

De computer in de wiskunde: breekijzer of oud schroot?

Syllabus Vakantiecursus 2017

Eindhoven, 25 en 26 augustus 2017

Amsterdam, 1 en 2 september 2017



De computer in de wiskunde: breekijzer of oud schroot?

Syllabus Vakantiecursus 2017

Eindhoven, 25 en 26 augustus 2017

Amsterdam, 1 en 2 september 2017

Programmacommissie

prof. dr. Frits Beukers (UU)

drs. Joke Blom (CWI)

drs. Swier Garst (PWN)

prof. dr. Wil Schilders (PWN, TU/e)

dr. Jeroen Spandaw (TUD)

dr. Marco Swaen (UvA)

drs. Kees Temme (Gymnasium Hilversum, UVA)

dr. Benne de Weger (TU/e) (eindredactie syllabus)

prof. dr. Jan Wiegerinck (UvA) (voorzitter)

e-mail: vakantiecursus@platformwiskunde.nl

Platform Wiskunde Nederland

Science Park 123, 1098 XG Amsterdam

Telefoon: 020-592 4006

Website: <http://www.platformwiskunde.nl>

Vakantiecursus 2017

De Vakantiecursus Wiskunde voor leraren in de exacte vakken in HAVO, VWO, HBO en andere belangstellenden is een initiatief van de Nederlandse Vereniging van Wiskundeleraren, en wordt georganiseerd door het Platform Wiskunde Nederland. De cursus wordt sinds 1946 jaarlijks gegeven op het Centrum Wiskunde en Informatica te Amsterdam, en later ook aan de Technische Universiteit Eindhoven.

Deze cursus wordt mede mogelijk gemaakt door een subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), en een bijdrage van 4TU.AMI, het toegepaste wiskunde-instituut van de 4 Nederlandse technische universiteiten. Organisatie vindt plaats in nauwe samenwerking met het Centrum voor Wiskunde en Informatica (CWI) en de Technische Universiteit Eindhoven (TU/e).

De presentaties van de sprekers zullen zo veel mogelijk beschikbaar komen op de PWN-website: <https://www.platformwiskunde.nl>.

Met dank aan

Ondersteuning PWN: Sjoukje Talsma.

Ondersteuning TU/e: Anita Klooster.

Historie

De eerste vakantiecursus wordt in het jaarverslag 1946 van het Mathematisch Centrum als volgt vermeld:

Op 29 en 31 Oct. '46 werd onder auspiciën van het M.C. een druk bezochte en uitstekend geslaagde vacantiecursus gehouden voor wiskundeleeraren in Nederland. Op 29 October stond de wiskunde, op 31 October de didactiek van de wiskunde op de voorgrond. De sprekers waren: Prof.Dr. O. Bottema, "De prismoïde", Dr. A. Heyting, "Punten in het oneindige", Mr. J. v. IJzeren, "Abstracte Meetkunde en haar betekenis voor de Schoolmeetkunde.", Dr. H.D. Kloosterman, "Ontbinding in factoren", Dr. G. Wielenga, "Is wiskunde-onderwijs voor alpha's noodzakelijk?", Dr. J. de Groot, "Het scheppend vermogen van den wiskundige" en Dr. N.L.H. Bunt, "Moelijkheden van leerlingen bij het beginnend onderwijs in de meetkunde".

Aan het einde van de vacantiecursus werden diverse zaken besproken die het wiskunde-onderwijs in Nederland betroffen. Een Commissie werd ingesteld, die het M.C. over de verder te organiseren vakantiecursussen van advies zou dienen. Hierin namen zitting een vertegenwoordiger van de Inspecteurs van het V.H. en M.O. benevens vertegenwoordigers van de lerarenverenigingen Wimecos en Liwenagel.

Ook werd naar aanleiding van "wensen" die tijdens de cursus naar voren gekomen waren ingesteld: "een colloquium over moderne Algebra, een dispuut over de didactiek van de wiskunde, beiden hoofdzakelijk bedoeld voor de leeraren uit Amsterdam en omgeving, terwijl tevens vanwege het M.C. een cursus over Getallenleer werd toegezegd te geven door de heeren v.d. Corput en Koksma. (Colloquium, dispuut en cursus zijn in 1947 gestart en verheugen zich in blijvende belangstelling).

Docenten

prof. dr. A.M. Cohen

Technische Universiteit Eindhoven, Faculteit Wiskunde en Informatica,
Postbus 513, 5600 MB Eindhoven
e-mail: a.m.cohen@tue.nl

dr. S.R. Dahmen

Vrije Universiteit, Faculteit der Bètawetenschappen, De Boelelaan 1085,
1081 HV Amsterdam
e-mail: s.r.dahmen@vu.nl

prof. dr. J. Hulshof

Vrije Universiteit, Faculteit der Bètawetenschappen, De Boelelaan 1085,
1081 HV Amsterdam
e-mail: j.hulshof@vu.nl

dr. H. Peters

Universiteit van Amsterdam, Korteweg-de Vries Instituut, Science Park
107, 1098 XG Amsterdam
e-mail: h.peters@uva.nl

dr. F. Wiedijk

Radboud Universiteit, Institute for Computing and Information Sciences,
Postbus 9102, 6500 HC Nijmegen
e-mail: freek@cs.ru.nl

prof. dr. W.H.A. Schilders

Technische Universiteit Eindhoven, Faculteit Wiskunde en Informatica,
Postbus 513, 5600 MB Eindhoven
e-mail: w.h.a.schilders@tue.nl

Programma

Vrijdag 25 augustus 2016 / 1 september 2016

15.00–15.30		<i>Ontvangst, koffie</i>
15.30–15.35	Wiegerinck	Introductie: 'De computer in de wiskunde: breekijzer of oud schroot?'
15.35–16.20	Cohen	Orakels en algoritmen voor getallen en integratie
16.20–16.45		<i>Pauze</i>
16.45–17.30	Dahmen	Het abc-vermoeden
17.30–18.30		<i>Diner</i>
18.30–19.15	Hulshof	Wat kun je allemaal zonder hulpmiddelen? En daarna?
19.15–19.45		<i>Pauze</i>
19.45–20.30		Practicum 1

Zaterdag 26 augustus 2016 / 2 september 2016

10.00–10.30		<i>Ontvangst, koffie</i>
10.30–11.15	Peters	Visualisatie en intuïtie in de complexe dynamica
11.15–12.00	Wiedijk	Efficiënt bollen stapelen
12.00–13.00		<i>Lunch</i>
13.00–13.45	Schilders	Moderne Wiskunde: MSO en CSE
13.45–14.30		Practicum 2
14.30		Afsluiting

1 Orakels en algoritmen voor reële getallen en integratie

Arjeh M. Cohen

Algoritmen zijn in het voortgezet onderwijs een onderbelicht onderwerp. De onderwezen wiskunde biedt goede kansen om hier aandacht aan te besteden. We bespreken twee van deze mogelijkheden: het begrip reëel getal en het primitiveren van een rationale functie, zoals $\frac{x^2 + 1}{x^3 + x + 1}$.

Voor de meeste leerlingen is een reëel getal een punt op de getallenlijn. Maar er is een concreet beeld te geven van een reëel getal als een orakel dat precies weet welk getal bedoeld wordt, en bij elke vraag een klein stukje van de sluier oplicht. Dat is de decimale ontwikkeling. Een mens kan hiervan slechts een eindig beginstuk, dat is een decimaal getal, te weten komen. Als je aanhoudend bent en meer geduld hebt, kan het orakel je een betere benadering geven (een decimaal getal dus met meer decimalen). Dit beeld sluit perfect aan bij de meetkundige opvatting van een reëel getal dat ingeklemd zit tussen rationale getallen en de meer formele definitie van een reëel getal als de equivalentieklasse van een rij in lengte afnemende geschakelde intervallen.

Als f een reële continue functie is, dan bestaat er een differentieerbare functie waarvan de afgeleide f is. Zo'n functie heet een primitieve of onbepaalde integraal van f . Primitiveren, het bepalen van die onbepaalde integraal voor een gegeven functie, is een uitdaging die vaak niet als het uitvoeren van een algoritme wordt ervaren. Er is een arsenaal aan technieken waarvan substitutie en partieel integreren de bekendste zijn. Maar er is een probleem dat belichaamd wordt door e^{-x^2} : deze functie heeft geen primitieve die in termen van bekende functies geschreven kan worden. Er bestaan algoritmen die voor bepaalde klassen functies een primitieve afleveren *als* die bestaat binnen die klasse (of een bepaalde iets grotere klasse)

Met dank aan Hans Sterk en Benne de Weger voor hun suggesties voor verbetering van de oorspronkelijke versie. Inspiratie voor het onderwerp heb ik gehaald uit mijn werk aan interactief wiskundig lesmateriaal voor SOWISO. Een kijkje op dat werk via het internet is mogelijk via sowiso.com.

en zo *niet* aangeven dat zo'n functie niet bestaat. We zullen ingaan op een methode om een rationale functie te primitiveren die dicht bij een algoritme komt (maar het net niet is).

Inhoud

- 1.1 Reële getallen als decimale ontwikkelingen
- 1.2 Decimale getallen
- 1.3 Rationale getallen
- 1.4 Berekenbare getallen en orakels
- 1.5 Integratie van rationale functies
- 1.6 Conclusie
- 1.7 Appendix
- 1.8 Oefeningen

1.1 Reële getallen en decimale ontwikkelingen

In deze paragraaf laten we zien dat de decimale ontwikkeling gebruikt kan worden als definitie van een reëel getal. We zullen bijna alle aandacht naar niet-negatieve reële getallen laten gaan. Negatieve getallen worden eenvoudig verkregen door positieve getallen van een minteken te voorzien. We beperken ons tot nul en positieve getallen om het minteken niet steeds in de beschouwingen mee te hoeven nemen. In plaats van de Nederlandse komma zullen we de Angelsaksische punt gebruiken.

Voor de definitie van een niet-negatief reëel getal zullen we gebruikmaken van naar rechts oneindig voortlopende rijen decimale cijfers, waar een decimale punt in voorkomt. We noemen zo'n rij een **decimale ontwikkeling**. De decimale ontwikkeling

$$a_{-k}a_{1-k} \cdots a_0.a_1a_2 \cdots$$

waarbij k een natuurlijk getal en elke a_i een decimaal cijfer is, staat voor

het (in de gebruikelijke wereld bekende) reële getal

$$\sum_{i=-k}^{\infty} a_i \cdot 10^{-i} \quad \text{ofwel} \quad \lim_{m \rightarrow \infty} \sum_{i=-k}^m a_i \cdot 10^{-i}.$$

Deze limiet bestaat omdat $\sum_{i=-k}^m a_i \cdot 10^{-i}$ voor $m = 1, 2, \dots$ een (zwak) stijgende reeks is (want elke term $a_i \cdot 10^{-i}$ is niet-negatief) en de reeks van boven begrensd is door de meetkundige reeks

$$\sum_{i=-k}^{\infty} 9 \cdot 10^{-i} = 9 \cdot 10^k \cdot \frac{1}{1 - \frac{1}{10}} = 10^{k+1}.$$

Twee onderdelen van de decimale ontwikkeling $a_{-k}a_{1-k} \cdots a_0.a_1a_2 \cdots$, die afhangen van een geheel getal m , geven we een naam:

- de **kop** met m decimalen is de uitdrukking $a_{-k}a_{1-k} \cdots a_0.a_1a_2 \cdots a_{m-1}a_m$;
- de **staart** vanaf de m -de decimaal is de uitdrukking $a_ma_{m+1} \cdots$.

Voor $m > 0$ zijn de definities geen verrassing. Als $-k \leq m \leq 0$, dan heeft de kop met m decimalen geen punt, en de staart wel. Als $m < -k$, dan is de kop 0 en is de staart gelijk aan de decimale ontwikkeling. Om de notatie eenvoudig te houden, schrijven we

- a voor de decimale ontwikkeling $a_{-k}a_{1-k} \cdots a_0.a_1a_2 \cdots$,
- $\text{kop}_m(a)$ voor de kop met m decimalen en
- $\text{staart}_m(a)$ voor de staart vanaf de m -de decimaal.

De kop met m decimalen van een decimale ontwikkeling is niet altijd gelijk aan de benadering van het reële getal bij die ontwikkeling tot op m decimalen: $\text{kop}_4(\pi) = 3.1415$ terwijl de benadering van π tot op 4 decimalen gelijk is aan 3.1416.

Niet elke decimale ontwikkeling stelt een uniek reëel getal voor. Er is zowel aan de kop als aan de staart een complicatie:

- Als de decimale ontwikkeling met een 0 begint, dan is het bijbehorende getal gelijk aan dat van de ontwikkeling zonder die 0. Dit geldt zelfs voor bijvoorbeeld $0.1212123451 \cdots$. We laten dus toe dat de ontwikkeling met een punt begint. De ontwikkeling $.00123121 \cdots$ begint niet met een 0 en kan dus niet worden ingekort.

- Als de decimale ontwikkeling een staart heeft die, op de punt na, uitsluitend uit negens bestaat, dan kunnen we een andere ontwikkeling voor hetzelfde getal vinden die niet die eigenschap heeft. Deze observatie berust op de gelijkheid $.999\ldots = 1$. Zo heeft het reële getal $\frac{33}{10}$ zowel de decimale ontwikkeling $3.3000\ldots$ als $3.2999\ldots$.

Wanneer we ons beperken tot decimale ontwikkelingen die niet met 0 beginnen en geen staart hebben die uitsluitend uit negens bestaat, dan kunnen we reële getallen en decimale ontwikkelingen met elkaar identificeren. Dergelijke decimale ontwikkelingen noemen we **standaard**. Elke decimale ontwikkeling is te **standaardiseren**:

- Als de decimale ontwikkeling met een 0 begint, dan schrappen we alle nullen vooraan, tot we op een ander karakter (de punt of een ander cijfer) komen.
- Als de decimale ontwikkeling een staart heeft die uitsluitend uit negens bestaat, dan zoeken we de eerste plaats van rechts waar een cijfer ongelijk 9 staat. Als die er niet is, dan schrijven we een 0 vóór de ontwikkeling. Vervolgens vervangen we het cijfer op die plaats door het opvolgende cijfer en vervangen we alle negens in de staart er achter door 0.

Zo wordt

- $\alpha = 0$ vertegenwoordigd door de decimale ontwikkeling $a = 0.000\ldots$ en de standaard $a = .000\ldots$,
- $\alpha = 10$ door $a = 9.999\ldots$ en de standaard $a = 10.000\ldots$, en
- $\alpha = \frac{1}{7}$ door de (standaard) ontwikkeling $a = .142857142857142857\ldots$

Conclusie: we kunnen niet-negatieve reële getallen definiëren als standaard decimale ontwikkelingen. We kunnen elke decimale ontwikkeling als reëel getal zien door haar te standaardiseren. De vraag rijst nu of we hiermee een goed kader opgezet hebben om algebraïsche bewerkingen (als optellen, vermenigvuldigen, aftrekken, delen, op grootte vergelijken) met de reële getallen uit te voeren. Maar om het juiste gereedschap daarvoor aan te dragen kijken we eerst naar rationale getallen in de nieuwe context van reële getallen.

1.2 Rationale getallen

We hebben al eerder vermeld dat de decimale ontwikkeling van $\frac{1}{7}$ gelijk is aan $a = .142857142857142857 \dots$. De drie gecentreerde punten $\dots$ suggereren hier een patroon. Om dat duidelijk te maken schrijven we

$$a = .\overline{142857}.$$

De horizontale streep geeft aan dat het rijtje cijfers 142857 ter lengte 6 steeds wordt herhaald. Een decimale ontwikkeling van deze vorm noemen we **periodiek met periode 6**. De periodiciteit slaat op de staart. De kop kan willekeurig zijn. De breuk $\frac{8}{3}$ heeft decimale ontwikkeling $2.\overline{3}$ en is dus periodiek met periode 1.

Stelling 1.2.1. *Een getal is dan en slechts dan rationaal als het een periodieke decimale ontwikkeling heeft.*

Bewijs. Om in te zien dat de decimale ontwikkeling van een positief rationaal getal $\alpha = \frac{p}{q}$, waarbij p en q positieve gehele getallen zijn, een periodieke decimale ontwikkeling oplevert, gebruiken we het feit dat de kop met m decimalen van α na weglating van de punt het getal

$$\left\lfloor \frac{p \cdot 10^m}{q} \right\rfloor$$

wordt. Dit getal wordt berekend door middel van een staartdeling. In die staartdeling wordt (vanaf een zeker moment) de rest van een deling door q steeds aangevuld met nullen, om weer gedeeld te worden door q . Omdat er niet meer dan q resten (verschillende getallen tussen 0 en $q - 1$) zijn, komt op een gegeven moment (als m groot genoeg is), dezelfde rest voor als eerder aan bod geweest is. Vanaf dat moment herhaalt het patroon zich en komen dezelfde cijfers in de decimale ontwikkeling voor als er eerder waren.

Andersom: stel dat een decimale ontwikkeling a periodiek is met periode N , zeg

$$a_{m+N \cdot k+1} a_{m+N \cdot k+2} \cdots a_{m+N \cdot (k+1)} = a_{m+1} a_{m+2} \cdots a_{m+N}$$

voor alle $k > 0$. Dan is het bijbehorende reële getal de som van een kop en de meetkundige reeks

$$\frac{a_{m+1} a_{m+2} \cdots a_{m+N}}{10^m} \cdot \sum_{i=0}^{\infty} \frac{1}{10^{N \cdot i}}$$

dus rationaal. Bijvoorbeeld

$$\overline{.142857} = 142857 \cdot \sum_{i=1}^{\infty} \frac{1}{10^{6 \cdot i}} = \frac{142857}{10^6 - 1} = \frac{1}{7}.$$

□

1.3 Bewerkingen met reële getallen

Van algebraïsche bewerkingen verwachten we dat ze in eindig veel stappen uitgevoerd kunnen worden. Twee reële getallen gegeven door decimale ontwikkelingen kunnen we dus niet zomaar optellen (zelfs niet standaardiseren want dit proces eindigt niet in eindige tijd). Om deze beperking zo goed mogelijk te omzeilen, werken we met decimale getallen.

Een **decimaal getal** is een breuk waarvan de teller een geheel getal is en de noemer een macht van 10. Zo'n getal (nog steeds verondersteld niet-negatief te zijn) is dus te schrijven als

$$\frac{a_{-k}a_{1-k} \cdots a_0a_1a_2 \cdots a_{m-1}a_m}{10^m}.$$

voor natuurlijke getallen k en m , waarbij a_i decimale cijfers zijn. Dit getal wordt vaak geschreven als

$$a_{-k}a_{1-k} \cdots a_0.a_1a_2 \cdots a_{m-1}a_m$$

We noemen m het **aantal decimalen** van het decimale getal. (Dit getal hangt strikt genomen niet van het decimale getal af, maar van de manier waarop de breuk geschreven is.) De kop (met m decimalen) van elke decimale ontwikkeling is een decimaal getal (met m decimalen).

Decimale getallen zijn precies de rationale getallen die geschreven kunnen worden als een breuk met een noemer die geen andere priemfactoren dan 2 en 5 heeft. De decimale ontwikkeling voor $\frac{1}{2^r \cdot 5^s}$ heeft een kop met $m = \max(r, s)$ decimalen, met daarachter alleen maar nullen.

Decimale getallen kunnen we optellen, aftrekken en vermenigvuldigen, met als resultaat decimale getallen. Het onderling vergelijken op grootte van twee decimale getallen is ook geen probleem. Om deze redenen kunnen we decimale getallen goed gebruiken om niet alleen reële getallen zelf, maar ook de resultaten van algebraïsche bewerkingen daarop te benaderen.

Het uitgangspunt is de benadering van een reëel getal door decimale getallen. Hiervoor maken we gebruik van de **geschakelde intervallen**, waarbij een reëel getal gedefinieerd wordt als een equivalentieklasse van geschakelde intervallen van rationale getallen. Onze definitie van een reëel getal als een standaard decimale ontwikkeling komt in deze context neer op de keuze van een unieke representant uit elke equivalentieklasse. Zo kunnen we begrippen als equivalentierelatie, die in het algemeen niet in het voortgezet onderwijs aan bod komen, vermijden (zie appendix 1.7.1). Decimale ontwikkelingen leiden als volgt tot geschakelde intervallen.

Stelling 1.3.1. *Laat α een niet-negatief reëel getal zijn en b_i een decimaal getal met i decimalen. Dan is b_i dan en slechts dan een kop van de standaard decimale ontwikkeling van α als*

$$b_i \leq \alpha < b_i + \frac{1}{10^i}.$$

De rij van koppen van een decimale ontwikkeling vormt een convergente reeks, waarvan de limiet het reële getal is dat bij de ontwikkeling hoort. Elke kop b_i geeft een interval $[b_i, b_i + 10^{-i})$ waarin het reële getal moet liggen. De rij koppen geeft ons een representant van de equivalentieklasse van geschakelde intervallen bij de gebruikelijke definitie van een reëel getal.

Om te laten zien hoe Stelling 1.3.1 kan helpen bij de bepaling van de kop van de decimale ontwikkeling, bepalen we één decimaal van $\alpha = \sqrt{2}$. Op veel rekenmachines is de decimale ontwikkeling van $\sqrt{2}$ op te vragen. Deze rekenmachines gebruiken ongetwijfeld efficiëntere methoden dan onze aanpak met de ongelijkheden uit de stelling, maar het gaat ons om het idee van de berekenbaarheid (ofwel *effectiviteit*). De wortel van twee is het unieke positieve getal α met $\alpha^2 = 2$. Omdat $1^2 = 1 \leq \alpha^2 = 2 < 4 = 2^2$, geldt $1 \leq \alpha < 2$, en geeft de stelling $\text{kop}_0(\sqrt{2}) = 1$. Laat c het cijfer zijn waarvoor $\text{kop}_1(\sqrt{2}) = 1.c$. Dan is c het grootste cijfer dat voldoet aan

$$\frac{c}{10} + 1 \leq \sqrt{2}.$$

We kwadrateren beide zijden van deze ongelijkheid en vermenigvuldigen ze vervolgens met 100:

$$(c + 10)^2 \leq 200.$$

De ongelijkheid is duidelijk correct voor $c = 0$. We berekenen de waarden van het linker lid voor oplopende cijfers c , totdat $(c + 10)^2 > 200$:

c	1	2	3	4	5
$(c + 10)^2$	121	144	169	196	225

De berekening laat zien dat $c = 4$ de grootste waarde van 1 is waarvoor het linker lid kleiner dan of gelijk aan 200 is. We concluderen met de stelling dat $\text{kop}_1(\sqrt{2}) = 1.4$.

We bespreken nog een decimale ontwikkeling van π die van de Taylorreeks voor arctangens met $0 \leq x \leq 1$ komt (zie appendix 1.7.2):

$$\arctan(x) = \sum_{k=0}^{n-1} \frac{(-1)^k \cdot x^{2k+1}}{2k+1} + E_{2n}(x)$$

met foutschatting $|E_{2n}(x)| \leq \frac{x^{2n+1}}{2n+1}$. De meest voor de hand liggende gelijkheid is $\pi = 4 \arctan(1)$, maar omdat de convergentiesnelheid (en dus de efficiëntie) hierbij te wensen overlaat, gebruiken we

$$\pi = 8 \arctan\left(\frac{1}{3}\right) + 4 \arctan\left(\frac{1}{7}\right).$$

In appendix 1.7.4 staat een afleiding van deze formule. Om de kop met 6 decimalen te bepalen, eisen we een foutschatting kleiner dan 10^{-7} . We kiezen daartoe n zo groot dat

$$\frac{8}{3^{2n+1} \cdot (2n+1)} + \frac{4}{7^{2n+1} \cdot (2n+1)} < 10^{-7}.$$

We kunnen volstaan met $n = 7$, aangezien het linker lid dan gelijk is aan

$$\frac{37980549475172}{1021834778744275234515} \approx 3.7 \cdot 10^{-8}.$$

Enig rekenwerk levert dat de som van de eerste 7 termen van de Taylorreeks voor $8 \arctan\left(\frac{1}{3}\right) + 4 \arctan\left(\frac{1}{7}\right)$ gelijk is aan

$$\frac{459056974189868332544096}{146122373360431358535645}.$$

Door de foutschatting hiervan af te trekken voor de ondergrens en er bij op te tellen voor de bovengrens, vinden we

$$\frac{91811393751729951518900}{29224474672086271707129} \leq \pi \leq \frac{346981844006868410804}{110447750083470414615}.$$

Zoals we hierboven zagen, kan staartdeling gebruikt worden om de kop van een rationaal getal te bepalen. Als we dit toepassen op het linker en rechter lid van bovenstaande ongelijkheden, krijgen we

$$3.141592 \leq \pi < 3.141593$$

zodat $\text{kop}_6(\pi) = 3.141592$.

Net zo kunnen we de Taylorreeks

$$e^x = \sum_{i=0}^n \frac{x^i}{i!} + F_n(x)$$

voor $x \geq 0$ met foutschatting $|F_n(x)| \leq \frac{e^x}{(n+1)!} \cdot x^{n+1}$ gebruiken om bijvoorbeeld af te leiden dat $\text{kop}_5(e) = 2.71828$.

Er blijven problemen bij de afronding als in de staart alleen negens te ontdekken zijn. Kijk eens naar de optelling

$$\frac{1}{9} + \frac{8}{9} = 1.$$

De decimale ontwikkelingen behorende bij de twee breuken in het linker lid zijn

$$.111\ldots \quad \text{en} \quad .888\ldots$$

Optellen van de koppen met 3 decimalen geeft .999, terwijl de kop met 3 decimalen van de som 1 gelijk is aan 1.000. We hebben dus wel de kop van een ontwikkeling van de som te pakken, maar niet de kop van de standaard decimale ontwikkeling. Wat erger is, als we de som

$$\frac{1}{9} + \frac{8}{9} + \frac{1}{10000} = \frac{10001}{10000}$$

op dezelfde manier willen benaderen, dan is de som van de drie koppen met drie decimalen nog steeds gelijk aan .999, maar niet langer de kop van een decimale ontwikkeling van het rechter lid. Bij benaderingen door decimale getallen kunnen we dus niet verwachten dat we altijd de kop van een decimale ontwikkeling vinden. Dit voorbeeld laat zien dat de equivalentieklasse bij de aanpak via geschakelde intervallen niet te klein gekozen moet worden. Daarom formuleren we de intervalschakeling in deze context als volgt.

Stelling 1.3.2. *Laat b_i ($i = 1, 2, 3, \dots$) een rij decimale getallen zijn, zodat $b_i = \text{kop}_i(b_{i+1})$. Dan is er precies één (standaard decimale ontwikkeling a met) reëel getal α , zodat $b_i \leq \alpha \leq b_i + 10^{-i}$ voor alle i .*

Bewijs. We gaan de definitie na van een reëel getal als rij niet-lege geschakelde gesloten intervallen met rationale grenzen. Hiertoe controleren we vier eigenschappen (zie appendix 1.7.1).

- niet-leeg: voor elke i geldt dat de ondergrens b_{i+1} kleiner is dan de bovengrens $b_{i+1} + 10^{-i-1}$. Het interval is dus niet leeg.
- rationale grenzen: de getallen b_i zijn decimaal, dus rationaal; dat geldt ook voor de som $b_i + 10^{-i}$ van twee decimale getallen.
- geschakeld: Het interval $[b_{i+1}, b_{i+1} + 10^{-i-1}]$ ligt binnen het interval $[b_i, b_i + 10^{-i}]$. Immers: $b_i \leq b_{i+1}$ is gegeven en de ongelijkheid $b_{i+1} + 10^{-i-1} \leq b_i + 10^{-i}$ kan als volgt afgeleid worden:

$$\begin{array}{rcl}
 b_{i+1} & < & b_i + 10^{-i} \\
 & & \text{gegeven} \\
 b_{i+1} \cdot 10^{i+1} & < & b_i \cdot 10^{i+1} + 10 \\
 & & \text{beide zijden vermenigvuldigd met } 10^{i+1} \\
 b_{i+1} \cdot 10^{i+1} + 1 & \leq & b_i \cdot 10^{i+1} + 10 \\
 & & \text{aan beide zijden staan gehele getallen} \\
 b_{i+1} + 10^{-i-1} & \leq & b_i + 10^{-i} \\
 & & \text{beide zijden gedeeld door } 10^{i+1}
 \end{array}$$

- lengte gaat naar 0: Verder nadert de lengte van het interval $[b_i, b_i + 10^{-i}]$ nul voor $i \rightarrow \infty$. Volgens de theorie van geschakelde intervallen is er precies één reëel getal dat in alle geschakelde intervallen $[b_i, b_i + 10^{-i}]$ ligt.

□

De koppen met i decimalen van een decimale ontwikkeling voldoen aan de voorwaarden van stelling 1.3.2, maar zijn daarin niet uniek: naast de constante rij 1 voldoet ook de rij $b_i = \underbrace{.999 \cdots 9}_i$ voor $\alpha = 1$.

Om dit verschijnsel verder te exploreren, gaan we in op de vermenigvuldiging van twee reële getallen α en β gegeven door hun standaard decimale ontwikkeling a , respectievelijk, b . We kiezen een natuurlijk getal n en stellen ons ten doel de kop met n decimalen te vinden van de decimale ontwikkeling van $\alpha \cdot \beta$.

Neem $m \geq n + 1 + \log_{10}(\max(\text{kop}_0(a) + 1, \text{kop}_0(b) + 1))$. We gaan als volgt te werk met de decimale getallen $x = \text{kop}_m(a)$ en $y = \text{kop}_m(b)$.

$$\begin{aligned}
\alpha \cdot \beta - x \cdot y &= (\alpha - x) \cdot (\beta - y) + x \cdot (\beta - y) + y \cdot (\alpha - x) \\
&\quad \text{in te zien door de haakjes weg te werken} \\
\alpha \cdot \beta - x \cdot y &< 10^{-2m} + x \cdot 10^{-m} + y \cdot 10^{-m} \\
&\quad 0 \leq \alpha - x < 10^{-m} \text{ en } 0 \leq \beta - y < 10^{-m} \\
&\quad \text{vanwege stelling 1.3.1} \\
x \cdot y &\leq \alpha \cdot \beta < x \cdot y + 3 \cdot 10^{-n-1} \\
&\quad x, y \leq 10^{m-n-1} \\
x \cdot y \cdot 10^n &\leq \alpha \cdot \beta \cdot 10^n < x \cdot y \cdot 10^n + \frac{3}{10} \\
&\quad \text{vermenigvuldigd met } 10^n \\
\lfloor x \cdot y \cdot 10^n \rfloor &\leq \lfloor \alpha \cdot \beta \cdot 10^n \rfloor \leq \lfloor x \cdot y \cdot 10^n \rfloor + 1 \\
&\quad \text{entier is een zwak stijgende functie} \\
\text{kop}_n(x \cdot y) &\leq \text{kop}_n(\alpha \cdot \beta) \leq \text{kop}_n(x \cdot y) + 10^{-n} \\
&\quad \text{gedeeld door } 10^n
\end{aligned}$$

Het lukt dus niet altijd om de kop met n decimalen van $\alpha \cdot \beta$ precies te bepalen (want daarvoor moet de laatste ongelijkheid strikt zijn). Maar er is wel een rij geschakelde intervallen met rationale grenzen te vinden die de equivalentieklasse van $\alpha \cdot \beta$ bepaalt, en daarmee een rij decimale getallen die aan stelling 1.3.2 voldoet. Ter illustratie berekenen we $\text{kop}_4(\pi \cdot e)$. Omdat

$$6 \geq 4 + 1 + \log_{10}(\max(\text{kop}_0(\pi) + 1, \text{kop}_0(e) + 1))$$

nemen we $m = 6$. We gebruiken kennis van de koppen van π en e :

$$\begin{aligned}
x = \text{kop}_6(\pi) &= 3.141592 \\
y = \text{kop}_6(e) &= 2.718281.
\end{aligned}$$

Vermenigvuldiging van de twee koppen geeft $x \cdot y = 8.539729843352$. De kop hiervan met 4 decimalen is $\text{kop}_4(x \cdot y) = 8.5397$. Volgens onze schattingen geldt dus

$$8.5397 \leq \text{kop}_4(\pi \cdot e) \leq 8.5397 + 10^{-4}.$$

We kunnen in dit geval wel afleiden dat de rechterongelijkheid strikt is: π is kleiner dan 3.141593 en e is kleiner dan 2.718282, dus het product is kleiner dan 8.539735703226. In het bijzonder is de kop van $\pi \cdot e$ met 4 decimalen niet 8.5398, en moet dus wel 8.5397 zijn. Verderop, in §1.4, bespreken we waarom je de kop van een product niet altijd zo kan afleiden. Degenen die bekend zijn met ϵ, δ -redeneringen zal bovenstaande niet onbekend voorkomen: de rol van ϵ is vervangen door 10^{-n} en van δ door 10^{-m} . De waarde van m als functie van n is daadwerkelijk te gebruiken bij berekeningen.

1.4 Berekenbare getallen en orakels

De decimale ontwikkeling heeft aftelbaar veel cijfers (zie appendix 1.7.3), maar we hebben niet de tijd om oneindig veel cijfers op te schrijven. De puntjes achterin de decimale ontwikkeling

$$a_{-k}a_{1-k} \cdots a_0.a_1a_2 \cdots$$

zijn ook erg suggestief. Bij de decimale ontwikkeling van $\frac{1}{7}$ suggereerden we er een patroon mee dat zich herhaalde. Dat impliceert dat er een eindig voorschrift is waarmee we alle decimalen vastgelegd kunnen worden. Omdat er aftelbaar veel voorschriften zijn (deze bestaan immers alle uit eindige boodschappen geschreven in een eindig alfabet), zijn er slechts aftelbaar veel reële getallen waarvan de volledige decimale ontwikkeling door een eindig voorschrift vastgelegd kan worden. In navolging van Turing [4] noemen we deze reële getallen **berekenbaar**. Aangezien er overaftelbaar veel reële getallen zijn, zullen er heel veel reële getallen zijn die niet berekenbaar zijn. Uit stelling 1.2.1 volgt dat rationale getallen berekenbaar zijn.

