CWI Syllabi

CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Telephone + 31 - 20 592 9333
Telefax + 31 - 20 592 4199
WWW page `http://www.cwi.nl`

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

Vakantiecursus 1999
Onbewezen vermoedens

45

Centrum voor Wiskunde en Informatica
CWI SYLLABUS

ISBN 90 6196 485 7
NUGI-code: 811

Ten Geleide

Het bewijs van De Laatste Stelling van Fermat door de Britse wiskundige Andrew Wiles is zonder enige twijfel de meest opzienbarende mathematische gebeurtenis van de jaren negentig geweest. De dramatische ontstaansgeschiedenis ervan, vol spanning en sensatie, die in 1995 uitmondde in de publikatie van Wiles bewijs in de *Annals of Mathematics*, is al vele malen verteld. Al die publiciteit heeft ook de schijnwerpers gericht op de belangrijke rol die vermoedens in de wiskunde spelen, en velen hebben zich inmiddels afgevraagd wat het volgende grote vermoeden zal zijn dat bedwongen gaat worden. Dat het daarbij snel kan gaan, bleek in 1998, toen Tom Hales bekendmaakte dat hij het *vermoeden van Kepler* bewezen had, dat neerkomt op de bewering dat de bekende kanonskogelstapeling de dichtst mogelijke bolstapeling in de ruimte is.

Het vermoeden van Fermat en het vermoeden van Kepler hebben gemeen dat de probleemstelling ervan (maar niet het bewijs!) vrij gemakkelijk uit te leggen is aan de geïnteresseerde leek. Dat geldt ook voor een aantal andere beroemde vermoedens die nog niet zijn opgelost, met name uit de getallentheorie. Maar andere vermoedens, die een minstens zo rijke geschiedenis hebben en die op wiskundigen een minstens zo grote aantrekkingskracht uitoefenen, zijn veel moeilijker aan buitenstaanders duidelijk te maken. Ook veel beroepswiskundigen zullen dikwijls niet direct een helder beeld hebben van de draagwijdte ervan.

Dit was voor de programmacommissie een van de redenen om voor de Vacantiecursus van 1999 als thema *Onbewezen vermoedens* te kiezen. Het is niet ondenkbaar, ja, gezien de enorme toename van de hoeveelheid wiskundig research in de afgelopen decennia zelfs tamelijk waarschijnlijk, dat we de komende jaren bewijzen zullen zien verschijnen van andere beroemde vermoedens. Het is dan alleszins gewenst dat elke wiskundige daar over mee kan praten. Maar belangrijker nog is het feit dat veel van die vermoedens laten zien dat de frontlijn van het wiskundig onderzoek toch altijd weer verbonden kan worden met eenvoudige resultaten en onderzoeksvragen, die alle wiskundig geïnteresseerden aan zullen spreken. De sprekers van deze Vacantiecursus hebben zich stuk voor stuk de taak gesteld om u niet alleen kennis te laten maken met een aantal beroemde Onbewezen Vermoedens, die men met recht kroonjuwelen van de wiskunde mag noemen, maar u ook de meer elementaire, maar daarom niet minder flonkerende edelstenen te tonen die in die gebieden ook voor de beginner al voor het oprapen liggen.

Gaarne wil ik deze inleiding besluiten met een woord van dank aan allen die hebben bijgedragen aan het welslagen van de Vacantiecursus 1999, waar-

van deze Syllabus de teksten bevat. In de eerste plaats natuurlijk de sprekers, die naast hun lezing ook een schriftelijke neerslag ervan hebben aangeleverd, waardoor deelnemers en andere belangstellenden een rijke bundel aan hun bibliotheek kunnen toevoegen. Het Centrum voor Wiskunde en Informatica te Amsterdam en de Technische Universiteit Eindhoven stelden zaalruimte beschikbaar, de administratieve en praktische organisatie was in handen van mw. Simone Panka en dr. Miente Bakker, die ook de inhoudelijke coördinatie van de Syllabus verzorgde.

Allen hartelijk dank!

Jan van de Craats

Vakantiecursus 1999

Inhoud

Vakantiecursus 1999 pp. 1 – 15

Over de Rol van Vermoedens in de Wiskunde

J. van de Craats
KMA, Postbus 90154, 4800 RG Breda
e-mail: jcr@euronet.nl

1. Inleiding

Veel beroemde stellingen in de wiskunde zijn hun bestaan als een vermoeden begonnen. Zonder de waarheid veel geweld aan te doen zou je de geschiedenis van de wiskunde zelfs in grote lijnen kunnen beschrijven als een geschiedenis van vermoedens en bewijzen. Kenmerkend voor de wiskunde is dat er een strikte scheiding bestaat tussen onbewezen vermoedens en bewezen stellingen. In principe is er over de status van belangrijke uitspraken in de wiskunde eigenlijk nooit discussie: is het een onbewezen uitspraak, dan is het een vermoeden, is het bewijs geleverd, dan is het vermoeden een stelling geworden. De enige slag die we om de arm moeten houden, is dat het controleren van een bewijsclaim een tijdrovende zaak kan zijn, waarbij een enkele keer wel eens details over het hoofd worden gezien. Maar in principe – ik herhaal die woorden nog maar eens voor de zekerheid – kan elke bekwame wiskundige elk bewijs verifiëren. Sommige wiskundigen zijn zelfs van mening dat die verificatie aan een computer zou kunnen worden overgelaten; het project Automath van N.G. de Bruijn is op deze overtuiging gestoeld.

Bij de andere wetenschappen spelen vermoedens en bewijzen een geheel andere rol. Meestal is er daar sprake van een glijdende schaal die loopt van wilde hypothesen tot algemeen aanvaarde theorieën. Bewijzen zijn daar niet meer dan argumenten die een theorie of een vermoeden in zekere mate ondersteunen. In de natuurwetenschappen toetst men een hypothese door experimenten uit te voeren en te bepalen of de uitslag ervan met die hypothese verenigbaar is. Uiteindelijk heeft het experiment daar het laatste woord. Bewijzen in de wiskunde daarentegen zijn absoluut: een bewijs van een stelling kan niet door de uitslag van een experiment onderuit worden gehaald. In geen enkele andere wetenschap is de zekerheid bereikbaar die een wiskundig bewijs verschaft. Wiskunde is dan ook geen experimentele wetenschap, maar een vrije schepping van de menselijke geest.

Toch bestaat er wel zoiets als experimentele wiskunde. De wiskunde ontwikkelt zich immers niet los van de werkelijkheid; ze laat zich voortdurend door haar inspireren. Getallen en getallenpatronen komen voort uit structuren en patronen die we in de werkelijkheid waarnemen. Hetzelfde geldt voor de meetkunde, de analyse, de topologie, de grafentheorie, de kansrekening, ja zelfs voor wezenlijk abstractere vakken als de algebra en de logica. Experimentele wiskunde kan nuttig zijn als middel om structuren en patronen te ontdekken en te onderzoeken. Als je speciale gevallen tekent of doorrekent, kun je vermoedens

testen of op het spoor komen. De computer speelt daarbij tegenwoordig een belangrijke rol.

2. Het $3n+1$ vermoeden

Een mooi voorbeeld hiervan is het nog steeds niet opgeloste $3n+1$ vermoeden, een vermoeden dat sinds de jaren vijftig onder tal van verschillende namen de ronde doet: het wordt ook wel het *Collatz-probleem*, het *Kakutani-probleem*, of het *Syracuse-probleem* genoemd. Het luidt als volgt. Begin met een willekeurig positief geheel getal n. Pas hierop de volgende transformatie $T(n)$ toe:

$$T(n) = \begin{cases} 3n+1 & \text{als } n \text{ oneven is} \\ n/2 & \text{als } n \text{ even is} \end{cases}$$

en blijf dit herhalen, zodat een rij

$$n,\ T(n),\ T^2(n) = T(T(n)),\ T^3(n) = T(T(T(n))), \ldots$$

ontstaat. Zodra de rij het getal 1 bereikt, wordt ze periodiek:

$$\ldots, 1, 4, 2, 1, 4, 2, 1, \ldots.$$

Het $3n+1$ vermoeden luidt dat dit voor iedere n na eindig veel stappen zal gebeuren, met andere woorden dat er bij elke n een $k = k(n)$ bestaat waarvoor $T^k(n) = 1$. De numerieke evidentie voor de juistheid van dit vermoeden is groot: in het boek *Getaltheorie voor beginners* van Frits Beukers lees ik dat het vermoeden in 1998 al geverifieerd was voor alle $n < 3.2 \times 10^{16}$, en inmiddels zal men waarschijnlijk al weer veel verder zijn.

Maar hoe veel waarden van n we ook testen, zolang we geen tegenvoorbeeld vinden, helpt de computer ons op die manier niet verder. Als het vermoeden juist is, zullen we er een bewijs voor willen vinden; numerieke evidentie is niet voldoende. Als waarschuwing kan dienen dat het overeenkomstige $5n+1$ vermoeden niet juist is: naast de cyclus

$$\ldots, 1, 6, 3, 16, 8, 4, 2, 1, \ldots$$

is er ook nog minstens één andere cyclus, namelijk

$$\ldots, 13, 66, 33, 166, 83, 416, 208, 104, 52, 26, 13, \ldots$$

Bij het $7n+1$ probleem is de situatie nog verwarrender: weliswaar levert 1 een eindige cyclus op, namelijk $1, 8, 4, 2, 1$, maar wat er met de geïtereerden van 3 gebeurt, is niet bekend: er zijn in de rij die met 3 begint al getallen gevonden van meer dan negenduizend cijfers!

Is het $3n+1$ vermoeden een belangrijk vermoeden in de wiskunde of is het alleen maar een curiositeit? Dat is moeilijk te zeggen. Het belang van een vermoeden zou men misschien willen afmeten aan de praktische toepasbaarheid ervan. In dat geval zijn we gauw uitgepraat: voor zover bekend is die er niet. Maar in de wiskunde wordt het belang niet in de eerste plaats bepaald door praktische toepassingen, maar door de vraag in hoeverre zo'n vermoeden

verbonden is met andere, substantiële delen van de wiskunde. Het gaat dus om de vraag hoe centraal de plaats is die zo'n vermoeden binnen de wiskunde inneemt. In dat opzicht is de kort geleden bewezen Laatste Stelling van Fermat van enorm belang gebleken: in zijn meer dan 350 jaar oude geschiedenis heeft ze meer dan eens geleid tot de ontwikkeling van nieuwe en veelbetekenende theorieën in de algebra, de getallentheorie en de algebraïsche meetkunde. Andrew Wiles wist de stelling uiteindelijk te bewijzen als een speciaal geval van een veel algemener vermoeden, dat van Taniyama en Shimura, dat twee op het eerste gezicht totaal verschillende terreinen binnen de wiskunde met elkaar verbindt: de theorie van de modulaire vormen en de theorie van de elliptische functies. De spannende geschiedenis ervan is de afgelopen jaren al herhaaldelijk beschreven. Voor de beginner is de beste gids het boek *Het laatste raadsel van Fermat* van Simon Singh.

