varepsilon$ met $\varepsilon \in \mathbb{R}$. Er geldt $\frac{\overline{z+i\varepsilon}-\overline{z}}{i\varepsilon} = \frac{\overline{z}-i\varepsilon-\overline{z}}{i\varepsilon} = \frac{-i\varepsilon}{i\varepsilon} = -1$, dus is de limiet gelijk aan -1 als we langs de imaginaire as lopen.

We kunnen zelfs een willekeurig complex getal op de eenheidscirkel als limiet produceren. Als we langs de lijn vanuit het getal $e^{i\varphi}$ naar 0 lopen, hebben we $h = e^{i\varphi} \cdot \varepsilon$ met $\varepsilon \in \mathbb{R}$. Dan is

$$\frac{\overline{z+h}-\overline{z}}{h} = \frac{\overline{z+e^{i\varphi}\varepsilon}-\overline{z}}{e^{i\varphi}\varepsilon} = \frac{\overline{z}+e^{-i\varphi}\varepsilon-\overline{z}}{e^{i\varphi}\varepsilon} = \frac{e^{-i\varphi}\varepsilon}{e^{i\varphi}\varepsilon} = e^{-2i\varphi},$$

dus is de limiet op dit traject gelijk aan $e^{-2i\varphi}$. Maar we kunnen elk getal op de eenheidscirkel als $e^{-2i\varphi}$ schrijven, want als φ van 0 naar $-\pi$ loopt, loopt $e^{-2i\varphi}$ een keer langs de eenheidscirkel.

Cauchy-Riemann differentiaalvergelijkingen

Het voorbeeld van de complexe conjugatie is een beetje verontrustend, want het is toch heel omslachtig om steeds te testen of de limiet voor alle mogelijke trajecten waarop h naar 0 gaat hetzelfde is. Gelukkig is dit echter niet nodig, er is een stelling die zegt dat het voldoende is om langs de reële en langs de imaginaire as te kijken. Deze stelling gaan we hier niet bewijzen, maar we zullen wel een (hopelijk overtuigende) motivatie geven.

We hebben al eerder gezien dat het handig is om apart naar de reële en imaginaire delen van een complexe functie te kijken, want hiervoor kunnen we 3-dimensionale plaatjes maken. Als we een complex getal $z \in \mathbb{C}$ in de vorm $z = x + iy$ met $x, y \in \mathbb{R}$ schrijven, kunnen we een complexe functie $f(z)$ zien als een functie van twee reële variabelen, namelijk van $\Re(z)$ en $\Im(z)$. We kunnen dus een complexe functie $f(z)$ beschrijven door twee reële functies van de twee reële variabelen $x = \Re(z)$ en $y = \Im(z)$, namelijk

$$f(z) = \Re(f(z)) + i \cdot \Im(f(z)) = u(x, y) + i \cdot v(x, y).$$

Als we nu ervan uitgaan dat $f(z)$ een complex differentieerbare functie met afgeleide $f'(z)$ is, kunnen we kijken wat dit voor de (reële) functies $u(x, y)$ en $v(x, y)$ betekent. In het bijzonder gaan we bekijken wat er gebeurt als we langs de reële as en langs de imaginaire as naar 0 lopen. Eerst lopen we langs de reële as met $h \to 0$ voor $h \in \mathbb{R}$, dan is

$$\begin{aligned} f'(z) &= \lim_{h\to 0} \frac{f(z+h) - f(z)}{h} \\ &= \lim_{h\to 0} \frac{(u(x+h, y) + i \cdot v(x+h, y)) - (u(x, y) + i \cdot v(x, y))}{h} \\ &= \lim_{h\to 0} \frac{u(x+h, y) - u(x, y)}{h} + i \cdot \lim_{h\to 0} \frac{v(x+h, y) - v(x, y)}{h} \\ &= \frac{\partial u(x, y)}{\partial x} + i \cdot \frac{\partial v(x, y)}{\partial x}. \end{aligned}$$

Het afleiden langs de reële as komt dus overeen met de partiële afgeleide naar het reële deel x van z, waarbij we y als een constante beschouwen.