Een reëel getal heet **algebraïsch** als het een nulpunt is van een niet-triviale veeltermvergelijking met rationale coëfficiënten. Denk aan $\sqrt{2}$, dat het positieve nulpunt is van $x^2 = 2$. Reële algebraïsche getallen zijn berekenbaar; zie bijvoorbeeld [1], [3]. Maar ook getallen als π en e zijn berekenbaar, zoals bleek uit de behandeling met gebruikmaking van de Taylorreeksen van $\arctan$ en e^x . Eigenlijk houden we ons in het algemeen voornamelijk met berekenbare getallen bezig.

Om niet-berekenbare reële getallen toch “benaderbaar” te maken, kun je aan de decimale ontwikkeling van stelling 1.3.1 of de decimale benaderingen als bedoeld in stelling 1.3.2 denken als uitkomsten van een orakel. Het **orakel** beschikt over meer kennis dan mensen. De mens kan vragen stellen aan het orakel en krijgt antwoorden, maar lang niet altijd het volledige antwoord. Dit geldt in het bijzonder voor orakels in de vorm van reële getallen. Op de vraag naar een benadering met precisie n van een reëel getal α geeft het orakel een decimaal getal b met n decimalen, zodat α in het interval $[b, b + 10^{-n}]$ ligt. Maar het orakel geeft het hele getal α niet bloot, ook niet bij herhaaldelijk vragen: de mens heeft maar eindige tijd en kan dus maar eindig veel vragen stellen, zodat we nooit een staart van α leren kennen. De notatie $\alpha = 0.11111111 \cdots$, bijvoorbeeld, suggereert dat we met $\frac{1}{9}$ te maken hebben, maar het orakel dat ons steeds een kop geeft waarvan alle decimalen 1 zijn, kan nog ergens, verder in de decimale

ontwikkeling van α dan we ooit met onze vragen gekomen zijn, een ander cijfer achter de hand hebben, waardoor α afwijkt van $\frac{1}{9}$.

Het voorbeeld van vermenigvuldiging geeft aan hoe een orakel gebouwd kan worden voor het product van twee getallen; dit orakel gebruikt op zijn beurt orakels voor die twee getallen. Het vermenigvuldigingsorakel geeft een benadering die goed te gebruiken is, maar kan onmogelijk de kop van de standaard decimale ontwikkeling aangeven. Denk maar aan de vermenigvuldiging met 3 van één van de getallen $\frac{1}{3}$ of $\frac{1}{3} + \frac{1}{10^k}$, waarbij k groter is dan alle lengtes van de koppen die opgevraagd zijn door het algoritme voor de berekening van de kop van het product voor de eerste 'zoveel' decimalen. De uitkomst van de berekening zal in alle gevallen $.999 \dots 9$ zijn. In het eerste geval is het product 1, dus is de uitkomst correct als we toelaten dat de kop van een decimale ontwikkeling komt die niet standaard is, maar in het tweede geval moet de uitkomst voor een kop met minder dan k decimalen wel $1.000000 \dots 0$ zijn.

1.5 Integratie van rationale functies

We hebben de arctangens gebruikt om een kop van de decimale ontwikkeling van π te vinden. De arctangens op het interval $[0, 1]$ is te definiëren door de bepaalde integraal

$$\arctan(x) = \int_0^x \frac{1}{t^2 + 1} dt.$$

De eerder gegeven Taylorreeks is hier eenvoudig uit af te leiden. We kunnen de logaritme gebruiken om machten als 2^π te berekenen met de efficiënte benadering van e^x door alweer een eerder gegeven Taylorreeks (kies $x = \pi \cdot \ln(2)$). De daarvoor benodigde decimale ontwikkeling van $\ln(2)$ kan (toegegeven: op weinig efficiënte wijze) bepaald worden aan de hand van Riemannsommen voor de integraal

$$\ln(x) = \int_1^x \frac{1}{t} dt.$$

In beide bepaalde integralen is de integrand een **rationale functie** in t , dat wil zeggen: een breuk $\frac{f(t)}{g(t)}$, waarbij $f(t)$ en $g(t)$ veeltermen in t zijn. Kennelijk zijn niet alle primitieven van rationale functies weer rationale functies. Dit verschijnsel doet zich vaker voor bij primitiveren,

zoals in de inleiding besproken. We laten in deze paragraaf zien hoe je een rationale functie kunt primitiveren, en welke functies je in het resultaat kunt verwachten.

We schrijven m en n voor de graad van $f(x)$, respectievelijk $g(x)$ en gaan te werk met reductie van het **gradenpaar** $[n, m]$ (eerst de graad van de noemer, dan de graad van de teller). We beschrijven acht stappen van een methode om de primitieve te vinden.

Veeltermregel Als $n = 0$ dan is de integrand (op een constante na) de veelterm $f(x) = a_m \cdot x^m + \dots + a_1 \cdot x + a_0$, zodat

$$\int f(x) dx = \frac{a_m}{m+1} \cdot x^{m+1} \dots + \frac{a_1}{2} \cdot x^2 + a_0 \cdot x + C.$$

Hier is C de integratieconstante. Omdat we ons hier concentreren op het vinden van een primitieve, is C niet van belang.

Vereenvoudiging van de breuk Het is denkbaar dat de breuk $\frac{f(x)}{g(x)}$ vereenvoudigd kan worden. Dit is na te gaan door de grootste gemene deler $h(x)$ van $f(x)$ en $g(x)$ uit te rekenen. Als $h(x)$ niet constant is, dan kunnen we teller en noemer door $h(x)$ delen, zodat

$$\frac{f(x)}{g(x)} = \frac{p(x)}{q(x)}$$

een onvereenvoudigbare breuk is met teller $p(x) = \frac{f(x)}{h(x)}$ en $q(x) = \frac{g(x)}{h(x)}$, waarbij het gradenpaar $[n, m]$ is afgenomen. Daarmee hebben we het probleem teruggebracht tot het primitiveren van een onvereenvoudigbare breuk. Voorbeeld: Stel we willen

$$\frac{x^4 + 3x^3 + 4x^2 + 3x + 1}{x^4 + x^3 - x - 1}$$

primitiveren. Dan vereenvoudigen we de breuk eerst als volgt:

$$\begin{aligned} \gcd(x^4 + 3x^3 + 4x^2 + 3x + 1, x^4 + x^3 - x - 1) &= x^3 + 2x^2 + 2x + 1 \\ \text{algoritme van Euclides} \\ \frac{x^4 + 3x^3 + 4x^2 + 3x + 1}{x^4 + x^3 - x - 1} &= \frac{\frac{x^4 + 3x^3 + 4x^2 + 3x + 1}{x^3 + 2x^2 + 2x + 1}}{\frac{x^4 + x^3 - x - 1}{x^3 + 2x^2 + 2x + 1}} \\ \text{teller en noemer gedeeld door de ggd} \\ &= \frac{x + 1}{x - 1}. \\ \text{deling in teller en noemer uitgevoerd} \end{aligned}$$

Logaritmische regel Als $m = n - 1$, dan zijn er een constante c en een veelterm h van graad kleiner dan of gelijk aan $n - 2$, zodat

$$\frac{f(x)}{g(x)} = c \cdot \frac{g'(x)}{g(x)} + \frac{h(x)}{g(x)}.$$

Bijgevolg geldt

$$\int \frac{f(x)}{g(x)} dx = c \cdot \ln(g(x)) + \int \frac{h(x)}{g(x)} dx$$

Voorbeeld:

$$\begin{aligned} \int \frac{2x+3}{x^2-8x+17} dx &= \int \frac{2x-8}{x^2-8x+17} dx + \int \frac{11}{x^2-8x+17} dx \\ &\quad \text{plitsing als boven} \\ &= \int \frac{1}{v} dv + \int \frac{11}{u^2+1} du \\ &\quad \text{substituties } u = x-4 \text{ en } v = x^2-8x+17 \\ &= \ln(v) + 11 \cdot \arctan(u) + C \\ &= \ln(x^2-8x+17) + 11 \cdot \arctan(x-4) + C. \end{aligned}$$

Deling met rest Als $m \geq n$, dan delen we $f(x)$ door $g(x)$ met rest: $f(x) = q(x) \cdot g(x) + r(x)$, waarbij $q(x)$ en $r(x)$ veeltermen zijn met graden $m - n$, respectievelijk, kleiner dan m . De integraal kan dan als volgt herschreven worden:

$$\int \frac{f(x)}{g(x)} dx = \int q(x) dx + \int \frac{r(x)}{g(x)} dx.$$

De eerste integraal kan berekend worden uit de veeltermregel, de tweede heeft een noemer van dezelfde graad n en een teller van graad kleiner dan n (en dus kleiner dan m). Voorbeeld: Stel we willen

$$\frac{x^4 + 3x^3 + 4x^2 + 4x + 1}{x^2 + 1}$$

primitiveren. Dan vervangen we de breuk als volgt door de som van een veelterm en een eenvoudiger breuk:

$$\begin{aligned} x^4 + 3x^3 + 4x^2 + 4x + 1 &= (x^2 + 3x + 3)(x^2 + 1) + x - 2 \\ &\quad \text{deling met rest} \\ \frac{x^4 + 3x^3 + 4x^2 + 4x + 1}{x^2 + 1} &= (x^2 + 3x + 3) + \frac{x - 2}{x^2 + 1} \\ &\quad \text{gedeeld door de noemer.} \end{aligned}$$

Breuksplitsing We ontbinden $g(x)$ in irreducibele factoren. Stel $g(x) = u(x) \cdot v(x)$ en $u(x)$ en $v(x)$ hebben geen gemeenschappelijke delers. Dan geeft breuksplitsen ons twee veeltermen $a(x)$ en $b(x)$, zodat $\frac{f(x)}{g(x)} = \frac{a(x)}{u(x)} + \frac{b(x)}{v(x)}$. Zo kun je doorgaan en krijg je een som van breuken waarvan de noemers machten zijn van irreducibele factoren. De graad van elke teller kan kleiner dan de graad van de irreducibele factor in de noemer gekozen worden. Voorbeeld:

$$\frac{3x^3 + 2x^2 - 4x + 2}{x^4 - x^3 - x + 1}.$$

Ontbinding in factoren van de noemer geeft $x^4 - x^3 - x + 1 = (x - 1)^2 \cdot (x^2 + x + 1)$. Breuksplitsing zou een uitdrukking van de vorm

$$\frac{cx + d}{x^2 + x + 1} + \frac{b}{x - 1} + \frac{a}{(x - 1)^2}$$

moeten opleveren, waarbij a , b , c en d constanten zijn. Onder één noemer brengen van deze breuken geeft de noemer van de oorspronkelijke breuk en de teller

$$(c + b)x^3 + (d - 2c + a)x^2 + (-2d + c + a)x + d - b + a.$$

Gelijkstellen van deze teller aan de oorspronkelijke teller $3x^3 + 2x^2 - 4x + 2$ geeft lineaire vergelijkingen in a , b , c en d , waarvan $a = 1$, $b = 2$, $c = 1$ en $d = 3$ de enige oplossing is. Zo vinden we

$$\frac{3x^3 + 2x^2 - 4x + 2}{x^4 - x^3 - x + 1} = \frac{x + 3}{x^2 + x + 1} + \frac{2}{x - 1} + \frac{1}{(x - 1)^2}.$$

Na het doorlopen van de vorige stappen moeten we ons nog buigen over het primitiveren van breuken $\frac{f(x)}{g(x)}$, waarbij $g(x) = q(x)^k$, met k een positief geheel getal, $q(x)$ een irreducibele veelterm is en m kleiner dan de graad van $q(x)$ is. Omdat elke irreducibele veelterm met reële coëfficiënten graad 1 of 2 heeft, komen we, eventueel na een lineaire substitutie, op een van de volgende drie typen uit:

$$\frac{f(x)}{g(x)} = \frac{1}{x^n}, \quad \frac{x}{(x^2 + 1)^r}, \quad \frac{1}{(x^2 + 1)^r}.$$

De drie gevallen worden behandeld in de drie regels hieronder.

Lineaire machtrege Als $f(x) = 1$ en $g(x) = x^n$, dan geldt

$$\int \frac{1}{x^n} dx = \begin{cases} \ln(x) + C & \text{als } n = 1 \\ \frac{1}{(1-n) \cdot x^{n-1}} + C & \text{als } n > 1 \end{cases}.$$

Met behulp van de substitutieregels kunnen we hiermee elke negatieve macht van een lineaire veelterm aan. Voorbeeld:

$$\begin{aligned} \int \frac{1}{(2x+3)^4} dx &= \int \frac{1}{2u^4} du \\ &\quad \text{substitutie } u = 2x+3 \\ &= \frac{-1}{6u^3} + C \\ &\quad \text{bekende primitieve} \\ &= \frac{-1}{6(2x+3)^3} + C \\ &\quad \text{substitutie } u = 2x+3. \end{aligned}$$

Kwadratische substitutie Laat r een positief geheel getal zijn met $r > 1$. Als $f(x) = 1$ en $g(x) = (x^2 + 1)^r$, dan geldt

$$\begin{aligned} \int \frac{x}{(x^2+1)^r} dx &= \int \frac{1}{2u^r} du \\ &\quad \text{substitutie } u = x^2 + 1 \\ &= \frac{-1}{2(k-1)u^{r-1}} + C \\ &= \frac{-1}{2(r-1)(x^2+1)^{r-1}} + C. \end{aligned}$$

Voor $r = 1$ is de logaritmeregel van toepassing, waarin overigens dezelfde substitutie, $u = x^2 + 1$, gebruikt wordt als hierboven.

Kwadratische machtrege Laat r een positief geheel getal zijn. Dan is $Q_r(x) = \int \frac{1}{(x^2+1)^r} dx$ bepaald door

$$Q_r(x) = \begin{cases} \arctan(x) + C & \text{als } r = 1 \\ \frac{2r-3}{2r-2} \cdot Q_{r-1}(x) + \frac{x}{(2r-2) \cdot (x^2+1)^{r-1}} & \text{als } m > 1 \end{cases}.$$

Met behulp van de substitutieregels en kwadraatafsplitsen kunnen we hiermee elke negatieve macht van een irreducibele kwadratische veelterm aan.

Voorbeeld:

$$\begin{aligned}
 \int \frac{2}{x^2 - 6x + 13} dx &= \int \frac{2}{(x-3)^2 + 4} dx && \text{kwadraat afgesplitst in noemer} \\
 &= \frac{1}{2} \int \frac{1}{\left(\frac{x-3}{2}\right)^2 + 1} dx && \text{noemer als in kwadratische machregel} \\
 &= \frac{1}{2} \int \frac{1}{u^2 + 1} du && \text{substitutie } u = \frac{x-3}{2} \\
 &= \arctan(u) + C && \text{kwadratische machregel} \\
 &= \arctan\left(\frac{x-3}{2}\right) + C && \text{substitutie } u = \frac{x-3}{2}.
 \end{aligned}$$

We hebben gezien dat om een primitieve van bepaalde rationale functies te vinden de functie $\ln(x)$ nodig kan zijn (vanwege de integrand $\frac{1}{x}$) evenals de functie $\arctan(x)$ (vanwege de integrand $\frac{1}{x^2 + 1}$). Deze ingrediënten zijn ook voldoende:

Stelling 1.5.1. *Elke rationale functie heeft een primitieve die een rationale uitdrukking is in x , $\ln(u(x))$ en $\arctan(u(x))$ voor veeltermen $u(x)$ in x .*

Er is nog één obstructie om het primitiveren van een rationale functie als hierboven te zien als een algoritme: we hebben niet aangegeven hoe je veeltermen met willekeurige reële coëfficiënten moet ontbinden. Voor veeltermen met rationale coëfficiënten is dit wel mogelijk, maar de irreducibele factoren over de rationale getallen kunnen dan graden hebben die groter dan 2 zijn. Hiervoor zijn wel oplossingen gevonden (zie [2]). Maar het is niet denkbaar dat niet-berekenbare getallen als coëfficiënten optreden in een algoritme voor ontbinding van veeltermen in irreducibele factoren.

1.6 Conclusie

De volledige behandeling van het begrip reëel getal en alle aspecten van het primitiveren van een rationale functie vormen ongetwijfeld te veel stof voor te weinig tijd in het voortgezet onderwijs. Maar hopelijk inspireren

deze voorbeelden om meer aandacht te besteden aan de drie getrapte E's in de wiskunde:

- *Existentie*: bestaat er een oplossing van het probleem (wat is een reëel getal, is er een primitieve binnen een zekere klasse van functies)? Zo ja:
- *Effectiviteit*: bestaat er ook een algoritme om de oplossing in elke gegeven instantie te bepalen (bereken of benader het getal, bepaal de primitieve van een gegeven rationale functie)? Zo ja:
- *Efficiëntie*: kun je die oplossing ook binnen een redelijke tijd vinden?

Bibliografie

- [1] Basu, S., R. Polack, and M.-F. Roy, *Algorithms in Real Algebraic Geometry*, ACM Series, 10, Second edition, Springer, 2006.
- [2] M. Bronstein, *Symbolic Integration I, Transcendental Functions*, Springer, 2005.
- [3] L. Lovász, *An Algorithmic Theory of Numbers, Graphs and Convexity*. SIAM, Philadelphia, 1986.
- [4] A.M. Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, Proc. London Math. Soc. **S2-42** (1936), 230–265.

1.7 Appendices

1.7.1 Appendix over geschakelde intervallen

De definitie van reële getallen waar in §1.3 naar verwezen wordt, luidt als volgt: Een reëel getal is een equivalentieklasse van rijen $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) van gesloten intervallen met de volgende vier eigenschappen:

- niet-leeg: voor $n \geq 1$ geldt $a_n < b_n$,
- rationale grenzen: alle a_n en b_n zijn rationale getallen,
- geschakeld: voor $n \geq 1$ is het interval $[a_{n+1}, b_{n+1}]$ bevat in $[a_n, b_n]$,
- lengtes gaan naar nul: $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$.

Twee rijen $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) en $[c_n, d_n]$ ($n = 1, 2, 3, \dots$) zijn hierbij **equivalent** als er voor elke m een p te vinden is, zodat $[c_p, d_p] \subseteq [a_m, b_m]$ en een q , zodat $[a_q, b_q] \subseteq [c_m, d_m]$. Het woord equivalent betekent hier, zoals overal in de wiskunde, dat aan de volgende drie eigenschappen voldaan is:

- reflexiviteit: de rij $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) is equivalent met zichzelf
- symmetrie: als $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) en $[c_n, d_n]$ ($n = 1, 2, 3, \dots$) equivalent zijn, dan zijn ook $[c_n, d_n]$ ($n = 1, 2, 3, \dots$) en $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) equivalent
- associativiteit: als $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) en $[c_n, d_n]$ ($n = 1, 2, 3, \dots$) equivalent zijn, en als $[c_n, d_n]$ ($n = 1, 2, 3, \dots$) en $[e_n, f_n]$ ($n = 1, 2, 3, \dots$) equivalent zijn, dan zijn ook $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) en $[e_n, f_n]$ ($n = 1, 2, 3, \dots$) equivalent

Een belangrijke eigenschap van een equivalentierelatie is dat de verzameling waarop de relatie gedefinieerd is, opgedeeld wordt in deelverzamelingen die elk geheel uit onderling equivalente elementen (hier dus: rijen van geschakelde intervallen) bestaan en waarbij geen tweetal uit verschillende deelverzamelingen equivalent is. (Ga na dat, andersom, als je een opdeling van een verzameling in deelverzamelingen hebt, de relatie “in dezelfde deelverzameling zitten” een equivalentierelatie is.) Deze deelverzamelingen heten **equivalentieklassen**. Elke rij geschakelde intervallen bepaalt een unieke equivalentieklasse; zo’n equivalentieklasse is per definitie een reëel getal. De bewerkingen op reële getallen worden dan gedefinieerd door ze te definiëren voor een willekeurig element uit elke betrokken equivalentieklasse en te laten zien dat het resultaat niet afhangt van de keuze van dat element binnen de equivalentieklasse. Zo kan de optelling van de equivalentieklasse van $[a_n, b_n]$ ($n = 1, 2, 3, \dots$) bij die van $[c_n, d_n]$ ($n = 1, 2, 3, \dots$) gedefinieerd worden als de equivalentieklasse van $[a_n + c_n, b_n + d_n]$ ($n = 1, 2, 3, \dots$).

1.7.2 Appendix over Taylorreeksen

Elke functie f die op een open omgeving van 0 gedefinieerd is en daar oneindig vaak differentieerbaar is, voldoet aan de volgende gelijkheid

$$f(x) = \sum_{k=0}^n \frac{1}{k!} f^{(k)}(0) \cdot x^k + E_n(x)$$

waarbij de restterm gegeven is door

$$E_n(x) = \int_0^x \frac{1}{n!} f^{(n+1)}(t) \cdot (x-t)^n dt.$$

De functie f wordt hier benaderd door een veelterm van graad n . De rij van veeltermen van graad n voor $n = 0, 1, 2, 3, \dots$ heet de **Taylorreeks** van f rond 0. De kwaliteit van de benadering wordt gegeven door de foutschatting

$$|E_n(x)| \leq \frac{\max_{y \in [0, x]} |f^{(n+1)}(y)|}{(n+1)!} \cdot |x|^{n+1}$$

als $x > 0$ en dezelfde uitdrukking met $y \in [x, 0]$ als $x < 0$. Uitgaande van de definitie van $E_n(x)$ als integraal, kan deze schatting bewezen worden met een variant van de middelwaardestelling. De schatting is goed bruikbaar voor het bepalen van de kop van de decimale ontwikkeling van veel reële getallen.

Als bijvoorbeeld $f(x) = e^x$, dan is f gelijk aan elke hogere afgeleide en monotoon stijgend, zodat $\max_{y \in [0, x]} f^{(n+1)}(y) = e^x$ en $|E_n(x)| \leq \frac{e^x}{(n+1)!} \cdot x^{n+1}$ voor $x > 0$. Deze formule is gebruikt in §1.3.

De afschatting voor de arctangens in §1.3 bij $x \geq 0$ is af te leiden door de integraalformule voor $E_{2n}(x)$ te vereenvoudigen tot $\int_0^x \frac{(-1)^n \cdot t^{2n}}{1+t^2} dt$, zodat:

$$\begin{aligned} |E_{2n}(x)| &= \left| \int_0^x \frac{(-1)^n \cdot t^{2n}}{1+t^2} dt \right| \\ &\leq \int_0^x \left| \frac{(-1)^n \cdot t^{2n}}{1+t^2} \right| dt \\ &= \int_0^x \frac{t^{2n}}{1+t^2} dt \\ &\leq \int_0^x t^{2n} dt \\ &= \frac{x^{2n+1}}{2n+1} \end{aligned}$$

1.7.3 Appendix over aftelbaarheid

De natuurlijke getallen zijn **aftelbaar**: ze zijn in een rij te zetten: $1, 2, 3, \dots$. Maar ook de rationale getallen zijn aftelbaar. Dat wil zeggen dat er een

afbeelding bestaat van de verzameling natuurlijke getallen naar de verzameling rationale getallen die zowel één-één (injectief) als op (surjectief) is. Met andere woorden: we kunnen de rationale getallen op een rij zetten, waarmee een identificatie tussen de natuurlijke en de rationale getallen gemaakt wordt. Om in te zien dat de positieve rationale getallen aftelbaar zijn, kun je de breuk $\frac{p}{q}$, als p en q positieve gehele getallen zijn zonder gemeenschappelijke delers, in het punt $[p, q]$ van het vlak intekenen en ze vervolgens aftellen door de tegendiagonalen op steeds grotere afstand van de oorsprong in het positieve kwadrant van het vlak af te lopen. De telling begint als volgt

$$\begin{array}{cccccccccccc} 1 & 2 & 1 & 3 & 1 & 1 & 2 & 3 & 4 & 1 & 5 \\ \frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{3}{1}, \frac{1}{3}, \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \frac{1}{5}, \frac{5}{1}, \dots \end{array}$$

De reële getallen zijn **overaftelbaar**: er is geen injectieve afbeelding van de verzameling natuurlijke getallen naar de verzameling reële getallen die surjectief is. Dit betekent dat de verzameling reële getallen groter is (een grotere **cardinaliteit** heeft) dan de verzameling natuurlijke getallen. In het bijzonder “passen” de reële getallen niet in de verzameling natuurlijke getallen.

1.7.4 Appendix over formules voor π met arctangens

We bespreken een afleiding van de benaderingsformule van π uit §1.3 met behulp van de arctangens. Uitgangspunt is de additieformule voor de tangens:

$$\tan(\alpha + \beta) = \frac{\tan(\alpha) + \tan(\beta)}{1 - \tan(\alpha) \cdot \tan(\beta)}.$$

Schrijf nu $u = \tan(\alpha)$. De verdubbelingsformule voor de tangens volgt uit de additieformule door $\beta = \alpha$ te kiezen:

$$\tan(2\alpha) = \frac{2u}{1 - u^2}.$$

Toepassing van de additieformule met $\tan\left(\frac{\pi}{4}\right) = 1$ geeft:

$$\tan\left(2\alpha - \frac{\pi}{4}\right) = \frac{\frac{2u}{1-u^2} - 1}{1 + \frac{2u}{1-u^2}} = -\frac{u^2 + 2u - 1}{u^2 - 2u - 1}.$$

Passen we aan beide zijden $\arctan$ toe, dan vinden we

$$2\alpha - \frac{\pi}{4} = -\arctan\left(\frac{u^2 + 2u - 1}{u^2 - 2u - 1}\right)$$

wat herschreven kan worden tot

$$8 \arctan(u) + 4 \arctan\left(\frac{u^2 + 2u - 1}{u^2 - 2u - 1}\right) = \pi.$$

Voor $u = \frac{1}{3}$ levert dit de in §1.3 gebruikte formule

$$8 \arctan\left(\frac{1}{3}\right) + 4 \arctan\left(\frac{1}{7}\right) = \pi.$$

Voor $u = \frac{1}{2}$ vinden we de variant

$$8 \arctan\left(\frac{1}{2}\right) - 4 \arctan\left(\frac{1}{7}\right) = \pi.$$

1.8 Oefeningen

1.8.1 Rationale getallen

Bereken het product van $0.\overline{846153}$ en $1.\overline{18}$.

1.8.2 Bewerkingen met reële getallen

Bereken de kop ter lengte 4 van de decimale ontwikkeling van $\frac{20 \cdot \pi}{231 \cdot e}$ met behulp van rationale boven- en ondergrenzen bepaald door de koppen ter lengte 5 van de decimale ontwikkelingen van π en e :

$$\text{kop}_5(\pi) = 3.14159, \quad \text{kop}_5(e) = 2.71828.$$

1.8.3 Riemannsommen voor benaderingen

Laat x een getal groter dan 1 zijn. Zoals besproken in §1.5 ligt het voor de hand om voor de decimale ontwikkeling van $\ln(x)$ de definitie $\ln(x) = \int_1^x \frac{dt}{t}$ te gebruiken. We klemmen $\ln(x)$ daarbij in tussen rationale getallen door gebruik te maken van Riemannsommen. De boven- en ondersom voor de functie $\frac{1}{t}$ op het interval $[1, x]$ bij een verdeling in n gelijke stukken geven de volgende afschattingen voor $\ln(x)$

$$\sum_{k=1}^n \frac{x-1}{n+k \cdot (x-1)} \leq \ln(x) \leq \sum_{k=0}^{n-1} \frac{x-1}{n+k \cdot (x-1)}.$$

We zullen hier zien dat de grenzen wel erg traag naar de goede waarde convergeren.

- (i) Laat zien dat het verschil tussen boven- en ondergrens gelijk is aan $\frac{(x-1)^2}{x \cdot n}$.
- (ii) Uit (i) volgt dat voor $n \rightarrow \infty$ het gesloten interval bepaald door de twee rationale grenzen inderdaad tot een punt convergeert. Om een benadering tot op 6 decimalen van $\ln(2)$ te kunnen maken, willen we n zo groot kiezen dat de lengte van het interval ten hoogste 10^{-7} is. Hoe groot moeten we n daartoe minstens kiezen?

1.8.4 Betere afschatting van π

Een betere afschatting van π dan in oefening 1.8.2 wordt bereikt door nog kleinere argumenten in de voorkomende arctangens-functies te vinden. Schrijf weer $\alpha = \tan(u)$ voor een variabele u .

- (i) Gebruik de verdubbelingsformule van appendix 1.7.4 voor de tangens om een rationale uitdrukking in u te vinden voor $\tan(4\alpha)$.
- (ii) Gebruik de additieformule voor de tangens nogmaals om een rationale uitdrukking in u te vinden voor $\tan\left(4\alpha - \frac{\pi}{4}\right)$. Noem deze uitdrukking v .
- (iii) Laat zien dat

$$16 \arctan(u) - 4 \arctan(v) = \pi.$$

- (iv) Kies $u = \frac{1}{5}$ en bepaal v . Tot welke concrete uitdrukking voor π leidt de formule uit het vorige onderdeel?
- (v) Als we in §1.3 deze formule gebruikt hadden in plaats van $\pi = 8 \arctan\left(\frac{1}{3}\right) + 4 \arctan\left(\frac{1}{7}\right)$, hoe groot hadden we het aantal termen n van de veelterm uit de Taylorreeks dan kunnen kiezen (in plaats van de daar gekozen waarde 7) om dezelfde precisie van 6 decimalen te bereiken?

1.8.5 Integraal van een rationale functie

Bepaal

$$\int \frac{x^2 + 1}{x^3 + 2x - 12} dx.$$

2 Het abc-vermoeden

Sander Dahmen

Het abc-vermoeden is een diep vermoeden over de interactie tussen optelling en vermenigvuldiging op de gehele getallen. Het werd in 1985 geformuleerd en heeft verreikende gevolgen voor onder andere Diophantische vergelijkingen. Dit vermoeden wordt gezien als één van de belangrijkste open problemen in de getaltheorie.

In deze voordracht zullen we verschillende formuleringen van het abc-vermoeden geven en enkele Diophantische gevolgen bespreken. Ook zullen we even ingaan op de vraag waarom we in het vermoeden zouden kunnen geloven en of er misschien al een bewijs voor is.



2.1 Optellen en vermenigvuldigen

Beschouw de verzameling van natuurlijke getallen

$$\mathbb{N} = \{1, 2, 3, 4, 5, 6, 7, \dots\}$$

en twee basisoperaties hierop: optellen en vermenigvuldigen.

Basiseigenschappen

Men zou kunnen zeggen dat het getal 1 het *atoom* voor optellen op $\mathbb{N}$ is:

$$1 = 1, \quad 2 = 1 + 1, \quad 3 = 1 + 1 + 1, \quad 4 = 1 + 1 + 1 + 1, \quad \dots$$

Oftewel, elk natuurlijk getal kan op unieke manier geschreven worden als som van énen. Wat in zekere zin een tautologie is. Een stuk interessanter wordt het als we naar vermenigvuldiging op $\mathbb{N}$ gaan kijken. Hiervoor zou men kunnen zeggen dat de *atomen* gegeven worden door de priemgetallen:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, \dots, 2017, \dots, 4207281, \dots, 2^{4207281} - 1, \dots$$

Dit komt tot uiting in de beroemde *hoofdstelling van de rekenkunde*, welke als volgt luidt.

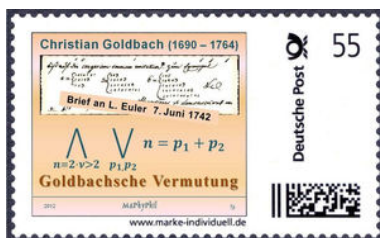
Stelling 1. *Elk natuurlijk getal kan geschreven worden als het product van priemgetallen. Deze schrijfwijze is uniek (op volgorde na).*

We hebben bijvoorbeeld

$$6 = 2 \cdot 3, \quad 12 = 2 \cdot 2 \cdot 3 = 2^2 \cdot 3, \quad 2016 = 2^5 \cdot 3^2 \cdot 7, \\ 2017 = 2017, \quad 2018 = 2 \cdot 1009, \quad 1 = \text{'leeg product'}.$$

Vermoedens en stellingen

De additieve en multiplicatieve structuur van de natuurlijke getallen kunnen afzonderlijk bestudeerd worden, maar vooral het samenspel tussen de basisoperaties geeft aanleiding tot zeer interessante vraagstukken die heel snel ontzettend moeilijk kunnen worden. We beginnen met twee klassieke voorbeelden over dit samenspel, een vermoeden en een stelling.



Vermoeden 1 (Goldbach 1742). *Elk even natuurlijk getal ≥ 4 is te schrijven als som van twee priemgetallen.*

We hebben bijvoorbeeld $4 = 2 + 2$, $6 = 3 + 3$, $8 = 3 + 5$, $10 = 3 + 7 = 5 + 5$, $2018 = 7 + 2011 = \dots = 1009 + 1009$. Het getal 2018 is zelfs op 28 manieren te schrijven als som van twee priemgetallen (zonder op de volgorde te letten). Met een computer is door Oliveira e Silva, Herzog en Pardi

bovenstaand vermoeden inmiddels geverifieerd voor getallen $\leq 4 \cdot 10^{18}$. Dit zou men kunnen interpreteren als een flinke aanwijzing dat het vermoeden waar is. Voorzichtigheid is echter geboden! Ter illustratie, beschouw de priemgetal-telfunctie $\pi(x) := (\text{aantal priemgetallen} \leq x)$ en de logaritmische integraal $\text{li}(x) := \int_0^x \frac{dx}{\log x}$, welke $\pi(x)$ asymptotisch benadert

in de zin dat $\lim_{x \rightarrow \infty} \frac{\pi(x)}{\text{li}(x)} = 1$ (een versie van de beroemde priemgetalstelling). Met de computer is door Büthe geverifieerd dat $\text{li}(x) > \pi(x)$ voor $2 \leq x \leq 10^{19}$ (en eerder waren er kleinere, maar nog steeds grote, bovengrenzen door anderen berekend), waaruit het wellicht verleidelijk is om het vermoeden uit te spreken dat $\text{li}(x) > \pi(x)$ voor alle $x \geq 2$. Er is, dankzij Littlewood, echter al meer dan een eeuw bekend dat de functie $\text{li}(x) - \pi(x)$ oneindig vaak van teken wisselt voor $x \geq 2$.