Zo'n centrale plaats als de Laatste Stelling van Fermat neemt het $3n + 1$ vermoeden in de wiskunde thans zeker nog niet in. Wel zijn er deelresultaten en bewijzen bekend onder aanname van zekere plausibel lijkende andere vermoedens, maar, in de woorden van Paul Erdös, de huidige wiskunde lijkt nog niet klaar te zijn om dit soort problemen aan te pakken.

3. Niet-euclidische meetkunde

Niet altijd wordt een belangrijk vermoeden met een bewijs bekroond. De geschiedenis van de wiskunde telt heel wat vermoedens die onjuist zijn gebleken, maar waarbij juist de ontdekking van de ongeldigheid ervan voor de wiskunde van groot belang is geweest. De oude Griekse meetkunde biedt een groot aantal voorbeelden. Zo is er de geschiedenis van het beroemde *parallellenaxioma* uit de Elementen van Euclides. Zoals bekend trachtte Euclides de meetkunde op strikt logische wijze af te leiden uit een beperkt aantal 'vanzelfsprekende' basisstellingen, de axioma's. Het volgende axioma nam daarbij een bijzondere plaats in:

> Als twee lijnen een derde lijn snijden, en de binnenhoeken aan één kant van die lijn zijn samen minder dan 180 graden, dan snijden die twee lijnen elkaar aan diezelfde kant van de derde lijn.

Al in de tijd van Euclides was men gefrappeerd door het, in vergelijking met de andere axioma's, gecompliceerde karakter van dit axioma. Het leek veel meer de gedaante te hebben van een stelling die uit de andere axioma's kan worden afgeleid. Het vermoeden dat dit parallellenaxioma een uit de andere axioma's afleidbare stelling is, heeft tot in de negentiende eeuw tal van wiskundigen beziggehouden – zonder succes. Dat wil zeggen, zonder dat men zo'n afleiding vond. Wel werden allerlei andere, equivalente formuleringen gevonden, zoals bijvoorbeeld

> Door een gegeven punt buiten een gegeven lijn gaat precies één lijn die de gegeven lijn niet snijdt.

of

De som van de hoeken van elke driehoek is gelijk aan 180 graden.

maar het afleiden van deze 'evidente' stellingen uit de overige axioma's leek net zo'n onmogelijke opgave te zijn als het bewijzen van het parallellenaxioma.

Een andere, veelbelovende aanpak leek die van het bewijs uit het ongerijmde: ga ervan uit dat het parallellenaxioma niet geldt, en probeer daaruit (en uit de overige axioma's) een ongerijmdheid af te leiden. Dat leidde onder andere tot de volgende uitspraken: als het parallellenaxioma niet geldt, dan

- is de som van de hoeken in elke driehoek kleiner dan 180 graden,
- bestaan er geen gelijkvormige niet-congruente figuren,
- bestaan er geen vierkanten of rechthoeken,
- geldt de Stelling van Pythagoras niet,
- bestaat er een absolute lengtemaat.

Hoezeer deze uitspraken ook in tegenspraak lijken met de 'dagelijkse ervaring' of de 'meetkundige intuïtie', het afleiden ervan uit de overige axioma's lukte niet, en daarmee mislukte ook het bewijs uit het ongerijmde.

Pas in het begin van de negentiende eeuw kwamen, onafhankelijk van elkaar, C.F. Gauss (1777-1855), J. Bolyai (1802-1860) en N.I. Lobachevsky (1793-1856) tot de overtuiging dat die bewijspogingen wel spaak móesten lopen, omdat de ontkenning van het parallellenaxioma helemaal niet leidt tot een tegenspraak, maar tot een andere, *niet-euclidische* meetkunde, die vanuit een logisch standpunt bekeken net zo veel bestaansrecht heeft als de gewone meetkunde van Euclides. In de vakantiecursus van vorig jaar heeft prof.dr. F. van der Blij hieraan een voordracht gewijd.

De ontdekking van de niet-euclidische meetkunde heeft overigens veel bijgedragen aan de opheldering van filosofische vragen omtrent de 'ware' meetkunde van de ons omringende werkelijkheid. Voor de oude Grieken was de euclidische meetkunde de beschrijving van de geïdealiseerde werkelijkheid. Geïdealiseerd, want meetkundige punten hebben geen afmetingen en meetkundige lijnen en vlakken hebben geen dikte. Toch waren de meetkundige punten, lijnen en vlakken in hun visie noodzakelijkerwijze een getrouwe afspiegeling van realiteit, de wereld om ons heen. Met de ontdekking van de niet-euclidische meetkunde werd het echter duidelijk dat de euclidische meetkunde slechts één van de oneindig veel andere mogelijke meetkundige modellen voor de werkelijkheid is. Later zou B. Riemann (1826-1866) het scala aan meetkundige modellen nog verder vergroten, waardoor het duidelijk werd dat ook de oude vertrouwde bolmeetkunde, naast de meetkunde van Gauss, Bolyai en Lobachevsky, als een niet-euclidische meetkundevorm kan worden gezien.

Voor dagelijks gebruik is de euclidische meetkunde natuurlijk nog steeds de eenvoudigste beschrijvingsvorm van de ons omringende ruimte – wat zouden we immers moeten beginnen zonder vierkante tegelvloeren? – maar op kosmologisch of subatomair niveau zijn andere modellen meer aangewezen. En de vraag welke meetkunde nu 'echt' de meetkunde van de werkelijkheid beschrijft, is inmiddels door veel wiskundigen en fysici als zinloos terzijde geschoven.

4. Passer-en-liniaalconstructies

In een *tour d'horizon* met als thema vermoedens uit de geschiedenis van de wiskunde mogen de klassieke Griekse problemen rond passer-en-liniaalconstructies niet ontbreken. Ook hier geldt weer dat ze hun belang niet ontlenen aan hun praktische toepasbaarheid, maar aan hun stimulerende invloed op de ontwikkeling van de wiskunde, met name die in de negentiende eeuw.

Zoals bekend, hanteerden de oude Grieken strikte regels omtrent het gebruik van passer en liniaal. Of liever gezegd, zij waren de eersten die zich afvroegen hoe ver je kunt komen als je jezelf zulke strikte regels oplegt. Ook voor de Grieken ging het daarbij niet om het oplossen van praktische vraagstukken, maar om een spel, dat volgens strikte regels gespeeld moest worden. Hun meetkunde-spel kent een geïdealiseerde liniaal en een geïdealiseerde passer, waarmee men uitgaande van een gegeven, uit punten, lijnen en cirkelbogen samengestelde vlakke meetkundige figuur, nieuwe punten, lijnen en cirkelbogen mag construeren. Daarbij gelden de volgende spelregels:

- Twee gegeven of geconstrueerde punten mogen door een lijn verbonden worden.
- Bij een gegeven of geconstrueerd punt M en twee gegeven of geconstrueerde punten P en Q mag men een cirkelboog tekenen met M als middelpunt en de afstand PQ als straal.
- Bij twee gegeven of geconstrueerde elkaar snijdende lijnen of cirkels mag men de snijpunten als geconstrueerd beschouwen.

Tal van meetkundige constructies kan men uitvoeren via een eindig aantal van zulke stappen, bijvoorbeeld het verdelen van een gegeven lijnstuk in een willekeurig aantal gelijke delen, het oprichten of neerlaten van loodlijnen vanuit een punt op een lijn, het halveren van een gegeven hoek, het construeren van een gelijkzijdige driehoek, vierkant of regelmatige vijfhoek met een gegeven zijdelengte. Maar er waren ook constructieopgaven die niet lukten, zoals de trisectie (het in drie gelijke delen verdelen) van een willekeurige hoek, de 'kubusverdubbeling', dat wil zeggen de constructie van de ribbenlengte van een kubus met een inhoud die twee maal zo groot is als die van een gegeven kubus, of de 'kwadratuur van de cirkel', dat wil zeggen het bepalen van de zijdelengte van een vierkant dat dezelfde oppervlakte heeft als die van een cirkel met een gegeven straal. Ook de constructie van regelmatige n-hoeken voor $n > 5$ gaf problemen, bijvoorbeeld voor $n = 7$ of $n = 9$.

Nogmaals zij erop gewezen dat dit geen praktische problemen waren, maar opgaven binnen de context van een meetkundig spel met strikte regels. Door versoepeling van de regels verdwijnen de problemen als sneeuw voor de zon. Zo wisten de oude Grieken bijvoorbeeld al dat de trisectie van de hoek gemakkelijk is wanneer men de liniaal van slechts twee merktekens mag voorzien. Maar dat was vals spelen.

Wat zijn bij deze problemen nu eigenlijk de vermoedens geweest? Het vermoeden dat zo'n constructie wel degelijk mogelijk was? Misschien, maar waarschijnlijk hebben ook in de Griekse oudheid al velen het idee gehad dat deze

constructieopgaven onuitvoerbaar zijn; zelfs in onze tijd is 'de kwadratuur van de cirkel' immers een metafoor voor een onmogelijke opgave. Toch blijken ook thans nog veel leken en amateurs niet te kunnen accepteren dat bepaalde constructies echt onmogelijk zijn. Meestal begint het er al mee dat zij de spelregels niet kennen, of zich zelfs helemaal niet realiseren dat het om een spel met strikte regels gaat. Vaak menen zij ook dat het om een brandende praktische kwestie gaat, en komen zij met benaderende oplossingen. En meestal reageren ze heel lakoniek op de mededeling dat in de negentiende eeuw reeds bewezen is dat die problemen onoplosbaar zijn. Zij hebben immers wel degelijk een oplossing gevonden, en dus moeten al die kamergeleerden er wel naast zitten!

Er kwam pas klaarheid in de passer-en-liniaalproblematiek toen men zich begon te realiseren dat zulke constructieopgaven corresponderen met algebraïsche vraagstukken. Gauss ontdekte al op 18-jarige leeftijd met algebraïsche middelen dat er een passer-en-liniaalconstructie bestaat voor elke regelmatige p-hoek waarvoor p een *Fermat-priemgetal* is, dat wil zeggen een priemgetal van de vorm $p = 2^m + 1$, waarbij $m = 2^k$ een macht is van 2. In het bijzonder zijn er dus constructies voor de regelmatige driehoek ($k = 0$) en de regelmatige vijfhoek ($k = 1$), zoals ook de oude Grieken al wisten. Maar dat het ook kan voor de regelmatige 17-hoek ($k = 2$), de regelmatige 257-hoek ($k = 3$), en de regelmatige 65537-hoek ($k = 4$) kwam als een volslagen verrassing. Bij Gauss heeft het vinden van dit resultaat in belangrijke mate bijgedragen aan zijn beslissing om zijn leven verder aan de wiskunde te wijden. Naast de genoemde Fermat-priemgetallen 3, 5, 17, 257 en 65537 zijn er geen andere Fermat-priemgetallen bekend. Het kleinste getal k waarvoor niet bekend is of $F_k = 2^{2^k} + 1$ priem is, is $k = 22$. Het door Fermat uitgesproken vermoeden dat alle Fermatgetallen priemgetallen zijn, werd reeds gelogenstraft door Euler, die ontdekte dat $F_5 = 641 \times 6700417$. Dit is overigens een van de heel zeldzame keren geweest dat Fermat een vermoeden uitsprak dat achteraf onjuist is gebleken!

