Wiskunde 2 voor kunstmatige intelligentie (BKI 316)

Bernd Souvignier

najaar 2005

Deel I

Voortgezette Analyse

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 die 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 $h(r, t) = \sin(x + t)\, e^{-t}$.

(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 maar 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 algemeen geval, waarbij de functie van meerdere parameters afhangt en ook meerdere waarden oplevert. Een voorbeeld van dit soort functies zijn de lineaire functies van de n-dimensionale vectorruimte naar de 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 (ii) 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

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

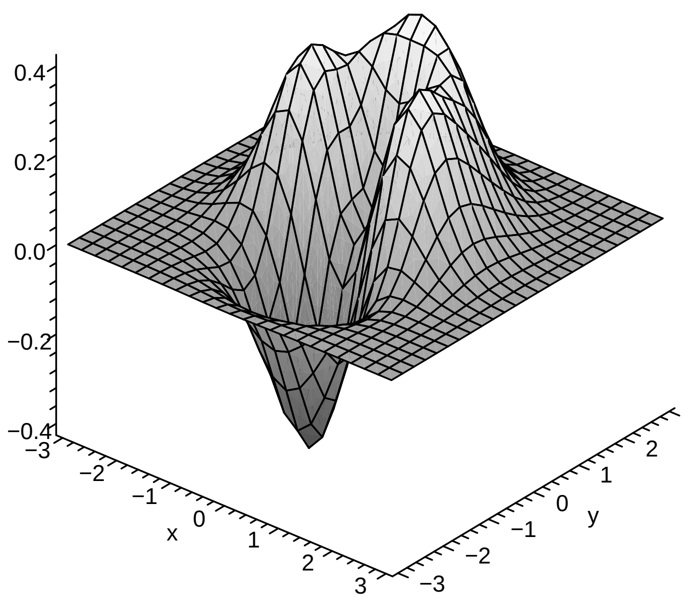
heeft de componenten

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

Door naar de m componenten $f_i(x_1,\ldots,x_n)$ te kijken kunnen we ons dus meestal beperken tot het geval (ii) 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 $(x,y,f(x,y))$ 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, waarvan de functiewaarden alle in het interval rond $f(x)$ liggen.

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 *cosinus*. 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 $e^{x^2+y^2}$ of $\sin(xy - z)$ 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 $\log(1 - x^2 - y^2)$ alleen maar continu op het gebied waar $f(x, y) = 1 - x^2 - y^2 > 0$ 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 de nulpunt precies gedraagt.

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 $f(x,y)$ met $f(x,x) := 2x$ tot de lijn met $x = y$ 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

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

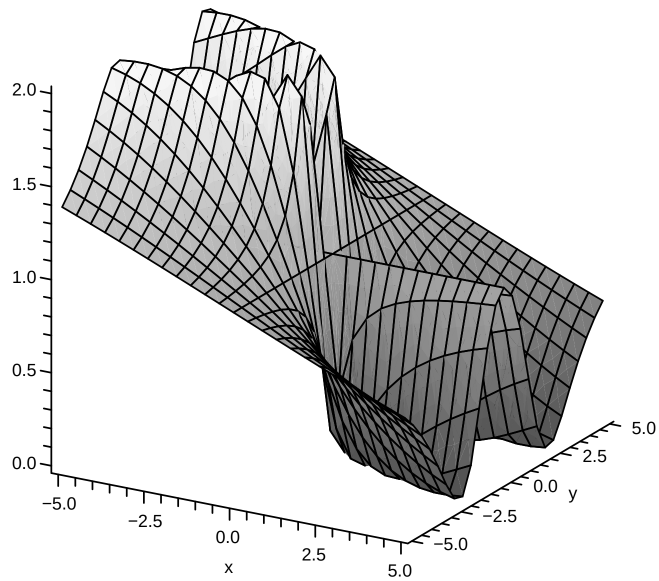
waarvan de grafiek in Figuur I.2 te zien valt.

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 $y = cx$ 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

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



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

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 hier verder geen aandacht meer aan besteden.

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 $(x, y, z = f(x, y))$ in de 3-dimensionale ruimte, net zo als we de grafiek van een gewone functie als de verzameling van punten $(x, y = f(x))$ 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).