Om nog even terug te komen op het vermoeden van Goldbach. Men gaat eenvoudig na dat dit het zogenaamde *oneven Goldbach-vermoeden* impliceert, welke zegt dat elk oneven natuurlijk getal ≥ 9 te schrijven is als som van drie oneven priemgetallen. Dit laatste vermoeden is recentelijk tot stelling gepromoveerd, toen Helfgott in 2013 het bewijs ervoor afmaakte.

Een andere beroemde uitspraak die optellen en vermenigvuldigen combineert, opgeschreven omstreeks 1637 door Pierre de Fermat en bewezen in 1994 door Andrew Wiles, is de *Laatste Stelling van Fermat*.



Stelling 2. Voor alle $n, x, y, z \in \mathbb{N}$ met $n \geq 3$ geldt: $x^n + y^n \neq z^n$.

In Sectie 2.3 verderop zullen we bekijken wat het abc-vermoeden hier mee te maken heeft. Ook kijken we daar naar een, nog onbewezen, generalisatie van de Laatste Stelling van Fermat en gaan we in op wat het abc-vermoeden over deze generalisatie impliceert.

2.2 Formuleringen van het abc-vermoeden

Het abc-vermoeden gaat simpelweg over drietallen (a, b, c) van natuurlijke getallen die voldoen aan

$$a + b = c \quad \text{en} \quad \text{ggd}(a, b, c) = 1. \quad (2.1)$$

Voor het gemak geven we zulke drietallen een naam.

Definitie 1. Een *ABC-drietal* is een $(a, b, c) \in \mathbb{N}^3$ dat voldoet aan (2.1).

Voorbeelden zijn $(1, 1, 2)$, $(2, 3, 5)$ en $(2019, 2021, 4040)$. Heel grofweg impliceert het abc-vermoeden bijvoorbeeld dat als voor zo'n drietal a en b beide deelbaar zijn door ‘veel’ ‘hoge’ machten van priemgetallen, dat dit dan niet meer voor c kan gelden. Dit is natuurlijk vaag, maar beschouw ter illustratie $a := 2^4$ en $b := 3^3$. Aangezien $\text{ggd}(a, b) = 1$ geldt dat met $c := a + b$ we een drietal (a, b, c) verkrijgen met (2.1). Nu is $c = 16 + 27 = 43$ niet deelbaar door een hoge macht van een priemgetal, het is zelfs priem.

Nog een opmerking over onze *terminologie*. In sommige teksten wordt bij de definitie van ABC-drietal nog de extra eis gesteld dat $\text{rad}(abc) < c$, in andere teksten vindt men de ordeningseis $a < b$, en in weer andere beide eisen. Een ABC-drietal (met onze definitie) dat aan deze beide extra eisen voldoet, zullen wij een *ABC-hit* noemen; zie Definitie 3 verderop.

Radicaal en formulering

Om het abc-vermoeden precies te formuleren, introduceren we het volgende fundamentele begrip.

Definitie 2. Het *radicaal* van $n \in \mathbb{N}$ is het product van zijn priemdelers:

$$\text{rad}(n) := \prod_{p|n, \text{ } p \text{ priem}} p. \quad (2.2)$$

Bijvoorbeeld

$$\begin{aligned} \text{rad}(2016) &= \text{rad}(2^5 \cdot 3^2 \cdot 7) = 2 \cdot 3 \cdot 7 = 42, \\ \text{rad}(2013) &= \text{rad}(3 \cdot 11 \cdot 61) = 3 \cdot 11 \cdot 61 = 2013, \\ \text{rad}(1) &= 1. \end{aligned}$$

Het abc-vermoeden, uit 1985 en afkomstig van Joseph Oesterlé en David Masser, kan nu als volgt geformuleerd worden.

Vermoeden 2 (abc-vermoeden). Voor alle $\epsilon > 0$ zijn er slechts eindig veel ABC-drietallen (a, b, c) zodat

$$c > \text{rad}(abc)^{1+\epsilon}. \quad (2.3)$$

Wellicht ten overvloede merken we op dat het aantal 0 ook ‘eindig veel’ is.



Joseph Oesterlé (links) en David Masser (rechts)

Laten we om een beetje gevoel te krijgen voor het abc-vermoeden de ϵ eens negeren en kijken of we ABC-drietallen $(a, b, c) \in \mathbb{N}^3$ kunnen vinden zodat $c > \text{rad}(abc)$. Wegens symmetrie kunnen we ons beperken tot $a < b$ (merk op dat $a = b$ in (2.1) impliceert dat $(a, b, c) = (1, 1, 2)$, hiervoor geldt $c = \text{rad}(abc)$). Zulke drietallen geven we een naam.

Definitie 3. Een *ABC-hit* is een drietal $(a, b, c) \in \mathbb{N}^3$ met

$$a + b = c, \quad \text{ggd}(a, b, c) = 1, \quad a < b \quad \text{en} \quad c > \text{rad}(abc).$$

Het drietal $(a, b, c) = (1, 8, 9)$ geeft

$$\text{rad}(abc) = \text{rad}(2^3 \cdot 3^2) = 2 \cdot 3 = 6 < 9 = c.$$

Dit blijkt de enige ABC-hit te zijn met $c < 32$. Voor $c \leq 100$ zijn er verder nog $(5, 27, 32)$, $(1, 48, 49)$, $(1, 63, 64)$, $(1, 80, 81)$ en $(32, 49, 81)$. In totaal zijn er echter 3043 ($\approx 100^2 \cdot \frac{3}{\pi^2} = 3039,6\dots$) ABC-drietallen (a, b, c) met $a < b$ en $c \leq 100$, waaronder dus slechts 6 ABC-hits. Het lijkt erop ABC-hits zeldzaam zijn, het is desalniettemin niet zo moeilijk om oneindige families van dergelijke drietallen te construeren. Het eenvoudigste voorbeeld is waarschijnlijk

$$(a_n, b_n, c_n) = (1, 9^n - 1, 9^n), \quad n \in \mathbb{N}.$$

Merk op dat $n = 1$ bovenstaand voorbeeld weer oplevert. Door haakjes uitwerken krijgen we dat $9^n = (1 + 8)^n = 1 + 8k$ voor een zekere $k \in \mathbb{N}$.

Hieruit volgt dus dat b_n deelbaar is door $8 = 2^3$ en derhalve $\text{rad}(b_n) \leq b_n/4 < c_n/4$. Zo krijgen we

$$\text{rad}(a_n b_n c_n) = \text{rad}(a_n) \text{rad}(b_n) \text{rad}(c_n) = 1 \cdot \text{rad}(b_n) \cdot 3 < \frac{3}{4} c_n < c_n.$$

Hiermee is aangetoond dat het simpelweg weghalen van de ϵ in het abc-vermoeden een uitspraak oplevert die niet waar is. Een logische vraag zou zijn of het mogelijk is de ϵ weg te halen ten koste van een multiplicatieve constante voor het radicaal, oftewel kan (2.3) vervangen worden door $c > C \text{rad}(abc)$ waarbij C een constante is? Dat dit niet mogelijk is (zonder een niet-ware uitspraak te krijgen), volgt uit de volgende opgave.

Opgave 1. Beschouw de familie van ABC-drietallen gegeven door

$$(a_n, b_n, c_n) = (1, 3^{2^n} - 1, 3^{2^n}), \quad n \in \mathbb{N}.$$

- (a) Bewijs dat voor alle $n \in \mathbb{Z}_{\geq 0}$ geldt dat $3^{2^n} - 1$ deelbaar is door 2^{n+1} . (Hint: pas herhaaldelijk de identiteit $x^2 - 1 = (x + 1)(x - 1)$ toe op $3^{2^n} - 1$; of gebruik Eulers totiëntstelling als men hier mee bekend is.)
- (b) Laat nu zien dat voor alle $n \in \mathbb{N}$ geldt

$$\text{rad}(a_n b_n c_n) < \frac{3}{2^n} c_n.$$

Opgave 2. Een kleine variatie op bovenstaande formulering, die vaak gebruikt wordt, is als volgt.

Voor alle $\epsilon > 0$ is er een $C_\epsilon > 0$ zodat voor alle drietallen $(a, b, c) \in \mathbb{N}^3$ met (2.1) geldt

$$c < C_\epsilon \text{rad}(abc)^{1+\epsilon}.$$

Bewijs dat bovenstaande formulering equivalent is met Vermoeden 2.

Kwaliteit

Om het abc-vermoeden wat inzichtelijker te maken, introduceren we nog een begrip. Voor ABC-drietal (a, b, c) beschouwen we het getal q met

$$c = \text{rad}(abc)^q.$$

Eliminatie van q hieruit (m.b.v. log), geeft aanleiding tot het volgende.

Definitie 4. De *kwaliteit* van een ABC-drietal (a, b, c) is

$$q(a, b, c) := \frac{\log(c)}{\log(\text{rad}(abc))}.$$

Hieronder een paar (semi-willekeurige) numerieke voorbeelden.

a	b	c	$\text{rad}(abc)$	$q(a, b, c)$
3^4	$5^2 \cdot 7$	$2^8 = 256$	$2 \cdot 3 \cdot 5 \cdot 7 = 210$	$\frac{\log(256)}{\log(210)} = 1,03704 \dots$
3^2	2^6	73	$2 \cdot 3 \cdot 73 = 438$	$\frac{\log(73)}{\log(438)} = 0,70541 \dots$
1	1	2	2	$\frac{\log(2)}{\log(2)} = 1$

Opgabe 3. Laat zien $(1, 1, 2)$ het enige ABC-drietal is waarvoor de kwaliteit een rationaal getal is.

Merk op dat voor een ABC-drietal (a, b, c) de conditie $c > \text{rad}(abc)$ direct vertaalt naar $q(a, b, c) > 1$ (in welk geval (a, b, c) of (b, a, c) een ABC-hit is). Verder vertaalt Vermoeden 2 zich direct naar de volgende formulering.

Vermoeden 3 (abc-vermoeden). *Voor elke $Q > 1$ zijn er slechts eindig veel ABC-drietalen met kwaliteit $> Q$.*

Als het abc-vermoeden waar is, dan bestaat er een grootse kwaliteit onder alle ABC-drietalen. Dit levert het volgende gevolg op.

Vermoeden 4 (zwak abc-vermoeden). *Er bestaat een $Q_{\max} > 1$ zodat alle ABC-drietalen kwaliteit $\leq Q_{\max}$ hebben.*

Hieronder de top drie van ABC-hits met hoogst bekende kwaliteit.

medaille	a	b	c	$q(a, b, c)$
goud	2	$3^{10} \cdot 109$	23^5	$1,62991$
zilver	11^2	$3^2 \cdot 5^6 \cdot 7^3$	$2^{21} \cdot 23$	$1,62599$
brons	$19 \cdot 1307$	$7 \cdot 29^2 \cdot 31^8$	$2^8 \cdot 3^{22} \cdot 5^4$	$1,62349$

Het gouden drietal werd ontdekt door Eric Reyssat in 1987. Het zilver is voor Benne de Weger, hij ontdekte zijn drietal in 1985 en had daarmee toen dus even goud in handen. Brons is nu voor Jerzy Browkin en Juliusz Brzezinski, met ontdekkingsjaar 1994.

Dit zijn alle tot op heden bekende ABC-hits met kwaliteit ten minste 1,6. Volgens het abc-vermoeden zijn er maar eindig veel van zulke ABC-hits en het is goed denkbaar dat bovenstaande 3 ze allemaal zijn, of dat ze ten minste de absolute top drie vormen. Er zijn verder nog 10 ABC-hits bekend met kwaliteit tussen 1,5 en 1,6, welke we voor de aardigheid geven.

plek	a	b	c	$q(a, b, c)$
4	283	$5^{11} \cdot 13^2$	$2^8 \cdot 3^8 \cdot 17^3$	1,58076
5	1	$2 \cdot 3^7$	$5^4 \cdot 7$	1,56789
6	7^3	3^{10}	$2^{11} \cdot 29$	1,54708
7	$7^2 \cdot 41^2 \cdot 311^3$	$11^{16} \cdot 13^2 \cdot 79$	$2 \cdot 3^3 \cdot 5^{23} \cdot 953$	1,54443
8	5^3	$2^9 \cdot 3^{17} \cdot 13^2$	$11^5 \cdot 17 \cdot 31^3 \cdot 137$	1,53671
9	$13 \cdot 19^6$	$2^{30} \cdot 5$	$3^{13} \cdot 11^2 \cdot 31$	1,52700
10	$3^{18} \cdot 23 \cdot 2269$	$17^3 \cdot 29 \cdot 31^8$	$2^{10} \cdot 5^2 \cdot 7^{15}$	1,52216
11	$13^{10} \cdot 37^2$	$3^7 \cdot 19^5 \cdot 71^4 \cdot 223$	$2^{26} \cdot 5^{12} \cdot 1873$	1,50943
12	$2^7 \cdot 23^8$	$19^9 \cdot 857^2$	$3^{22} \cdot 13 \cdot 47^2 \cdot 263$	1,50332
13	239	$5^8 \cdot 17^3$	$2^{10} \cdot 37^4$	1,50284

Ook op ABC-drietallen met kwaliteit ten minste 1,4 wordt nog gejaagd; deze worden *goed* genoemd, ook al is de grens van 1,4 tamelijk arbitrair. Benne de Weger stelde als eerste een lijst op met bekende goede ABC-drietallen en momenteel is de meest recente lijst te vinden op de website van Bart de Smit, met ook o.a. info over ontdekkers:

<http://www.math.leidenuniv.nl/~desmit/abc/index.php?set=2>

Hierop staan nog 227 ABC-hits met kwaliteit tussen 1,4 en 1,5. Er zijn dus 240 goede ABC-hits bekend. Volgens het abc-vermoeden zijn er slechts eindig veel ABC-hits, maar het lijkt onwaarschijnlijk dat deze 240 ze allemaal zijn. Alle goede ABC-hits (a, b, c) met $c < 10^{20}$ staan echter wel gegarandeerd in de lijst. Recentelijk, namelijk op 15 mei 2017, is de laatste erbij gekomen:

$$(a, b, c) = (3^{13} \cdot 37^5 \cdot 44939^2, 5^5 \cdot 7^2 \cdot 3 \cdot 19 \cdot 463 \cdot 863 \cdot 2^{20} \cdot 53^8 \cdot 61 \cdot 113^4)$$

ontdekt door Fank Rubin. Met een kwaliteit van 1,4206 staat dit drietal (voorlopig) op de 135e plek. Opmerkelijker is de grootte, c heeft namelijk maar liefst 30 cijfers en is tevens de grootste c behorende bij een goede ABC-hit. Er zijn nog 2 andere goede ABC-hits bekend met een c van 30 cijfers, beide ook gevonden door Rubin.

Computerinzet

Het moge duidelijk zijn dat veel van het bovenstaande berust op gebruik van de computer. Alleen al het vinden van de priemfactoren zou in veel gevallen al een echte uitdaging zijn om met de hand te doen. Ook het produceren van een (bewezen) complete lijst van goede ABC-hits (a, b, c) met $c < 10^{20}$ is natuurlijk ondoenbaar zonder computer. Het algoritme dat hiervoor gebruikt is, is ook ingezet om een (bewezen) complete lijst van

alle ABC-hits (a, b, c) met $c < 10^{18}$ te vinden. Dit laatste is uitgevoerd middels het gedistribueerd computerproject *ABC@Home*, opgezet door het Mathematisch Instituut in Leiden in 2006, als onderdeel van het educatieve project *Reken mee met abc*. Hierbij hebben wereldwijd duizenden mensen computertijd beschikbaar gesteld om zo stukjes van de benodigde zoekruimte af te speuren. In 2011 werd de grens van 10^{18} bereikt; in het bijzonder is nu bekend dat er precies 14 482 065 ABC-hits (a, b, c) met $c < 10^{18}$ zijn. Dit rekenproject heeft hierna nog een aantal jaren verdere data over ABC-hits verzameld totdat het in 2015 beëindigd werd.

Tot slot geven we nog heel kort twee andere indicaties voor het berekenen van goede ABC-drietallen. De eerste is met behulp van zogenaamde *kettingbreuken*. Zonder dit verder precies uit te leggen, willen we hierover zegen dat dit *goede* (in de zin van relatief kleine noemer) rationale benaderingen geeft van irrationale getallen. Als voorbeeld geven we dat de eerste 5 kettingbreukbenaderingen van $\sqrt[5]{109}$ achtereenvolgens de breuken $2, 3, 5/2, 23/9, 1787864/699599$ opleveren. Noemer (en teller) ‘exploderen’ in de laatste breuk, wat aangeeft dat de breuk ervoor al (relatief) een hele goede benadering is. Dus $\sqrt[5]{109} \approx 23/3^2$, oftewel $3^{10} \cdot 109 - 23^5$ is in absolute waarde ‘klein’. Uitrekenen geeft dat dit -2 is, wat ons brengt op $2 + 3^{10} \cdot 109 = 23^5$, waarmee we de gouden ABC-hit terugvinden. Dit zou in principe zelfs met de hand gedaan kunnen worden, al is het de vraag hoe je er dan op komt om met $\sqrt[5]{109}$ te starten. Volgens Reyssat zelf heeft hij trouwens zijn drietal met ‘brute kracht’ gevonden. Voor het semi-systematisch zoeken naar grotere ABC-hits d.m.v. kettingbreuken is het zeker wel handig om een (eenvoudige) computer te gebruiken. En op deze wijze zijn er inderdaad enkele goede ABC-hits gevonden. Een nadeel van deze methode is dat het type ABC-hits (a, b, c) hierbij altijd ‘kleine’ a heeft, wat een significante beperking is. Iets algemener kan men beginnen met $A, B, C \in \mathbb{N}$ die allen ‘klein’ radicaal hebben en $\text{ggd}(A, B, C) = 1$. Vervolgens is het zoeken geblazen naar in absolute waarde ‘zo klein mogelijke’ $\alpha, \beta, \gamma \in \mathbb{Z}$ ongelijk 0, zodat $\alpha A + \beta B + \gamma C = 0$, dit kan gedaan worden met het beroemde *LLL-algoritme*. We kunnen natuurlijk aannemen dat $\text{ggd}(\alpha A, \beta B, \gamma C) = 1$. Middels $\text{rad}(|\alpha A \beta B \gamma C|) \leq |\alpha \beta \gamma| \text{rad}(A) \text{rad}(B) \text{rad}(C)$ krijgen we zo, als α, β, γ daadwerkelijk ‘klein genoeg’ in absolute waarde zijn, dat een eventuele permutatie van $(|\alpha A|, |\beta B|, |\gamma C|)$ dan een ABC-drietal oplevert waarvan de kwaliteit ‘hoog’ is. Op deze wijze zijn ook een aantal goede ABC-hits gevonden.

Bovengenoemde data verschaft ons zeker inzicht in het abc-vermoeden. Maar hoe ver we zulke grenzen ook oprekken met nog snellere en slimmere

algoritmen en computers, het abc-vermoeden kan er niet mee worden bewezen of weerlegd. Hiervoor moet geredeneerd worden over oneindig veel ABC-drietallen middels een fantastisch nieuw idee. (Heel misschien is dat er al, zie de laatste paragraaf ...)

Opgave 4.

- (a) Probeer zelf (zonder in een tabel te spieken) een ABC-drietal te vinden met zo hoog mogelijke kwaliteit.
- (b) Probeer een nieuwe goede ABC-hit te vinden.
(Waarschuwing: de auteur is dit (zoals verwacht) niet gelukt binnen een halve dag; er is zelfs geen 100% garantie dat het mogelijk is...)

2.3 Gevolgen

De kracht en het belang van het abc-vermoeden komen o.a. tot uiting door de vele interessante gevolgen die er uit kunnen worden afgeleid. We noemen er hier slechts enkele, gerelateerd aan de Laatste Stelling van Fermat. Voor een langer lijstje van gevolgen, zie bijvoorbeeld:

<http://www.math.unicaen.fr/~nitaj/abc.html#Consequences>

Merk op dat sommige uitspraken hierop volgen uit bepaalde generalisaties van het abc-vermoeden (bv. met meer dan drie termen of in grotere getalsystemen dan de natuurlijke of gehele getallen).

De Fermat-vergelijking

Om te beginnen impliceert het abc-vermoeden een asymptotische versie van de Laatste Stelling van Fermat.

Stelling 3. *Neem aan: alle ABC-drietallen hebben kwaliteit $\leq \alpha$. Dan geldt voor alle $x, y, z, n \in \mathbb{N}$ met $n \geq 3\alpha$ dat*

$$x^n + y^n \neq z^n.$$

Bewijs. Laat $x_0, y_0, z_0, n \in \mathbb{N}$ met $x_0^n + y_0^n = z_0^n$. Definieer $g := \text{ggd}(x_0, y_0, z_0)$ en $(x, y, z) := (x_0/g, y_0/g, z_0/g)$. Dan geldt $x^n + y^n = z^n$ en $\text{ggd}(x, y, z) = 1$. Dit geeft een ABC-drietal $(a, b, c) = (x^n, y^n, z^n)$ met

$$\text{rad}(abc) = \text{rad}(x^n y^n z^n) = \text{rad}(xyz) \leq xyz < z^3.$$

Voor de kwaliteit krijgen we dan

$$q(a, b, c) = \frac{\log(c)}{\log(\text{rad}(abc))} > \frac{\log(z^n)}{\log(z^3)} = \frac{n}{3}.$$

Samen met de aanname $\alpha \geq q(a, b, c)$ geeft dit $\alpha > n/3$, oftewel

$$n < 3\alpha.$$

□

De gegeneraliseerde Fermat-vergelijking

We kunnen ook naar een generalisatie van Fermat's probleem kijken. Voor exponenten $p, q, r \in \mathbb{N}$ beschouwen we de Diophantische vergelijking

$$x^p + y^q = z^r, \quad \text{ggd}(x, y, z) = 1 \quad (2.4)$$

die we willen oplossen met $x, y, z \in \mathbb{N}$. Dit wordt de *gegeneraliseerde Fermat-vergelijking* genoemd. Als de exponenten p, q, r klein zijn, in de zin dat $1/p + 1/q + 1/r \geq 1$, dan zijn alle oplossingen voor (2.4) al geclassificeerd. Voor $1/p + 1/q + 1/r < 1$ zijn er enkele speciale gevallen opgelost, maar in het algemeen is het oplossen van (2.4) een groot open probleem. De enige oplossingen die gevonden zijn voor $1/p + 1/q + 1/r < 1$ worden gegeven door de Catalan oplossing $1^p + 2^3 = 3^2$ (met $p \in \mathbb{Z}_{\geq 7}$) en de 9 sporadische oplossing

$$\begin{aligned} 2^5 + 7^2 &= 3^4, & 3^5 + 11^4 &= 122^2, & 9262^3 + 15312283^2 &= 113^7, \\ 2^7 + 17^3 &= 71^2, & 17^7 + 76271^3 &= 21063928^2, & 1414^3 + 2213459^2 &= 65^7, \\ 43^8 + 96222^3 &= 30042907^2, & 33^8 + 1549034^2 &= 15613^3, & 7^3 + 13^2 &= 2^9. \end{aligned}$$

Een vermoeden is dat dit alle oplossingen zijn. Merk op dat bij bovenstaande oplossingen altijd een exponent gelijk aan 2 voorkomt. Dit geeft aanleiding tot een directe vermoedelijke generalisatie van de Laatste Stelling van Fermat.

Vermoeden 5. Voor $x, y, z, p, q, r \in \mathbb{N}$ met $p, q, r \geq 3$ geldt nooit (2.4).

Het is iets zwakker dan het eerder genoemde vermoeden, maar nog steeds een groot open probleem. Voor het bewijzen of weerleggen van Vermoeden 5 heeft de Amerikaanse bankier en zakenman Andrew Beal een miljoen dollar beschikbaar gesteld. Het abc-vermoeden impliceert weer een asymptotische versie van dit vermoeden, maar ook een asymptotische versie van het sterkere vermoeden dat over alle p, q, r met $1/p + 1/q + 1/r < 1$ gaat.

Vermoeden 6. *Er bestaan slechts eindig veel drietallen zuivere machten (x^p, y^q, z^r) met $x, y, z, p, q, r \in \mathbb{N}$ en $1/p + 1/q + 1/r < 1$ zodat (2.4) geldt.*

Opgave 5.

- (a) Laat $p, q, r \in \mathbb{N}$ met $1/p + 1/q + 1/r < 1$. Laat zien dat $1/p + 1/q + 1/r \leq 1/2 + 1/3 + 1/7 = 41/42$.
(Opmerking: dit is enigszins ‘priegelen’; het resultaat kan ook aangenomen worden, zodat met het volgende, veel interessantere, onderdeel snel verder gegaan kan worden.)
- (b) Bewijs dat Vermoeden 6 geïmpliceerd wordt door het abc-vermoeden.

2.4 Is het abc-vermoeden waar?

We kunnen deze vraag niet eenduidig beantwoorden, maar we zullen wel een aantal aanwijzingen in de richting van het abc-vermoeden geven. Het liefst zouden we natuurlijk een echt bewijs beschikbaar hebben. Mochizuki claimt zo’n bewijs gevonden te hebben, hier gaan we heel kort op in.

Aanwijzingen

De meeste getaltheoretici denken dat het abc-vermoeden waar is. We geven een aantal aanwijzingen hiervoor.

- ABC-drietallen met grote kwaliteit vinden lijkt lastig.
- Het impliceert veel ware en aannemelijke uitspraken (en tot op heden geen uitspraken waarvan we weten dat ze onwaar zijn).
- Er is een probabilistische heuristiek (zoals net doen of een groot natuurlijk getal n een ‘kans’ van $1/\log n$ heeft om priem te zijn) die het abc-vermoeden verklaart. Zie bijvoorbeeld de blog van Terry Tao: <https://terrytao.wordpress.com/2012/09/18/the-probabilistic-heuristic-justification-of-the-abc-conjecture/>
- Er zijn getsystemen met analoge eigenschappen aan de gehele getallen waarvoor een corresponderend abc-vermoeden bewezen is.

Dit laatste zullen we nog wat preciezer maken. Ten eerste kunnen we het abc-vermoeden iets symmetrischer maken door het voor de gehele getallen $\mathbb{Z}$ te formuleren. Elke $n \in \mathbb{Z}$ ongelijk 0 is op unieke manier (op volgorde na) te schrijven als een *eenheid*, d.w.z. 1 of -1 , keer een product van priemgetallen. Het radicaal van zo’n n kunnen we dus ook door for-

mule (2.2) definiëren. (We hadden ook eenvoudigweg $\text{rad}(n) := \text{rad}(|n|)$ kunnen zeggen.) Het is eenvoudig in te zien dat we het abc-vermoeden als volgt kunnen herformuleren.

Vermoeden 7. *Voor alle $\epsilon > 0$ geldt voor alle behalve eindig veel drietallen $(a, b, c) \in (\mathbb{Z} - \{0\})^3$ met $a + b + c = 0$ en $\text{ggd}(a, b, c) = 1$ dat*

$$\max(\log |a|, \log |b|, \log |c|) < (1 + \epsilon) \log |\text{rad}(abc)|.$$

Beschouw nu alle polynomen in 1 variabele met complexe coëfficiënten, genoteerd als $\mathbb{C}[x]$. Deze *ring* (een ‘getalsysteem’ met optelling en vermenigvuldiging) lijkt qua structuur enigszins op $\mathbb{Z}$. In beide kan men goed ‘deling met rest’ uitvoeren en (daarom) is elk element ongelijk 0 te schrijven als een eenheid maal een product van priemelementen. In de ring $\mathbb{C}[x]$ zijn de eenheden de constante polynomen ongelijk 0 en als priemelementen nemen we de polynomen van de vorm $x - \alpha$ met $\alpha \in \mathbb{C}$. Het radicaal $\text{rad}(f)$ van een polynoom $f \in \mathbb{C}[x]$ ongelijk 0 is nu per definitie weer het product van zijn priemdelers. Nemen we bv. $f = 2x^5 - 2x^4 - 2x + 2$, dan kunnen we f factoriseren als

$$f = 2(x-1)(x^4-1) = 2(x-1)(x^2-1)(x^2+1) = 2(x-1)^2(x+1)(x-i)(x+i).$$

Derhalve $\text{rad}(f) = (x-1)(x+1)(x-i)(x+i) = (x^2-1)(x^2+1) = x^4-1$. De functie $f \mapsto \deg f$ (die de graad van een polynoom geeft) op $\mathbb{C}[x] - \{0\}$ en de functie $n \mapsto \log(|n|)$ op $\mathbb{Z} - \{0\}$ lijken op elkaar, ze hebben in ieder geval gemeen dat de functiewaarde van een product van 2 elementen gelijk is aan de som van de afzonderlijke functiewaarden van de 2 elementen. Om het analogon van het abc-vermoeden voor $\mathbb{C}[x]$ te krijgen, kunnen we nu in Vermoeden 7 gelijktijdig $\mathbb{Z}$ voor $\mathbb{C}[x]$ vervangen en elke $\log |\cdot|$ voor $\deg(\cdot)$. Dit geeft ons een uitspraak die niet waar is, wat komt door het flauwe geval dat alle drie de elementen eenheden zijn (iets wat voor het $\mathbb{Z}$ -geval niet mogelijk is). Sluiten we dit uit, dan blijken we iets waars te krijgen. Verder kunnen we ook nog zowel de ϵ als de ‘behalve eindig veel’ weggooien.

Stelling 4. *Voor alle drietallen $(a, b, c) \in (\mathbb{C}[x] - \{0\})^3$ met a, b, c niet allen constant, $a + b + c = 0$ en $\text{ggd}(a, b, c) = 1$ geldt dat*

$$\max(\deg a, \deg b, \deg c) < \deg \text{rad}(abc).$$

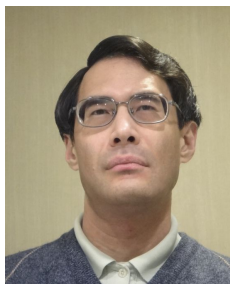
Het is relatief elementair om deze stelling, vaak vernoemd naar Mason en Stothers, te bewijzen. We zullen dit hier echter niet doen, maar willen enkel nog opmerken dat het in feite berust op het kunnen differentiëren

van polynomen en zo $\deg \text{rad } f$ uit te drukken als $\deg f - \deg \text{ggd}(f, f')$. Uit Stelling 4 kunnen we weer interessante gevolgen voor $\mathbb{C}[x]$ afleiden. Bijvoorbeeld de Laatste Stelling van Fermat voor polynomen.

Stelling 5. *Laat $n \in \mathbb{Z}_{\geq 3}$ en $f, g, h \in \mathbb{C}[x]$ niet allen constant en met $\text{ggd}(f, g, h) = 1$. Dan geldt $f^n + g^n \neq h^n$.*

Opgave 6. Bewijs Stelling 5. (Stelling 4 mag natuurlijk gebruikt worden.)

Een bewijs?



Shinichi Mochizuki

De Japanse wiskundige Shinichi Mochizuki, die in het verleden al grote faam verworven heeft door het zogenaamde Grothendieck-vermoeden over anabelse meetkunde te bewijzen, beweert een bewijs te hebben voor het abc-vermoeden. Op 30 augustus 2012 publiceerde hij 4 artikelen ‘Inter-universal Teichmüller Theory I,II,III,IV’ op zijn eigen website, die ruim 500 pagina’s exotische (inter-universele?) meetkunde bevatten. Het abc-vermoeden zou een direct gevolg van de algemene theorie zijn. Hij heeft er meer dan 10 jaar aan gewerkt en het berust op nog eens vele honderden pagina’s ander eigen werk. Verder lijkt er een duidelijke overzichtsstrategie van het bewijs te ontbreken. Het enige dat de auteur erover durft te zeggen, is dat Mochizuki lijkt te mikken op het *Szpiro vermoeden*, een vermoeden over elliptische krommen waarvan bekend is dat het het abc-vermoeden impliceert (dit is ook gerelateerd aan de ontstaansgeschiedenis van het abc-vermoeden). Er zijn inmiddels al wel enige conferenties gehouden om het werk van Mochizuki te begrijpen, maar consensus binnen de wiskundige gemeenschap over Mochizuki’s bewijsclaim lijkt echter nog ver weg.

3 Wat kun je allemaal zonder hulpmiddelen? En daarna?

Joost Hulshof

Integraalrekening

Woord vooraf: ik verwijs af en toe naar het groene boekje “Wiskunde in je Vingers” met Ronald Meester [HM] (VU University Press, 2015). Onderstaande tekst bevat net als [HM] geen plaatjes. Het is verstandig en leerzaam om die zelf te maken, op blanco papier of op ruitjespapier¹.

Wat ik schrijf is onderdeel van een hoofdstuk over integraalrekening uit een nieuw boek/dictaat dat ik met permissie voorlopig maar “Basisboek Analyse” noem², en dat op het web wel te vinden is. Het verwerken van kritiek van collega’s (met name René Swarttouw) wissel ik af met het lezen in [Eves], de zesde editie van “An Introduction to the History of Mathematics” van (vader Howard) Eves (en zoon Jamie voor de cultural connections) (McGraw-Hill, 2006).

Dat hoofdstuk over integraalrekening komt na een algemene inleiding over wat nu vaak gezien wordt als de fundamenteën van de wiskunde, en begint met formules voor sommen van machten de facto bij Archimedes. Voor bewijzen (met volledige inductie) van die formules zie eventueel ook Hoofdstuk 6 van [HM], maar Opgaven 3.1.5, 3.1.6, 3.1.7 bieden een leuk en wat constructiever alternatief waarbij iets minder geraden wordt.

Terzijde: integraalrekening gaat in het nieuwe boek aan differentiaalrekening vooraf, zoals ik dat voor het eerst zag in de calculusboeken van Apostol. Jan Aarts wees me daar destijds op en Eves beschrijft het in zijn boek ook als de historische volgorde. In het nieuwe boek is de differentiaalrekening echter vanaf het begin gebaseerd op lineaire benaderingen, die volgens niet alleen Eves gebruikt³ werden voor de benadering van $\sqrt{2}$ op kleitablet

¹ Met vierkante ruitjes.

² Met als subtitel “not another ε -book”.

³ Dat wist ik nog niet toen we [HM] schreven.

YBC7289. Dat kleitablet is ouder dan het werk van Archimedes. Wat de natuurlijke volgorde is kan dus onderwerp van discussie blijven.