Ook het trisectieprobleem en het probleem van de kubusverdubbeling werden in het begin van de negentiende eeuw met algebraïsche middelen opgelost: het bleken inderdaad onmogelijke constructieopgaven te zijn. De kwadratuur van de cirkel bood meer weerstand: de onmogelijkheid daarvan was pas bewezen toen F. Lindemann in 1882 aantoonde dat het getal π transcendent is.

Als we de geschiedenis van de passer-en-liniaalconstructies bekijken, zien we dat het belang ervan voor de wiskunde eigenlijk helemaal niet heeft gelegen in het oplossen van de oorspronkelijke constructieopgaven, maar veeleer in de verbindingen die er bleken te liggen met nieuwe algebraïsche vraagstukken. Dat ze nog zo lang in het elementaire meetkundeonderwijs zijn blijven voortleven, is niet meer dan een teken van de hardnekkigheid waarmee tradities zich in het onderwijs weten te handhaven, ook als ze vanuit een hoger standpunt bekeken hun bestaansrecht allang verloren hebben.

5. Vermoedens in de klas

Ontegenzeggelijk is de betekenis van de experimentele wiskunde de laatste jaren enorm toegenomen. De mogelijkheden om met behulp van computers structu-

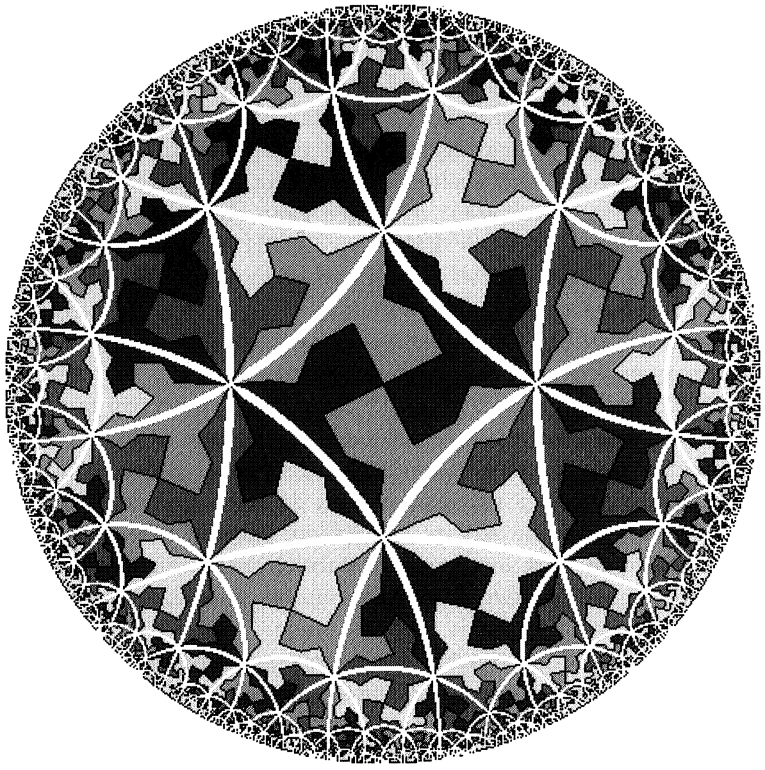
ren te onderzoeken en vermoedens te toetsen zijn thans ongekend. Het zal niet lang meer duren voordat de computer ook in de wiskundeles een onmisbaar instrument is geworden. Ongetwijfeld zal het karakter van het wiskundeonderwijs daardoor radicaal veranderen. Ik ben er echter van overtuigd dat de nieuwe mogelijkheden om leerlingen met de computer actief bij de wiskunde te betrekken een overwegend positieve invloed op het onderwijs zullen uitoefenen. Leerlingen zullen zelf *what if* vragen gaan stellen zodra we ze over de drempel van programma's als DERIVE en CABRI hebben geholpen. Daarbij zullen sommige leerlingen vanzelf allerlei wiskundige patronen gaan ontdekken – de jonge wiskundige onderzoeker is geboren. Met name het meetkundeprogramma CABRI biedt daarvoor prachtige mogelijkheden; er zijn op dat gebied al experimenten geweest met heel positieve ervaringen.

Waar het om gaat, is dat de leraar de leerling steeds op een positieve wijze stuurt en stimuleert. Door uitdagende onderzoeksvragen te stellen en interressante opdrachten te formuleren. Door op de juiste momenten de noodzaak van het zoeken naar een bewijs te signaleren en de leerlingen dan bij hun moeizame eerste stappen op die weg te begeleiden. Zijn er geen gevaren op dat pad? Natuurlijk wel! Dat leerlingen de subtiele redeneringen die nodig zijn voor een wiskundig bewijs niet zelf kunnen vinden, en zich verstrikken in drogredenen en cirkelredeneringen, is nog het minste gevaar. Ook in de huidige onderwijssituatie is dat immers het geval; met de computer erbij hebben we echter veel meer mogelijkheden om absurde consequenties van redeneerfouten te signaleren. Een reëler gevaar lijkt me kans dat leerlingen met veel te moeilijke problemen worden geconfronteerd. Met vermoedens waarvan de oplossing volledig buiten hun bereik ligt. Het kan voor de beginner heel frustrerend zijn om veel te tijd steken in een onderzoek dat niet op een bevredigende manier kan worden afgesloten. Daarom is het nodig dat je als leraar al in grote lijnen weet waar je je leerlingen aan laat beginnen. Belangrijk lijkt me dat elk onderzoeksproject wordt bekroond met een werkstuk waar je mee voor de dag kunt komen.

6. Eschers prent *Cirkellimiet* III – een vermoeden

Tegen het einde van mijn verhaal komend, kan ik de verleiding toch niet weerstaan om nog een keer terug te keren naar de wondere wereld van de niet-euclidische meetkunde van Gauss, Bolyai en Lobachevsky, en wel aan de hand van twee van Eschers beroemde cirkellimietprenten. In Figuur 1 ziet u een gestileerde computerversie van *Cirkellimiet* III. De vier kleuren geel, groen, bruin en blauw zijn door grijstinten vervangen en van de vissen zijn alleen de contouren getekend. Maar u heeft natuurlijk allemaal wel een of meer Escherboeken in de kast staan waarin de echte prent in kleur is afgedrukt. Als excuus voor mijn uitwijding over deze prent zal ik er een vermoeden aan vast knopen, namelijk een vermoeden omtrent de manier waarop ik denk dat Escher de prent *Cirkellimiet* III geconstrueerd heeft. Een bewijs van dat vermoeden geef ik niet: de titel van de vacantiecursus luidt immers 'Onbewezen vermoedens'.

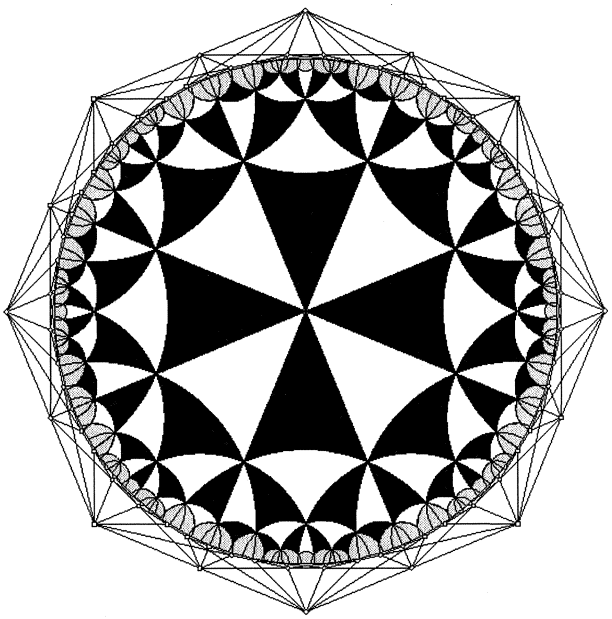
Bekend is dat de cirkellimietprenten van Escher (hij maakte er vier) hun ontstaan te danken hebben aan de Canadese wiskundige H.S.M. Coxeter, die



FIGUUR 1. Een gestileerde versie van *Cirkellimiet* III

Escher in 1958 een artikel van zijn hand stuurde met daarin een afbeelding van een betegeling met driehoeken van het cirkelmodel van Poincaré van het niet-euclidische vlak. Dat is dus een vlak waarin de bovenvermelde vreemde meetkundige wetten heersen: het parallellenaxioma is er niet geldig, de som van de hoeken van een driehoek is altijd minder dan 180 graden, er zijn geen vierkanten of rechthoeken en er zijn geen gelijkvormige niet-congruente figuren. In het model van Poincaré is het gehele niet-euclidische vlak afgebeeld binnen een cirkelschijf Ω (de rand van Ω doet niet mee). De lijnen zijn de cirkelbogen binnen Ω die loodrecht op Ω staan, inclusief de middellijnen van Ω. In dit model worden niet-euclidische hoeken 'op ware grootte' afgebeeld, maar de niet-euclidische afstanden worden naar de rand toe steeds meer verkleind. Vergelijk het maar met de bekende Mercatorprojectie uit de aardrijkskunde, waar gebieden rond de evenaar vrijwel onvervormd worden afgebeeld, maar waarbij de vervormingen naar de polen toe steeds groter wordt.