Nu lopen we langs de imaginaire as met $ih \to 0$ voor $h \in \mathbb{R}$, dan geldt

$$\begin{aligned} f'(z) &= \lim_{h\to 0} \frac{f(z+ih) - f(z)}{ih} = (-i) \cdot \lim_{h\to 0} \frac{f(z+ih) - f(z)}{h} \\ &= (-i) \cdot \lim_{h\to 0} \frac{(u(x, y+h) + i \cdot v(x, y+h)) - (u(x, y) + i \cdot v(x, y))}{h} \\ &= (-i) \cdot \lim_{h\to 0} \frac{u(x, y+h) - u(x, y)}{h} + (-i) \cdot i \cdot \lim_{h\to 0} \frac{v(x, y+h) - v(x, y)}{h} \\ &= -i \cdot \frac{\partial u(x, y)}{\partial y} + \frac{\partial v(x, y)}{\partial y}. \end{aligned}$$

Maar als $f(z)$ differentieerbaar is, weten we dat de twee limieten gelijk moeten zijn, daarom hebben we de noodzakelijke voorwaarde dat de reële en imaginaire delen van de limieten hetzelfde zijn, dus

$$\frac{\partial u(x, y)}{\partial x} = \frac{\partial v(x, y)}{\partial y} \quad \text{en} \quad \frac{\partial u(x, y)}{\partial y} = -\frac{\partial v(x, y)}{\partial x}.$$

Deze twee noodzakelijke voorwaarden heten de *Cauchy-Riemann differentiaalvergelijkingen*.

Het belangrijke (en misschien iets verrassende) punt is nu dat deze voorwaarde ook voldoende is, d.w.z. een complexe functie waarvoor de reële en

imaginaire delen aan de Cauchy-Riemann differentiaalvergelijkingen voldoen is complex differentieerbaar.

Stelling: Een complexe functie $f(z) = f(x+iy) = u(x,y) + iv(x,y)$ is complex differentieerbaar dan en slechts dan als $u(x,y)$ en $v(x,y)$ continu differentieerbaar zijn en voldoen aan de Cauchy-Riemann differentiaalvergelijkingen

$$\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y} \text{ en } \frac{\partial u(x,y)}{\partial y} = -\frac{\partial v(x,y)}{\partial x}.$$

Omdat we de limiet van $\frac{f(z+h)-f(z)}{h}$ langs de reële as en langs de imaginaire as al in de partiële afgeleiden van $u(x,y)$ en $v(x,y)$ hebben uitgedrukt, kunnen we de waarde van de afgeleide $f'(z)$ expliciet aangeven, namelijk door

$$\begin{aligned} f'(z) &= \frac{\partial u(x,y)}{\partial x} + i \cdot \frac{\partial v(x,y)}{\partial x} \\ &= \frac{\partial v(x,y)}{\partial y} - i \cdot \frac{\partial u(x,y)}{\partial y} \\ &= \frac{\partial u(x,y)}{\partial x} - i \cdot \frac{\partial u(x,y)}{\partial y} \\ &= \frac{\partial v(x,y)}{\partial y} + i \cdot \frac{\partial v(x,y)}{\partial x}. \end{aligned}$$

Voorbeeld: We passen het criterium van de Cauchy-Riemann differentiaalvergelijkingen op de complexe exponentiële functie toe. We hebben gezien dat voor $z = x + iy$ geldt dat

$$\exp(z) = \exp(x)\cos(y) + i \cdot \exp(x)\sin(y),$$

dus hebben we:

$$\exp(x+iy) = u(x,y) + i \cdot v(x,y) \text{ met}$$

$$u(x,y) = \exp(x)\cos(y) \quad \text{en} \quad v(x,y) = \exp(x)\sin(y).$$

Om de Cauchy-Riemann differentiaalvergelijkingen te testen moeten we nu de partiële afgeleiden van $u(x,y)$ en $v(x,y)$ berekenen. Er geldt:

$$\begin{aligned} \frac{\partial u(x,y)}{\partial x} &= \exp(x)\cos(y), \quad \frac{\partial u(x,y)}{\partial y} = -\exp(x)\sin(y), \\ \frac{\partial v(x,y)}{\partial x} &= \exp(x)\sin(y), \quad \frac{\partial v(x,y)}{\partial y} = \exp(x)\cos(y) \end{aligned}$$

en we zien dat inderdaad $\frac{\partial u(x,y)}{\partial x} = \frac{\partial v(x,y)}{\partial y}$ en $\frac{\partial u(x,y)}{\partial y} = -\frac{\partial v(x,y)}{\partial x}$. We concluderen dat de complexe exponentiële functie complex differentieerbaar is met afgeleide $\exp'(z) = \frac{\partial u(x,y)}{\partial x} + i \cdot \frac{\partial v(x,y)}{\partial x} = \exp(x)\cos(y) + i \cdot \exp(x)\sin(y) = \exp(z)$.