Zowel differentiaalrekening (met lineaire benaderingen) als integraalrekening (met benaderende eindige sommen) kan beginnen met simpele voorbeelden, bijvoorbeeld voor de functies $x \rightarrow x^n$ met $n = 0, 1, 2, 3, \dots$, met expliciete uitdrukkingen waaraan je ziet hoe het werkt en wat er uit komt, en in bijzonder de observatie dat de hoofdstelling van de integraalrekening voor monomen geldt. De stap van monomen naar polynomen ligt dan voor de hand, en vervolgens is een snelle keuze voor het generaliseren naar machtreeksen een optie, nog voor continuïteit⁴ van functies aan de orde komt.

Een opmerking van Jan Wiegerinck bracht me op de gedachte om⁵ een minder gangbare convergentiestelling voor integralen van $f_n : [a, b] \rightarrow \mathbb{R}$ te formuleren, waarbij monotonie in zowel $n \in \mathbb{N}$ als $x \in [a, b]$ wordt aangenomen. Het bewijs van die stelling is een variant op het eenvoudige bewijs dat monotone functies Riemann integreerbaar zijn. De stelling⁶ kan worden gebruikt om de stap van polynomen $a_0 + a_1x + a_2x^2 + \dots + a_Nx^N$ naar machtreeksen

$$a_0 + a_1x + a_2x^2 + \dots$$

voor integraalrekening snel te maken, eerst voor het geval dat $0 \leq a \leq b$ en alle $a_n \geq 0$, en vervolgens door splitsen in negatieve en positieve termen in het algemene geval.

Helemaal expliciet kan integraalrekening voor machtreeksen natuurlijk ook worden gedaan maar dat is wat bewerkelijk. Voor differentiaalrekening is de stap van polynomen naar machtreeksen (met lineaire benaderingen en expliciete foutafschattingen) veel makkelijker, zoals ik leerde uit het boek van Conway over *complex functions of one variable*. Komt in deze vakantiecursus niet aan de orde, maar is te vinden in Hoofdstuk 10 van [HM].

Bij integraalrekening kan de analyse zich aanvankelijk beperken tot de uitspraak dat monotone (niet-dalend of niet-stijgend) begrensde rijen in $\mathbb{R}$ convergent zijn (met hun kleinste bovengrens of grootste ondergrens als limiet), en de observatie dat iedere rij in $\mathbb{R}$ een deelrij heeft die monotoon is. Daaruit volgt de in de analyse zo belangrijke stelling dat iedere begrensde rij in $\mathbb{R}$ een convergente deelrij heeft. Han Peters wees me op deze wat minder gebruikelijke maar wel zo efficiënte manier om het bestaan van

⁴ Secties 3.4 en 3.5 gaan over continuïteit en limieten, maar zijn over te slaan.

⁵ Hier nu al in de laatste subsectie.

⁶ Kan ook worden gebruikt voor het bewijs in [HM] van de formule van Stirling.

convergente deelrijen te bewijzen.

De al genoemde hoofdstelling⁷ van de integraalrekening legt het verband tussen differentiaalrekening en integraalrekening door de b in

$$\int_a^b f(x) dx$$

variabel te maken en de variatie in de integraal op de voor de hand liggende manier lineair te benaderen. Lineaire benaderingen gaan terug tot kleitablet YBC7289. Oppervlakteberekeningen met polygoonbenaderingen gaan minstens terug tot Archimedes. Eves speculeert dat formules voor sommen van kwadraten al eerder bekend waren. Je vraagt je af welke vraagstelling destijds al tot het ontdekken van het verband had kunnen leiden. De notaties van nu hadden ze natuurlijk nog niet⁸.

3.1 Terug naar Archimedes met simpele voorbeelden

Voor functies $f : \mathbb{R} \rightarrow \mathbb{R}$ begint integraalrekening met de vraag naar de oppervlakte van verzamelingen in het x, y -vlak begrensd door de x -as en de grafiek van f . Die grafiek wordt gegeven door $y = f(x)$. Om niet meteen over onbegrensd gebieden te praten kijken we eerst naar verzamelingen van de vorm

$$A_+ = \{(x, y) : x, y \in \mathbb{R}, 0 < y < f(x), a < x < b\}$$

en

$$A_- = \{(x, y) : x, y \in \mathbb{R}, f(x) < y < 0, a < x < b\},$$

met $a, b \in \mathbb{R}$ en $a < b$, waarbij we opmerken dat de “grotere” verzamelingen die ontstaan als we hierin elke $<$ door $\leq$ vervangen dezelfde oppervlakte zouden moeten hebben. Althans bij een zinvolle definitie van het begrip oppervlakte, een definitie die we hier echter achterwege laten.

Dat lijnstukken oppervlakte nul hebben, en grafieken van niet al te lelijke functies ook, lijkt vanzelfsprekend, maar komt hier nog niet aan de orde. In wat hieronder volgt gaat het alleen over welomschreven benaderingen van oppervlakten van verzamelingen waarvan we kunnen aannemen dat die op

⁷ Meestal als twee stellingen geformuleerd.

⁸ En ook geen rekenmachines.

de een of andere manier meetbaar zijn, bijvoorbeeld met een gereedschapskist vol rechthoekjes waarvan de oppervlakte als lengte en breedte buiten kijf staat. De wiskunde is uiteindelijk ook correct ook zonder dat aan het begrip oppervlakte wordt gedacht. Uitgangspunt is wel de consensus dat de oppervlakte tussen de lijnen $x = 0$, $x = 1$, $y = 0$, $y = 1$ in het xy -vlak gelijk is aan $1 \times 1 = 1 \cdot 1 = 1$, let wel, zonder fysische eenheden erbij.

We beginnen met het geval dat $A_- = \emptyset$: we nemen aan dat $f(x) \geq 0$ voor $x \in [a, b]$. Het geval $[a, b] = [0, 1]$ en $f(x) = x^n$ met $n \in \mathbb{N}$ is een leerzaam eerste voorbeeld. We rekenen wat we de oppervlakte van de verzameling

$$A_n = \{(x, y) : 0 \leq y \leq x^n \leq 1\}$$

zouden willen noemen expliciet uit. Als notatie voor de oppervlakte⁹ van een verzameling $A \subset \mathbb{R}^2$ gebruiken we $|A|$. Als uitgangspunt nemen we dat over de oppervlakte van een rechthoek geen twijfel bestaat, waarbij we zonder eenheden (dus geen vierkante centimeters of zo) werken. De oppervlakte van een rechthoek die parallel ligt aan de assen wordt eenvoudig bepaald door de schaalverdeling op de assen¹⁰.

Voor $n = 1$ gaat het om het gebied in het xy -vlak begrensd door

$$x = 1, \quad y = 0 \quad \text{en} \quad y = x,$$

een driehoek met oppervlakte $\frac{1}{2}$, precies de helft van de oppervlakte van het vierkant met hoekpunten $(0, 0)$, $(1, 0)$, $(1, 1)$ en $(0, 1)$. Er zal dus wel gelden dat $|A_1| = \frac{1}{2}$.

Hoe zit het met $|A_2|$? Omdat $A_2 \subset A_1$ is het duidelijk dat moet gelden dat $|A_2| \leq |A_1|$, nog voor je de vraag stelt hoe en of de oppervlakte eigenlijk precies gedefinieerd is. Het is ook duidelijk¹¹ dat $|A_2|$ niet kleiner kan zijn dan de oppervlakte van een rechthoek die geheel in A_2 ligt, bijvoorbeeld de rechthoek met hoekpunten $(\frac{1}{2}, 0)$, $(1, 0)$, $(1, \frac{1}{4})$ en $(\frac{1}{2}, \frac{1}{4})$. We concluderen hieruit dat $|A_2|$ minimaal $\frac{1}{2} \times \frac{1}{4}$ moet zijn.

Opgave 3.1.1. Laat met rechthoekjes zien dat ook moet gelden dat

$$\frac{1}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{4}{9} < |A_2| < \frac{1}{3} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{4}{9} + \frac{1}{3} \cdot \frac{9}{9}.$$

Opgave 3.1.2. Laat zien dat voor elke $N \in \mathbb{N}$ moet gelden dat

$$\frac{1^2 + 2^2 + \dots + (N-1)^2}{N^3} < |A_2| < \frac{1^2 + 2^2 + \dots + N^2}{N^3}.$$

⁹ Of elke $A \subset \mathbb{R}^2$ een oppervlakte $|A|$ heeft is natuurlijk wel de vraag.

¹⁰ Voor $n = 0 \notin \mathbb{N}$ volgt dus $|A_0| = 1$.

¹¹ Maak een plaatje.

Als je even nadenkt over hoe je de grenzen in Opgave 3.1.2 hebt gevonden dan is duidelijk dat het verschil precies $\frac{1}{N}$ is, en dat kunnen we zo klein maken als we willen door N groot te nemen. Bovendien kunnen we deze onder- en bovengrenzen gewoon uitrekenen:

Opgave 3.1.3. Laat zien dat voor elke $N \in \mathbb{N}$ geldt dat

$$\sum_{n=1}^N n^2 = 1^2 + 2^2 + \dots + N^2 = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}.$$

We gebruiken hier de somnotatie waarin de index n aangeeft dat we de som nemen van de kwadraten van n over alle n in de verzameling

$$\{n \in \mathbb{N} : 1 \leq n \leq N\}.$$

In [Eves] lezen we op pagina 47 dat de formule in Opgave 3.1.3 al door Archimedes gevonden is. Het rechterlid factoriseert als

$$\frac{N(N+1)(2N+1)}{6}$$

maar het gaat ons in het vervolg alleen om de coëfficiënt van N^3 in het rechterlid. De bovengrens voor $|A_2|$ in Opgave 3.1.2 is nu immers gelijk aan

$$\frac{1}{3} + \frac{1}{2N} + \frac{1}{6N^2}$$

en wordt kleiner als N groter genomen wordt. De ondergrens is (niet opnieuw uitrekenen maar gewoon $\frac{1}{N}$ eraf)

$$\frac{1}{3} - \frac{1}{2N} + \frac{1}{6N^2},$$

en dat wordt juist groter als N groter genomen wordt (al zie je dat misschien niet meteen). De onvermijdelijke conclusie is dat $|A_2| = \frac{1}{3}$. In termen van de integralen¹² die je natuurlijk al eerder gezien hebt is hiermee de uitspraak dat

$$\int_0^1 x^2 dx = \frac{1}{3}$$

gedaan zonder dat er een functie geïntegreerd is!

Wat aardig en nuttig is om op te merken is dat we van het derdegraadspolynoom in Opgave 3.1.3 alleen de kopcoëfficiënt in het antwoord terugzien.

¹² Maar denk ook aan volumes van piramiden, zie Hoofdstuk 2 van [Eves].

Het is een leerzame oefening dat het voor bijvoorbeeld¹³ A_{42} precies hetzelfde gaat, je bewijst en gebruikt

$$1^{42} + 2^{42} + \dots + N^{42} = \frac{N^{43}}{43} + \dots$$

om via oppervlakten van rechthoekjes op $|A_{42}| = \frac{1}{43}$ uit te komen.

Opgave 3.1.4. Doe $|A_1|$ ook op deze manier. Hint: $1 + 2 + \dots + N$ is¹⁴ een som die Gauss ooit als strafwerk zou hebben gekregen.

Opgave 3.1.5. Schrijf de differenties $(n+1)^2 - n^2$, $(n+1)^3 - n^3$, $(n+1)^4 - n^4$, enzovoorts uit.

Opgave 3.1.6. De ad hoc notaties

$$N_1 = \sum_{n=1}^N n, \quad N_2 = \sum_{n=1}^N n^2, \quad N_3 = \sum_{n=1}^N n^3, \dots$$

definiëren sommen die voor elke index $1, 2, 3, \dots$ van $N \in \mathbb{N}$ afhangen¹⁵. Gebruik nu Opgave 3.1.5: laat met de differenties van n^2 zien dat

$$(N+1)^2 = 1 + N + 2N_1,$$

met de differenties van n^3 dat

$$(N+1)^3 = 1 + N + 3N_1 + 3N_2,$$

met de differenties van n^4 dat

$$(N+1)^4 = 1 + N + 4N_1 + 6N_2 + 4N_3,$$

enzovoorts.

Opgave 3.1.7. Laat voor $k = 1, 2, 3, \dots$ zien¹⁶ dat N_k een polynoom is van graad $k+1$ met kopcoëfficiënt $\frac{1}{k+1}$. Hint: los uit de in Opgave 3.1.6 gevonden indentiteiten achtereenvolgens $N_1, N_2, N_3, \dots$ op; voor elke volgende k alle vorige resultaten gebruiken en goed kijken wat de graad van elke term is.

¹³ Amusant in [Eves, §3.16]: bewijs meetkundig dat $1^3 + \dots + N^3 = (1 + \dots + N)^2$.

¹⁴ Met $N = 100$ meen ik.

¹⁵ Met eventueel

$$N_0 = \sum_{n=1}^N 1.$$

¹⁶ Index k omdat de index n al in gebruik is, het geval $k = 42$ is al genoemd.

Nog voor dat we aan precieze definities zijn toegekomen is de onvermijdelijke conclusie dat wel moet gelden dat

$$|A_n| = \int_0^1 x^n dx = \frac{1}{n+1},$$

en een kleine variant hier op is:

Opgave 3.1.8. Voor $b > 0$ en $n \in \mathbb{N}$ is de oppervlakte van

$$\{(x, y) : 0 \leq y \leq x^n \leq b^n\}$$

gelijk aan het quotiënt van b^{n+1} en $n+1$. Laat dit zien door met schaling van de eenheden op de assen de vraag te herleiden tot het geval $b = 1$.

Lezend in [Eves] is het duidelijk dat Archimedes deze conclusie al had kunnen trekken en dat wellicht ook gedaan heeft voor $n = 2$ en $n = 3$. In moderne notatie:

$$\int_0^b x^n dx = \frac{b^{n+1}}{n+1}.$$

Opgave 3.1.9. Voor $0 \leq a < b$ en $n \in \mathbb{N}$ is de oppervlakte van

$$\{(x, y) : a \leq x \leq b, 0 \leq y \leq x^n\}$$

(in moderne notatie) gelijk aan

$$\int_a^b x^n dx = \left[\frac{x^{n+1}}{n+1} \right]_a^b = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1}.$$

De functie

$$x \rightarrow \frac{x^{n+1}}{n+1}$$

noemen we tegenwoordig de primitieve van de functie

$$x \rightarrow x^n$$

omdat de afgeleide van de eerste functie gelijk is aan de tweede. Het is niet waarschijnlijk dat de lezer¹⁷ de begrippen primitieve en afgeleide functies nog niet in de een of andere vorm gezien heeft maar het advies is om een en ander bij verder lezen in het achterhoofd te houden.

¹⁷ Ik zal de lezer weleens tutoyeren, sorry.

3.2 Integralen van monotone functies

De integraal van $x = 0$ tot $x = 1$ van x^2 in de inleiding hierboven is bij afspraak de nog niet wiskundig gedefinieerde oppervlakte van de verzameling¹⁸

$$A = \{(x, y) : x, y \in \mathbb{R}, 0 \leq y \leq f(x), a \leq x \leq b\}, \quad (3.1)$$

met in dit voorbeeld $a = 0$, $b = 1$ en $f(x) = x^2$. Voor $a, b \in \mathbb{R}$ met $a < b$ en een functie $f : [a, b] \rightarrow \mathbb{R}$ met $f(x) \geq 0$ voor $a \leq x \leq b$, beoogt de wiskundige definitie van

$$\int_a^b f(x) dx \quad (3.2)$$

wiskundig betekenis te geven aan de oppervlakte $|A|$ van A in (3.1).

Daartoe worden zogenaamde ondersommen en bovensommen gedefinieerd, waarvan het evident is dat ze bij de beoogde definitie van de oppervlakte van A kleiner respectievelijk groter¹⁹ dan $|A|$ zijn. Als vervolgens blijkt dat er precies één reëel getal is dat groter is dan alle ondersommen en kleiner dan alle bovensommen, dan is dit getal de enige kandidaat voor $|A|$, en vervolgens wordt dit getal de integraal van f over $[a, b]$ genoemd, d.w.z. (3.2). Het is natuurlijk niet verplicht om dan vervolgens te zeggen dat deze integraal gelijk is aan de oppervlakte $|A|$ van A .

We bekijken deze aanpak eerst voor het geval dat $f : [a, b] \rightarrow \mathbb{R}$ een niet-dalende niet-negatieve functie is.

Definitie 3.2.1. Als $a, b \in \mathbb{R}$ met $a < b$ en $f : [a, b] \rightarrow \mathbb{R}$, dan heet f niet-negatief als $f(x) \geq 0$ voor alle $x \in [a, b]$; f heet niet-dalend als de implicatie

$$x_1 \leq x_2 \implies f(x_1) \leq f(x_2)$$

geldt voor alle $x_1, x_2 \in [a, b]$.

Voor zo'n niet-dalend niet-negatieve functie ligt het voor de hand om de oppervlakte $|A|$ van A in (3.1) van onder te benaderen met in dit geval *ondersommen* van de vorm

$$\underline{S}_N = \sum_{k=1}^N f(x_{k-1})(x_k - x_{k-1}) \quad (3.3)$$

$$= f(x_0)(x_1 - x_0) + \cdots + f(x_{N-1})(x_N - x_{N-1}),$$

¹⁸ Nu maar met $\leq$ i.p.v. $<$.

¹⁹ Precies gezegd: groter dan of gelijk aan.

waarin

$$a = x_0 \leq x_1 \leq \dots \leq x_N = b \quad \text{met} \quad N \in \mathbb{N}. \quad (3.4)$$

Zo'n keuze van $x_1, \dots, x_{N-1}$ noemen we een *partitie* P van $[a, b]$ en $\underline{S}_N$ is niets anders dan de som van de oppervlakten van de rechthoekjes²⁰

$$R_k = [x_{k-1}, x_k] \times [0, f(x_{k-1})] \quad (k = 1, \dots, N).$$

Hoe $|A|$ verder ook gedefinieerd wordt, je kunt je zelf er makkelijk van overtuigen dat bij iedere zinvolle definitie van $|A|$ wel moet gelden dat $\underline{S}_N \leq |A|$. Teken maar een plaatje met grafiek en rechthoekjes.

Evenzo maken de oppervlakten van de rechthoekjes

$$R_k = [x_{k-1}, x_k] \times [0, f(x_k)] \quad (k = 1, \dots, N)$$

de *bovensom*

$$\bar{S}_N = \sum_{k=1}^N f(x_k)(x_k - x_{k-1}). \quad (3.5)$$

Voor een zinvolle definitie van de oppervlakte van A moet dus wel gelden dat

$$\underline{S}_N \leq |A| \leq \bar{S}_N. \quad (3.6)$$

Slordig²¹ gezegd, de gezochte waarde $|A|$ zitten *tussen*²² de ondersom $\underline{S}_N$ en de bovensom $\bar{S}_N$, en het verschil van die twee is

$$\bar{S}_N - \underline{S}_N = \sum_{k=1}^N (f(x_k) - f(x_{k-1}))(x_k - x_{k-1}).$$

Als we nu de partitie P in (3.4) equi-distant kiezen, d.w.z.

$$x_k - x_{k-1} = \frac{b-a}{N} \quad (k = 1, \dots, N),$$

dan volgt

$$\bar{S}_N - \underline{S}_N = \sum_{k=1}^N (f(x_k) - f(x_{k-1})) \frac{b-a}{N} = (f(b) - f(a)) \frac{b-a}{N}$$

en dat kunnen we willekeurig klein maken door N groot te kiezen²³. De onvermijdelijke conclusie is dan dat er maar één getal tussen alle ondersommen en bovensommen kan liggen.

²⁰ Zo'n rechthoekje kan ook een lijnstuk of een punt zijn.

²¹ Het kan gebeuren dat $\underline{S}_N = |A|$ en/of $|A| = \bar{S}_N$.

²² Als $\underline{S}_N < |A| < \bar{S}_N$ spreken we van *strict tussen*.

²³ Anders zou $\mathbb{N}$ naar boven begrensd zijn in $\mathbb{R}$.

Opgave 3.2.2. Met $N = 1, 2, 4, 8, 16 \dots$ vormen de ondersommen $\underline{S}_1, \underline{S}_2, \underline{S}_4, \underline{S}_8, \dots$ een niet-dalende rij getallen. Laat dit zien.

Evenzo vormen de bovensommen $\bar{S}_1, \bar{S}_2, \bar{S}_4, \bar{S}_8, \dots$ een niet-stijgende rij. Omdat

$$\underline{S}_1 \leq \underline{S}_2 \leq \underline{S}_4 \leq \underline{S}_8 \leq \dots \leq \bar{S}_8 \leq \bar{S}_4 \leq \bar{S}_2 \leq \bar{S}_1,$$

zijn beide rijen begrensd.

Definitie 3.2.3. Als $a \leq M \in \mathbb{R}$ respectievelijk $a \geq M$ voor alle $a \in A \subset \mathbb{R}$ dan heet M een bovengrens respectievelijk een ondergrens voor A , en A naar boven respectievelijk beneden begrensd.

De kleinste bovengrens²⁴ $\underline{J}$ van deze rij ondersommen kan niet groter zijn dan een bovensom en is daarmee een ondergrens voor alle bovensommen in deze rij bovensommen. Voor de grootste ondergrens²⁵ $\bar{J}$ van alle bovensommen geldt dus $\bar{J} \geq \underline{J}$. Omdat

$$\bar{J} - \underline{J} \leq \bar{S}_{2^n} - \underline{S}_{2^n} = (f(b) - f(a)) \frac{b-a}{2^n}$$

en 2^n willekeurig groot gekozen kan worden volgt $\bar{J} - \underline{J} = 0$. Zo definiëren we een uniek getal $J = \bar{J} = \underline{J}$ dat vervolgens genoteerd wordt als

$$J = \int_{[a,b]} f = \int_a^b f = \int_a^b f(x) dx,$$

sprek naar keuze uit als: de *integraal* van f over $[a, b]$, respectievelijk de integraal van f of $f(x)$ van a of $x = a$ tot b of $x = b$. In de notatie is x een *dummy* variabele, die door elke andere letter vervangen kan worden, maar bij voorkeur niet door a of b , en liever ook niet door d of f .

Opgave 3.2.4. Voor niet-stijgende niet-negatieve functies is het verhaal hierboven op details na hetzelfde. Schrijf dat uit.

Opgave 3.2.5. Niet-stijgende functies en niet-dalende functies worden monotone functies genoemd. Ook als deze functies zowel positieve als negatieve waarden aannemen, definiëren eindige sommen als in (3.3) en (3.5) via $N = 2^n$ een uniek getal J dat de integraal van f over $[a, b]$ genoemd wordt. Wat is de interpretatie van deze integraal in termen van de gebieden in het x, y -vlak begrensd door $x = a$, $x = b$, $y = 0$ en $y = f(x)$? Leg uit waarom.

²⁴ Ander woord voor kleinste bovengrens: supremum. Waarom bestaat die/dat?

²⁵ Ander woord voor grootste ondergrens: infimum. Waarom bestaat die/dat?

3.3 Andere partities

De integraal van een monotone functie f is nu gedefinieerd aan de hand van equi-distante partities van het interval $[a, b]$. Hoe zit het met andere partities en de ordening van onder- en bovensommen? Een vraag om algemeen te stellen maar nu eerst nog even het voor specifieke geval van monotone functies.

Opgave 3.3.1. Neem weer het geval dat f niet-dalend is. Neem een partitie P gedefinieerd door (3.4) en een partitie Q gedefinieerd door

$$a = y_0 \leq y_1 \leq \dots \leq y_M = b \quad \text{met} \quad M \in \mathbb{N}.$$

Neem²⁶ bij P de bovensom $\bar{S}$ in (3.5) en bij Q de ondersom

$$\underline{S} = \sum_{l=1}^M f(y_{l-1})(y_l - y_{l-1})$$

Laat zien dat $\underline{S} \leq \bar{S}$. Hint: eerst zelf proberen, maar kijk eventueel al naar (3.11).

Definitie 3.3.2. Laat A en B twee deelverzamelingen zijn van $\mathbb{R}$. We zeggen dat $A \leq B$ als $a \leq b$ voor alle $a \in A$ en alle $b \in B$. In dat geval zeggen we dat $J \in \mathbb{R}$ tussen A en B ligt als $A \leq \{J\} \leq B$.

Al onze onder- en bovensommen maken twee zulke verzamelingen A en B . De volgende opgave is nu van belang.

Opgave 3.3.3. Laat A en B twee niet-lege deelverzamelingen zijn van $\mathbb{R}$ met de eigenschap dat $A \leq B$. Bewijs dat er een getal $J \in \mathbb{R}$ is dat tussen A en B ligt. Hint: zie de redenatie onder Definitie 3.2.3.

Stelling 3.3.4. Laat $f : [a, b] \rightarrow \mathbb{R}$ een monotone functie zijn. Dan bestaat er precies één getal J dat tussen de verzameling van alle sommen van de vorm (3.3) en de verzameling van alle sommen van de vorm (3.5) ligt. Dit getal J is per definitie de integraal van $f(x)$ van $x = a$ tot $x = b$, notatie

$$J = \int_a^b f(x) dx.$$

Bewijs. Het werk is hierboven gedaan, tenzij Opgave 3.3.1 niet gelukt is, maar lees dan toch maar verder. Zonder beperking der algemeenheid²⁷ nemen we f niet-dalend. De ondersommen vormen een niet-lege verzameling

²⁶ Even zonder index in de notatie.

²⁷ Afkorting: zbda.

A en de bovensommen een niet-lege verzameling B . Er geldt dat $A \leq B$, dus er bestaat een J tussen A en B . Met equi-distante partities zien we dat er maar één zo'n J kan zijn. Klaar.

Monotone functies zijn dus integreerbaar. Merk op dat gegeven een partitie P zoals in (3.4), en iedere keuze van zogenaamde *tussenvolpunten*

$$x_{k-1} \leq \xi_k \leq x_k \quad (k = 1, \dots, N),$$

ook

$$S = \sum_{k=1}^N f(\xi_k)(x_k - x_{k-1}) \quad (3.7)$$

een voor de hand liggende benadering van $\int_a^b f(x) dx$ zou moeten zijn, ook als $f : [a, b] \rightarrow \mathbb{R}$ niet monotoon is, maar we kunnen dan niet meer op een natuurlijke manier van boven- of ondersommen spreken.

Als de waardenverzameling²⁸

$$\{f(x) : a \leq x \leq b\} = \bigcup_{x \in [a, b]} \{f(x)\}$$

echter begrensd is, d.w.z. als er $m, M \in \mathbb{R}$ bestaan zodat $m \leq f(x) \leq M$ voor alle $x \in [a, b]$, dan is op elke deelinterval $[x_{k-1}, x_k]$ de waardenverzameling ook begrensd en bestaan er bijbehorende ondergrenzen m_k en bovengrenzen M_k met $m \leq m_k \leq M_k \leq M$, waarmee we

$$\underline{S} = \sum_{k=1}^N m_k(x_k - x_{k-1}) \quad \text{en} \quad \bar{S} = \sum_{k=1}^N M_k(x_k - x_{k-1}) \quad (3.8)$$

weer als onder- en bovensommen voor zowel (3.7) als de te definiëren integraal $\int_a^b f(x) dx$ kunnen introduceren. Bij dezelfde partitie P geldt dan onmiddellijk dat $\underline{S} \leq \bar{S}$.

3.4 Integralen van uniforme continue functies

Wat hebben we nu nodig om het bewijs van Stelling 3.3.4 over te schrijven? Twee dingen goed begrijpen nu. Ten eerste, de verzamelingen van ondersommen A en bovensommen B moeten weer voldoen aan $A \leq B$. Ten

²⁸ Het bereik B_f van $f : [a, b] \rightarrow \mathbb{R}$, door sommigen ook het codomein van f genoemd.

tweede, we moeten de uniciteit van het getal J tussen A en B aantonen, en daartoe nemen we voor m_k en M_k natuurlijk de grootst respectievelijk kleinst mogelijk waarde, teneinde het verschil

$$\bar{S} - \underline{S} = \sum_{k=1}^N (M_k - m_k)(x_k - x_{k-1})$$

zo klein mogelijk²⁹ te krijgen.

Kies dus

$$M_k = \sup_{[x_{k-1}, x_k]} f = \sup\{f(x) : x_{k-1} \leq x \leq x_k\} \quad \text{en} \quad m_k = \inf_{[x_{k-1}, x_k]} f.$$

Hierin staat $\sup$ voor *supremum* van de verzameling in kwestie. Iedere niet-lege verzameling $A \subset \mathbb{R}$ die begrensd is naar boven, middels een *bovengrens* $M \geq a$ voor alle $a \in A$, heeft een kleinste bovengrens die het supremum van A genoemd wordt. Dit is één van de vanzelfsprekende eigenschappen die $\mathbb{R}$ moet³⁰ hebben. Evenzo staat $\inf$ voor *infimum* of grootste ondergrens.

Opgave 3.4.1. Laat zien dat bij het *verfijnen* van de partitie, d.w.z. extra punten toevoegen in (3.4), ondersommen niet kleiner en bovensommen niet groter kunnen worden.

Als nu voor alle $k = 1, \dots, N$ geldt dat

$$M_k - m_k \leq \varepsilon \tag{3.9}$$

(voor een $\varepsilon > 0$), dan volgt dat

$$\bar{S} - \underline{S} = \sum_{k=1}^N (M_k - m_k)(x_k - x_{k-1}) \leq \varepsilon(b - a). \tag{3.10}$$

Wat we dus nodig hebben voor uniciteit van J is dat elke $\varepsilon > 0$ realiseerbaar is in (3.9), middels een geschikte keuze van (3.4). In dat geval geldt voor iedere andere $\tilde{J}$ tussen A en B dat

$$|\tilde{J} - J| \leq \bar{S} - \underline{S} \leq \varepsilon(b - a),$$

en omdat $\varepsilon > 0$ dan zo klein gekozen mag worden als we maar willen volgt³¹ $\tilde{J} = J$.

²⁹ Hopelijk willekeurig klein.

³⁰ Anders $\mathbb{R}$ terugsturen naar de verzamelingsgeleerden!

³¹ Ook dit gebruikt weer dat $\mathbb{N}$ niet begrensd is in $\mathbb{R}$ via het kiezen van $\varepsilon = \frac{1}{n}$.

De definitie van (meteen maar) *uniforme continuïteit* brengt nu uitkomst. Dit is de eerste keer dat we een uitspraak formuleren waarin voor *alle* $\varepsilon > 0$ iets moet gelden dat met die ε te maken heeft, en zo'n uitspraak meteen gebruiken om een fundamentele stelling te bewijzen. Merk op dat je zo'n uitspraak alleen maar *echt gebruikt* als je meerdere (de facto oneindig veel) keren een ε kiest³².

Definitie 3.4.2. Als $a, b \in \mathbb{R}$ met $a < b$ en $f : [a, b] \rightarrow \mathbb{R}$, dan heet f uniform continu als er voor alle $\varepsilon > 0$ een $\delta > 0$ bestaat zodanig dat voor alle $x, y \in [a, b]$ geldt dat

$$|x - y| \leq \delta \implies |f(x) - f(y)| \leq \varepsilon.$$

Stelling 3.4.3. Laat $f : [a, b] \rightarrow \mathbb{R}$ een uniform continue functie zijn. Dan bestaat er precies één getal J dat tussen de verzameling van alle ondersommen en de verzameling van alle bovensommen ligt, waarbij we onder de onder- en bovensommen de sommen in (3.8) verstaan. Dit getal J is per definitie de integraal van $f(x)$ van $x = a$ tot $x = b$, notatie

$$J = \int_a^b f(x) dx.$$

Bewijs. Wederom is vrijwel al het werk al gedaan. De ondersommen vormen een niet-lege verzameling A en de bovensommen een niet-lege verzameling B . Met $\delta > 0$ bij $\varepsilon > 0$ zoals in Definitie 3.4.3 volgt³³ (3.9) als de partitie P in (3.4) zo gekozen wordt dat $x_k - x_{k-1} \leq \delta$ voor alle $k = 1, \dots, N$.

Als we nu ook nog laten zien dat $A \leq B$ zijn we klaar. Neem daartoe P als in (3.4) en Q net als in Opgave 3.3.1, en vorm de *gemeenschappelijke* verfijning

$$a = z_0 \leq z_1 \leq \dots \leq z_K = b, \quad (3.11)$$

door $x_1 \leq \dots \leq x_{N-1}$ en $y_1 \leq \dots \leq y_{M-1}$ op één gemeenschappelijke volgorde te zetten. Dus $K - 1 = M - 1 + N - 1$ en iedere z_i is een x_k of een y_l , waarmee met

$$m_l = \inf_{[y_{l-1}, y_l]} f, \quad \tilde{m}_i = \inf_{[z_{i-1}, z_i]} f \leq \sup_{[z_{i-1}, z_i]} f = \tilde{M}_i, \quad M_k = \sup_{[x_{k-1}, x_k]} f$$

volgt dat

$$\sum_{l=1}^M m_l (y_l - y_{l-1}) \leq \sum_{i=1}^K \tilde{m}_i (z_i - z_{i-1}) \leq \sum_{i=1}^K \tilde{M}_i (z_i - z_{i-1}) \leq \sum_{k=1}^N M_k (x_k - x_{k-1}).$$

³² Houd het met $\varepsilon = \frac{1}{n}$ bij voorkeur simpel, en je slaagt met vlag en wimpel.

³³ Ga dit na, al volgt aanstonds dat m_k en M_k aangenomen worden in $[x_{k-1}, x_k]$.

Waarom? Omdat de eerste ongelijkheid de ondersommen betreft van een partitie en een verfijning³⁴ daarvan, en de derde ongelijkheid de bovensommen van een andere partitie naar dezelfde verfijning. De tweede ongelijkheid spreekt vanzelf en Stelling 3.4.3 is nu voor eens en altijd met onder- en bovensommen bewezen.

3.5 Toch maar wat meer over continuïteit en limieten

Uniform continue functies $f : [a, b] \rightarrow \mathbb{R}$ zijn dus integreerbaar met dezelfde aanpak als monotone functies. Maar hoe weet je of een functie uniform continu is? Wel, bijvoorbeeld als je een schatting van de vorm

$$|f(x) - f(y)| \leq C|x - y|^\alpha \quad (3.12)$$

hebt voor alle $x, y \in [a, b]$ met een vaste $\alpha > 0$ en een vaste $C > 0$. In het geval dat $\alpha \in (0, 1)$ spreekt men van uniform *Hölder continu* met exponent α , en als $\alpha = 1$ van uniform *Lipschitz continu*. Het geval $\alpha > 1$ is niet heel relevant³⁵.