Coxeters tekening was zoiets als Figuur 2: het Poincaré-model met daarin een patroon van cirkelbogen (dat wil dus zeggen niet-euclidische lijnen) die tezamen een betegeling van het vlak vormen met niet-euclidische driehoeken. Het steigerwerk eromheen wordt gebruikt bij de constructie van de cirkelbogen: de punten zijn de middelpunten van de cirkelbogen, en twee punten zijn met



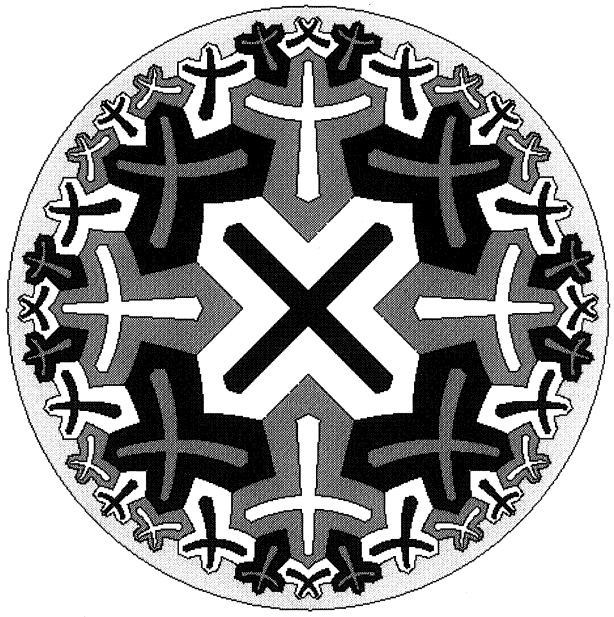
FIGUUR 2. Een variant op de betegeling uit Coxeters artikel

elkaar door een lijn in het steigerwerk met elkaar verbonden als de betreffende cirkelbogen elkaar snijden. Na voltooiing van de constructie kun je het steigerwerk weer verwijderen, maar voor Escher was het prettig dat dit niet was gebeurd: zo kon hij de constructie ervan goed bestuderen.

Omdat hoeken tussen lijnen op ware grootte worden afgebeeld, kunnen we de hoeken van de driehoeken uit de betegeling direct uit de figuur aflezen: allemaal zijn ze 45 graden. De som van de hoeken van zo'n driehoek is dus in dit geval inderdaad minder dan 180 graden, namelijk 135 graden. U ziet ook onmiddellijk dat het parallellenaxioma niet geldt: er zijn situaties waarin twee lijnen een derde lijn snijden en waarbij de som van de hoeken aan een kant van de snijlijn minder is dan 180 graden, maar waarbij die twee lijnen elkaar toch niet snijden. Ook alle andere merkwaardige eigenschappen van de niet-euclidische meetkunde laten zich aan de hand van Figuur 2 direct illustreren.

Coxeter gebruikte de Poincaré-illustratie in zijn artikel om er allerlei stellingen over symmetrieën en betegelingen in het niet-euclidische vlak mee te illustreren. Escher zag in Coxeters tekening echter geen niet-euclidisch vlak – hij wist helemaal niet wat dat was, en hij kon Coxeters explicaties erover ook niet volgen – maar een geheel nieuwe oplossing van een probleem waar hij al jaren mee worstelde. Dat probleem was: hoe kun je een vlakvulling maken waarin gelijkvormige figuurtjes steeds maar kleiner worden, terwijl alles toch binnen een begrensde figuur besloten blijft. In een filosofisch jasje gegoten: hoe breng je het oneindige op een harmonieuze wijze binnen handbereik?

In Coxeters prent zag Escher een fraaie oplossing in beeld gebracht: een cir-



FIGUUR 3. Een gestileerde versie van *Cirkellimiet* II

kelschijf met daarbinnen 'gelijkvormige driehoeken' die naar buiten toe steeds kleiner worden, waarbij de randcirkel als een soort limietfiguur optreedt. Onmiddellijk toog hij aan de slag om Coxeters driehoeken te vervormen tot herkenbare dierenfiguren zoals hij dat ook al zo vaak met gewone regelmatige vlakvullingen had gedaan. Met wat extra hulp van Coxeter kreeg hij de constructiegeheimen van zulke betegelingen al snel volledig onder de knie.

Eschers eerste poging was *Cirkellimiet* I, een houtsnede met primitief getekende zwarte en witte vissen waarin Coxeters oorspronkelijke tekening nog duidelijk herkenbaar is. In *Cirkellimiet* II (zie Figuur 3) heeft Escher de motieven tot kruisen vervormd, en is het aantal kleuren drie geworden: zwart, wit en rood (hier als grijs afgebeeld). Eschers meesterwerk is echter *Cirkellimiet* III. In Eschers eigen woorden: '[Hierin] zijn de gebreken [van *Cirkellimiet* I] grotendeels verholpen. Er zijn nu alleen nog maar series "met doorgaand verkeer": alle vissen van dezelfde serie hebben ook dezelfde kleur en zwemmen elkaar, kop aan staart, achterna langs een cirkelvormige baan van rand tot rand. Hoe dichter zij het centrum naderen, hoe groter zij worden. Vier kleuren zijn nodig opdat elke rij in haar geheel met de omgeving contrasteert. Geen enkele component van al deze reeksen die van oneindig ver als vuurpijlen loodrecht uit de limiet opstijgen en er weer in teloorgaan, bereikt de grenslijn ooit.'

De rijen vissen waar Escher het over heeft, worden extra benadrukt door de witte streep die over hun rug loopt. Samen vormen die witte strepen een patroon van cirkels die het vlak verdelen in driehoeken en vierhoeken. Binnen elke driehoek komen drie linkervinnen bij elkaar, en binnen elke vierhoek vier

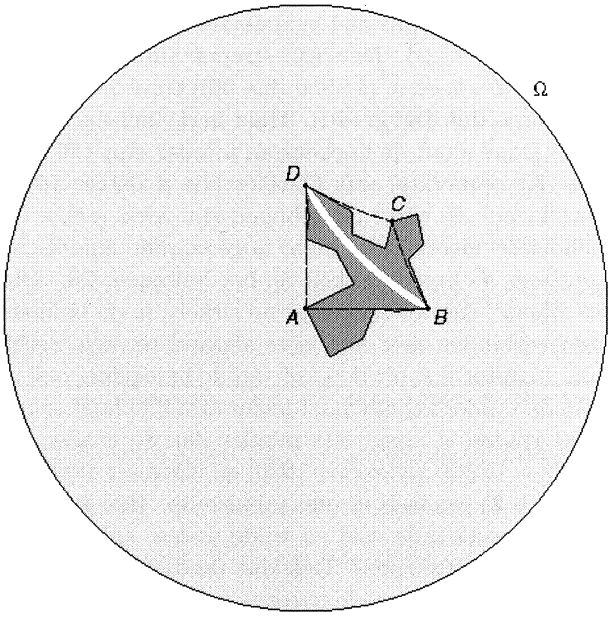
rechtervinnen. De vissen zijn dus niet symmetrisch! Toch is er met die vlakvulling iets vreemds aan de hand. De witte strepen snijden elkaar onder hoeken van 60 graden, en de driehoeken hebben dus blijkbaar drie hoeken van 60 graden. Hun hoekensom is dan 180 graden. Maar in de niet-euclidische meetkunde is dat onmogelijk! Daar moet de hoekensom minder dan 180 graden zijn.

De eerste die dat opmerkte, was Coxeter (zie [Coxeter 1979]). Hij kwam ook met een verklaring: de witte cirkelbogen zijn *geen* rechte lijnen in de zin van de niet-euclidische meetkunde, maar zogenaamde equidistantielijnen, die vergelijkbaar zijn met de breedtecirkels op het boloppervlak. Dat zijn immers ook geen 'rechte lijnen', dat wil zeggen grote cirkels, in de bolmeetkunde, maar lijnen die punten verbinden met een vaste afstand tot een 'rechte lijn', in dit geval de equator. Coxeter kon uit de aard van de betegeling ook afleiden dat de witte cirkelbogen de randcirkel allemaal onder dezelfde hoek ω moeten snijden, en dat ω geen 90 graden is, maar iets minder dan 80 graden. Om precies te zijn: $\cos\omega = \frac{1}{2}(\sqrt[4]{2} - 1/\sqrt[4]{2})$ ([Coxeter 1979] en [Coxeter 1996-97]).

Maar dat maakte de raadsels er niet minder op. Hoe kon Escher, die niets van niet-euclidische meetkunde wist en wilde weten, zo'n subtiele vlakvulling met equidistantielijnen ontwerpen? Ook dat raadsel bracht Coxeter dicht bij een oplossing door het onderliggende patroon van cirkelbogen die de randcirkel wel degelijk onder hoeken van 90 graden snijden, aan de oppervlakte te brengen. Escher had dat patroon op listige wijze onzichtbaar gemaakt. Wat Coxeter nergens expliciet opmerkt, is het geheim van de precieze vissenvorm, en de subtiele, typisch Escheriaanse wijze waarop die vorm uit een simpel vlakvullingmotief kan zijn afgeleid. Ik wil daar hier een bescheiden, plausibel vermoeden over uitspreken.

In Figuur 4 ziet u zo'n vissencontour, met daaronder in stippellijnen een vlieger $ABCD$ met hoeken van 90, 60, 120 en 60 graden. Die vlieger is samengesteld uit niet-euclidische rechte lijnen: twee delen van onderling loodrechte middellijnen van de randcirkel Ω en twee delen van cirkelbogen die Ω loodrecht snijden en die elkaar snijden onder een hoek van 120 graden. De congruente lijnstukken AB en AD heeft Escher vervormd tot congruente contourdelen van de rechterhelft van de vis, en de congruente cirkelbogen CB en CD zijn congruente contourdelen van de linkerhelft van de vis geworden. De rest (de ogen en de strepen op de vinnen en de staart) is slechts opvulling en verfraaiing.

Figuur 5 laat zien hoe *Cirkellimiet* III eruit zou zien met vliegers in plaats van vissen. Saaier, maar heel helder van structuur. De witte cirkelbogen heb ik laten staan, maar nu hebben ze geen andere functie meer dan het benadrukken van monochromatische kleurenrijen. Je ziet dat er nu ook weer spiegelsymmetrieën zijn; door Eschers vervormingstruc waren die verdwenen. Je kunt de vliegers in groepjes van vier samennemen tot regelmatige achthoeken met hoeken van 120 graden, zoals in de linkerhelft van Figuur 6. Dat is weer een andere betegeling van het niet-euclidische vlak. In elk hoekpunt komen drie achthoeken samen. De *duale* betegeling ontstaat door de middelpunten van de achthoeken te nemen, en die door een lijn te verbinden wanneer de bijbehorende achthoeken een zijde gemeen hebben. In de rechterhelft van Figuur 6 is te zien hoe dat gaat. De resulterende betegeling is er een met regelmatige

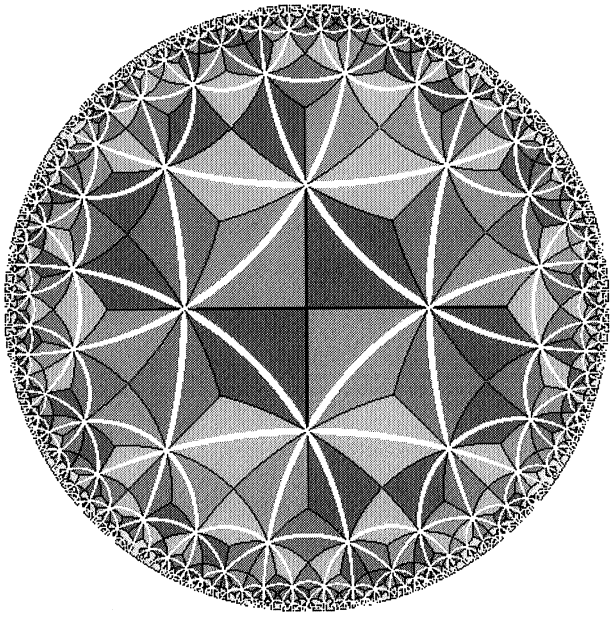


FIGUUR 4. Een tot vis vervormde vlieger

driehoeken, acht rond elk hoekpunt. En u raadt het al, dat is niets anders dan de betegeling van Figuur 2.