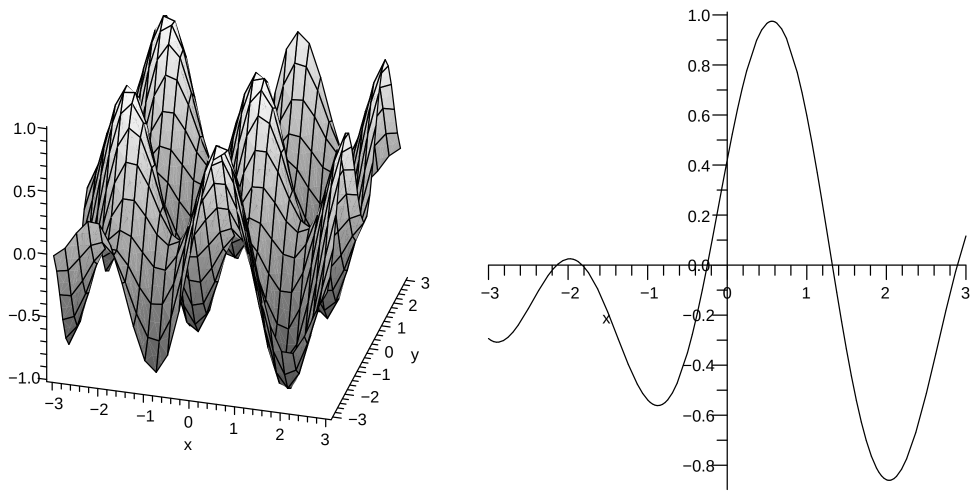
De plaatjes in Figuur I.3 laten links de functie $f(x, y) := \sin(2x+y)\cos(\frac{x}{2} - y)$ en rechts de doorsnede door deze grafiek voor $y = \frac{1}{2}$ zien, dus de functie $g(x) := f(x, \frac{1}{2}) = \sin(2x + \frac{1}{2})\cos(\frac{x}{2} - \frac{1}{2})$.

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 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



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

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

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

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 goedachtige 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 $f(x, y, z) := z\, e^{xy} + yz^3 x^2$ 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:

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

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

Merk op: 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 inzicht 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

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

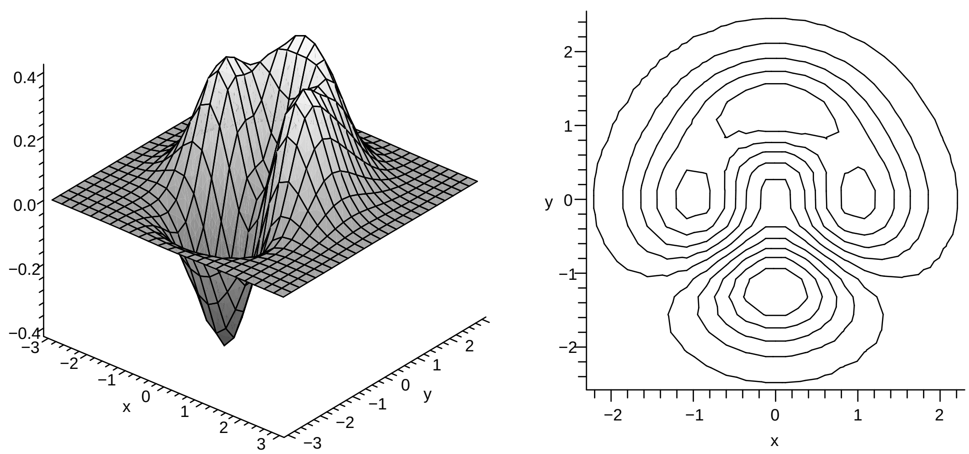
waarbij φ de hoek tussen de vectoren ∇f en $\mathbf{v}$ is. Maar omdat de richtingsvectoren $\mathbf{v}$ alle dezelfde lengte $\|\mathbf{v}\| = 1$ hebben, geldt $\nabla f \cdot \mathbf{v} = \|\nabla f\| \cdot \cos(\varphi)$ en dit is maximaal als $\cos(\varphi) = 1$, dus als $\varphi = 0$. Het inproduct is dus maximaal als $\mathbf{v}$ precies in de richting van ∇f wijst. Omgekeerd is het inproduct minimaal als $\cos(\varphi) = -1$, dus als $\varphi = 180°$ 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:

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

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 0. 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 stapsgrote 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 stapsgrote $\Delta\mathbf{x}$ 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. De 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 $\left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)(\mathbf{v}) = (\nabla f)^{tr}(\mathbf{v})$ 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 $f(x,y) := (xy\,e^{xy}, x\sin(y), 5xy^2)$.