Uniforme continuïteit betreft functies op hun hele domein, hierboven steeds $[a, b]$. *Gewone continuïteit* betreft het gedrag van een functie in de buurt van een gegeven punt.

Definitie 3.5.1. Als we in Definitie 3.4.2 de y vast nemen en gegeven iedere $\varepsilon > 0$, altijd een $\delta > 0$ kunnen vinden waarvoor de implicatie geldt voor alle x met $x \in [a, b]$, dan heet f continu in y . We zeggen dan ook dat

$$f(y) = \lim_{x \rightarrow y} f(x),$$

of ook wel

$$f(x) \rightarrow f(y)$$

als $x \rightarrow y$ (steeds met $x \in [a, b]$). Als deze uitspraak geldt met uitsluiting van $x = y$, en de functiewaarde $f(y)$ vervangen door een $L \in \mathbb{R}$, dan zeggen we dat L de limiet is van $f(x)$ voor $x \rightarrow y$.

³⁴ Zie Opgave 3.4.1.

³⁵ Waarom niet? Leuke opgave: bewijs dat zulke functies constant zijn.

In Definitie 3.5.1 kan $[a, b]$ vervangen worden door een willekeurige $I \subset \mathbb{R}$ en in de uitspraak $f(x) \rightarrow L$ hoeft y dan niet per se in I te liggen.

De voor de hand liggende vraag is natuurlijk of een $f : [a, b] \rightarrow \mathbb{R}$ die continu is in elke $y \in [a, b]$ ook uniform continu is op heel $[a, b]$. Daartoe herformuleren we nu eerst de definitie van $f(x) \rightarrow L$ als $x \rightarrow y$ in termen van *convergente rijen*.

Definitie 3.5.2. Een door $n \in \mathbb{N}$ genummerde³⁶ rij reële getallen x_n heet convergent als er een (limiet) $\bar{x} \in \mathbb{R}$ bestaat zodanig dat er voor alle $\varepsilon > 0$ een $N \in \mathbb{N}$ bestaat waarvoor de implicatie

$$n \geq N \implies |x_n - \bar{x}| \leq \varepsilon$$

geldt voor alle $n \in \mathbb{N}$. Zonder woorden luidt deze uitspraak

$$\exists \bar{x} \in \mathbb{R} \forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n \in \mathbb{N} : n \geq N \implies |x_n - \bar{x}| \leq \varepsilon.$$

We zeggen dat $x_n \rightarrow \bar{x}$ (als $n \rightarrow \infty$).

Opgave 3.5.3. Laat zien dat een convergente rij reële getallen x_n maar één limiet heeft. Hint: stel niet dan zijn er minstens twee limieten, zeg $\bar{x}$ en $\hat{x}$; gebruik

$$|\bar{x} - \hat{x}| = |\bar{x} - x_n + x_n - \hat{x}| \leq |\bar{x} - x_n| + |x_n - \hat{x}|$$

en kies één geschikte ε om een tegenspraak af te leiden met Definitie 3.5.2.

Opgave 3.5.4. Laat zien dat iedere convergente rij reële getallen x_n begrensd is, d.w.z. er is een $M \geq 0$ waarvoor geldt dat $|x_n| \leq M$ voor alle n in $\mathbb{N}$. Hint: pas Definitie 3.5.2 toe met één $\varepsilon > 0$ naar vrije keuze en kies M zo dat $M \geq |x_n|$ voor alle n kleiner dan de bijbehorende N en $M \geq |x_N| + \varepsilon$. Geef een voorbeeld van een begrensde rij die niet convergent is en maak die rij convergent door voldoende veel termen van de rij weg³⁷ te laten.

Opgave 3.5.5. Laat $f : I \rightarrow \mathbb{R}$ en $y, L \in \mathbb{R}$. De uitspraak $f(x) \rightarrow L$ als $x \rightarrow y$ is equivalent met de uitspraak dat voor elke rij x_n in I de implicatie

$$x_n \rightarrow y \implies f(x_n) \rightarrow L$$

geldt. Bewijs dit. Hint: rijgen van definities, noem de ε in de uitspraak dat $x_n \rightarrow y$ daartoe maar δ .

³⁶ Of door $n \in k + \mathbb{N} = \{k + j : j \in \mathbb{N}\}$, met $k \in \mathbb{Z}$.

³⁷ Grapje: denk aan ongewenste data.

NB. Het kan gebeuren in Opgave 3.5.5 dat $y \notin I$ en dat er geen rijen x_n in I zijn met $x_n \rightarrow y$. In dat geval zijn beide uitspraken zinloos³⁸ voor elke $L \in \mathbb{R}$. Zo'n punt $y \notin I$ moeten we dan maar *geen limietpunt* van I noemen.

Opgave 3.5.6. De enige manier waarop $f : [a, b] \rightarrow \mathbb{R}$ niet uniform continu kan zijn is als er een $\varepsilon > 0$ bestaat waarvoor geen bijbehorende $\delta > 0$ te vinden is waarmee de uitspraak in Definitie 3.4.2 geldt. Kies nu een willekeurige rij $0 < \delta_n \rightarrow 0$. Voor iedere $n \in \mathbb{N}$ zijn er dan $x_n, y_n \in [a, b]$ met wel $|x_n - y_n| \leq \delta_n$ maar niet $|f(x_n) - f(y_n)| \leq \varepsilon$. Neem nu aan³⁹ dat $y_n \rightarrow \eta \in [a, b]$ en leid een tegenspraak af met de continuïteit van f in η . Hint: laat zien dat ook $x_n \rightarrow \eta$; wat geldt er dus voor $f(x_n)$ en $f(y_n)$?

Stelling 3.5.7. Laat $f : [a, b] \rightarrow \mathbb{R}$ continu zijn in elk punt van $[a, b]$. Dan is $f : [a, b] \rightarrow \mathbb{R}$ uniform continu.

Bewijs. Stel dat $f : [a, b] \rightarrow \mathbb{R}$ niet uniform continu is. Redeneer dan als in Opgave 3.5.6: als de rij y_n niet convergent is dan kan de rij convergent gemaakt worden door voldoende elementen van de rij weg te laten (met andere woorden, neem een deelrij) zodanig dat de rij convergent wordt (en de andere rij dus ook). Gebruik hiertoe de STELLING die zegt dat *iedere begrensde rij in $\mathbb{R}$ een convergente deelrij heeft*. Noem de limiet van de deelrij η , laat zien dat $\eta \in [a, b]$ en het werk is gedaan.

De in het bewijs gebruikte STELLING is belangrijk genoeg om een eigen hoofdstuk te krijgen. Het is de eerste stelling die we formuleren over het bestaan van een limiet van een (in dit geval deel-)rij zonder die limiet al op de een of andere manier bepaald te hebben. Een poging om het convergent zijn van een rij x_n te formuleren *zonder* de limiet in de definitie te gebruiken gaat terug tot Cauchy en correspondeert met wat tegenwoordig een Cauchyrij noemen. Vergelijk de volgende definitie daarom nauwkeurig met de eerste zin in Definitie 3.5.2.

Definitie 3.5.8. Een door $n \in \mathbb{N}$ genummerde rij reële getallen x_n heet een Cauchyrij als er voor alle $\varepsilon > 0$ een $N \in \mathbb{N}$ bestaat waarvoor de implicatie

$$m, n \geq N \implies |x_n - x_m| \leq \varepsilon$$

geldt voor alle $m, n \in \mathbb{N}$.

³⁸ Maar wel waar.

³⁹ Absurde aanname?

In Definitie 3.5.8 is $m, n \geq N$ te lezen als $m \geq N$ en ook $n \geq N$. Evenzo is $m, n \in \mathbb{N}$ te lezen als $m \in \mathbb{N}$ en ook $n \in \mathbb{N}$. In het volgende hoofdstuk volgt het convergent zijn van Cauchyrijen via de STELLING uit het begrensde zijn van Cauchyrijen.

Opgave 3.5.9. Laat $f : [a, b] \rightarrow \mathbb{R}$ uniform continu zijn. Bewijs met de STELLING dat f begrensd is op $[a, b]$.

Opgave 3.5.10. Laat $f : [a, b] \rightarrow \mathbb{R}$ continu zijn in elk punt van $[a, b]$. Bewijs dat f een maximum M en een minimum m op $[a, b]$ heeft. Hint: omdat f uniform continu is op $[a, b]$, is f begrensd op $[a, b]$. Laat $M = \sup\{f(x) : a \leq x \leq b\}$. Dan bestaat er een rij x_n in $[a, b]$ met $f(x_n) \rightarrow M$. Neem daarvan een convergente deelrij.

3.6 Riemann integreerbaarheid

We weten tot nu toe alleen dat voor monotone en voor continue $f : [a, b] \rightarrow \mathbb{R}$ met behulp van onder- en bovensommen een ondubbelzinnige betekenis kan worden gegeven aan de *integreerbaarheid* van f en de waarde van de integraal

$$\int_a^b f(x) dx.$$

Voor iedere begrensde functie kunnen we echter bij een partitie (3.4)

$$M_k = \sup_{[x_{k-1}, x_k]} f \quad \text{en} \quad m_k = \inf_{[x_{k-1}, x_k]} f$$

definiëren en de bijbehorende onder- en bovensom. Net als in het bewijs van Stelling 3.4.3 volgt nu dat de verzameling van alle ondersommen A en de verzameling van alle bovensommen B voldoen aan $A \leq B$. Op grond van Opgave 3.3.3 is er dus tenminste één getal J dat tussen A en B ligt.

Definitie 3.6.1. Een begrensde functie $f : [a, b] \rightarrow \mathbb{R}$ heet Riemann integreerbaar als er precies één getal J tussen de verzameling A van alle ondersommen en de verzameling B van alle bovensommen ligt. In dat geval is per definitie

$$J = \int_a^b f(x) dx.$$

Een functie $f : [a, b] \rightarrow \mathbb{R}$ kan alleen maar Riemann integreerbaar zijn als f begrensd is. Maar welke begrensde functies integreerbaar zijn is een

vraag die we nog niet gesteld hebben. Zonder bewijs: het antwoord is dat daartoe de verzameling $\mathcal{N}$ van x -waarden waarin f niet continu is een zogenaamde nulverzameling moet zijn. Dat is een verzameling die voor elke $\varepsilon > 0$ overdekt kan worden met intervallen $[a_n, b_n]$ genummerd door $n \in \mathbb{N}$, waarbij voor elke $N \in \mathbb{N}$ geldt dat

$$\sum_{n=1}^N (b_n - a_n) \leq \varepsilon.$$

Overdekken betekent hier dat

$$\mathcal{N} \subset \bigcup_{n \in \mathbb{N}} [a_n, b_n],$$

en de overdekking hangt af van ε . Intuïtief: met een willekeurig klein intervalletje opgeknipt in aftelbaar veel stukjes is $\mathcal{N}$ te bedekken.

3.7 Sommen en limieten van integralen

We zijn begonnen met integralen van de functies $x \rightarrow x^n$ en hebben daarna laten zien dat monotone functies $f : [a, b] \rightarrow \mathbb{R}$ integreerbaar zijn in de zin van Definitie 3.6.1. Het is niet moeilijk om rechtstreeks vanuit de definitie te laten zien dat met $f : [a, b] \rightarrow \mathbb{R}$ en $g : [a, b] \rightarrow \mathbb{R}$ ook de somfunctie $f + g : [a, b] \rightarrow \mathbb{R}$ weer integreerbaar is met de integraal die je verwacht: de functie $x \rightarrow f(x) + g(x)$ is integreerbaar met

$$\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$$

Ook veelvouden van integreerbare functies zijn weer integreerbaar. Ook het rijgen van integralen, d.w.z.

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

voor $c \in (a, b)$ volgt eenvoudig: als de integraal links bestaat dan bestaan de integralen recht en omgekeerd, met de gelijkheid in de formule. Met Opgave 3.1.9 volgt zo eenvoudig de hoofdstelling van de integraalrekening voor polynomen: de functie

$$x \rightarrow p(x) = a_0 + a_1x + \cdots + a_Nx^N = \sum_{n=0}^N a_nx^n$$

is integreerbaar op ieder begrens interval $[a, b]$ met

$$\int_a^b p(x) dx = P(b) - P(a),$$

waarbij

$$P(x) = a_0x + \frac{a_1}{2}x^2 + \cdots + \frac{a_N}{N+1}x^{N+1} = \sum_{n=0}^N \frac{a_n}{n+1}x^{n+1}$$

Neem nu even aan dat $0 \leq a \leq b$ en dat $a_n \geq 0$ voor alle $n \in \mathbb{N} \cup \{0\}$ gegeven zijn. Wat kunnen we zeggen over $f : [a, b] \rightarrow \mathbb{R}$ gedefinieerd door

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + \cdots = \lim_{N \rightarrow \infty} \sum_{n=0}^N a_n x^n \quad (3.13)$$

als we aannemen dat deze formule in ieder geval $f(b)$ als limiet van een stijgende rij getallen definieert? De volgende opgave is voldoende om te concluderen dat

$$\int_a^b f(x) dx = F(b) - F(a),$$

met

$$F(x) = a_0x + \frac{a_1}{2}x^2 + \cdots = \sum_{n=0}^{\infty} \frac{a_n}{n+1}x^{n+1}.$$

Het bewijs borduurt voort op de bewijzen in Sectie 3.2.

Opgave 3.7.1. Gegeven functies $f_n : [a, b] \rightarrow \mathbb{R}$ met

$$\forall x \in [a, b] \quad 0 \leq n_1 \leq n_2 \implies f_{n_1}(x) \leq f_{n_2}(x),$$

$$\forall n \in \mathbb{N} \cup \{0\} \quad a \leq x_1 \leq x_2 \leq b \implies f_n(x_1) \leq f_n(x_2),$$

en $f_n(b)$ begrensd, bestaat

$$f(x) = \lim_{n \rightarrow \infty} f_n(x)$$

voor alle x . Er geldt

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \int_a^b f_n(x) dx.$$

Bewijs dit rechtstreeks vanuit de definitie met onder- en bovensommen. Hint: gebruik de aanpak in Sectie 3.2, het is voldoende om equidistante partities te gebruiken.

Er zijn nog drie versies van deze stelling. Welke drie?

De conclusie nu is dat we machtreeksen als in (3.13), ook zonder de beperkingen op de tekens van a_n, a, b , kunnen integreren als de somfunctie gedefinieerd⁴⁰ is, precies zoals we dat met polynomen doen, zonder dat we over integraalrekening voor continue functies hoeven te spreken. Het enige limietbegrip dat we nodig hebben beperkt zich tot monotone begrensde rijen en sommen en verschillen van die rijen. De volgende open opgave vat dit samen.

Opgave 3.7.2. Voor $0 \leq a \leq b$ kan de aanname dat $a_n \geq 0$ worden wegelaten door de som in (3.13) te splitsen in positieve termen en negatieve termen. Wat moet je aannemen om tot dezelfde conclusies te komen? Hoe kom je vervolgens van de aanname dat $0 \leq a \leq b$ af?

3.8 Discussie

Een vraag om te stellen is wat van het bovenstaande al (vroeg) op school zou kunnen worden behandeld. Ik beperk me voor deze vraag tot Sectie 3.1. Elke formule waarin behalve getallen ook nog een letter in voorkomt, vraagt een abstractieniveau dat aanvankelijk nog afwezig is. De eerste letter die in Opgave 3.1.1 en daarna voorkomt is N . Die kun je natuurlijk eerst 10, 100, 1000 en groter nemen, en via Opgave 3.1.4 leidt dat tot de oppervlakte die we later de integraal van x over $[0, 1]$ noemen, en die we al kenden als de oppervlakte van de driehoek A_1 .

Opgave 3.1.6 is geformuleerd met een algemene N en is gebaseerd op de differenties in Opgave 3.1.5, waarin natuurlijk de driehoek van Pascal te herkennen is. Zodra die driehoek herkend of ontdekt is in het uitwerken van $(n+1)^2$, $(n+1)^3 = (n+1)(n+1)^2$, $(n+1)^4 = (n+1)(n+1)^3$, enzovoorts, liefst met de recurrente betrekkingen die daar bij horen om steeds de volgende rij te maken, is een aanpak als in Opgave 3.1.6 te overwegen.

Met differenties $n^2 - (n-1)^2 = 2n-1$, $n^3 - (n-1)^3 = 3n^2 - 3n + 1, \dots$, worden de uitdrukkingen in Opgave 3.1.6 dan net wat prettiger, met links steeds een macht van N in plaats van een macht van $N+1$, en rechts dezelfde uitdrukkingen maar met plussen en minnen. De formules voor de machten van N volgen dan stap voor stap in termen van de sommen van de lagere machten van n , ook als voor N een macht van 10 gekozen

⁴⁰ Als verschil van twee limieten van niet-negatieve rijen functies.

wordt. Opgave 3.1.7 wordt vervolgens uitvoerbaar. Een alternatief is om tabellen⁴¹ van de machten te maken en alles met getallen zonder letters uit te schrijven, minder abstractie, maar meer gereken (en wellicht daarom juist nuttig). Als voor de oppervlakte die we later de integraal van x^2 over $[0, 1]$ noemen benaderingen verschijnen waarin we $\frac{1}{3}$ herkennen is de eerste stap gemaakt. Piramides kunnen hier natuurlijk bij helpen. Daarna is het eigenlijk voor elke macht hetzelfde, en dat kun je verder wel geloven dan op weg naar Opgave 3.1.9.

⁴¹ Die vinden we ook al terug op kleitabletten.

4 Computers in theoretisch wiskundeonderzoek

Han Peters

4.1 Wiskunde zonder computers.

Ik onderzoek theoretische wiskunde, voornamelijk in de complexe dynamica. Als ik vertel dat ik onderzoek doe naar nieuwe wiskunde wordt mij vaak gevraagd: “O, dan zul je wel veel met computers doen”. Tot voor kort antwoordde ik hierop dat mijn onderzoek niets met computers te maken had. Niet zonder enige trots vertelde ik dat ik voor mijn onderzoek enkel gebruik maakte van pen, papier en gezond verstand. Computers gebruikte ik enkel voor emails, en als tekstverwerker.

Helaas kan ik dit opschepperige antwoord niet langer geven. Steeds vaker wordt de computer bij mijn onderzoeksprojecten betrokken, en de opkomst van computers heeft een enorme rol gespeeld in mijn vakgebied. In deze cursus wil ik een idee geven van de manier waarop computers gebruikt kunnen worden in theoretisch wiskundeonderzoek. In plaats van me te richten op de beroemde wiskundigen en hun stellingen wil ik vooral voorbeelden die ik in mijn onderzoek ben tegengekomen gebruiken als illustraties.

4.2 De computer als bewijzer.

In de aankondiging werd gesproken over een toekomst waarin computers onze stellingen zullen bewijzen. Ik zie het nog niet zo snel gebeuren. Het hangt misschien van het onderzoeksgebied af, maar voor mij is dit de minst belangrijke rol van computers in wiskundig onderzoek. Ik kan me slechts één project herinneren waarbij we de computer op een essentiële manier in een bewijs gebruikten, en dat was om te laten zien dat een collectie van n vectoren $v_1, \dots, v_n$ in n variabelen onafhankelijk was.

Herinner je dat deze vectoren onafhankelijk zijn als de vergelijking

$$a_1 \cdot v_1 + \dots + a_n \cdot v_n = 0$$

als enige oplossing $a_1 = \dots = a_n = 0$ heeft. Onafhankelijkheid is equivalent aan de eis dat de *matrix* met rijvectoren v_j *maximale rang* heeft. Omdat het aantal vectoren precies gelijk is aan het aantal variabelen, is maximale rang equivalent aan het inverteerbaar zijn van deze matrix, en dit kun je testen door de determinant uit te rekenen.

Probleem: *Wat als een matrix zelfs voor de computer te groot is om de determinant uit te rekenen?*

Dit probleem kwam naar voren in een artikel uit 2016 genaamd *Polynomial completion of symplectic jets and surfaces containing involutive lines*, met co-auteurs Wold en Løw uit Noorwegen en Pereira uit Brazilië. We wilden weten of een matrix van grootte 6188×6188 inverteerbaar was. Lastige bijkomendheid was dat de getallen in de matrix allemaal gehele getallen van ongeveer 50 cijfers waren, wat de berekening niet eenvoudiger maakte. Misschien dat een supercomputer dit geval nog wel aankon, maar mijn computer was in ieder geval niet in staat om de determinant te berekenen. Dankzij een eenvoudige truc kon de computer toch bepalen of de matrix inverteerbaar was. Enig idee wat we gedaan hebben?

4.3 De computer als tester.

Stel we hebben een plan voor het bewijs van een grote stelling. Als we drie kleinere lemmaatjes kunnen bewijzen, dan kunnen we die aan elkaar knopen en krijgen we de stelling. Wiskunde kan echter verraderlijk zijn, wat eerst lijkt op een eenvoudig lemma kan veel lastiger blijken dan verwacht, soms zelfs lastiger dan de originele stelling. En soms zijn de lemma's gewoon helemaal niet waar.

Zou het niet fijn zijn als we, voordat we gaan proberen om de lemma's één voor één te bewijzen, even kunnen testen of de lemma's wel waar zijn? Soms kan dat, met de computer.

Hier is een eenvoudig voorbeeld uit de praktijk. We schrijven $D_r(a)$ voor de open schijf met middelpunt a en straal r , oftewel

$$D_r(a) = \{z \in \mathbb{C} : |z - a| < r\}.$$

We kennen allemaal de functie $\exp(z) = e^z$, gegeven door

$$e^{x+iy} = e^x \cdot (\cos(y) + i \sin(y)).$$

Klein Lemma. *Het beeld*

$$\exp D_r(a) := \{e^z, z \in D_r(a)\}$$

is convex dan en slechts dan als $r \leq 1$.

Herinner je dat een verzameling $A \subset \mathbb{R}^n$ convex wordt genoemd als voor elk paar $z, w \in A$ geldt dat het interval tussen z en w bevat is in A . Om te bewijzen dat $\exp D_r(a)$ convex is moet je dus bewijzen dat voor elke twee punten $z, w \in \exp D_r(a)$ geldt dat voor alle $t \in [0, 1]$ het punt $t \cdot z + (1-t) \cdot w$ ook in $\exp D_r(a)$ ligt. Dit kun je controleren door te laten zien dat er een tak van de logaritme is waarvoor

$$|\log(t \cdot z + (1-t) \cdot w) - a| < r.$$

Die formule ziet er helemaal niet lekker uit. Is er, voordat we beginnen, niet een eenvoudige manier om te testen dat het kleine lemma waar is?

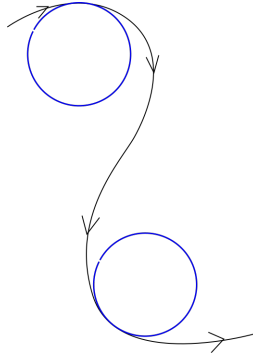
Opgave 1. *Schrijf een computerprogramma (in geogebra, mathematica, matlab, ...) om het beeld van een schijf met straal r en middelpunt a onder de functie $\exp(z)$ te tekenen. Wat gebeurt er als $r = 1$? Wat gebeurt er als $r = \pi$?*

Nu dat we met de computer gezien hebben dat het lemma waar zou moeten zijn, is het de moeite waard om er meer tijd in te stoppen en het te bewijzen. De computer speelt geen enkele rol meer in het bewijs.

Het idee van bewijs is om te laten zien dat voor $r < 1$ de rand van de verzameling $\exp D_r(a)$ een *positieve kromming* heeft. Hiervoor hoeven geen moeilijke logaritmische vergelijkingen te worden uitgerekend. Het enige wat we over de exponentiele functie hoeven te weten is dat de afgeleide weer gelijk is aan de functie zelf. Dit geldt dan dus ook voor de tweede afgeleide. De Taylorexpansie rond een punt $b \in \mathbb{C}$ wordt dus gegeven door

$$e^z = e^b + e^b \cdot (z - b) + \frac{e^b}{2}(z - b)^2 + \dots$$

We bespreken kort de kromming van een gladde geparametriseerde kromme $\gamma : \mathbb{R} \rightarrow \mathbb{C}$, zie het internet voor meer informatie. We nemen aan dat γ minstens twee maal differentieerbaar is, en dat γ' nergens 0 is. In dat geval is er voor ieder punt $\gamma(t)$ een unieke cirkel die dezelfde tweede orde benadering heeft in het punt $\gamma(t)$, wat wil zeggen dat een parametrisatie van de cirkel dezelfde eerste en tweede afgeleide heeft.



Figuur 4.1: Eerst negatieve, dan positieve kromming

We definiëren de kromming in $\gamma(t)$ als de inverse van de straal van deze cirkel. Als het middelpunt van de cirkel links van de kromme ligt (wat afhangt van de richting van de parametrisatie van de kromme) dan zeggen we dat de kromming positief is, en als het middelpunt rechts ligt dan zeggen we dat de kromming negatief is. Als de tweede afgeleide van de kromme nul is (in welk geval we zouden kunnen zeggen dat de cirkel een rechte lijn is, met straal oneindig), dan spreken we af dat de kromming nul is.

De kromming is, op het teken na, onafhankelijk van de parametrisatie. Nemen we bijvoorbeeld een parametrisatie van een cirkel $C_r(a)$ met straal r die de cirkel tegen de klok in doorloopt, dan is de kromming constant gelijk aan $\frac{1}{r}$.

Stelling 1. *Een domein met gladde rand is convex dan en slechts dan als de kromming van de rand overal hetzelfde teken heeft.*

Ons doel is dus om te bewijzen dat voor $r < 1$ de kromming van de rand van $\exp D_r(a)$ overal hetzelfde teken heeft. We nemen een parametrisatie die afkomstig is van een parametrisatie die de cirkel $C_r(a)$ tegen de klok in doorloopt, en zullen bewijzen dat de kromming van het beeld hiervan overal positief is.

Omdat we ons alleen maar met het teken van de kromming zullen bezighouden kunnen we gebruik maken van translaties $z \mapsto z + \alpha$ en dilataties $z \mapsto \beta \cdot z$. Dilataties veranderen de kromming met een constante, maar veranderen niet het teken. Door een compositie met een translatie te nemen

mogen we aannemen dat $0 \in C_r(a)$.

Door een compositie met een translatie en een dilatatie te nemen mogen we aannemen dat onze functie 0 op 0 afbeeldt, en daar afgeleide 1 heeft. De functie heeft nu dus de vorm

$$z \mapsto 0 + z + \frac{1}{2}z^2 + \dots$$

De hogere afgeleiden kunnen we negeren, want die hebben geen invloed op de kromming in de oorsprong.

Tot slot kunnen we nog composities nemen met rotaties, zodat a op de positieve imaginaire as komt te liggen. Onze functie krijgt nu de vorm

$$z \mapsto 0 + z + \frac{e^{i\theta}}{2}z^2.$$

De cirkel met straal r en middelpunt $+ri$ wordt gegeven door de uitdrukking

$$x^2 + (y - r)^2 = r^2.$$

In de buurt van de oorsprong komt dit neer op

$$y = r - \sqrt{r^2 - x^2},$$

De x -coördinaat geeft deze grafiek een parametrisatie met positieve kromming. Omdat we alleen geïnteresseerd zijn in eerste en tweede afgeleiden kunnen we dit herschrijven tot

$$y = r - r \cdot \sqrt{1 - \frac{x^2}{r^2}} = r - r \cdot (1 + \frac{x^2}{2r^2} + \dots) = \frac{x^2}{2r} + \dots$$

De vraag is dus wat het beeld van de parabool $y = \frac{x^2}{2r}$ is onder de functie

$$z \mapsto z + \frac{e^{i\theta}}{2}z^2.$$

Opgave 2. *Schrijf het beeld lokaal als grafiek, en laat zien dat de kromming strikt positief is als $r < 1$. Wanneer gaat het mis als $r > 1$?*

4.4 De computer als verkenner.

In de complexe dynamica bestudeer je de *iteratie* van complex differentieerbare functies. Voorbeelden van complex differentieerbare functies zijn

zo ongeveer alle functies die je op de middelbare school tegenkomt: polynomen, rationale functies, trigonometrische functies, exponentiële functies, logaritmen, enzovoort. In vergelijking met de middelbare school verschilt alleen het domein van deze functies: we beschouwen ze op het complexe vlak. Hoewel dit een lastige stap is als je niet vaak met complexe getallen werkt, blijkt het eigenlijk veel eenvoudiger te zijn.

De iteratie van polynomen wordt veel beter begrepen over de complexe getallen dan over de reële getallen. Echt verrassend is dit niet. Periodieke punten spelen bijvoorbeeld een belangrijke rol bij het begrijpen van iteratieproblemen. Deze punten zijn oplossingen van de vergelijking

$$f^{\circ n}(z) = f \circ \dots \circ f(z) = z.$$

Als f een polynoom is van graad $d \geq 2$, dan is $f^{\circ n}$ een polynoom van graad d^n . Het volgt uit de Hoofdstelling van de Algebra dat er d^n punten zijn die voldoen aan de vergelijking

$$f^{\circ n}(z) - z = 0,$$

ten minste, over de complexe getallen. Over de reële getallen weet je alleen dat er *hoogstens* d^n zijn.

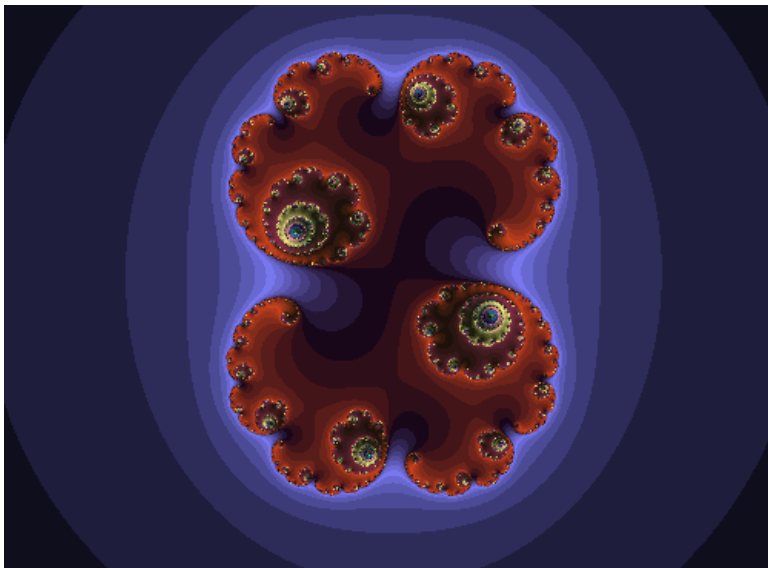
Iteratieproblemen in het complexe vlak werden voor het eerst bestudeerd in de tweede helft van de negentiende eeuw. Een van de oudste stellingen in het vakgebied is die van de Duitser Koenigs, die in 1870 liet zien dat het gedrag in de buurt van een periodiek punt $z = f^{\circ n}(z)$ precies kan worden beschreven aan de hand van de afgeleide $(f^{\circ n})'(z)$ (behalve als deze afgeleide norm 1 heeft).

Een van de belangrijkste wiskundigen uit die tijd, Henri Poincaré, raakte bijzonder geïnteresseerd in complexe dynamica. In de Franse school werd het vakgebied zeer populair in het begin van de twintigste eeuw, met belangrijke contributies van met name Gaston Julia en Pierre Fatou, naar wie de twee belangrijkste objecten in het vakgebied nu zijn vernoemd.

Voor een polynoom f definiëren we de *gevulde Juliaverzameling* K , die bestaat uit alle punten wiens baan begrensd blijft:

$$K = \{z \in \mathbb{C} : |f^{\circ n}(z)| < R \text{ voor alle } n \geq 0\}.$$

Hier kan de straal R onafhankelijk van het punt z gekozen worden, omdat voor $|z|$ groot genoeg het gedrag van f vooral door de term van hoogste graad wordt bepaald.



Figuur 4.2: De Juliaverzameling van $z \mapsto z^2 + 0.29 + 0.01i$.

De *Juliaverzameling* J is de rand van K :

$$J = \partial K,$$

en de *Fatouverzameling* F is het complement hiervan:

$$F = \mathbb{C} \setminus J.$$

Fundamentele eigenschappen van J en F werden door Fatou en Julia beschreven. We noemen enkele voorbeelden.

Stelling 2. *De Juliaverzameling is overaftelbaar en heeft geen geïsoleerde punten. Afstotende periodieke punten liggen dicht in de Juliaverzameling.*

Stelling 3. *De Juliaverzameling is samenhangend dan en slechts dan als de baan van alle kritieke punten begrensd is.*

Herinner je dat een punt z kritiek is als $f'(z) = 0$. Voor een tweedegraads polynoom is er precies 1 kritiek punt, want de afgeleide is een lineaire functie. Door te conjugeren met een translatie mogen we aannemen dat

dit kritieke punt in de oorsprong ligt. Door daarna te conjugeren met een dilatatie kunnen we er voor zorgen dat f van de vorm $z \mapsto z^2 + c$ is.