Het is heel waarschijnlijk dat Escher inderdaad begonnen is met de betegeling van Figuur 2 en daar de duale betegeling van achthoeken bij heeft geconstrueerd. Diezelfde betegelingen liggen namelijk ten grondslag aan de houtsnede *Cirkellimiet* II, zoals ook Coxeter heeft opgemerkt [Coxeter 1981, p. 207]. U zult weinig moeite hebben om aan de hand van de Figuren 3 en 6 de structurele overeenkomst tussen de achthoekenbetegeling en *Cirkellimiet* II te achterhalen. Escher heeft eenvoudig de achthoeken tot brede kruisen vervormd, en de gehele tekening over $22\frac{1}{2}$ graad gedraaid. De binnenste, smalle kruisen laten met elkaar nog een rudiment van de duale betegeling van driehoeken met hoeken van 45 graden zien.

De volgende stap in de constructie van *Cirkellimiet* III was waarschijnlijk dat Escher de achthoekenbetegeling in vliegers onderverdeeld heeft, vier vliegers per achthoek, en gezocht heeft naar een kleuringswijze met zoveel mogelijk symmetrie. Vier kleuren bleken nodig te zijn om elke vlieger met zijn buren te laten contrasteren. Escher ontdekte dat er bij zo'n kleuring vanzelf 'stromen' ontstonden van gelijkgekleurde vliegers. Het vervormen van een vlieger tot een vissenfiguur was voor Escher daarna een kolfje naar zijn hand. Dat zo'n vis asymmetrisch moest worden, zal hij met plezier hebben geconstateerd; dat maakte de prent alleen maar intrigerender. Maar de aardigste vondst moet toch de ontdekking zijn geweest dat de vissen van een ruggengraat konden worden voorzien die uit doorlopende cirkelbogen bestaat. Juist daarmee kon hij zijn

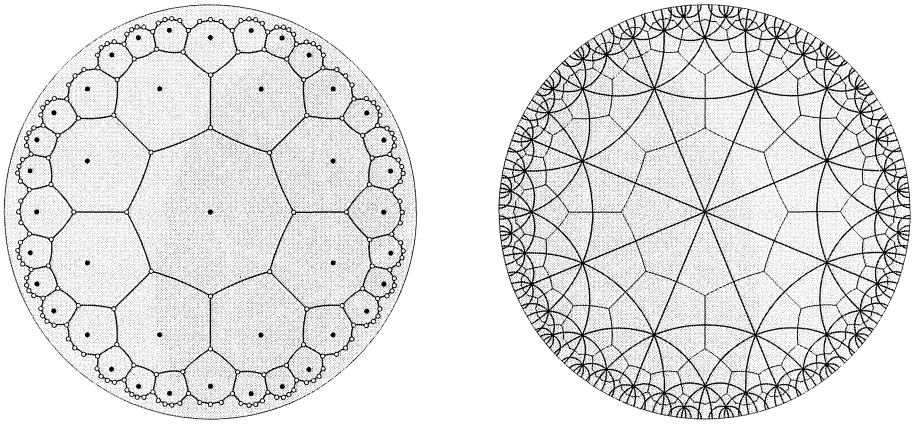


FIGUUR 5. Een versie met vliegers van *Cirkellimiet* III

oorspronkelijke schema volledig aan het oog onttrekken, en alle oppervlakkige beschouwers zand in de ogen strooien. Zelfs Bruno Ernst, die in zijn boek *De Toverspiegel van M.C. Escher* op bladzijde 109 over *Cirkellimiet* III opmerkt: 'Het netwerk daarvan is een vrije variatie op het oorspronkelijke netwerk. Behalve cirkelbogen die loodrecht op de omtrek staan (zoals het behoort) zijn er ook cirkelbogen die dat niet doen.' Van 'vrije variatie' is in *Cirkellimiet* III echter geen sprake: de gehele prent is met een volmaakte mathematische precisie geconstrueerd; je kunt hem met een computertekenprogramma waarin je de bijbehorende niet-euclidische rotaties inbouwt, volledig reproduceren. Er is in *Cirkellimiet* III ook geen enkele cirkelboog meer te zien die loodrecht op de omtrek staat. Alle zichtbare bogen snijden de omtrek onder dezelfde hoek $\omega \approx 80°$.

Eschers cirkellimietprenten zijn daarom zo interessant, omdat je er op twee manieren naar kunt kijken. Je kunt ze, net als Escher deed, zien als euclidische cirkelprenten waarin 'gelijkvormige' figuren naar de randcirkel Ω toe steeds kleiner worden. Maar die gelijkvormigheid is eigenlijk geen gelijkvormigheid in de strikte, euclidische betekenis van het woord: rechte lijnen worden tot cirkelbogen vervormd met een steeds veranderende kromtestraal. Toch blijven we die figuren wel als 'gelijkvormig' herkennen. De meetkundige transformaties die erachter zitten, zijn eigenlijk inversies in cirkels en samenstellingen daarvan.

Veel duidelijker wordt de zaak als we er anders tegenaan kijken. Als we, met Poincaré, de loodcirkels en middellijnen van Ω opvatten als 'rechte lijnen' in een nieuw soort meetkunde, en de inversies in die 'rechte lijnen' opvatten als



FIGUUR 6. Betegeling met regelmatige achthoeken (links) en dezelfde betegeling (rechts) met daarop gesuperponeerd de duale betegeling

spiegelingen. In Figuur 2 zie je dat zulke spiegelingen de witte en zwarte driehoeken verwisselen, en in Figuur 5 kun je zien hoe zulke spiegelingen telkens twee kleuren verwisselen en de andere twee onveranderd laten. Via opeenvolgingen van spiegelingen kun je elke vlieger in elke andere vlieger transformeren, en het is dus geen gek idee om je in te denken dat al die vliegers in deze nieuwe meetkunde onderling *congruent* (en dus niet gelijkvormig!) moeten zijn.

De opeenvolging van twee spiegelingen in assen die elkaar onder een hoek α snijden, geeft een rotatie over een hoek $2 \times \alpha$, net zoals in de gewone, euclidische meetkunde. In de betegelingen die we hebben afgebeeld, is dit ook duidelijk te zien. Door de asymmetrische vissenvorm heeft Escher in *Cirkellimiet* III alle spiegelingen verwijderd, en alleen de rotaties in stand gehouden. De rotatiecentra bevinden zich in de tips van de rechtervinnen (90 graden) en de linkervinnen (120 graden), en in de punten waar drie koppen en drie staarten samenkomen (eveneens 120 graden). Diezelfde rotaties zijn ook in *Cirkellimiet* II terug te vinden, maar daar zorgt de symmetrie van de kruisvorm dat sommige spiegelingen nog wel aanwezig zijn.

In de euclidische meetkunde is de opeenvolging van twee spiegelingen in assen die elkaar niet snijden een translatie, en in de niet-euclische meetkunde is het al niet anders. In Figuur 1 zwemmen de vissen in 'translatiestromen' achter elkaar aan, en aan de hand van Figuur 5 kunt u gemakkelijk uitvinden hoe zo'n translatie tot stand komt door spiegelingen in twee assen die elkaar niet snijden, achter elkaar te schakelen.

Zo zouden we nog lang door kunnen gaan. Eigenlijk ken ik geen betere introductie tot de niet-euclidische meetkunde dan via Eschers cirkellimietprenten. Ze vormen een rijke bron van voorbeelden en illustratiemateriaal, die ook voor niet-wiskundigen goed toegankelijk is. Het is alleen jammer dat de wiskundewereld er destijds niet voldoende in geslaagd is Escher ervan te overtuigen

hoezeer hij daarmee deel is gaan uitmaken van onze gemeenschap, en hoe goed we zijn prenten kunnen gebruiken om ook aan leken uit te leggen waar het in ons vak nu eigenlijk om draait.

7. Vermoedens als uitdagingen

We keren terug naar ons algemene thema: de rol van vermoedens in de wiskunde. Zoals gezegd, men zou de geschiedenis van de wiskunde kunnen beschrijven als een geschiedenis van bewezen en onbewezen vermoedens. Elke stelling zonder bewijs is een vermoeden; zodra er een bewijs is, wordt het vermoeden een stelling. Toch is die visie op de wiskunde natuurlijk nog wat te beperkt. Stellingen en vermoedens zijn slechts kristallisatiepunten in het wiskundig onderzoek. Vermoedens zijn bakens waarop het onderzoek zich richten kan. Terreinen die volledig in kaart gebracht zijn, tellen geen onbewezen vermoedens meer. Maar zulke terreinen zijn er in de wiskunde slechts weinig. Vrijwel in elk gebied zijn er nog witte plekken op de kaart te vinden. Dat kunnen plekken zijn waar niemand komt omdat niemand ze op dit moment nog interessant vindt, of ondoordringbare gebieden waar juist al heel veel bezoekers tevergeefs naar toegangspoorten hebben gezocht. Die laatste zijn dan meestal ook de plaatsen waar beroemde onbewezen vermoedens het landschap markeren. De onbewezen vermoedens die in deze vacantiecursus ter sprake zullen worden gebracht, behoren vrijwel allemaal tot die laatste categorie. Het zijn vermoedens met een rijke historie, vermoedens die met verschillende takken van de wiskunde verbonden zijn en die weerstand hebben geboden aan de meest uiteenlopende aanvallen van bekwame wiskundigen. Die vermoedens behoren tot de grote uitdagingen voor de wiskunde van de volgende eeuw. En stuk voor stuk zullen ze degenen die ze weet te bedwingen wereldroem bezorgen – in elk geval binnen de wereld van de wiskunde.

8. Literatuur

Over het $3n+1$ vermoeden:

Frits Beukers, *Getaltheorie voor Beginners*, Utrecht, Epsilon Uitgaven, 1999.

Stan Wagon, The Collatz Problem, *The Mathematical Intelligencer* **7**-1, 1985, 72-76.

Over de Laatste Stelling van Fermat:

Simon Singh, *Het laatste raadsel van Fermat*, Amsterdam, De Arbeiderspers, 1997.