OPDRACHT 25 *Ga na dat* $\sin(z)$ *en* $\cos(z)$ *aan de Cauchy-Riemann differentiaalvergelijkingen voldoen en dus complex differentieerbaar zijn. (Hint: De reële en imaginaire delen van de complexe sinus en cosinus functies hebben we bepaald, voor de afgeleiden van de reële hyperbolische functies geldt* $\cosh'(x) = \sinh(x)$ *en* $\sinh'(x) = \cosh(x)$*.)*

Merk op: Met onze kennis over functies van meerdere veranderlijken kunnen we de Cauchy-Riemann differentiaalvergelijkingen ook op een iets andere manier afleiden: Als we het complexe getal $z = x + iy$ als 2-dimensionale vector $z = \begin{pmatrix} x \\ y \end{pmatrix}$ schrijven, wordt de vermenigvuldiging van z met een complex getal $a + ib$ beschreven door de 2×2-matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, want $(a + ib)(x + iy) = (ax - by) + i(bx + ay)$ en er geldt

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix}.$$

Vermenigvuldiging met een complex getal wordt dus beschreven door de speciale 2×2-matrices van de vorm $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$.

We kunnen nu een complexe functie $f(z)$ opvatten als functie van de twee reële variabelen x en y met $z = x+iy$, die gegeven is door de twee componenten $u(x, y) = \Re(f(z))$ en $v(x, y) = \Im(f(z))$, dus als functie van de vorm

$$f(x, y) = (u(x, y), v(x, y)).$$

Maar voor zo'n functie hadden we gezien, dat de afgeleide van $f(x, y)$ een lineaire afbeelding $\mathbb{R}^2 \to \mathbb{R}^2$ gegeven door een 2×2-matrix is, namelijk door de Jacobi matrix J met de partiële afgeleiden:

$$J = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}.$$

In het algemeen is J een willekeurige lineaire afbeelding $\mathbb{R}^2 \to \mathbb{R}^2$, en alleen maar in het geval dat J van de vorm $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ is, is de afgeleide de vermenigvuldiging met een complex getal en dit is juist het geval als $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ en $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$, dus als $u(x, y)$ en $v(x, y)$ aan de Cauchy-Riemann differentiaalvergelijkingen voldoen.

Niet elke functie $u(x, y)$ kan reëel of imaginair deel van een holomorfe functie $f(z)$ zijn. Door toepassen van de Cauchy-Riemann differentiaalvergelijkingen op $u(x, y) = \Re(f(z))$ vinden we namelijk: $\frac{\partial^2 u(x,y)}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial u(x,y)}{\partial x} = \frac{\partial}{\partial x}\frac{\partial v(x,y)}{\partial y} = \frac{\partial^2 v(x,y)}{\partial x \partial y}$. Volgens de stelling van Schwarz mogen we partiële afgeleiden verruilen, dus is $\frac{\partial^2 v(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\frac{\partial v(x,y)}{\partial x} = -\frac{\partial}{\partial y}\frac{\partial u(x,y)}{\partial y} = -\frac{\partial^2 u(x,y)}{\partial y^2}$. We hebben dus $\frac{\partial^2 u(x,y)}{\partial x^2} + \frac{\partial^2 u(x,y)}{\partial y^2} = 0$. Met een analoge berekening vinden we dezelfde relatie ook voor $v(x, y) = \Im(f(z))$, dus $\frac{\partial^2 v(x,y)}{\partial x^2} + \frac{\partial^2 v(x,y)}{\partial y^2} = 0$.

Algemeen heten functies met de eigenschap $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$ *harmonische functies* of, gemotiveerd door de natuurkunde, *potentieelfuncties*. We hebben dus gezien dat alleen maar harmonische functies reëel of imaginair deel van een holomorfe functie $f(z)$ kunnen zijn.