Gevolg 4. *De Juliaverzameling van een polynoom $f(z) = z^2 + c$ is samenhangend dan en slechts dan als de baan van de oorsprong begrensd is.*

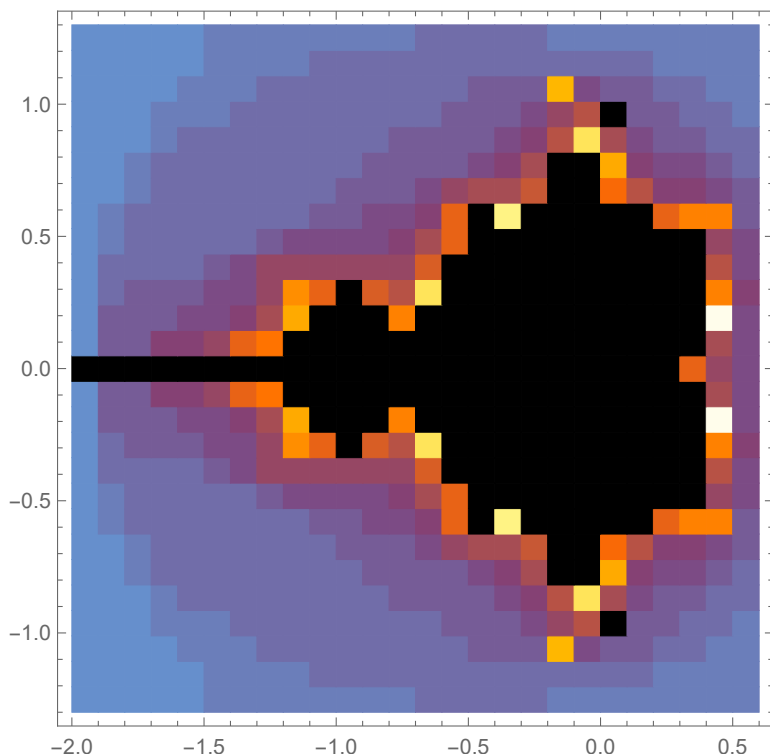
Deze stellingen waren in de jaren twintig bekend, maar toch kwam het vakgebied min of meer tot stilstand. De openstaande vragen werden te moeilijk gevonden, en men wist niet goed welke andere vragen interessant konden zijn. Hier kwam begin jaren tachtig plotseling verandering in. Één reden was dat opnieuw een aantal van de allerbeste wiskundigen zich in het vakgebied begonnen te interesseren. Er werden stellingen bewezen waarbij gebruik werd gemaakt van technieken die in de jaren twintig nog niet bestonden. Het gaat te ver om deze resultaten hier te bespreken.

Een andere reden voor de nieuwe golf van interesse was de opkomst van computerplaatjes. Deze plaatjes zorgden er niet alleen voor dat Juliaverzamelingen plots teruggevonden werden op t-shirts en posters, ze zorgden ook voor veel nieuwe ideeën, zowel voor vraagstellingen als voor mogelijke routes om vermoedens op te lossen. Ze zorgden trouwens ook voor veel irritaties bij wiskundigen, die het maken van mooie computerplaatjes niet als serieuze wiskunde zagen. Misschien dat deze irritatie minder sterk was geweest als Benoît Mandelbrot zichzelf niet zo op de borst bleef kloppen.

Mandelbrot was bijzonder geïnteresseerd in illustraties van complexe dynamische systemen, en had bij IBM de beschikking over de allermooiste computers. Nou moet je je niet al te veel voorstellen bij de plaatjes uit die tijd, misschien dat een van de eerste plaatjes van Mandelbrot er ongeveer uitzag als in Figuur 4.3.

Wat wordt er hier afgebeeld? In Gevolg 4 hierboven hebben we gezien dat een Juliaverzameling samenhangend is dan en slechts dan als de baan van 0 begrensd blijft. Dit laatste kunnen we met de computer aardig testen: we nemen een “pixel” c en berekenen de baan van 0 voor de functie $f(z) = z^2 + c$ en voor een groot aantal iteraties. Is $f^{on}(z)$ dan nog steeds begrensd (zeg kleiner dan 5) dan kleuren de we de betreffende pixel zwart. Raakt de baan van 0 op een gegeven moment buiten de cirkel met straal 5 dan geven we de pixel een kleur, afhankelijk van het aantal iteraties dat nodig was. De zwarte pixels geven dus een “kaart” van alle waarden van c waarvoor de Juliaverzameling van $z \mapsto z^2 + c$ samenhangend is:

$$\mathcal{M} := \{c \in \mathbb{C} : J_c \text{ is samehangend} \}.$$



Figuur 4.3: Afbeelding van $\mathcal{M}$ met lage resolutie.

Mandelbrot keek naar “zijn” verzameling en vroeg zich af:

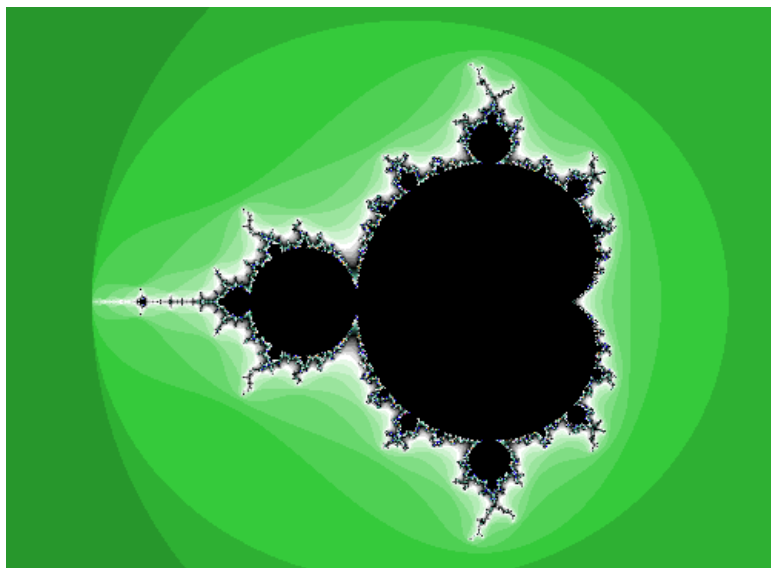
Vraag. *Wat gebeurt er bij die twee mysterieuze zwarte eilandjes? Zijn er oneindig veel eilandjes?*

Deze vraag werd al snel beantwoord door de Douady-Hubbard en door Sibony:

Stelling 5. *De verzameling $\mathcal{M}$ is samenhangend.*

Er zijn dus helemaal geen mysterieuze eilandjes.

Tegenwoordig is het eenvoudig om met computerprogramma's (bijvoorbeeld met mathematica, of het programma Xaos, dat je gratis kunt downloaden) plaatjes van de Mandelbrotverzameling te maken van zeer hoge



Figuur 4.4: Hogere resolutie afbeelding van $\mathcal{M}$.

resultatie. Als je naar deze verzameling kijkt, hoe zou je deze beschrijven? Kun je bewijzen dat de Mandelbrotverzameling deze eigenschappen ook werkelijk heeft?

5 Een wiskundig bewijs correct bewezen

Freek Wiedijk

Op de volgende pagina's is een artikel afgedrukt dat verscheen in het Nieuw Archief voor Wiskunde, September 2016.

F. Wiedijk
Faculteit FNWI, iCIS
Radboud Universiteit
Postbus 9010
6500 GL Nijmegen

f.wiedijk@cs.ru.nl

H. Geuvers
Faculteit FNWI, iCIS
Radboud Universiteit
Postbus 9010
6500 GL Nijmegen

h.geuvers@cs.ru.nl

J. Urban
Czech Institute of Informatics, Robotics,
and Cybernetics
Czech Technical University
Zikova 4, 16636 Praha 6
Czech Republic
josef.urban@gmail.com

Een wiskundig bewijs correct bewezen: De meest efficiënte manier om bollen op te stapelen

Sommige bewijzen van wiskundige stellingen zijn zo bewerkelijk dat ze alleen met computerondersteuning geverifieerd kunnen worden. Een voorbeeld is het bewijs van de 4-kleurenstelling, die zegt dat iedere landkaart met vier kleuren ingekleurd kan worden zonder dat aangrenzende landen dezelfde kleur krijgen. Een recent voorbeeld is het bewijs van het vermoeden van Kepler, dat zegt dat de meest voor de hand liggende stapeling van bollen in de ruimte (zoals de groenteman sinaasappelen stapelt) ook de meest efficiënte is. Johannes Kepler beschreef dit vermoeden in 1611 en in 1998 kondigde de Amerikaan Tom Hales aan het bewezen te hebben. Zijn bewijs reduceert het probleem tot een kleine 20.000 mogelijke tegenvoorbeelden, die vervolgens verworpen worden. Daarvoor worden meer dan 23.000 niet-lineaire vergelijkingen opgelost en wordt aangetoond dat meer dan 43.000 lineaire programma's onoplosbaar zijn. Dat stuk van het bewijs is gedaan met computerprogramma's, en dit deel werd door de wiskundige reviewers niet geaccepteerd. Als reactie daarop is Hales het "Flyspeck" project begonnen, met als doel het gehele bewijs met een bewijsassistent te verifiëren. Een bewijsassistent is een computerprogramma waarmee een gebruiker interactief een wiskundig bewijs construeert, dat vervolgens door het programma gecheckt wordt. Omdat bewijsassistenten een buitengewoon hoge betrouwbaarheid hebben zijn er daarna geen referenten meer nodig. In 2014 is het Flyspeck project voltooid. In dit artikel geven we een overzicht van het wiskundige deel van het bewijs en leggen uit op welke manier de computer bij de verificatie is gebruikt.



FIGUUR 1. Efficiënt opgestapelde sneeuwballen

1. Het Kepler-vermoeden

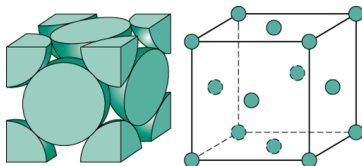
Stel we hebben oneindig veel even grote bollen, en we willen die zo dicht mogelijk tegen elkaar in de driedimensionale ruimte plaatsen. Wat is de manier om dat met de grootste mogelijke dichtheid te doen?

Een voor de hand liggende manier is om de rangschikking te kiezen die een groenteboer voor sinaasappelen gebruikt, en die ook gebruikt wordt om kanonskogels naast een kanon op te stapelen. Dit is de zogenaamde *kubische vlakgecentreerde stapeling* (in het Engels: *face centered cubic* of *FCC packing*). In deze stapeling is

$$\frac{\pi}{\sqrt{18}} = 74,0480 \dots \%$$

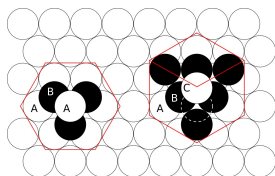
van de ruimte gevuld. Bij deze stapeling liggen de middelpunten van de bollen zowel op de hoekpunten als in het midden van de zijvlakken van de kubussen uit een driedimensionaal kubisch rooster. De roosterpuntafstand is $2r\sqrt{2}$ als r de straal van de bollen is. Zoals het plaatje laat zien passen er precies 4 bollen in een kubus met ribbe

$2r\sqrt{2}$, dus het deel van de kubus dat opgevuld wordt door de bollen is precies $\pi/\sqrt{18}$.



FIGUUR 2. De FCC stapeling

Een stapeling met deze dichtheid is overigens niet uniek. Ook de *hexagonale dichtste stapeling* (hexagonal close-packed of HCP packing) heeft deze dichtheid. Deze wordt bereikt door in het platte vlak een bol te omsluiten door 6 andere bollen en zo verder het platte vlak vol te leggen. Zo ontstaat een hexagonaal rooster, als de cellen van een honingraat. Vervolgens wordt hier op dezelfde wijze een tweede laag bovenop gelegd. De derde laag kan nu weer op dezelfde wijze gelegd worden als de eerste, en de vierde weer op dezelfde wijze als de tweede en zo voort. Dit levert HCP, de hexagonale dichtste stapeling. Daarbij blijft het grondvlak altijd zichtbaar: we kunnen door de stapeling heen kijken. De derde laag kan ook gelegd worden over de openingen die er na de eerste twee lagen nog zijn. Als we vervolgens de weer bollen leggen volgens de posities in laag 1, en vervolgens weer in de posities van laag 2, laag 3 enzovoort, dan is er sprake van FCC, de kubische vlakgecentreerde stapeling. Als we in de FCC stapeling een bol op het derde vlak beschouwen met de 6 bollen daar direct onder in het tweede vlak, zien we één uitstekende punt van de kubus van het eerste figuur van de FCC stapeling.



FIGUUR 3. Laagjes bollen in de HCP en FCC stapelingen

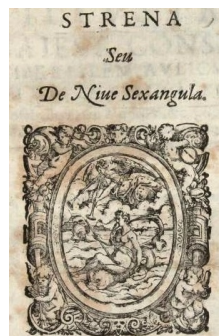
Er bestaan zelfs overaftelbaar veel stapelingen met de optimale dichtheid. Ze worden allemaal gevormd door oneindig veel vlakke lagen van bollen op elkaar te plaatsen. Binnen zo'n laag liggen de bollen op een hexagonaal rooster en voor iedere volgende laag zijn er dan twee manieren om die op de laag eronder te plaatsen. Als je hierbij alterneert tussen twee posities (loodrecht gezien) dan krijg je de hexagonale dichtste stapeling. Als je telkens dezelfde afstand opschuift krijg je de kubische vlakgecentreerde stapeling.



FIGUUR 4. Johannes Kepler

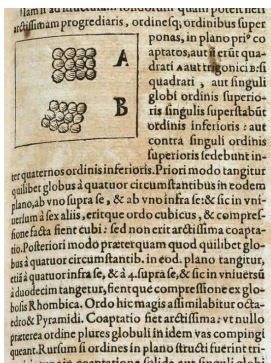
Een dichtheid van $\pi/\sqrt{18}$ kan dus op veel manieren bereikt worden. Maar kan het ook beter?

In 1611 schreef Johannes Kepler (1571–1630) in zijn boekje *Strena Seu De Nive Sexangula* (Over de zeshoekige sneeuwvlok) [9] dat dit niet het geval was:



FIGUUR 5. Detail van de titelpagina van het boek *Strena Seu De Nive Sexangula*

Posteriori modo praeterquam quod quilibet globus a quatuor circumstantibus in eodem plano tangitur, etiam a quatuor infra se et a quatuor supra se, et sic in universum a duodecim tangetur, fientque compressione ex globosis rhombica. Ordo hic magis assimilabitur octaedro et pyramidi. Coaptatio fiet arctissima: ut nullo praeterea ordine plures globuli in idem vas compingi queant.



FIGUUR 6. De claim van Kepler

Dit kan vertaald worden als:

In deze rangschikking raakt elke bol niet alleen zijn vier buren in hetzelfde vlak, maar ook vier onder hem, en vier erboven, zodat alle bollen door twaalf andere worden aangeraakt, en de bollen worden samengeperst tot rombische dodecaëders. Deze rangschikking heeft meer de structuur van een octaëder en een pyramide. Deze stapeling is de dichtst mogelijke: in geen enkele rangschikking kunnen meer bollen worden geplaatst in dezelfde ruimte.

Hoewel Kepler claimt dat deze stapeling het dichtst is, geeft hij geen bewijs. De uitspraak dat de kubische vlakgecentreerde stapeling de grootst mogelijke dichtheid heeft van alle bolstapelings is daarom bekend geworden als *het Kepler-vermoeden*.

Optimaliteit bewijzen in het geval dat de bollen gerangschikt zijn volgens een rooster was al gedaan door Carl Friedrich Gauss (1777–1855). Evenwel was het lange tijd een open probleem of de kubische vlakgecentreerde stapeling ook optimaal was zonder deze restrictie.

2. Een bewijs te ingewikkeld om te beoordelen

In 1953 liet László Fejes Tóth (1915–2005) zien dat het bewijzen van het Kepler-vermoeden kon worden gereduceerd tot een groot maar eindig aantal berekeningen [2].

Gebaseerd op deze ideeën gaf Tom Hales (1958–) samen met zijn student Samuel Ferguson een bewijs voor het Kepler-vermoeden. Hij kondigde de voltooiing van dit bewijs in augustus 1998 aan.



FIGUUR 7. Tom Hales

Een bijzonderheid van dit bewijs is dat het is gebaseerd op de resultaten van een divers en groot aantal computerberekeningen. Om het bewijs te vertrouwen moet je er dus op kunnen vertrouwen dat de gebruikte software geen bugs heeft en dat de computer tijdens het rekenen nergens een storing heeft gehad. We zullen in Sectie 5 de structuur van het bewijs schetsen.

Hales probeerde zijn bewijs gepubliceerd te krijgen in de *Annals of Mathematics*. Wat hij ter beoordeling instuurde was een document van 250 pagina's, vergezeld van 3 gigabytes aan computerbestanden. De *Annals* stelde vervolgens een leesgroep van twaalf referenten in met als opdracht de correctheid van het bewijs te controleren. In 2003, na vier jaar werk, gaf deze groep de opdracht terug, zonder de correctheid te hebben vastgesteld. De groep had geen fouten in het bewijs kunnen vinden. Maar aan de andere kant was het bewijs zo complex dat de groep niet in staat was geweest het te beoordelen. De groep meldde dat ze "99% zeker" waren van de correctheid van het bewijs, maar dat ze er niet volledig waren uitgekomen.

Er werd vervolgens besloten om het bewijs in tweeën te hakken. Het tekstuele deel werd geaccepteerd in de *Annals* in 2005 [4], terwijl het computergedeelte werd gepubliceerd in twee artikelen in *Discrete and Computational Geometry* in 2006 [7]. Het *Annals*-artikel bevatte een verwijzing naar de computerhelft van het bewijs, maar gaf daarbij de disclaimer van de editors van het tijdschrift dat het hun niet gelukt was de correctheid hiervan te beoordelen.

3. Het Flyspeck-project

Computers zijn bezig het karakter van de wiskunde te veranderen. Als je een hulpmiddel hebt gebruik je het ook, en computers zijn voor wiskunde een buitengewoon krachtig hulpmiddel.

Er zijn allerlei manieren waarop je computers kunt gebruiken om wiskunde te doen:

- Ten eerste kun je computers gebruiken om te rekenen. Hierbij kan het gaan om getalsmatige berekeningen, maar ook om grote hoeveelheden gevallen

langs te gaan. Dit is het soort computergebruik dat de basis is voor het bewijs van Hales en Ferguson.

Computers rekenen gewoonlijk met eindige precisie (met *floating point numbers* of drijvendekommagetallen), en maken dus afrondfouten, maar dit hoeft geen reden te zijn dat berekeningen niet wiskundig hard bruikbaar zijn. Zo kun je kiezen naar welke kant je getallen laat afronden, en in plaats van alleen met benaderingen kun je met intervallen rekenen. Bij *interval-aritmetiek* zorg je ervoor dat het reële getal waarover je een uitspraak doet gegarandeerd binnen de intervalbenadering ligt. Interval-aritmetiek is een essentieel onderdeel van het Hales-Ferguson-bewijs.

- Wiskundige berekeningen hoeven niet numeriek te zijn, maar kunnen ook *symbolisch* zijn. Als ik het getal $\sqrt{2}$ beschouw, dan kan ik maar eindig veel cijfers achter de komma opschrijven. Evenwel, door dit getal symbolisch te beschouwen heb ik toch maar eindig veel inkt nodig om het in volledige precisie te kunnen manipuleren.

Een dergelijk onderscheid bestaat ook in de computer. De *computer algebra* systemen als Mathematica en Maple werken niet met getallen maar met symbolische expressies. Ook deze systemen worden vaak gebruikt als handig hulpmiddel in wiskundig onderzoek.

- Bij de vorige twee manieren om computers te gebruiken gaat het over rekenen, en het *redeneren* over wat de berekeningen voorstellen wordt aan de mens overgelaten. Er zijn ook computersystemen die helpen bij het redeneren [3]. Hierbij kun je onderscheid maken tussen het vinden van redeneringen versus het verifiëren van redeneringen. In het eerste geval gaat het om ATP systemen (automated theorem provers), en in het tweede geval om ITP systemen (interactive theorem provers). Een andere term voor een ITP systeem is een *bewijsassistent*. Hoewel bij ITP het bewijs door de mens wordt aangedragen bieden deze programma's toch een heleboel hulp bij het redeneren.

De eerste twee categorieën computergebruik binnen de wiskunde hebben een hoge vlucht genomen. De laatste is momenteel voornamelijk nog een researchonderwerp voor informatici, maar was precies wat Hales nodig had.

In januari 2003 besloot Hales zijn bewijs met een ITP systeem te gaan controleren. Omdat deze systemen een buitengewoon hoge betrouwbaarheid hebben, zouden er daarna geen referenten meer nodig zijn. Hij had met ITP-specialisten gecorrespondeerd (waaronder met ons), en de schatting was dat het ongeveer twintig manjaar zou kosten om het bewijs van het Kepler-vermoeden in een ITP-systeem in te voeren en te verifiëren. Hales besloot dat dit een doenlijk project was. Als naam voor het project

[6] koos hij *Flyspeck*, door met het patroon **f*p*k** (voor “formal proof of Kepler”) in een lijst woorden te zoeken. In het Engels betekent “to flyspeck” ook zoiets als “in extreem detail op gebreken controleren”, dus deze naam was zeer toepasselijk. Als systeem voor het project koos hij het HOL Light systeem [8] van John Harrison. Harrison is op *interactive theorem proving* (ITP) gepromoveerd in Cambridge, en is als ITP-specialist werkzaam bij Intel in Portland in de VS. Hoewel hij in zijn baan correctheidsbewijzen voor Intel verifieert, die dus over informatica gaan, heeft hij in zijn vrije tijd een indrukwekkende hoeveelheid wiskundige bewijzen in zijn systeem geformaliseerd.



FIGUUR 8. John Harrison, de ontwikkelaar van de HOL Light bewijsassistent

Het HOL Light systeem is vooral voor wiskunde één van de sterkere systemen, maar wordt in de informatica minder gebruikt. Andere, in de informatica bekendere ITP-systemen zijn het Coq systeem [1] uit Frankrijk en het Isabelle [11] systeem uit Engeland/Duitsland. De gebruikers van deze systemen bleken ook zeer in Flyspeck geïnteresseerd (“eindelijk een echte wiskundige die onze systemen nodig heeft!”) en hebben ook aan Flyspeck gewerkt. Een deel van het werk met Isabelle is ook in het uiteindelijke Flyspeck-bewijs terecht gekomen (zie Sectie 6), maar het overgrote deel van het Flyspeck-bewijs is geformaliseerd in het HOL Light systeem.

Hales besloot Flyspeck een extra impuls te geven door het coderen van het bewijzen te “outsourcen”, namelijk in Vietnam. Hij organiseerde daar een groep Vietnamese wiskundigen, leerde ze HOL Light, en zette ze aan het coderen van het bewijs van het Kepler-vermoeden. Tegelijk werkte hij aan een boek dat het bewijs minutieus gedetailleerd weergeeft, het zogeheten “blueprint” boek, want de subtitel is “a blueprint for formal proofs”. Dit boek [5] telde uiteindelijk 334 bladzijden.

Op 10 augustus 2014 werd bekend gemaakt dat het Flyspeck-project was voltooid. Het bewijs was in de tussentijd aangepast en gestroomlijnd (ook om verificatie met

de computer te vergemakkelijken), en dit aangepaste bewijs is dus 100% zeker correct. Het Kepler-vermoeden kan sinds deze datum als bewezen worden beschouwd.

4. Zekerheid dat een bewijs correct is

Een wiskundige gaat achter een computer met HOL Light zitten, en claimt: dankzij dit systeem weet ik 100% zeker dat deze stelling bewijsbaar is. Hoe kan hij zo zeker van zijn zaak zijn? Het oorspronkelijke Kepler-bewijs van Hales en Ferguson bevatte een groot aantal computerberekeningen, en het was daardoor moeilijk vast te stellen of het allemaal correct was. Het Flyspeck-bewijs heeft net zo goed een groot aantal computerberekeningen (door HOL Light) nodig om te verwerken. Waarom zou dit anders zijn?

Een ITP-systeem als HOL Light implementeert een redeneersysteem uit de mathematische logica dat extreem eenvoudig is. Alle complexiteit van de redenering in het zeer ingewikkelde Kepler-bewijs wordt teruggebracht tot een gigantisch aantal (miljarden) zeer elementaire redeneerstapjes. Stuk voor stuk zijn deze stapjes uiterst simpel. Het gebruikte redeneersysteem heet HOL (voor “higher order logic”, hogere orde logica), en bestaat uit maar tien zeer eenvoudige redeneerregels. Ieder stapje past één van die regels toe.

Nu komt de controle van de correctheid van het bewijs uitsluitend neer op het controleren van deze redeneerstapjes. Maar omdat het redeneersysteem zo eenvoudig is, is het maar een heel klein deel van de gebruikte software dat dit doet. Deze *logische kern* is ongeveer vierhonderd programmeerregels lang (geschreven in de functionele programmeertaal OCaml). Door het gebruik van de techniek van *abstracte datatypen* kan er geen wiskundige incorrectheid optreden als deze kern van het programma foutloos is. En bij een programma van vierhonderd regels is het goed te doen om alle fouten eruit te halen.

In andere woorden: in plaats van het belachelijk ingewikkelde Kepler-bewijs op correctheid te moeten beoordelen, hoeven de referenten nu alleen vierhonderd regels OCaml-code op correctheid te controleren, en de computer doet dan de rest.

Nu zijn er ook wel programmeerfouten in kernen van ITP-systemen gevonden, dus we zouden nog een stap verder willen gaan. In de informatica bestaat er technologie – ook gebaseerd op ITP – om van programma’s (in het bijzonder als het geschreven is in een functionele programmeertaal) vast te stellen dat er *nul* fouten in zitten. Deze technologie hebben eerst John Harrison en vervolgens Ramana Kumar (een andere promovendus uit Cambridge) gebruikt om formeel te bewijzen dat de kern van HOL Light foutloos is [10]. Zelfs hier hoeven de referenten zich dus geen zorgen meer over te maken.

Er is hier natuurlijk wel sprake van een kip-en-ei probleem, en tevens volgt uit de Gödelstelling (die zegt dat

een consistent logisch systeem niet van zichzelf kan bewijzen dat het consistent is) dat een systeem niet de correctheid van zijn eigen broncode kan vaststellen. Maar al met al betekent al deze technologie dat de betrouwbaarheid van ITP-systemen, en daardoor van de gecontroleerde bewijzen, extreem hoog is. Onvoorstelbaar veel hoger dan die van bewijzen die alleen door mensen geverifieerd zijn.

Als de implementatie van het ITP-systeem volledig vertrouwd wordt kan er uiteraard in een ander onderdeel van de computer een softwarefout of hardwarefout zitten, waardoor een incorrect bewijs geaccepteerd wordt door de computer. Het is uiterst onwaarschijnlijk dat, bijvoorbeeld, een fout in het besturingssysteem ervoor zou zorgen dat een incorrect bewijs juist wel door de ITP-kern geaccepteerd zou worden. Dit punt wordt ondervangen door de bewijsverificatie op verschillende computers met verschillende software opnieuw te draaien.

5. Het Hales-Ferguson-bewijs

Het bewijs van het Kepler-vermoeden bestaat uit ruwweg acht stappen, waarbij we hier het bewijs vertellen van het resultaat terugwerkend naar eerdere lemma’s. De spil van het bewijs is stap (d), de “lokale annulus-ongelijkheid”.

(a) De precieze formulering van het Kepler-vermoeden.

Het Kepler-vermoeden gaat over oneindigheid: wat er mogelijk is als we een oneindige ruimte vullen met oneindig veel bollen met vaste straal r_0 . Dit betekent dat we het over over een limietproces hebben, want de dichtheid van de bollen is de limiet van de dichtheid van deze bollen binnen een groeiende bol met straal r , waarbij r naar oneindig gaat. Nu hoeft deze limiet niet te bestaan, maar de $\limsup$ bestaat natuurlijk altijd wel. Wat we dus moeten bewijzen is dat dit voor iedere mogelijke configuratie van de bollen nooit groter zal zijn dan $\pi/\sqrt{18}$.

Het is duidelijk dat de $\limsup$ niet afhangt van de keuze van r_0 , dus vanaf nu zullen we $r_0 = 1$ nemen. We kijken dus naar een stapeling van eenheidsbollen.

(b) De uitspraak die bewezen wordt in het Flyspeck-project.

Flyspeck kijkt niet van naar het volume van de doorsnede van de bollen met de grote bol met straal r , maar telt de middelpunten van de kleine bollen die zich binnen de grote bol bevinden. Omdat het volume van de grote bol evenredig is met r^3 en wat we zo verwaarlozen correspondeert met een volume ten hoogste evenredig met r^2 , maakt dit voor de $\limsup$ van de dichtheid niet uit.

Voorts heeft Flyspeck het niet over de $\limsup$, maar bewijst het de volgende uitspraak:

Voor iedere stapeling S bestaat een c zodat voor iedere $r \geq 1$ geldt dat:

$$\text{card}(V_S \cap B(0, r)) \leq \frac{\pi}{\sqrt{18}} r^3 + cr^2$$

Hierin is V_S de verzameling met de middelpunten van de kleine bollen behorend bij stapeling S , en $B(0, r)$ is een

groeïende (open) bol met straal r . De eis $r \geq 1$ is duidelijk nodig voor het geval dat de oorsprong in V_S zit. De constante c hangt in deze formulering af van de stapeling, maar Hales claimt dat deze ongelijkheid ook voor een vaste c kan worden bewezen. Deze verfijning zit evenwel niet in Flyspeck.

In de computercode die de input voor het HOL Light systeem is, is deze uitspraak gecodeerd als:

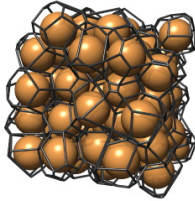
```
!V. packing V ==>
  (?c. !r. &1 <= r ==>
    &(CARD(V INTER ball(vec 0,r))) <=
      pi * r pow 3 / sqrt(&18) + c * r pow 2))
```

Hierin is “!” de universele kwantor $\forall$, “?” de existentiële kwantor $\exists$, en “&” de functie die een natuurlijk getal afbeeldt op het corresponderende reële getal.

(c) Verwaarloosbare FCC-compatibele functies.

We hebben het ook in deze herformulering nog steeds over een oneindig probleem, en we willen dit graag reduceren naar een eindig probleem, en zo een aanpak mogelijk maken die analyseert wat er rond een enkele bol kan gebeuren.

Hier toe nemen we een stapeling S en we verdelen de ruimte in de *Voronoi-cellen* van de verzameling van bolmiddenpunten van S , V_S . Voor ieder punt $v \in V_S$ is de Voronoi-cel $\Omega(V_S, v)$ de verzameling punten die dichterbij v ligt dan bij ieder ander punt in V_S .



FIGUUR 9. Voronoi-cellen van een bolstapeling

Een reëelwaardige functie $G(v)$ op de punten van V_S heet *verwaarloosbaar* als er een c_1 bestaat zodat voor alle $r \geq 1$:

$$\sum_{v \in V_S(0,r)} G(v) \leq c_1 r^2.$$

De functie $G(v)$ heet *FCC-compatibel* als voor iedere $v \in V_S$

$$4\sqrt{2} - \text{vol}(\Omega(V_S, v)) \leq G(v).$$

Het getal

$$4\sqrt{2} = 5,55685 \dots$$

is het volume van de Voronoi-cellen van FCC, de kubische vlakgecentreerde stapeling.

Als de stapeling S echt beter is dan FCC, dan zijn de Voronoi-cellen van S kleiner dan die van FCC. De definitie van een verwaarloosbare FCC-compatibele functie $G(v)$ betekent dat de Voronoi-cellen van S niet significant kleiner zijn dan die van FCC: het verschil is maximaal van orde r^2 , terwijl het aantal punten in $V_S(0, r)$ van orde r^3 is. Dus in de limiet is de afwijking verwaarloosbaar. Hieruit volgt duidelijk het Kepler-vermoeden.

Het bestaan van zo'n verwaarloosbare FCC-compatibele functie $G(v)$ is stap (f) van deze bewijsschets.

(d) De lokale annulus-ongelijkheid.

Dit is het centrale lemma van het bewijs, dat ook voor andere stellingen toepasbaar is gebleken.

Definieer het magische getal h_0 exact als

$$h_0 = 1,26$$

Bekijk nu de “schil” van punten v om een eenheidsbol met middelpunt in de oorsprong met de eigenschap

$$1 \leq h(v) \leq h_0$$

waarbij $h(v)$ de helft van de afstand van punt v tot de oorsprong is. Deze “schil” heet de *annulus* van de centrale bol. We gaan kijken wat er gebeurt als we bollen om deze centrale bol plaatsen, zodanig dat de middelpunten zich allemaal in deze annulus bevinden.

Voorts tellen we deze bollen gewogen: als ze verder naar buiten zijn tellen ze minder mee. (In dat geval is er meer ruimte tussen de bollen, en we willen een ongelijkheid die het aantal begrenst.) Precies gezegd definiëren we een functie $L(h)$ door

$$L(h) := \begin{cases} h_0 - h & \text{if } 1 \leq h \leq h_0 \\ h_0 - 1 & \text{if } 1 \leq h \leq h_0 \\ 0 & \text{if } h_0 \leq h \end{cases}$$

Deze functie vervalt lineair van 1 op het binnenoppervlak van de annulus naar 0 op het buitenoppervlak. De lokale annulus-ongelijkheid is nu:

$$\sum_{v \in V_S} L(h(v)) \leq 12$$

Het centrale lemma zegt dat deze ongelijkheid geldt voor iedere stapeling S .

De geldigheid van deze ongelijkheid zal volgen uit het resultaat in stap (g), en is één van de ingrediënten voor het bestaan van een verwaarloosbare FCC-compatibele functie in stap (f).

Dat er maximaal twaalf bollen aan een centrale bol kunnen raken was Kepler al duidelijk. Door de bollen verder naar buiten te bewegen passen er mogelijk meer, maar als je ze minder meetelt, blijft het “aantal” dus toch hoogstens twaalf.

(e) Marchal-cellen en de cel-cluster ongelijkheid

Nu wordt het bewijs nog technischer, en we worden we hier nog schetsmatiger in onze beschrijving van het bewijs dan we toch al waren.

De Voronoi-cellen worden verdeeld in *Rogers-simplices* en deze worden weer onderverdeeld in fragmentjes die *Marchal-cellen* worden genoemd. Dit alles wordt gedaan om te komen tot de zogenaamde *cel-cluster ongelijkheid*. Die heeft te maken met configuraties waarin twee bollen een afstand in een specifiek interval hebben (zo'n configuratie wordt een 'cel-cluster' genoemd). De cel-cluster ongelijkheid blijkt de cruciale stap om te laten zien dat de functie $G(v)$, die we in de volgende sectie definiëren, verwaarloosbaar is.

(f) Met wat we nu hebben kunnen we, gegeven een stapeling S , een verwaarloosbare FCC-comptibele functie definiëren:

$$G(v) = -\text{vol}(\Omega(V_S, v)) + 8m_1 - \sum 8m_2 L(h(v))$$

Hierin:

$$\text{sol}_0 = 3 \arccos(1/3) - \pi = 0,551285 \dots$$

$$\tau_0 = 4\pi - 20\text{sol}_0 = 1,54065 \dots$$

$$m_1 = \text{sol}_0 2\sqrt{2}/\tau_0 = 1,01208 \dots$$

$$m_2 = (6\text{sol}_0 - \pi)\sqrt{2}/(6\tau_0) = 0,0254145 \dots$$

De som in de definitie van $G(v)$ loopt over alle bolmiddenpunten in V_S uitgezonderd de oorsprong. Als v te ver van de oorsprong ligt geldt $L(h(v)) = 0$, dus dit is een eindige som.