Over Eschers Cirkellimiet-prenten

H.S.M. Coxeter, The non-Euclidean symmetry of Escher's picture 'Circle Limit III', *Leonardo* **12**, 1976, 19-26.

H.S.M. Coxeter, Angels and Devils, in: David A. Klarner (ed.), *The Mathematical Gardner*, Belmont, California, Wadsworth International, 1981, 197-209.

H.S.M. Coxeter, The Trigonometry of Escher's Woodcut 'Circle Limit III', *The Mathematical Intelligencer* **18**-4, 1996, 42-46, correctie in **19**-1, 1997.

PRIEMGETALLEN

PETER STEVENHAGEN

ABSTRACT. De verdeling van de priemgetallen over de verzameling van de natuurlijke getallen is een bron van eenvoudig te formuleren vermoedens die voor een groot deel nog onbewezen zijn. We gaan in op een aantal klassieke vermoedens.

1. INLEIDING

Een *priemgetal* is een geheel getal $n > 1$ dat geen andere delers heeft dan 1 en zichzelf. Merk op dat 1 per definitie geen priemgetal is – we willen namelijk niet dat priemgetallen elkaar delen. Priemgetallen zijn getallen zonder 'echte' delers, en het is niet moeilijk in te zien dat willekeurige getallen $n \geq 1$ altijd als een product van priemgetallen te schrijven zijn:

$$1998 = 2 \cdot 3^3 \cdot 37, \qquad 1999 = 1999, \qquad 2000 = 2^4 \cdot 5^3, \qquad 2001 = 3 \cdot 23 \cdot 29.$$

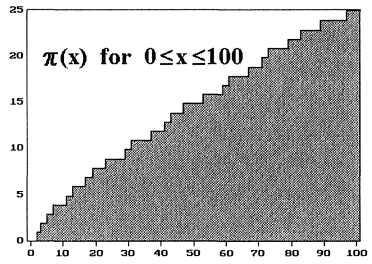
Minder evident, maar eveneens waar is dat een dergelijke schrijfwijze van n uniek is: het is de *priemfactorontbinding* of *factorisatie* van n. Als 'multiplicatieve bouwstenen' van de gehele getallen zijn de priemgetallen van fundamenteel belang in de getaltheorie.

Opgave 1. Laat zien dat de factorisatie van n *niet* uniek is als we 1 als priemgetal toelaten.

Het begin van de lijst van priemgetallen ziet er uit als

$$\mathcal{P} = \{2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ 37, \ldots\}. \tag{1.1}$$

Zoals de suggestieve puntjes in (1.1) al aangeven zijn er *oneindig veel* priemgetallen. Anders gezegd: de enigszins onregelmatig springende *priemgetal-telfunctie* $\pi(x)$, die het aantal priemgetallen tot aan x telt, groeit onbegrensd voor $x \to \infty$. Onderstaand plaatje geeft het gedrag van $\pi(x)$ voor $x \leq 100$.



Naast het groeigedrag van $\pi(x)$, waarop we in sectie 2 uitgebreid ingaan, zijn er nog diverse andere vragen die al op grond van het kleine lijstje in (1.1) in het oog springen. laten we bijvoorbeeld eens kijken naar de decimale eindcijfers van de priemgetallen in ons lijstje. Het is duidelijk 0, 4, 6 en 8 niet als eindcijfer voorkomen, en dat de eindcijfers 2 en 5 alleen voor de priemgetallen 2 en 5 zelf voorkomen. De 4 overige mogelijke eindcijfers 1, 3, 7 of 9 komen in ons lijstje elk twee of drie maal voor. Zouden ze 'gemiddeld' alle vier 'even vaak' voorkomen? Ook als we weten dat er oneindig veel priemgetallen zijn, is het in het geheel niet duidelijk waarom er oneindig veel priemgetallen op een 3 zouden moeten eindigen. Sectie 3 geeft aan met welke methoden men zo'n vraag kan benaderen.

De overige secties gaan in op vragen over priemgetallen waarop tot op de dag van vandaag geen bevredigend antwoord gevonden is. Het blijkt dat er heel erg veel van dit soort vragen zijn, en dat hun formulering vaak uiterst eenvoudig is. Ons lijstje (1.1) geeft bijvoorbeeld al vijf voorbeelden van *priemgetaltweelingen*, zoals $(17, 19)$ en $(29, 31)$. Er zijn ook grotere tweelingparen, zoals $(1997, 1999)$ en het bijna onopschrijfbaar grote 11755-cijfer-paar

$$(361700055 \cdot 2^{39020} - 1, \quad 361700055 \cdot 2^{39020} + 1).$$

Of er oneindig veel van zulke paren zijn is niet bekend, maar er zijn eenvoudige *heuristieken* gebaseerd op het (bewezen) groeigedrag van de functie $\pi(x)$. Dergelijke heuristieken 'werken' erg goed in een veelheid van situaties waar het priemen van een speciale vorm betreft. Hiermee bedoelen we dat ze in overeenstemming zijn met de beschikbare numerieke gegevens – ze *bewijzen* helemaal niets. We besteden in het bijzonder aandacht aan populaire priemgetallen als de *Fermat-* en *Mersenne-priemen*, waarvan we nog steeds niet weten of er oneindig veel zijn.

Ons afsluitende open probleem, het *Artin-vermoeden* voor primitieve wortels, neemt een bijzondere plaats in; hier zijn er slechts bewezen resultaten onder aanname van de *gegeneraliseerde Riemann-hypothese*, die in het artikel van Tijdeman in deze bundel nader wordt uitgelegd.

2. De priemgetalstelling

De eenvoudigste kwantitatieve stelling betreffende het aantal priemgetallen noemden we al in de inleiding.

Stelling 1. *Er zijn oneindig veel priemgetallen.*

Reeds bij Euclides (300 v. Chr.) vinden we een *bewijs uit het ongerijmde* voor stelling 1, als volgt. Stel dat er maar n verschillende priemgetallen zijn, namelijk $p_1, p_2, \ldots, p_n$. Neem nu het grote getal $N = p_1 p_2 \cdots p_n + 1$ verkregen door 1 op te tellen bij het product van de priemgetallen, en schrijf N als product van priemgetallen. Omdat N rest 1 heeft bij deling door elk van de priemgetallen p_i, is geen van de priemgetallen p_i een deler van N. De priemdelers van N komen dus niet voor in ons lijstje: klaar!

Een iets ander bewijs, dat meer analytisch van aard is, maakt gebruik van de somformule voor de meetkundige reeks

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 \ldots \qquad \text{voor} \quad |x| < 1. \tag{2.1}$$

Laten we weer aannemen dat er slechts eindig veel priemgetallen $p_1, p_2, \ldots, p_n$ zijn. Dan is ieder geheel getal $k \geq 1$ te schrijven als $k = \prod_{i=1}^{n} p_i^{e_i}$ met $e_i \in \{0, 1, 2, 3, \ldots\}$. Vermenigvuldig nu de identiteiten

$$\frac{1}{1 - \frac{1}{p_i}} = 1 + \frac{1}{p_i} + \frac{1}{p_i^2} + \frac{1}{p_i^3} + \frac{1}{p_i^4} + \ldots$$

voor $i = 1, 2, \ldots, n$. Het product van de linkerleden is een rationaal getal, waarvan de waarde ons niet interesseert. Het rechterlid wordt na uitvermenigvuldigen een som van oneindig veel termen $\frac{1}{k}$, waarbij k loopt over alle gehele getallen die te schrijven zijn als $k = \prod_{i=1}^{n} p_i^{e_i}$ met $e_i \in \{0, 1, 2, 3, \ldots\}$. Wegens onze aanname zijn *alle* $k \geq 1$ zo te schrijven, dus de bekende harmonische reeks $\sum_{k=1}^{\infty} \frac{1}{k}$ heeft een eindige limiet: tegenspraak.

Het tweede bewijs, waarin we voor de eenvoud de convergentieproblemen die bij het vermenigvuldigen van oneindige sommen op kunnen treden onder de mat geveegd hebben, heeft een aardig gevolg. Omdat het product $\prod_p (1 - \frac{1}{p})^{-1}$ over alle priemgetallen naar $+\infty$ divergeert, moet de logaritme van dit product, die gelijk is aan $\sum_p -\log(1 - \frac{1}{p})$, ook naar $+\infty$ divergeren. Nu geldt als p groot is $-\log(1 - \frac{1}{p}) \approx \frac{1}{p}$, en dit geeft het volgende resultaat.

Stelling 2. *De som $\sum_p \frac{1}{p}$ over alle priemgetallen p is divergent.*

Stelling 2 is al iets meer 'kwantitatief' dan stelling 1. Hij zegt dat er bijvoorbeeld meer priemgetallen dan kwadraten zijn; immers, de som $\sum_{n \geq 1} \frac{1}{n^2}$ over de inversen van de kwadraten is wel convergent.

De priemgetalstelling vertelt ons nog veel preciezer hoeveel priemgetallen er zijn: hij geeft het precieze groeigedrag van de priemgetal-telfunctie $\pi(x)$ voor $x \to \infty$. Men kan deze stelling, net als twee eeuwen geleden de vijftienjarige Gauss, zelf ontdekken aan de hand van numeriek materiaal. Een tabelletje van het aantal priemgetallen tot aan x, dat ik gekopieerd heb uit een zeer lezenswaardige voordracht van Zagier [6] en dat inmiddels tot aan 10^{20} uitgebreid is [2], ziet er namelijk zo uit:

x	$\pi(x)$	$x/\pi(x)$
10	4	2.5
100	25	4.0
1000	168	6.0
10 000	1 229	8.1
100 000	9 592	10.4
1 000 000	78 498	12.7
10 000 000	664 579	15.0
100 000 000	5 761 455	17.4
1 000 000 000	50 847 534	19.7
10 000 000 000	455 052 511	22.0

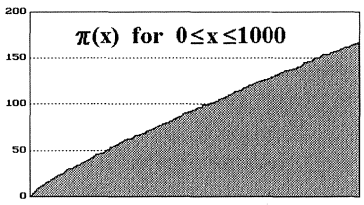
Opgemerkt zij dat Gauss, die met de hand de eerste paar waarden in de tabel berekende, niet al deze moderne computerwaarden tot zijn beschikking had! We zien:

iedere keer als x met een factor 10 toeneemt, neemt de fractie $x/\pi(x)$ van het aantal getallen tot aan x dat priem is af met een factor in de buurt van $2.3 \approx \log 10$. De 'dichtheid' $\pi(x)/x$ van de verzameling priemgetallen loopt dus heel regelmatig terug voor $x \to \infty$, namelijk als $\frac{1}{\log x}$. (Merk op dat $\log x$ voor ons altijd de *natuurlijke* logaritme is, die ook wel met $\ln x$ aangegeven wordt.)