Opdracht 26

(i) In welke punten $z \in \mathbb{C}$ is $f(z) := \Re(z)^2 + i \cdot \Im(z)^2$ complex differentieerbaar? Bepaal in deze punten $f'(z)$.

(ii) In welke punten $z \in \mathbb{C}$ is $f(z) := \overline{z}(3z^2 + \overline{z}^2)$ complex differentieerbaar? Bepaal in deze punten $f'(z)$.

Belangrijke begrippen in deze les

- complexe exponentiële functie $\exp(z)$
- $2\pi i$-periodiciteit van de complexe exponentiële functie
- complexe sinus en cosinus functies $\sin(z)$ en $\cos(z)$
- reële en imaginaire delen van $\exp(z)$, $\sin(z)$, $\cos(z)$
- complexe logaritme $\log(z)$
- termsgewijs afleiden van Taylor reeksen
- complexe differentieerbaarheid
- holomorfe functies
- Cauchy-Riemann differentiaalvergelijkingen

Opgaven

54. Bepaal voor de afbeelding $f(z) := z^2$ de beelden van de lijnen

    (i) $L_1 := \{z = x + iy \in \mathbb{C} \mid x = 2y\}$,
    (ii) $L_2 := \{z = x + iy \in \mathbb{C} \mid x = 2\}$,
    (iii) $L_3 := \{z = x + iy \in \mathbb{C} \mid y = -1\}$.

    Teken de beelden van de lijnen in het complexe vlak.

55. Bepaal het beeld van de rechthoek $R = \{x + iy \in \mathbb{C} \mid x \in [-1, 1], y \in [\frac{1}{2}, 1]\}$ onder de complexe exponentiële functie. Maak een schets. Kan je algemeen aangeven wat het beeld van een rechthoek $\{z \in \mathbb{C} \mid \Re(z) \in [a, b], \Im(z) \in [c, d]\}$ met $a, b, c, d \in \mathbb{R}$ is?

56. Laat zien dat de nulpunten van $\sin(z)$ alle reëel zijn, d.w.z. dat $\sin(z) \neq 0$ als $\Im(z) \neq 0$. Ga na dat hetzelfde ook voor $\cos(z)$ geldt.

57. Laat zien dat voor $z = x+iy$ geldt dat $\Re(\cosh(z)) = \cosh(x)\cos(y)$ en $\Im(\cosh(z)) = \sinh(x)\sin(y)$.

    Bepaal ook $\Re(\sinh(z))$ en $\Im(\sinh(z))$.

58. Vind de oplossingen in $\mathbb{C}$ voor de volgende vergelijkingen:

    (i) $e^z = i$, (ii) $e^z = 1 + i$ (iii) $\cos(z) = -3$.

59. We bekijken de afbeelding $f(z) := e^{iz}$.

    (i) Bepaal voor een vaste $w \in \mathbb{C}$ de waarden van z met $f(z) = w$.

    (ii) Bepaal een deel $D \subseteq \mathbb{C}$ van het complexe vlak zo dat $f(z)$ op D een omkeerfunctie heeft. Geef de omkeerfunctie aan.

60. Gebruik de relaties $\cos(z) = \frac{e^{iz}+e^{-iz}}{2}$ en $\sin(z) = \frac{e^{iz}-e^{-iz}}{2i}$ om de afgeleiden $\cos'(z) = -\sin(z)$ en $\sin'(z) = \cos(z)$ rechtstreeks uit de afgeleide van $\exp(z)$ te berekenen (zonder Taylor reeksen of partiële afgeleiden). Let op dat volgens de kettingregel $(e^{iz})' = i \cdot e^{iz}$.

61. De *arcustangens* functie heeft in $z_0 = 0$ de Taylor reeks

$$\arctan(z) = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{2n+1} = z - \frac{z^3}{3} + \frac{z^5}{5} - \frac{z^7}{7} + \ldots.$$

    Laat zien dat $\arctan'(z) = \frac{1}{1+z^2}$.

62. Schrijf voor $z = x + iy$ de volgende functies in de vorm $f(z) = u(x, y) + iv(x, y)$ waarbij $u(x, y)$ en $v(x, y)$ reële functies zijn:

    (i) $f(z) = z^2 + 2iz$;

    (ii) $f(z) = \frac{z}{3+z}$;

    (iii) $f(z) = \exp(z^2)$;

    (iv) $f(z) = \log(1 + z)$.