Uit de lokale annulus-ongelijkheid volgt nu dat deze functie FCC-compatibel is:

$$\begin{aligned} 4\sqrt{2} - \text{vol}(\Omega(V_S, v)) \\ \leq -\text{vol}(\Omega(V_S, v)) + 8m_1 - 12 \cdot 8m_2 \\ \leq -\text{vol}(\Omega(V_S, v)) + 8m_1 - \sum 8m_2 L(h(v)) \\ = G(v) \end{aligned}$$

Dat deze functie ook verwaarloosbaar is volgt dus uit de cel-cluster ongelijkheid.

(g) Het niet-bestaan van een **tegenvoorbeeldstapeling**.

We hebben nog niet beschreven hoe de lokale annulus-ongelijkheid en de cel-cluster ongelijkheid bewezen worden. We gaan nu eerst kijken hoe het bewijs van de eerste gaat. Over het bewijs van de tweede hebben we het in Sectie 6.

Een verzameling bollen S rond een centrale bol waarvoor de lokale annulus-ongelijkheid niet geldt noemen we een *tegenvoorbeeldstapeling*. In feite eisen we voor een tegenvoorbeeldstapeling nog wat meer, zoals dat het aantal bollen ten hoogste veertien is, en dat de waarde van $\sum_{v \in V_S} L(h(v))$ maximaal is.

We willen laten zien dat er geen tegenvoorbeeldstapeling bestaat, want dan geldt de lokale annulus-ongelijkheid voor alle stapelingen.

(h) **Tamme grafen** zijn de grafen van tegenvoorbeeldstapelingen.

We definiëren nu bij een tegenvoorbeeldstapeling een graaf door twee middelpunten van bollen te verbinden als

de afstand kleiner is dan $2h_0$. Deze grafen voldoen aan een reeks eigenschappen, en alle grafen die aan die eigenschappen voldoen noemen we *tamme* grafen. Deze worden gedefinieerd met een lijst van elf eigenschappen, waarvan de ingewikkeldste is dat er een functie van de vlakken in de graaf naar de reële getallen moet bestaan die aan drie verdere eigenschappen moet voldoen.

Het bewijs wordt nu afgemaakt door eerst te bewijzen dat de graaf van een tegenvoorbeeldstapeling altijd tam is. Dan wordt bewezen dat iedere tamme graaf in een lange lijst van 19.715 grafen voorkomt. En tenslotte wordt voor iedere graaf uit die lijst gekeken of het mogelijk is om de bollen zo te plaatsen dat de lokale annulus-ongelijkheid niet geldt. Dit correspondeert met het zoeken van een oplossing in een hele reeks lineaire programma's. Doordat dat voor geen enkele tamme graaf lukt, volgt hieruit de lokale annulus-ongelijkheid.

6. De computer helpt

Zoals gezegd rust het Hales-Ferguson-bewijs op een aantal computerberekeningen:

(i) **Een programma dat de lijst van 19.715 grafen genereert waar alle tamme grafen bij zitten.**

Dit programma was oorspronkelijk in Java geschreven. De versie van dit programma in Fyyspeck is geschreven in de taal van de bewijsassistent Isabelle en vervolgens geëxporteerd naar een programma in de programmeertaal ML. De verificatie van dit programma is het enige onderdeel van Fyyspeck dat niet in HOL Light is gedaan maar in Isabelle. In de HOL Light-verificatie is het resultaat hiervan, inclusief de lijst grafen, als axioma aangenomen.

(j) **Een computerberekening die verifieert dat 43.078 lineaire programma's onoplosbaar zijn.**

Ieder lineair programma in deze lijst heeft rond de duizend variabelen en een vergelijkbaar aantal vergelijkingen. Deze lineaire programma's waren de hoofdmoot van de 3 gigabyte data uit het oorspronkelijke bewijs. In het HOL Light bewijs worden ze *on the fly* gegenereerd en de onoplosbaarheid ervan wordt vervolgens bewezen.

(k) **Een computerverificatie dat 23.242 niet-lineaire ongelijkheden met hoogstens zes variabelen gelden.**

Dit is de verificatie waarbij in het oorspronkelijke bewijs interval-aritmetiek wordt gebruikt. Deze ongelijkheden worden op allerlei plaatsen in het bewijs gebruikt. Om

een idee te geven: het gaat hierbij om te verifiëren dat ongelijkheden van de volgende vorm gelden binnen een bepaald interval:

$$\sqrt{4x_2 \left(\begin{array}{l} x_2x_4(-x_2+x_1+x_3-x_4+x_5+x_6)+ \\ x_1x_5(x_2-x_1+x_3+x_4-x_5+x_6)+ \\ x_3x_6(x_2+x_1-x_3+x_4+x_5-x_6) \\ -x_1x_3x_4-x_2x_3x_5-x_2x_1x_6-x_4x_5x_6 \end{array} \right)} < \tan\left(\frac{\pi}{2} - 0.74\right)$$

Hoe dit allemaal aan elkaar hangt is in het diagram in onderstaande figuur weergegeven. De rechthoeken met dubbele randen zijn de rekenintensieve taken waarvoor speciale programma's zijn geschreven. De pijlen geven de afhankelijkheden tussen de verschillende onderdelen.

Het moeilijkste in de HOL Light-verificatie was het doenlijk krijgen van de verificatie van de computerberekeningen, omdat deze ook allemaal in hele kleine redeneerstapjes moesten worden afgebroken en door de HOL Light

kern geverifieerd. Om dit efficiënt genoeg te doen was een heleboel slinheid nodig. Dit was het onderwerp van het proefschrift van Alexey Solov'yev [12], een promovendus van Hales.

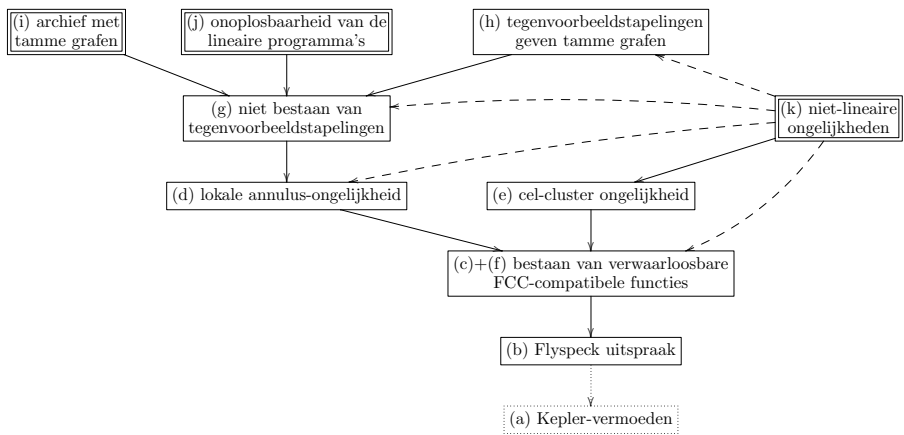


FIGUUR 11. Alexey Solov'yev

En zelfs hierna kon het project pas worden afgerond toen Microsoft genereus rekentijd in de cloud beschikbaar stelde. Evenwel werd vrij kort daarna de hele verificatie nog eens overgedaan op één van de clusters van de Radboud Universiteit in Nijmegen.

Referenties

- [1] Coq development Team. The Coq proof assistant, <https://coq.inria.fr/>.
- [2] L. Fejes Tóth. *Lagerungen in der Ebene auf der Kugel und im Raum*, first edition. Springer-Verlag, Berlin-New York, 1953.
- [3] Georges Gonthier, Thomas C. Hales, John Harrison, and Freek Wiedijk. Formal proof. *Notices of the American Mathematical Society*, 55(11):1370 – 1414, 2008.
- [4] Thomas C. Hales. A proof of the Kepler conjecture. *Annals of Mathematics*, 162:1063–1183, 2005.
- [5] Thomas C. Hales. *Dense Sphere Packings: a blueprint for formal proofs*, volume 400 of *London Math Soc. Lecture Note Series*. Cambridge University Press, 2012.
- [6] Thomas C. Hales, Mark Adams, Gertrud Bauer, Dat Tat Dang, John Harrison, Truong Le Hoang, Cezary Kaliszyk, Victor Magron, Sean McLaughlin, Thang Tat Nguyen, Truong Quang Nguyen, Tobias Nipkow, Steven Obua, Joseph Pleso, Jason Rute, Alexey Solov'yev, An Hoai Thi Ta, Trung Nam Tran, Diep Thi Trieu, Josef Urban, Ky Khac Vu, and Roland Zunkeller. A formal proof of the Kepler conjecture. *CoRR*, abs/1501.02155, 2015.
- [7] Thomas C. Hales and Samuel P. Ferguson. Historical overview of the Kepler conjecture; A formulation of the Kepler conjecture; Sphere packing III, Extremal cases; Sphere packing IV, Setailed bounds; Sphere packings VI, Tame graphs and linear programs. *Discrete & Computational Geometry*, 36(1):5–20; 21–69; 71–110; 111–166; 205–265, 2006.
- [8] J. Harrison. The HOL Light theorem prover, <http://www.cl.cam.ac.uk/~jrh13/hol-light/>.
- [9] J. Kepler. *The Six-Cornered Snowflake*. Oxford: Clarendon Press, 1966.
- [10] Ramana Kumar, Rob Arthan, Magnus O. Myreen, and Scott Owens. Self-formalisation of higher-order logic. *Journal of Automated Reasoning*, 56(3):221–259, 2016.
- [11] L. Paulson, T. Nipkow, and M. Wenzel. The Isabelle proof assistant, <https://isabelle.in.tum.de/>.
- [12] A. Solov'yev. *Formal Computations and Methods, Doctoral Dissertation*. PhD thesis, University of Pittsburgh, 2013.



FIGUUR 10. Het bewijs van het Kepler-vermoeden

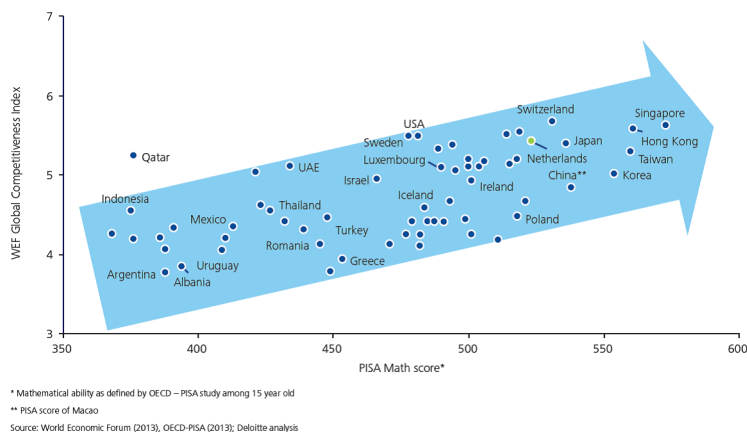
6 Moderne wiskunde: MSO en CSE

Wil Schilders

Inleiding

Wiskunde is vandaag de dag alom tegenwoordig. Dit blijkt al uit het door Platform Wiskunde Nederland in 2015 gepubliceerde, goed leesbare, boekje “Succesformules”, bedoeld om scholieren een beeld te geven van wat je zoal met wiskunde kunt doen. Daarnaast zijn er in vele landen boeken en boekjes gepubliceerd met zogenaamde “success stories” van wiskunde: in Frankrijk, Duitsland, Engeland, en ook in Europa als geheel: “European success stories in industrial mathematics”. Daarnaast zijn er in het Verenigd Koninkrijk, Nederland en Frankrijk rapporten verschenen, opgemaakt door accountants, over de economische waarde van wiskunde. Citerend uit het Nederlandse rapport (verkrijgbaar via de PWN website): *“Het blijkt dat het equivalent van 900.000 hoger opgeleide Nederlanders wiskunde, statistiek en besliskunde gebruikt in hun dagelijks werk. Zij variëren van wetenschappers die doorlopend met wiskunde bezig zijn, tot bankiers die wiskunde een deel van hun tijd gebruiken, bijvoorbeeld om activa te waarderen, en artsen die soms wiskunde nodig hebben om de resultaten van medische onderzoeken te kunnen interpreteren. Deze 900.000 banen genereren niet alleen inkomsten voor de betrokkenen zelf, maar ook bij bedrijven die producten of diensten leveren aan de gebruikers van wiskunde of aan de organisaties waar zij werken. Een standaard input-output analyse van de Nederlandse economie indiceert dat deze zogenaamde indirecte en afgeleide effecten nog eens 1,4 miljoen additionele banen opleveren. Daarmee dragen gebruikers van wiskunde, statistiek en besliskunde bij aan een kwart van de Nederlandse werkgelegenheid en, omdat zij relatief goed betaald worden, aan circa 30% van het bruto nationaal product.”* Dat is behoorlijk, 30% van het BNP! Ook in de andere landen kwam men tot dergelijke cijfers.

Daarnaast constateerde men ook het volgende, grafisch weergegeven in figuur 6.1: *“Verder blijkt dat wiskundige kennis sterk correleert met de internationale concurrentiekracht van een land. Een goede basis in wis-*



Figuur 6.1: Relatie tussen wiskundige kennis en concurrentievermogen van landen

kunde, statistiek en besliskunde is een belangrijke pijler onder het succes van de Nederlandse economie. Met de explosieve toename van data, onder andere door het internet, met steeds snellere computers en met betere rekenmethodes zal het belang van deze pijler de komende jaren alleen maar toenemen.” Meer en beter gebruik van wiskunde zorgt er dus voor dat Nederland zijn positie als een van de meest competitieve landen ter wereld kan behouden en dat ons welvaartspeil onverminderd hoog blijft.

Wiskunde is dus erg belangrijk, en dat wordt onder andere mooi verwoord in een statement van de bekende Nederlandse wiskundige Lex Schrijver: “Wiskunde is als zuurstof. Als het er is, merk je het niet. Als het er niet zou zijn, merk je dat je niet zonder kunt.” De vergelijking met zuurstof gaat ook in een andere zin op, omdat de meeste mensen zich niet realiseren wat het belang is van wiskunde. Ik werkte zelf vele jaren in de industrie, we maakten mathematische software die vervolgens door de ingenieurs werd gebruikt om nieuwe ontwerpen te maken. Zij kregen de schouderklopjes, de wiskundigen niet; wij leverden feitelijk een onzichtbare bijdrage aan een zichtbaar succes.

Gelukkig begint de situatie snel te veranderen. Wiskunde speelt zowel in de industrie als in andere wetenschappen een in toenemende mate belangrijke rol, en men begint zich ook goed het belang van wiskunde te realiseren. Veel producten worden tegenwoordig ontworpen achter een beeldscherm, fysieke prototypes worden nauwelijks nog gemaakt. Dit geldt zelfs voor het ontwerp van vliegtuigen: het is nu mogelijk om een nieuw type vliegtuig

geheel achter het computerscherm te ontwerpen, zonder dat er 1 windtunneltest nodig is. En het vliegt nog ook! Dit vereist natuurlijk wel dat de software die gebruikt wordt, nauwkeurig kan voorspellen hoe producten zich in de werkelijke wereld gedragen. En dat impliceert op zijn beurt dat de wiskundige modellen welke worden gebruikt, de werkelijkheid goed beschrijven. En voorts dat de algoritmen welke oplossingen berekenen, nauwkeurig zijn, bestand tegen allerlei invloeden, en liefst heel snel tot resultaten leiden. Last but not least wil men ook graag nog eens het onderste uit de kan halen, en ontwerpen optimaal krijgen, in een tevoren bepaalde zin. Dit kunnen zeer diverse criteria zijn, bijvoorbeeld gericht op bepaalde eigenschappen van het product, of het totale kostenplaatje. Kortom, het wiskundige werk bestaat uit het **modelleren** van het product en zijn eigenschappen, vervolgens het **simuleren** ervan op allerlei mogelijke manieren en onder verschillende scenarios, en tenslotte het **optimaliseren**. We vatten dit samen onder de term MSO: (mathematical) modeling, simulation and optimisation.

MSO wordt tegenwoordig veelvuldig gebruikt, zowel binnen de industrie als ook in andere wetenschappen. Traditioneel wordt wiskunde natuurlijk al veel gebruikt in de fysica en de chemie, maar het werkveld van de wiskundige wordt steeds meer uitgebreid naar andere wetenschappelijke disciplines. Denk bijvoorbeeld aan het gebruik van wiskunde binnen de economische wetenschappen, financiële wiskunde heeft een grote vlucht genomen. Of het toenemende gebruik van wiskunde binnen de alfa en gamma wetenschappen. Ook biologische processen worden tegenwoordig meer en meer benaderd met wiskundige methoden, en daarom wordt de biologie ook wel eens ‘the new physics’ genoemd. Allerlei processen welke te maken hebben met het leven op aarde kunnen tegenwoordig met wiskunde worden beschreven, gemodelleerd dus. Vaak wordt er voor deze wiskundige activiteit binnen andere wetenschappen een andere afkorting gebruikt: CSE, ofwel computational science and engineering. Het is een bloeiende tak van wetenschappelijke sport, zo organiseert de Amerikaanse organisatie SIAM (Society for Industrial and Applied Mathematics) sinds de eeuwwisseling een conferentie over CSE, en dit is uitgegroeid tot hun grootste en meest succesvolle conferentie in de afgelopen jaren. De meest recente editie vond plaats in Atlanta eind februari-begin maart dit jaar, ik was een van de organisatoren. De website met alle presentaties is: <https://www.siam.org/meetings/cse17/>.

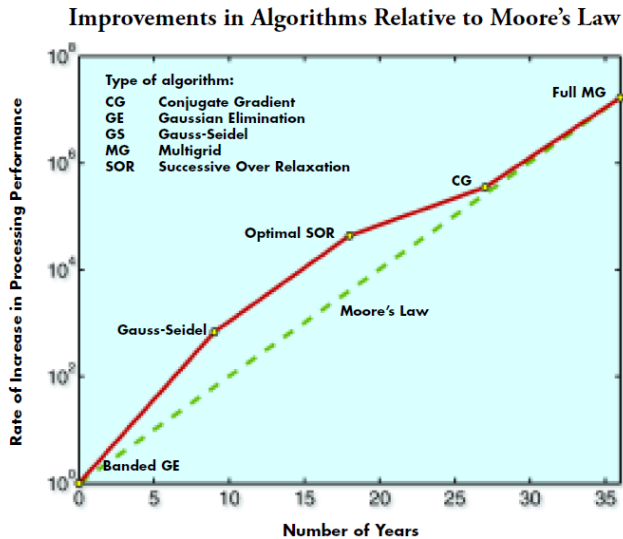
MSO en CSE zijn belangrijke nieuwe begrippen in de wiskundewereld, en feitelijk is dit dus de moderne wiskunde van deze tijd! We zullen hier wat dieper ingaan op wat dit alles betekent, zowel voor de wiskundigen als voor de wetenschappers en ingenieurs die dankbaar gebruik van maken van deze

6.1 Wiskunde en computers

Alvorens we dieper ingaan op de moderne wiskunde, vormgegeven door MSO en CSE, even naar het thema van deze vakantiecursus, namelijk de relatie tussen wiskunde en computers. En dat is een erg interessante relatie, zeker als je het bekijkt vanuit de ‘public relations’ kant. Iedereen weet dat computers steeds sneller worden, althans, dat was tot voor kort zeker het geval. Dit wordt beschreven door de zogenaamde ‘wet van Moore’, waarvan verschillende versies bestaan, maar door de bank genomen betekent het dat chips elke 18 maanden 2x zo snel worden. Veel mensen zijn zich bewust van deze wetmatigheid, die overigens enkel geobserveerd wordt en niet bewezen kan worden. Dus stelt men vaak de aankoop van een laptop of PC uit, wachtend op de nieuwe generatie welke 2x zo snel is. Het algemeen gangbare idee is dan ook dat alles wat we tegenwoordig kunnen met simulaties op computers, met MSO en CSE, te danken is aan de wet van Moore en ontwikkelingen in de elektronische industrie, Niets is echter minder waar! Wiskundige algoritmen leveren hier een essentiële bijdrage aan, en zelfs zodanig dat ze de snelheidswinst van chips op allerlei manieren verslaan. Voor bepaalde problemen kan men aantonen dat wiskundige algoritmen ook een verloop hebben als beschreven door de wet van Moore, en in bepaalde vakgebieden is het zelfs zo dat wiskundige algoritmen de snelheidswinst van chips ruim verslaan, met vele factoren. Laten we om te beginnen eens kijken naar algoritmen voor het oplossen van lineaire stelsels, $Ax = b$. Dan krijgen we de volgende figuur:

Maar ook voor het vinden van optimale oplossingen van bepaalde optimaliseringsproblemen blijken wiskundige algoritmen zeer effectief te zijn, effectiever dan computers:

Dit zijn cijfers van een tijd geleden, toch tonen deze aan dat wiskundige algoritmen heel krachtig zijn, en vaak een betere performance vertonen dan de karakteristieken van computers en chips. Dit is nog veel meer het geval als we kijken naar algoritmen in het gebied van de zogenaamde ‘mixed integer programming’, optimalisering waarbij een menging ontstaat van optimaliseren met reële getallen en ook gehele getallen, integers in het Engels. Hierbij optimaliseer je dus een probleem waarvan je weet dat bepaalde eigenschappen enkel met integer waarden beschreven kunnen worden. Dit is een uitermate belangrijk onderzoeksgebied tegenwoordig. In Figuur 6.4 zien we hoe de algoritmen op dit gebied zich verhouden met de snelheidswinst van computers.



Figuur 6.2: Ontwikkeling van numerieke algoritmen voor het oplossen van $Ax=b$ in vergelijking met de wet van Moore

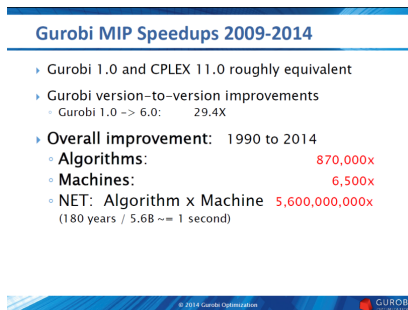
Progress in LP: 1988—2004

(Operations Research, Jan 2002, pp. 3—15, updated in 2004)

- Algorithms (*machine independent*):
Primal versus best of Primal/Dual/Barrier 3,300x
- Machines (workstations → PCs): 1,600x
- NET: Algorithm × Machine 5,300,000x
(2 months/5300000 ≈ 1 second)

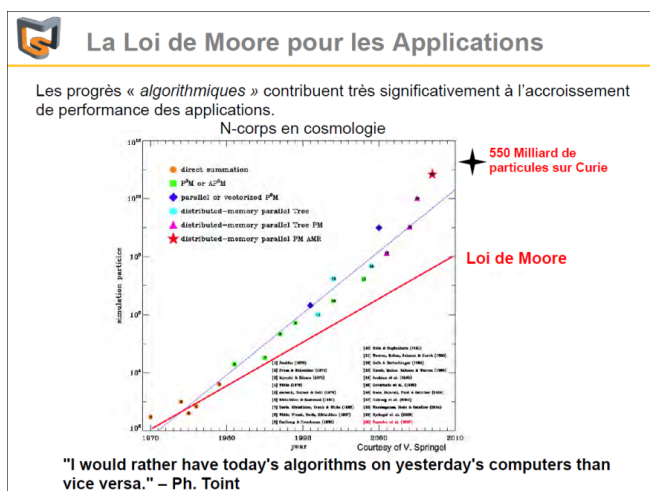
Figuur 6.3: Snelheidswinst van algoritmen voor ‘lineair programmeren’ in vergelijking met de snelheidswinst van computers

Tenslotte kijken we naar algoritmen welke gebruik maken van de simulatie van deeltjes, zogenaamde ‘particle simulations’. Veel van de simulaties binnen MSO en CSE gebruiken modellen die beschreven worden door gewone of partiële differentiaalvergelijkingen. Er zijn echter ook methoden die op een nog elementairder niveau opereren en deeltjes volgen. Voorbeelden zijn de methoden bekend als ‘lattice Boltzmann’ of ‘smoothed particle hydrodynamics’ (SPH). Ook voor deze deeltjesmethoden blijkt dat de wis-



Figuur 6.4: Snelheidswinst van zogenaamde ‘mixed integer programming’ algoritmen in vergelijking met computers

kundige algoritmen vele malen meer snelheidswinst oplevereh dan die van computers:



Figuur 6.5: Snelheidswinst voor deeltjes-methoden in vergelijking met de wet van Moore

Kortom, we zien dat wiskundige algoritmen vaak een veel grotere invloed hebben op de snelheidswinst die door computers wordt behaald. Het probleem is dat iedereen de snelheidswinsten van computers kent, maar weinigen ook de winst die behaald wordt door wiskundige algoritmen. En dat is vervelend, het ondermijnt het belang van wiskunde. We zouden er

voor moeten zorgen dat meer mensen zich bewust worden van het feit dat juist wiskundige algoritmen er voor gezorgd hebben dat we tegenwoordig in staat zijn al die ingewikkelde simulaties te doen. En dat het dus niet de computers zijn die ons in staat stellen om die ingewikkelde simulaties te bewerkstelligen. Philippe Toint verwoordde het als volgt: “I would rather have today’s algorithms on yesterday’s computers than vice versa.” Inderdaad, als we enkel zouden vertrouwen op de snelheidswinst van computers, dan zouden we nu de simulaties kunnen doen welke we reeds in 1990 deden.

Als we het dus hebben over een breekijzer, zoals de titel van deze editie van de Vakantiecursus aangeeft, dan is het eerder de wiskunde dan de computer. Als je er even over nadenkt, dan is dat ook niet zo vreemd. De snelheidswinst van computers hangt vooral af van de mogelijkheden van elektronische chips, en het is noeste arbeid om de miniaturisering hiervan vol te houden en elke 2 jaar weer te zorgen voor een nieuwe generatie die nog sneller is. Maar om algoritmen sneller te maken is vaak gezond verstand en veel inzicht nodig. Na bestudering van het gedrag van een probleem is het vaak mogelijk om een nieuw algoritme te bedenken dat gebruik maakt van bepaalde eigenschappen en daardoor vele malen sneller kan worden, een factor 2, een factor 10 of zelfs nog meer. Het is ook noeste arbeid, maar op een andere manier en niet behept met de wetten van de fysica.

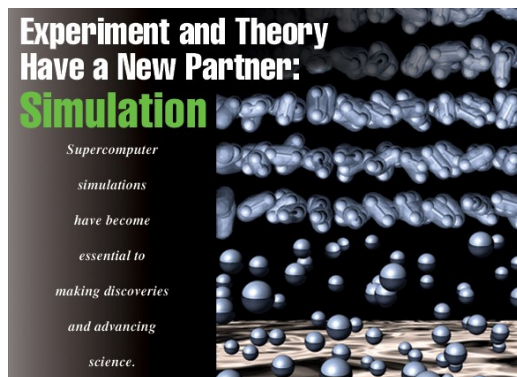
Na dit uitstapje gemaakt te hebben, over de relatie tussen wiskundige algoritmen en computers, is het tijd om terug te gaan naar CSE en MSO, de moderne wiskunde van onze tijd. Daartoe beginnen we met een kleine historische terugblik.

6.2 De derde discipline

Honderden jaren geleden dacht Sir Isaac Newton na over allerlei natuurverschijnselen, en bestudeerde onder andere vallende voorwerpen, zoals appels; althans, dat zegt de overlevering. Newton probeerde deze verschijnselen te verklaren door een theorie hiervoor te bedenken. Het leidde onder andere tot de universele wetten van de zwaartekracht. Newton was niet de enige die zo te werk ging, er waren veel wetenschappers die experimentele observaties probeerden te vangen in wetmatigheden. En als men eenmaal dacht de juiste theorie gevonden te hebben, dan kon deze middels nieuwe experimenten weer op de proef gesteld worden. Zo vond een wisselwerking plaats tussen theorie en experiment. Op deze manier zijn vele beroemde natuurwetten ontstaan: de Navier-Stokes vergelijkingen voor

stromingen die zelfs turbulent kunnen zijn, de wetten van Maxwell voor het verklaren van electromagnetische verschijnselen, en ook bijvoorbeeld de in Nederland ontdekte Korteweg-de Vries vergelijking waarmee zogenaamde ‘solitary waves’ kunnen worden verklaard.

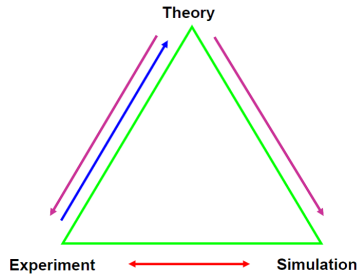
Met de komst van de computer werd een component toegevoegd aan deze eeuwenoude traditie van theorie en experiment: het computer-experiment, ofwel de simulatie. In Figuur 6.6 het omslag van een populair wetenschappelijk tijdschrift dat deze nieuwe trend mooi illustreert. Omdat men theorie en experiment als de 2 basisdisciplines beschouwde, wordt “simulatie” ook wel “de 3^e discipline” genoemd.



Figuur 6.6: Introductie van de 3^e discipline

Figuur 6.7 toont hoe een en ander in de praktijk werkt. Zoals hierboven gezegd was er in eerste instantie enkel een interactie tussen theorie en experiment, waarbij experimenten leiden tot een theorie en die theorie op zijn beurt weer getest kan worden door meer experimenten uit te voeren. Nu komt er simulatie bij als discipline, en kan men met behulp van de ontwikkelde theorie computer-experimenten uitvoeren. Als de theorie klopt, en de rekenmethoden nauwkeurig genoeg zijn, dan zullen deze computer-experimenten net zo nauwkeurig zijn als de fysieke experimenten. Maar een en ander kan veel sneller worden uitgevoerd vaak, er hoeven geen meetopstellingen gemaakt te worden, een druk op de knop is voldoende. De simulaties kunnen op hun beurt weer getoetst worden op daadwerkelijke experimentele resultaten, waarmee een extra check van de theorie plaats kan vinden. De 3^e discipline speelt een belangrijke rol in de hedendaagse wetenschap, en is ook niet meer weg te denken.

Sinds de eeuwwisseling wil men ook nog wel eens spreken van een 4^e disci-



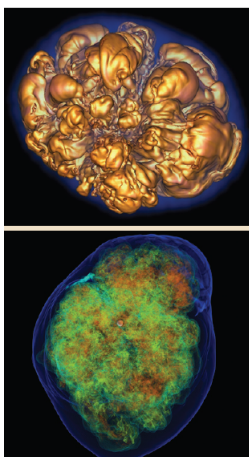
Figuur 6.7: De 3 disciplines theorie, experiment en simulatie

pline, waarmee men “data” bedoelt, en dan met name grote hoeveelheden data. Of, zo men wil, “data science”. Er worden gigantische hoeveelheden data geproduceerd tegenwoordig, vanwege alle sociale netwerken, sensoren die overal in zitten, enzovoorts. Men probeert uit deze grote hoeveelheden data wetmatigheden te destilleren en begrip te kweken over de structuur van deze data. Wij zullen hier verder niet op ingaan, maar het is wel een belangrijk nieuw ingrediënt binnen de wetenschap, waar veel aandacht aan wordt besteed.

6.3 CSE: computational science and engineering

De 3^e discipline heeft er voor gezorgd dat gedurende de afgelopen twee decennia het gebied van de computational science and engineering (CSE) doorgedrongen is zowel in fundamenteel als toegepast onderzoek binnen universiteiten, de industrie en laboratoria, teneinde ontdekkingen te stimuleren, systemen te optimaliseren en beslissingen te ondersteunen. Geinformeerd door eeuwen van theorie en experiment voert CSE computer-experimenten uit om vragen te beantwoorden waar zowel theorie noch experiment alleen een antwoord op kan geven. Op wiskunde gebaseerde geavanceerde rekenmethoden zijn een krachtig middel geworden voor onderzoek en ontdekkingen in vrijwel alle wetenschapsgebieden inclusief ingenieurswetenschappen, technologie en de maatschappij. De CSE gemeenschap zit in de kern van deze transformatie. Zoals in de vorige sectie al aangegeven zijn er nieuwe ontwikkelingen die de scope van CSE voortdurend veranderen, zoals bijvoorbeeld de ‘data revolutie’ en de wens om ‘exascale computing’ (rekenen op topsnelheid) te kunnen doen. Computational science and engineering is dus een erg belangrijk onder-

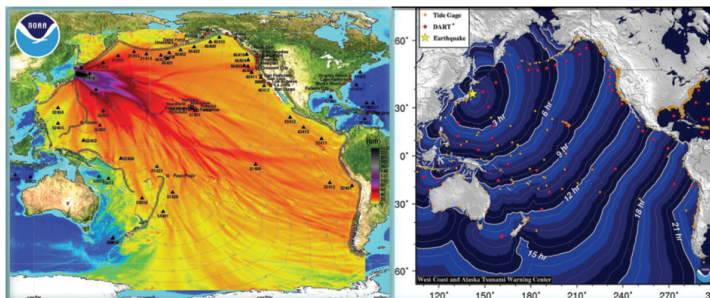
zoeksgebied tegenwoordig. Zoals eerder aangehaald is de SIAM conferentie over CSE uitgegroeid tot hun grootste en meest succesvolle conferentie. Maar niet alleen dat. CSE is ontzettend belangrijk in heel veel verschillende wetenschappelijke disciplines. Overal wordt tegenwoordig 'gerekend' aan allerlei fenomenen, er zijn modellen die de werkelijkheid beschrijven en hiermee kunnen dan vervolgens simulaties worden gedaan. Dit geldt zowel in de industrie als in veel verschillende wetenschappen. Een voorbeeld is het zogenaamde 'folding' (vouwen) en 'unfolding' van proteïnen. Dit leidt op een natuurlijke wijze tot zogenaamde 'multiscale' simulaties, simulaties dus die rekening houden met de fors verschillende tijdschalen waarop biologische processen plaatsvinden. Maar bijvoorbeeld ook de simulatie van het gedrag van supernovas valt hieronder (zie Figuur 6.8).....