Priemgetalstelling. *Voor $x \to \infty$ geldt de asymptotische gelijkheid $\pi(x) \sim \frac{x}{\log x}$.*

Met 'asymptotische gelijkheid' van twee functies voor $x \to \infty$ bedoelen we dat hun quotiënt voor $x \to \infty$ de limietwaarde 1 heeft.

De priemgetalstelling zegt dat de groei van $\pi(x)$ uiterst regelmatig is. De trapfunctie $\pi(x)$ ziet er op een iets langer interval dan $[0, 100]$ dan ook al snel 'glad' uit.



Een *bewijs* van de priemgetalstelling werd pas in 1896, een eeuw na de ontdekking van de stelling, onafhankelijk gevonden door de Franse wiskundigen Hadamard en de la Vallée-Poussin. Het bewijs, dat niet gemakkelijk is, begint ermee de truc van de divergente harmonische reeks om te smeden tot een identiteit

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p} \frac{1}{1 - \frac{1}{p^s}} \tag{2.2}$$

voor de *convergente* reeks $\sum_{n=1}^{\infty} \frac{1}{n^s}$. Deze identiteit, waarin voor s een willekeurig reëel getal groter dan 1 genomen kan worden, is een soort 'analytische vertaling' van de stelling van de eenduidige priemfactorontbinding. De waarde van (2.2) als functie van s wordt met $\zeta(s)$ aangegeven, en men noemt de functie ζ de *Riemann-zeta-functie*. De priemgetalstelling blijkt een gevolg te zijn van het gedrag van deze functie voor *complexe* waarden van s – een wonder dat tot op de dag van vandaag voor onopgeloste problemen zorgt. We verwijzen naar Tijdeman's artikel voor nadere details.

3. Priemen in rekenkundige rijen

De priemgetalstelling geeft ons scherpe informatie over de verzameling van *alle* priemgetallen. Indien we verzamelingen van priemgetallen bestuderen die aan een of andere extra eigenschap voldoen, bijvoorbeeld het hebben van een decimaal eindcijfer 3, staan we direct met lege handen. Numeriek is al snel duidelijk dat de *rekenkundige rij*

$$3, \quad 13, \quad 23, \quad 33, \quad 43, \quad 53, \quad 63, \quad , \ldots$$

oneindig veel priemgetallen bevat. Algemener ligt het voor de hand te verwachten dat, indien we de priemgetallen niet modulo 10 maar modulo een willekeurig getal n

bekijken, alle 'redelijke' restklassen modulo n oneindig veel priemgetallen bevatten. Modulo $n = 4$ zijn 1 en 3 de redelijke restklassen: de restklassen 0 en 2 bevatten immers alleen even getallen. In het geval $n = 10$ zijn het de restklassen 1, 3, 7 en 9 die alle priemen behalve 2 en 5 bevatten. Voor algemene n zijn alleen die restklassen $a \bmod n$ met $a \in \{0, 1, 2, 3, \ldots, n-1\}$ redelijk waarvoor a *onderling ondeelbaar* is met n. Immers, als a en n een gemeenschappelijke factor $d > 1$ hebben, dan zijn ook alle getallen van de vorm $a + kn$ deelbaar door d.

Stelling van Dirichlet. *Laat $n > 1$ een geheel getal zijn en $a \geq 1$ een getal dat onderling ondeelbaar is met n. Dan bevat de rekenkundige rij*

$$a, \quad a+n, \quad a+2n, \quad a+3n, \quad a+4n, \quad \ldots$$

oneindig veel priemgetallen.

In incidentele gevallen is het mogelijk Euclides' bewijs van stelling 1 aan te passen aan de situatie van Dirichlet's stelling. Willen we bijvoorbeeld bewijzen dat er oneindig veel priemgetallen $p \equiv 3 \bmod 4$ zijn, dan kunnen we voor ieder n-tal priemgetallen $p_1, p_2, \ldots, p_n$ het getal $N = 4(p_1 p_2 \ldots p_n)^2 - 1$ vormen. Omdat N congruent is met 3 modulo 4 kunnen niet alle priemdelers van N congruent zijn met 1 modulo 4. Bovendien is N niet door de priemgetallen p_i deelbaar. Er is dus een priemgetal $p \equiv 3 \bmod 4$ buiten ieder voorgegeven n-tal priemgetallen. Dit impliceert dat er oneindig veel priemgetallen congruent met 3 modulo 4 zijn.

[k] **Opgave 2.** Laat zien dat alle priemdelers van $N = 4(p_1 p_2 \ldots p_n)^2 + 1$ congruent met 1 modulo 4 zijn. Concludeer dat er oneindig veel priemgetallen congruent met 1 modulo 4 zijn.

Voor de meeste n, zoals $n = 10$, kan men de stelling van Dirichlet niet zo eenvoudig bewijzen. Dirichlet's bewijs is een slimme generalisatie van het bewijs dat we van stelling 2 gaven, en we schetsen het voor het geval $n = 4$.

We willen laten zien dat de sommen $\sum_{p \equiv 1 \bmod 4} \frac{1}{p}$ en $\sum_{p \equiv 3 \bmod 4} \frac{1}{p}$ allebei divergent zijn. Hiertoe laten we zien dat het 'verschil' van beide sommen begrensd is. Met andere woorden, als we de functie χ voor priemgetallen p definiëren door

$$\chi(p) = \begin{cases} 0 & \text{als } p = 2; \\ 1 & \text{als } p \equiv 1 \bmod 4; \\ -1 & \text{als } p \equiv -1 \bmod 4, \end{cases}$$

dan willen we laten zien dat de som $\sum_p \frac{\chi(p)}{p}$ een *convergente* som is. Uit stelling 2 volgt dan direct dat er niet maar eindig veel priemen p zijn waarvoor de waarde van $\chi(p)$ gelijk is aan, zeg, $+1$. Immers, de priemen met $\chi(p) = -1$ zouden dan de som naar $-\infty$ laten divergeren.

Om de som $\sum_p \frac{\chi(p)}{p}$ te analyseren maken we gebruik van de benadering

$$x \approx -\log(1-x) = \log(\frac{1}{1-x})$$

voor reële getallen x in de buurt van 0. We vinden dat $\sum_p \frac{\chi(p)}{p}$ niet ver afligt van de logaritme van het oneindige product

$$P = \prod_p \frac{1}{1 - \frac{\chi(p)}{p}}.$$

We gaan bewijzen dat het product P een positieve reële waarde aanneemt, zodat $\log P$ ook begrensd blijft. De somformule (2.1) voor de meetkundige reeks geeft

$$\frac{1}{1-\frac{\chi(p)}{p}} = \begin{cases} 1+\frac{1}{p}+\frac{1}{p^2}+\frac{1}{p^3}+\frac{1}{p^4}+\dots & \text{als } p \equiv 1 \bmod 4; \\ 1-\frac{1}{p}+\frac{1}{p^2}-\frac{1}{p^3}+\frac{1}{p^4}+\dots & \text{als } p \equiv -1 \bmod 4; \end{cases}$$

Uitvermenigvuldigen geeft nu het analogon

$$P = \prod_p \frac{1}{1-\frac{\chi(p)}{p}} = \sum_{n=1}^{\infty} \frac{\chi(n)}{n} \tag{3.1}$$

van (2.2). Hierbij breiden we χ door 'multiplicatieve voortzetting' uit tot een functie op de verzameling van alle positieve gehele getallen: $\chi(\prod_{i=1}^n p_i^{e_i}) = \prod_{i=1}^n \chi(p_i)^{e_i}$. De definitie van χ wordt heel eenvoudig:

$$\chi(n) = \begin{cases} 0 & \text{als } n \text{ even is;} \\ 1 & \text{als } n \equiv 1 \bmod 4; \\ -1 & \text{als } n \equiv -1 \bmod 4. \end{cases}$$

Nu blijkt dat het oneindige product P wel degelijk eindig is: het is gelijk aan de alternerende reeks

$$P = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots,$$

die convergent is. Wie de machtreeksontwikkeling van de arctangens-functie kent, weet zelfs dat de limiet gelijk is aan $\frac{\pi}{4} = \arctan 1$. Einde bewijs.

Dirichlet's bewijs laat een beetje meer zien dan we vroegen. Niet alleen zijn er oneindig veel priemgetallen congruent met 1 dan wel 3 modulo 4, maar de priemen in elk van beide klassen houden elkaar in evenwicht in de zin dat de sommen $\sum_p \frac{1}{p}$ voor beide klassen 'even snel' naar oneindig gaan. Men zegt wel dat de beide klassen gelijke *Dirichlet-dichtheid* hebben.

Voor willekeurige n wordt het bewijs iets ingewikkelder, omdat er meer dan twee restklassen zijn waarvan men de sommen $\sum_p \frac{1}{p}$ geschikt moet combineren tot een convergente som. De functies χ die men hierbij nodig heeft heten *Dirichlet-karakters*. Voor $n = 10$, waar niet twee maar vier interessante restklassen zijn, hebben de bijbehorende karakters als waarden niet ± 1 maar de machten van i, de complexe vierde eenheidswortel die aan $i^2 = -1$ en $i^4 = 1$ voldoet. Het zijn de 'multiplicatieve afbeeldingen' van de *vermenigvuldiggroep* $(\mathbf{Z}/10\mathbf{Z})^* = \{\overline{1}, \overline{3}, \overline{7}, \overline{9}\}$ van restklassen modulo 10 naar de *vermenigvuldiggroep* $\mathbf{C}^* = \mathbf{C} \setminus \{0\}$ van de complexe getallen.

Opgave 3. Laat zien dat er precies 4 afbeeldingen $\chi : (\mathbf{Z}/10\mathbf{Z})^* \to \mathbf{C}^*$ zijn die de vermenigvuldiging respecteren.

Het argument van Dirichlet laat zien dat de Dirichlet-dichtheid van de vier klassen van priemen eindigend op 1, 3, 7 en 9 voor elk van de vier klassen even groot is. In het bijzonder bevat elke klasse oneindig veel priemgetallen. Hetzelfde geldt als we 10 door een willekeurig getal n vervangen. Met de technieken van het bewijs van de priemgetalstelling kan men natuurlijker dichtheidsresultaten voor priemgetallen bewijzen, zoals

$$\lim_{x\to\infty} \frac{\#\{p \equiv 1 \bmod 4 : \quad p \le x\}}{\#\{p \equiv 3 \bmod 4 : \quad p \le x\}} = 1.$$

4. Priemgetallen van speciale vorm

De stelling van Dirichlet is één van de weinige stellingen die ons vertelt dat er oneindig veel priemgetallen zijn van een voorgeschreven vorm. Zonder direct een recept te geven voor het maken van een priemgetal eindigend op, bijvoorbeeld, de cijfercombinatie 123456789, vertelt hij ons dat zulke priemgetallen zeker bestaan.