Figuur 6.8: Simulatie van het gedrag van supernovas

..... en de simulatie van tsunamis of aardbevingen (zie Figuur 6.9).

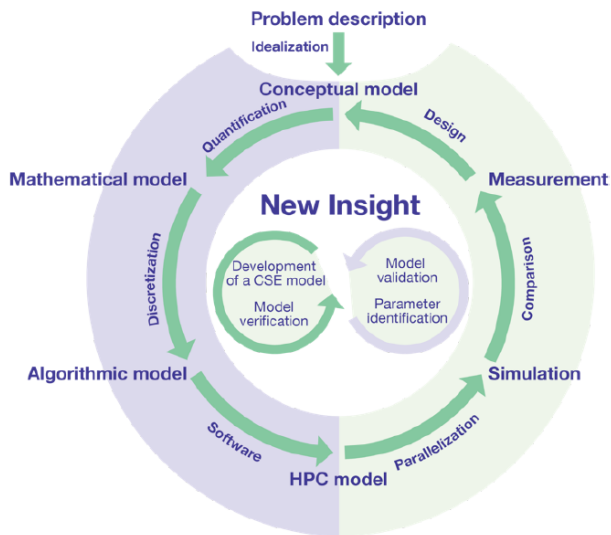
Wat is CSE dan eigenlijk, is de vraag? Dit wordt schematisch weergegeven in figuren 6.10 en 6.11. Computational science and engineering (CSE) is een multidisciplinair gebied van zowel onderzoek als onderwijs dat ligt in de doorsnijding van toegepaste wiskunde, statistiek, informatica, en kerndisciplines uit de wetenschap (Figuur 6.10). CSE bouwt op al deze disciplines, en probeert kennis uit de verschillende gebieden te integreren teneinde nieuwe ideeën te kunnen ontwikkelen op de snijvlakken van deze disciplines. In deze zin is CSE zelf een geheel nieuw gebied, het houdt zich bezig met het ontwikkelen en gebruiken van rekenkundige methoden voor wetenschappelijke ontdekkingen in alle takken van de wetenschap.



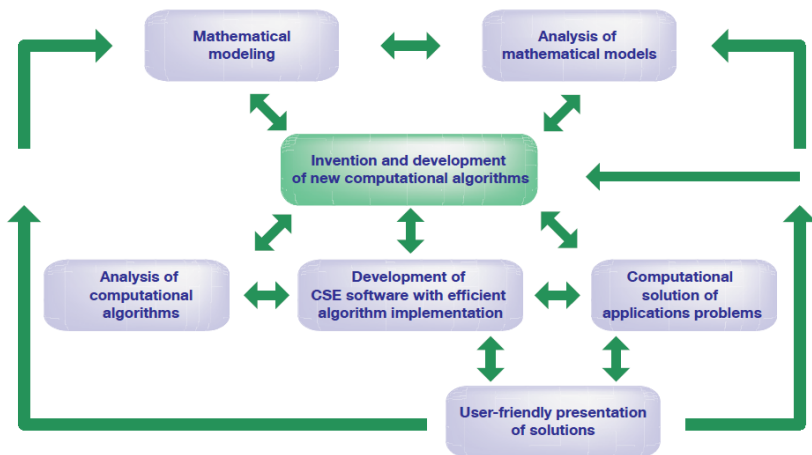
Figuur 6.9: Simulaties van aardbevingen en tsunamis

Het is een breed en zeer belangrijk gebied, waarin ook methoden uit de zogenaamde “high-performance computing” (HPC) een rol spelen. CSE heeft tegenwoordig een eigen agenda welke het nastreeft. Het gebied wordt tegenwoordig alom gerespecteerd als een essentiële bouwsteen voor vooruitgang in wetenschappelijk en technologisch opzicht, samen met de traditionele disciplines van theorie en experiment. Betrouwbare voorspellingen kunnen tegenwoordig vaak enkel gedaan worden middels simulaties, en dat is dan ook wat CSE vooral nastreeft: het kunnen doen van zeer betrouwbare voorspellingen van het gedrag van processen en systemen. We zien daarom overal zogenaamde virtuele ontwerpomgevingen verschijnen. Ontwerp gebeurt tegenwoordig achter het scherm. In de elektronische industrie, bijvoorbeeld, worden nieuwe chips geheel ontworpen achter een beeldscherm. Geavanceerde software zorgt er voor dat de voorspellingen van het gedrag van de chips heel nauwkeurig en betrouwbaar zijn, en dus kunnen ontwerpers volstaan met het uitvoeren van vele simulaties die tenslotte leiden tot het gewenste ontwerp. Het maken van prototypes is niet meer nodig, en dat is maar goed ook, want dat kost door de bank genomen minimaal 3 tot 6 maanden. Zoveel tijd heeft de elektronische industrie niet, zeker niet met de toenemende concurrentie en de grote variëteit van chips nodig om onze maatschappij te voorzien van alle nieuwe technologische snufes en producten. En zo hebben vele takken van de industrie dergelijke virtuele ontwerpomgevingen. Daarnaast zijn er soortgelijke systemen voor beleidsmakers om beslissingen te nemen op grond van vele observaties. Kortom, veel van ons werk speelt zich tegenwoordig af achter een beeldscherm, mogelijk gemaakt door ontwikkelingen binnen het vakgebied CSE.

Figuren 6.10 en 6.11 laten zien wat er allemaal nodig is hiervoor. Belangrijk is dat er een goed model is van het proces waar we berekeningen voor



Figuur 6.10: De CSE pijplijn



Figuur 6.11: Interactie tussen verschillende componenten van CSE

willen uitvoeren. Als we daar een verkeerd model voor gebruiken, of een model dat enkel geldig is onder bepaalde voorwaarden, dan zullen de simu-

laties ook niet een nauwkeurige weergave kunnen geven van de praktijk. Alles hangt dus in eerste instantie af van het achterliggende model. Dat kan een model zijn uit de fysica, de chemie, of een andere wetenschappelijke discipline, vertaald naar een wiskundig model. Vervolgens dienen er berekeningen te worden uitgevoerd, dat betekent vaak een vertaalslag van het oorspronkelijke wiskundige model naar een model dat door computers gehanteerd kan worden. Immers, computers kunnen slechts een eindige hoeveelheid informatie verwerken, terwijl de wiskundige modellen ons in principe in staat stellen om overal waar we zouden willen het probleem op te lossen (op oneindig veel plaatsen dus). De volgende stap is dan het daadwerkelijk simuleren, waarvoor wiskundige algoritmen nodig zijn. En als er dan tenslotte uitkomsten zijn, dan moeten deze aan de gebruiker, de ontwerper dus, getoond worden op een voor hem/haar inzichtelijke manier. Ook dat vergt nogal wat werk, men dient na te denken over welke visualiseringstechnieken geschikt zijn, wat een ontwerper zou willen zien aan afgeleide grootheden. Kortom, het geheel is een ingewikkeld proces waarbij veel verschillende aspecten een rol spelen. De kern van de hele zaak is sterk gerelateerd aan MSO, en dat zullen we dan ook in de volgende sectie uitgebreider bespreken.

CSE is inmiddels verworden tot een niet meer weg te denken vakgebied, essentieel voor onze huidige maatschappij. Er zijn veel succesverhalen, een waarvan we laten zien in Figuur 6.12. Een geweldig mooie verzameling van succesverhalen, vooral van de wiskundige kern binnen CSE, is te vinden in het rapport “Fueling Innovation and Discovery: The Mathematical Sciences in the 21st Century”, te vinden op de volgende link:

https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/NACAD_US/N120711F.pdf.

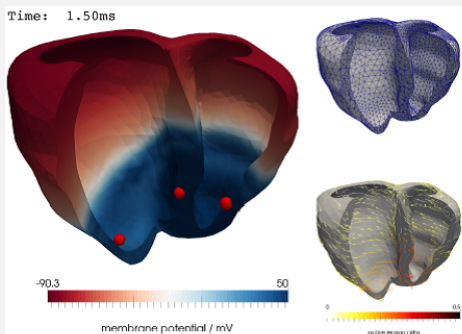
6.4 MSO: mathematical modelling, simulation and optimisation

In Figuur 6.11 zagen we een uitwerking van de ingrediënten van CSE, en daarin speelt wiskunde een uitermate belangrijke rol. Deze rol wordt goed beschreven door het relatief nieuwe acronym “MSO”, dat de hele keten van wiskundige activiteiten binnen CSE samenvat:

- **M:** Eerst moet er een mathematisch model gevormd worden dat de te beschrijven processen of problemen voldoende nauwkeurig beschrijft
- **S:** Vervolgens kunnen er middels dit model simulaties worden uitgevoerd; dit leidt tot een hele keten van wiskundige activiteit, in

Computational medicine has always been at the frontier of CSE: the virtual design and testing of new drugs and therapies accelerate medical progress and reduce cost for development and treatment. For example, CSE researchers have developed elaborate models of the electromechanical activity of the human heart.¹ Such complex processes within the human body lead to elaborate multiscale models. Cardiac function builds on a complicated interplay between different temporal and spatial scales (i.e., body, organ, cellular, and molecular levels), as well as different physical models (i.e., mechanics, electrophysiology, fluid mechanics, and their interaction).

CSE advances in computational medicine are helping, for example, in placing electrodes for pacemakers and studying diseases such as atrial fibrillation. Opportunities abound for next-generation CSE advances: The solution of inverse problems can help identify suitable values for material parameters, for example, to detect scars or infarctions. Using uncertainty quantification, researchers can estimate the influence of varying these parameters or varying geometry.



Figuur 6.12: Een succesverhaal van CSE: berekeningen binnen de medische wetenschappen

vaktermen: vermazen, discretiseren, oplossen van niet-lineaire en lineaire stelsels

- **O:** Tenslotte is er vaak ook behoefte om zaken te optimaliseren, statistische analyses uit te voeren, onzekerheden mee te nemen

Modelleren is een vak apart, en hiervoor is ook veel inzicht nodig uit de betreffende onderliggende wetenschappelijke disciplines. In het verleden waren dat vaak fysische processen, en daarvoor heeft men veel begrip nodig van hoe zaken zich gedragen, hoe natuurverschijnselen werken. Tegenwoordig gebeurt dit niet meer alleen voor fysische processen, maar ook voor bijvoorbeeld chemische en biologische processen. Vaak kost deze modelleerstep erg veel tijd, en even zo vaak is er ook een evolutie in de ontwikkelde modellen. Het hangt er ook helemaal van af wat men precies wil, welke nauwkeurigheid vereist is voor de uit te voeren simulaties. Wil je alle details zien, of volstaat het om een ruw idee te krijgen van het gedrag? Vaak is er wel een model dat in principe de werkelijkheid goed beschrijft, maar waar dan vervolgens nog parameters in zitten die van invloed zijn op het precieze verloop van het proces, of het precieze gedrag.

Een voorbeeld is het gedrag van transistoren, de bouwstenen van elektro-

nische schakelingen en chips. Om het gedrag hiervan te beschrijven wordt vaak gebruik gemaakt van een stelsel van 3 partiële differentiaalvergelijkingen, het zogenaamde ‘drift-diffusion model’: het beschrijft hoe deeltjes bewegen onder invloed van externe krachten (een elektrisch veld), de ‘drift’ component, en onder invloed van elkaar, de diffusie-component. Het model bevat echter een aantal parameters die nog nader moeten worden bepaald, zoals de mobiliteit van de deeltjes (elektronen en zogenaamde ‘gaten’) en de recombinatie (beschrijvend hoe deeltjes botsen en wat er vervolgens gebeurt). Men kan beginnen met constante mobiliteit en nul-recombinatie, en dan al berekeningen uitvoeren waaruit eerste inzichten gedestilleerd kunnen worden. Voor gebruik in industriële toepassingen is dit niet voldoende, en dus zijn er in de loop der jaren geavanceerde mobiliteits- en recombinatiemodellen ontwikkeld welke de werkelijkheid heel nauwkeurig beschrijven. Deze modellen zijn erg complex, en zorgen er mede voor dat allerlei verschijnselen heel nauwkeurig worden beschreven door het model. Voor bepaalde toepassingen, en meer in het algemeen heden ten dage bij de voortschrijdende miniaturisering, blijken er grenzen te zijn aan het drift-diffusie model, en kunnen zelfs de meest geavanceerde modellen voor mobiliteit en recombinatie de processen niet meer nauwkeurig beschrijven. Daarom ziet men de afgelopen jaren vaak een uitbreiding van het drift-diffusie model met een aantal extra differentiaalvergelijkingen, om zodoende ook bijvoorbeeld de effecten van zogenaamde ‘hete elektronen’ te kunnen beschrijven.

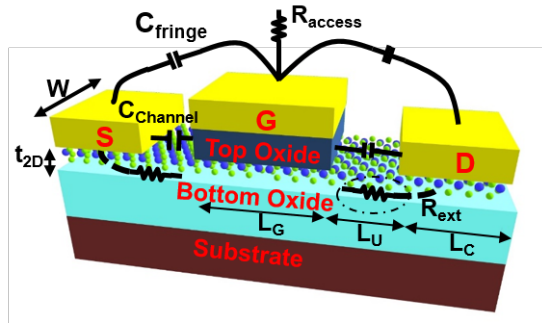
Als men uiteindelijk denkt een voldoende nauwkeurig model te hebben, gestaafd door vergelijking met experimentele resultaten bijvoorbeeld, dan is het proces van modelleren nog niet voorbij. Vaak zijn er bepaalde natuurwetten of eigenschappen van processen die speciale eisen stellen aan het wiskundige model. Zo weten we bijvoorbeeld dat de zogenaamde BH-curve, welke hysteresis-verschijnselen beschrijft, altijd een positief stijgende functie is. Met andere woorden, de afgeleide van de grootte B naar de grootte H moet positief zijn. Heeft ons model die eigenschap niet, dan zullen we een andere BH-curve dienen te modelleren. Ook bij transistoren is er een dergelijk probleem: we kunnen voor het drift-diffusie model wiskundig aantonen dat er oplossingen bestaan als de afgeleide van de mobiliteit naar het elektrische veld positief is. In de loop der jaren werden wel eens ooit mobiliteitsmodellen voorgesteld die niet deze eigenschap hadden, en daar liepen de simulaties op een bepaald moment mis. Dit alles betekent onder andere dat we, nadat het model is vastgesteld, een gedetailleerde wiskundige analyse moeten uitvoeren om te laten zien dat het model inderdaad bepaalde gewenste eigenschappen heeft, zoals existentie van oplossingen, in veel gevallen uniciteit van oplossingen (maar

niet altijd, soms kunnen er wel meerdere oplossingen bestaan), en andere eigenschappen die er vaak voor zorgen dat de simulatie-algoritmen zich op een gunstige manier gedragen. Een voorbeeld van dat laatste is het aantonen dat een bepaalde oplossingscurve convex is, waardoor het vinden van minima gemakkelijk is en ook leidt tot een uniek minimum.

Wiskundig modelleren is een vak apart. Er zijn wel wat boeken over geschreven, maar de algemene consensus is toch dat men het enkel kan leren door het te doen. Daarom organiseren universiteiten, maar ook Europese organisaties als ECMI (<https://ecmi-indmath.org>), zogenaamde modelleer-cursussen en workshops. Onder leiding van ervaren onderzoekers wordt dan voor een aantal problemen wiskundig gemodelleerd, en komen alle bovenstaande aspecten aan de orde. Vaak zie je ook dat de wiskundigen die zich bezig houden met de “M” in MSO, andere mensen zijn dan degenen die zich met de “S” of de “O” bezighouden. Ieder zijn vak, en dat geldt ook voor de differentiatie binnen de wiskunde. Wiskundig modelleren vereist een bepaalde expertise, ook kennis van andere gebieden, en dat is niet alle wiskundigen gegeven.

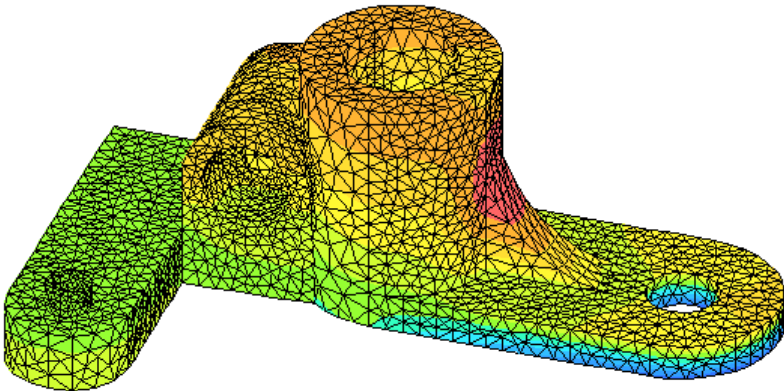
Als het wiskundige model beschikbaar is, dan treedt de “S” van MSO in werking, ofwel we gaan proberen oplossingen te genereren van het wiskundige model. Ook dit is weer een tijdrovende bezigheid, waarin allerlei beslissingen genomen dienen te worden over te gebruiken methoden. En vaak dienen er ook nieuwe algoritmen ontwikkeld te worden, zowel aanpassingen aan bestaande technieken als volledig nieuwe methoden. De eerste stap duiden we vaak aan met ‘discretisatie’, en heeft te maken met de eindige hoeveelheid informatie die een computer kan opslaan. Een wiskundig model beschrijft een bepaald proces vaak op oneindig veel plaatsen. Denk bijvoorbeeld terug aan een transistor, zoals in Figuur 6.13. In principe zouden we op elk punt binnen in de transistor bijvoorbeeld de concentratie van elektronen kunnen berekenen (overigens is dat ook een additioneel voordeel van simulatie boven experiment: bij simulaties kun je overal in het inwendige kijken zonder het gedrag te verstoren, bij experimenten kan dat niet). Maar door de eindigheid van computers dienen we een keuze te maken, en een eindig aantal punten te selecteren waar we de oplossing gaan berekenen. Overigens is er geen man overboord, want middels zogenaamde interpolatie kunnen we toch schatten wat de waarde is op punten die we niet tevoren hebben uitgekozen.

Het proces van discretisatie bestaat vaak uit 2 stappen. De eerste is om een zogenaamde ‘vermazing’ te creëren. Als de structuur geometrisch eenvoudig is, dan is dat redelijk simpel: we maken dan vaak een rechthoekig rooster door horizontaal en verticaal, of in 3D ook nog in de 3^e richting, lijntjes te trekken op gelijke afstanden of wisselende afstanden. Of we de-



Figuur 6.13: Driedimensionaal model van een MOS transistor

len alle rechthoekjes vervolgens ook nog op in driehoekjes. Maar als de geometrie van het te simuleren voorwerp ingewikkelder is, dan wordt vermazen toch een probleem en hebben we gespecialiseerde software nodig om dit voor elkaar te krijgen. En aan de basis hiervan ligt natuurlijk weer een boel mooie wiskunde..... In Figuur 6.14 laten we een voorbeeld zien van zo'n ingewikkelde geometrie met daarin geconstrueerd een vermazing.



Figuur 6.14: Vermazing van een gecompliceerd object

Overigens is vermazing veel meer dan een selectie van punten waarop we de oplossing willen berekenen. Immers, anders zouden we gewoon willekeurig

punten kunnen strooien en zeggen dat we daar de oplossing willen bepalen. Maar verbazing is slechts de eerste stap in het proces voor discretisatie, de tweede stap is zo mogelijk nog belangrijker maar is wel erg gerelateerd aan de stap van verbazing. En dus kunnen we bij verbazing niet zomaar wat doen; vaak is het zaak om de onderliggende geometrische structuur, ook wel ‘het probleemgebied’ genoemd, op te delen in kleine elementjes, vaak rechthoekjes, driehoekjes of vierhoekjes. En vanwege wat we aan eigenschappen nodig hebben later in het simulatieproces is het vaak zo dat deze kleine elementjes aan bepaalde voorwaarden dienen te voldoen. In het geval van driehoekjes wordt vaak verlangd dat deze een zogenaamde Delauney-verbazing vormen, hetgeen betekent dat stomphoekige driehoeken enkel mogen voorkomen als er een driehoekje tegen aan ligt zodanig dat de som van de stompe en de er tegenover liggende scherpe hoek in de buurdriehoek samen maximaal 180 graden bedraagt.

De tweede stap in het discretisatieproces is de feitelijke discretisatie, ofwel het eindig maken van de differentiaalvergelijkingen in het wiskundige model (voor het gemak gaan we er even van uit dat het model gegeven wordt door differentiaalvergelijkingen; dit is de meest gangbare vorm). Meestal betekent dit dat we afgeleiden moeten gaan vervangen door zogenaamde ‘gedeelde differenties’. Dit geldt zowel voor de afgeleiden naar de tijd als afgeleiden in de plaats-richtingen. Een tijdafgeleide wordt dan lokaal vervangen door het verschil van twee functiewaarden, gedeeld door de tijdstap. Feitelijk is dit dus terug naar hoe we ooit afgeleiden uitgelegd kregen: eerst neem je verschillen van functiewaarden gedeeld door een eindige tijdstap, en vervolgens laat je die tijdstap naar nul gaan. Nu willen we dus juist dat die tijdstap ongelijk is aan 0, en een eindige waarde heeft. Voor de plaatsafgeleiden wordt een dergelijk proces uitgevoerd. Er zijn ook meer ingewikkelde methoden, waarin het berekenen van de afgeleiden minder expliciet gaat als in het voorgaande beschreven, en impliciet worden gevormd. Dit is bijvoorbeeld zo bij de zogenaamde eindige elementen methode (‘finite element method’). Deze methode maakt juist expliciet gebruik van de kleine elementen die we geformeerd hebben bij de verbazingsstap. Het voert te ver om hier deze methoden verder uit te gaan leggen, er is veel literatuur over te vinden voor geïnteresseerden. Belangrijk is om zich te realiseren dat uiteindelijk, nadat de discretisatie is voltooid, er een groot stelsel vergelijkingen is ontstaan, meestal met als onbekenden de waarden van de oplossing op de hoekpunten van de elementjes.

De discretisatiestap vormt dus het oorspronkelijke wiskundige model om in een discreet model dat door de computer kan worden behapt. Het is vaak een groot systeem van zogenaamde algebraïsche vergelijkingen met

even zoveel onbekenden, namelijk de waarden van de oplossing. Helaas zijn deze vergelijkingen in veel gevallen niet-lineair in de onbekenden, en dus zal er een oplosproces voor niet-lineaire stelsels van stal gehaald dienen te worden. In veel gevallen is dit de methode van Newton, maar er zijn ook andere mogelijkheden. Ook hier verwijzen we naar de omvangrijke literatuur. De methode van Newton is aantrekkelijk, omdat deze kwadratische convergentie vertoont: het is een iteratief proces, waarin aan het einde van het proces de afstand tot de werkelijke oplossing kwadratisch naar beneden gaat.

De methode van Newton bestaat dus uit een eindig aantal iteraties, waarin we hopen steeds dichterbij de oplossing te komen. In elke stap moet een lineair stelsel worden opgelost, immers, de methode van Newton betekent feitelijk dat men het stelsel vergelijkingen ('F') lineariseert rondom een bepaalde tussenoplossing x :

$$F(x + d) = F(x) + dF'(x) + O(d^2) \quad (6.1)$$

Het is duidelijk dat het hier om zeer grote lineaire stelsels gaat, vaak met honderdduizenden of zelfs miljoenen vergelijkingen en onbekenden. Het oplossen daarvan is een kunst op zich, ofwel een vakgebied binnen de wiskunde. Feitelijk kun je het vergelijken met het bekende puzzeltje dat vaak in familiekring wordt gebezigd: "Vader is momenteel $4x$ zo oud als Jantje, over 5 jaar is vader $3x$ zo oud als Jan(tje). Hoe oud is Jantje nu." Wiskundig onderlegde familieleden schrijven natuurlijk meteen de vergelijkingen op:

$$\begin{aligned} v - 4j &= 0 \\ (v + 5) - 3(j + 5) &= 0 \end{aligned}$$

Anders opgeschreven:

$$\begin{aligned} v - 4j &= 0 \\ v - 3j &= 10 \end{aligned}$$

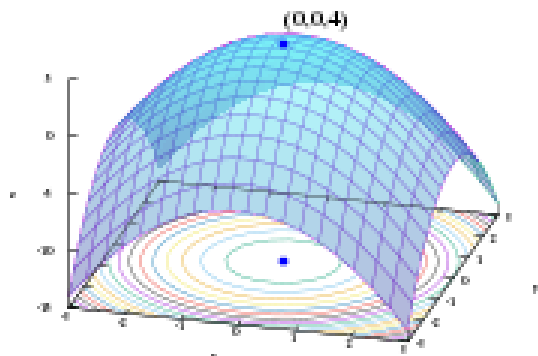
Door eenvoudige eliminatie van v vinden we dan dat $j=10$. Maar stel je nu eens voor dat je van alle 17 miljoen Nederlanders informatie opschrijft over relaties tussen hun leeftijden, en dan moet proberen om de leeftijden van allen te achterhalen. In principe is dat mogelijk door variabelen te elimineren, maar dat zou tot een zeer tijdrovend proces leiden. Je moet het dus handig en slim aanpakken, en daartoe zijn allerlei methoden ontwikkeld. Bekend is de methode van Gauss eliminatie, en deze ligt in het verlengde van de eliminatie die we hierboven deden. De methode behoort tot de categorie "directe methoden", methoden welke met een van tevoren bekend aantal rekenkundige operaties tot een oplossing komt. Er zijn

allerlei subtiliteiten, zoals het veranderen van de volgorde van vergelijkingen, maar dat laten we even buiten beschouwing.

Naast de directe oplosmethoden zijn er ook de tegenwoordig zeer populaire iteratieve methoden, die beginnen met een schatting van de oplossing (dit kan een zeer slechte schatting zijn, of bijvoorbeeld een startoplossing bestaande uit enkel nullen), en deze vervolgens iteratief verbeteren. Deze methoden zijn populair omdat er vaak veel minder operaties nodig zijn om tot een goede benadering van de oplossing van het lineaire stelsel te komen. Bekend is de methode van geconjugeerde gradiënten ('conjugate gradient method'), welke al in 1952 werd voorgesteld door de heren Hestenes en Stiefel, maar na aanvankelijk enthousiasme toch in het verdomhoekje raakte. Totdat in 1977 de Nederlanders Henk van der Vorst en Koos Meijerink de populariteit van de methode nieuw leven inbliezen door er een cruciaal element aan toe te voegen: preconditionering. Dit behelst het vermenigvuldigen van de matrixvergelijking met een andere matrix, die enigszins lijkt op de inverse van de stelselmatrix. Hiermee worden allerlei zeer mooie eigenschappen verkregen en loopt het aantal iteraties fors terug. De methode staat bekend onder de naam ICCG ('incomplete Choleski conjugate gradients'), en is feitelijk een combinatie van een iteratieve en directe oplosmethode. ICCG wordt wereldwijd in heel veel simulatiesoftware gebruikt. Na 1977 is erg veel onderzoek gedaan naar iteratieve methoden en preconditioneringen, en zijn methoden als GMRES, biCGstab en CGS ontwikkeld. Het is een bloeiend onderzoeksterrein, en er zijn zelfs speciale conferenties welke zich concentreren op het onderwerp 'preconditioning', zie bijvoorbeeld <http://www.win.tue.nl/precon2015/> voor informatie over de Eindhovense editie uit 2015, en <http://www.cs.ubc.ca/~greif/precon17/> voor de recente editie in Vancouver.

Het Newtonproces, inclusief de per iteratie benodigde oplossingen van de lineaire stelsels, moet dan uiteindelijk leiden tot convergentie en dat levert dan de oplossing op van het discrete wiskundige model. Dit is echter niet de oplossing van het oorspronkelijke wiskundige model, maar slechts een benadering. Onderweg zijn allerlei foutjes gemaakt, die hopelijk binnen gestelde marges blijven: de discretisatie levert fouten op, en ook de oplossing van de niet-lineaire en lineaire stelsels levert bijdragen aan de approximatiefout. Ook kan het zo zijn dat de data, of waarden van parameters, uit het oorspronkelijke model niet helemaal nauwkeurig bekend zijn. Alles bij elkaar dienen we dus altijd kritisch te blijven over de gevonden oplossingen, het zou kunnen zijn dat we ergens onderweg onacceptabele fouten hebben gemaakt. Daar dient dus nauwkeurig naar gekeken te worden, en ook dit is de taak van een wiskundige: wat kunnen we uiteindelijk zeggen over de

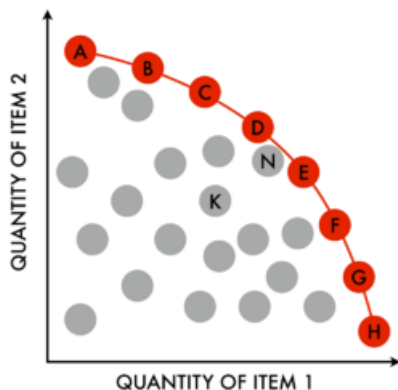
nauwkeurigheid van gevonden oplossingen. Hierover verwacht de gebruiker wel een uitspraak, hij/zij wil weten wat de waarde is van de berekeningen. Al met al behelst de “S” van simulatie toch wel het een en ander. Allerlei disciplines in de wiskunde zijn in dit proces betrokken. Dat geldt ook voor de laatste letter in MSO, de “O” van optimalisatie. Ontwerpers willen vaak het onderste uit de kan, en dat betekent dat er geoptimaliseerd dient te worden. Vraag is dan allereerst natuurlijk: welke eigenschap dient er geoptimaliseerd te worden? Wiskundig geformuleerd betekent het dat we een kostenfunctie moeten opstellen welke we vervolgens willen minimaliseren. Als deze kostenfunctie mooie eigenschappen heeft, zoals in Figuur 6.15, dan is een extremum prima te bepalen (in dit geval een maximum). Maar soms is de kostenfunctie erg grillig, heeft allerlei lokale extrema, en dus moet er van alles uit de kast gehaald worden om het optimalisatieproces goed te laten verlopen.



Figuur 6.15: Voorbeeld van een kostenfunctie

Ook deze stap binnen MSO kent verschillende kanten. Soms is er geen eenduidig optimum te vinden, en zal men moeten kiezen wat men het belangrijkste vindt. In dat geval wordt vaak gekozen voor zogenaamde Pareto optimalisatie. Figuur 6.16 illustreert dit voor het geval van 2 parameters. Afhankelijk van wat je wilt zou je de punten A-H kunnen kiezen als optimum. Bij Pareto wordt dan meestal een punt rond D-E gekozen, omdat je dan voor beide grootheden in ieder geval gelijkelijk verdeeld een opti-

imum vindt. Maar je kunt dus ook de meer extreme optima A of H kiezen, bijvoorbeeld als economische motieven belangrijker zijn dan andere motieven.



Figuur 6.16: Voorbeeld van Pareto optimaliteit

Binnen MSO wordt ook vaak nog met andere aspecten rekening gehouden, zoals eerder aangegeven. Vanwege allerlei onzekerheden in bijvoorbeeld het productieproces dienen we rekening te houden met onzekerheden in de parameters van het model. Je kunt wel harde waarden aannemen voor bijvoorbeeld diktes en breedtes, maar als het product wordt gemaakt dan zullen de daadwerkelijke maten afwijken van de ideale maten (denk aan het ‘e’ kenmerk op allerlei producten die we kopen). En dus is het goed om met deze onzekerheden rekening te houden. Het relatief nieuwe vakgebied ‘uncertainty quantification’ staat dan ook volop in de belangstelling. Daarnaast worden statistische analyses gedaan, wordt er gekeken naar sensitiviteit van bepaald parameters, enzovoorts. Kortom, MSO is een heel erg uitgebreid gebied. En dan hebben we het hierboven enkel gehad over traditionele simulaties van fysische of aanverwante processen. Maar er zijn ook totaal andere processen, zoals bijvoorbeeld wachttijden in een ziekenhuis. Ook dat wordt middels wiskundige modellen beschreven, en vervolgens proberen we ook daarin te optimaliseren, namelijk om de wachttijden zo kort mogelijk te maken. Het vakgebied van de ‘operations research’ komt hier om de hoek kijken, met ingrediënten als LP en MIP (zie figuren 6.3 en 6.4).

6.5 Conclusie

Het is duidelijk dat de komst van computers onze wereld geheel heeft veranderd, en ook de wiskundige wereld. Numerieke wiskunde is een nog relatief jong vak, en ontstond toen er met computers berekeningen gedaan konden worden, begin jaren 50 van de vorige eeuw. Tegenwoordig duidt men het vakgebied ook wel aan met ‘scientific computing’, niet te verwarren met CSE. Er ontstaat dus nieuwe wiskunde, moderne wiskunde, ingegeven door ontwikkelingen om ons heen. Dit stelt vervolgens ook hoge eisen aan wiskundigen, zij dienen flexibel te zijn en nieuwe wiskunde te ontwikkelen om in te kunnen springen op de behoeften van aanpalende wetenschappen, van de industrie en van de maatschappij. Computational Science and Engineering (CSE) neemt een grote vlucht en is niet meer weg te denken uit allerlei disciplines, het is DE manier van onderzoek doen tegenwoordig. Wiskundig modelleren, simuleren en optimaliseren, ofwel MSO, is een sleuteldiscipline in dit geheel. Feitelijk omvat MSO het gehele wiskundige bouwwerk, uit allerlei wiskundige deelgebieden zijn methoden nodig om tot de gewenste oplossingen te komen. Natuurlijk spelen ook computers, informatica en software een grote rol in dit geheel, maar de basis is toch gelegen in de wiskunde. En in die zin is wiskunde het breekijzer waarmee we allerlei problemen kunnen aanpakken.



Voor wie is PWN interessant?

Beroepswiskundigen

Wiskundeleraren

Bedrijven

Leerlingen en studenten

Breed publiek

Platform Wiskunde Nederland is hét landelijke loket voor alles wat met wiskunde te maken heeft.

PWN behartigt de belangen van, en fungeert als spreekbuis voor, de gehele Nederlandse wiskunde.

Platform Wiskunde Nederland | Science Park 123 | kamer L013 | 1098 XG Amsterdam | 020 592 40 06

Ga voor meer informatie naar:
www.platformwiskunde.nl



platform
wiskunde nederland