Opgave 4. Zijn er oneindig veel priemen die in decimale notatie met 123456789 beginnen?

Eigenschappen van priemgetallen die verband houden met hun decimale representatie geven bijna altijd aanleiding tot onbeantwoordbare vragen. Dat er oneindig veel priemen zijn waarin het decimale cijfer 3 niet voorkomt is bijvoorbeeld een vermoeden dat zonder twijfel waar, maar met de huidige technieken volstrekt onbewijsbaar is. Door veel wiskundigen worden dergelijke eigenschappen van priemgetallen vanwege hun weinig intrinsieke karakter als minder interessant gezien. Hier staat tegenover dat iedere wiskundige zeer nieuwsgierig zou zijn naar een *methode* om dergelijke rare vermoedens te bewijzen.

Een in wiskundige zin 'natuurlijke' generalisatie van het door Dirichlet opgeloste probleem voor priemen in rekenkundige rijen krijgt men door de getallen $a + kn$ met $k \geq 0$ op te vatten als de waardenverzameling van het lineaire polynoom $a+nx$ voor positieve gehele x.

Opgave 5. Laat zien dat de waardenverzameling van een lineair polynoom met gehele coëfficiënten geen, één of oneindig veel priemgetallen bevat.

Als we polynomiale uitdrukkingen met gehele coëfficiënten van graad groter dan 1 nemen, dan is de stelling van Dirichlet niet van toepassing. Er zijn polynomiale uitdrukkingen, zoals $x^2 + x$ en $13x^2 + 91$, die maar heel weinig priemwaarden aannemen. Immers, $x^2 + x = x(x + 1)$ is voor alle $x > 1$ een samengesteld getal, en $13x^2 + 91 = 13(x^2 + 7)$ is altijd deelbaar door 13. We moeten ons dus beperken tot uitdrukkingen die als polynoom geen factoren hebben verschillend van ± 1.

Opgave 6. Laat zien dat $x^2 + x + 2$ geen polynomiale factoren heeft, maar voor geen enkele positieve gehele waarde van x priem is.

Men vermoedt dat iedere polynomiale uitdrukking $f(x)$ die niet zoals hierboven om een 'evidente reden' voor bijna alle x samengesteld is, voor oneindig veel verschillende waarden van x priem is. Anders gezegd: een polynoom met gehele coëfficiënten dat *irreducibel* is en twee verschillende priemwaarden aanneemt, neemt *oneindig* veel priemwaarden aan. Zo hebben we bijvoorbeeld de volgende bekende vermoedens voor graad 2 en 4.

Vermoeden 1. *Er zijn oneindig veel priemgetallen p van de vorm $p = x^2 + 1$. Sterker nog: er zijn oneindig veel priemgetallen p van de vorm $p = x^4 + 1$.*

Men kan verlangen dat niet een enkele, maar meerdere polynomiale uitdrukkingen in x *gelijktijdig* priem zijn. Neemt men hiervoor x en $x + 2$, dan krijgen we het al in de inleiding genoemde vermoeden omtrent *priemgetaltweelingen.*

Vermoeden 2. *Er zijn oneindig veel priemgetallen p waarvoor $p + 2$ priem is.*

Er zijn eindeloos veel variaties op bovenstaand vermoeden. Men kan priemgetallen zoeken waarvoor $p + 4$ priem is, of $p^2 + 1$. Priemgetallen p waarvoor $2p + 1$ priem

is heten *Sophie Germain-priemen*, naar aanleiding van een toepassing die Sophie Germain hiervoor vond in verband met de laatste stelling van Fermat. Ook populair zijn de *priemgetalkwartetten* als $(11, 13, 17, 19)$ en

$$10^{99} + 349781731, \quad 10^{99} + 349781733, \quad 10^{99} + 349781737, \quad 10^{99} + 349781739.$$

Van al deze varianten vermoedt men dat er oneindig veel voorbeelden zijn. Niemand weet echter op dit moment hoe dergelijke vermoedens bewezen kunnen worden.

Bij gebrek aan bewijzen bedient men zich wel van *heuristische* argumenten om de beschikbare numerieke gegevens te 'verklaren'. Deze 'argumenten' zijn vaak zeer overtuigend, en niemand verwacht dan ook dat de boven genoemde vermoedens *fout* zijn. Sterker nog, ook aan de generalisaties van deze vermoedens naar collecties van k polynomiale uitdrukkingen wordt niet getwijfeld: 'alles wat niet evident fout is zou waar moeten zijn'. De volgende opgaven geven een indruk van wat 'evident fout' betekent.

Opgave 7. Bepaal *alle* priemgetallen p waarvoor $p^2 + 1$ priem is.

Opgave 8. Bepaal *alle* priemgetallen p waarvoor $p^2 + 2$ priem is.

Opgave 9. Bepaal *alle* priemgetallen p waarvoor $p + 2$ en $p + 4$ *beiden* priem zijn.

De grondgedachte achter de meeste heuristieken is dat de priemgetalstelling ons vertelt dat een groot getal x priem is met 'kans' $1/\log x$. Letterlijk genomen is dat een onzinnige mededeling, zoals duidelijk wordt door voor x maar eens 10001 te nemen: de kans dat 10001 priem is, is nul, want er geldt $10001 = 73 \cdot 137$. Wat wel waar is, is dat voor grote x een aselect getrokken geheel getal tussen 1 en x priem is met kans ongeveer $1/\log x$.

Wil men nu bijvoorbeeld weten hoeveel priemgetallen $p < X$ er zijn waarvoor $p + 2$ ook weer priem is, dan schat men de kans dat voor een willekeurig gekozen getal $x \in [1, X]$ zowel x als $x + 2$ priem is op $1/\log^2 X$. Deze productkans is in feite de kans dat twee *onafhankelijk* gekozen getallen tussen 1 en X allebei priem zijn. Men veronderstelt dus dat het priem zijn van x weinig invloed zal hebben op de primaliteit van $x + 2$. Is $1/\log^2 X$ inderdaad de kans dat voor een willekeurig gekozen getal x tussen 1 en X zowel x als $x+2$ priem is, dan moet het totale aantal van waarden $x \in [1, X]$ waarvoor dit optreedt ongeveer gelijk zijn aan $\frac{X}{\log^2 X}$. In het bijzonder krijgt men oneindig veel priemgetaltweelingen door de limiet $X \to \infty$ te nemen. Een dergelijke natte-vinger-argument, dat niet veel met een bewijs te maken heeft, heet in iets respectabelere bewoordingen een heuristisch argument.

Opgave 10. Hoeveel priemen p van de vorm $p = x^2 + 1$ verwacht je op bovenstaande heuristische gronden tot aan X, voor X een groot getal?

De boven geschetste heuristiek is zeer grof, en niet in staat om eenvoudige obstructies tegen het bestaan van priemwaarden te onderkennen. Zo werkt dezelfde heuristiek voor het aantal priemgetallen p tot aan X waarvoor $p + 1$ priem is!

Een betere heuristiek krijgt men door voor alle kleine priemgetallen ℓ de distributie van de waarden modulo ℓ te bekijken. Voor priemgetaltweelingen kan men bijvoorbeeld opmerken dat, indien $x \in [1, X]$ priem is, de kans dat $x + 2$ priem is niet meer $1/\log X$ is. Immers, als $x > 2$ priem is weten we al direct dat $x+2$ oneven is, en dus niet deelbaar door 2. Dit maakt de kans dat $x + 2$ priem is 'twee keer

zo groot' als bij een willekeurig gekozen getal. Hier staat weer tegenover dat, als x priem is, de kans dat $x+2$ door 3 deelbaar is extra groot is. Immers, daar voor de helft van de priemen p de congruentie $p \equiv 1 \bmod 3$ geldt, is de kans dat $p+2$ *niet* door 3 deelbaar slechts $1/2$ in plaats van $2/3$. Dit leidt tot een 'correctiefactor' $3/4$ in het aantal priemgetaltweelingen tot aan X. Soortgelijke overwegingen modulo andere priemgetallen ℓ leiden ertoe dat het aantal priemgetaltweelingen tot aan X voor grote X geschat moet worden op $C\frac{X}{\log^2 X}$, waarbij de priemgetaltweeling-constante C gedefinieerd is door

$$C = 2 \prod_{\ell>2 \text{ priem}} \left(1 - \frac{1}{(\ell-1)^2}\right) \approx 1.3203236316$$

Deze schatting blijkt in redelijke overeenstemming te zijn met de beschikbare numerieke gegevens. Met zogenaamde *zeeftechnieken* kan men bewijzen dat het aantal priemgetaltweelingen tot aan X niet heel veel *groter* is – dat er oneindig veel zijn als we X naar oneindig laten gaan, blijft echter een vermoeden.

Naast polynomiale uitdrukkingen voor priemgetallen zijn er ook exponentiële uitdrukkingen voor priemgetallen die uitgebreid bestudeerd zijn. Ook hier zijn er slechts onbewezen vermoedens, en de heuristieken moeten met veel meer zorg worden toegepast. We beperken ons tot de twee bekendste voorbeelden.

Vermoeden 3. *Er zijn oneindig veel priemgetallen p van de vorm $p = 2^n - 1$.*

Priemgetallen van de vorm $M_n = 2^n - 1$ heten *Mersenne-priemgetallen*, naar de Franse monnik Marin Mersenne, die in de zeventiende eeuw een (niet geheel correct) lijstje van waarden van n gaf waarvoor $2^n - 1$ priem is. De kleinste waarden zijn $M_2 = 3$, $M_3 = 7$ en $M_5 = 31$. Het grootste bekende Mersenne-priemgetal

$$M_{3021377} = 2^{3021377} - 1 \qquad (909526 \text{ decimale cijfers})$$

heeft de eer het op dit moment (juli 1999) het grootste bekende priemgetal zijn.

Enigszins overmoedig zouden we kunnen denken dat $2^n - 1$ priem is met kans $1/\log(2^n - 1) \approx 1/(n \log 2)$, zodat we door alle n tot aan X af te lopen ongeveer

$$\sum_{n<X} \frac{1}{n \log 2} \approx \frac{\log X}{\log 2}$$

Mersennepriemen tegen moeten komen. Deze gedachte blijkt in het geheel niet in overeenstemming te zijn met de werkelijkheid: tot aan $M_{3021377}$ heeft men niet $\log M_{3021377}/\log 2 \approx 3021377$ maar slechts 38 Mersennepriemen gevonden. Onze heuristiek is dus kennelijk 'fout'.

De oorzaak van ons falen ligt in de speciale vorm van het 'grote getal' $2^n - 1$. Het blijkt namelijk dat $2^n - 1$ slechts priem kan zijn als n het is.

Opgave 11. Stel dat n deelbaar is door d. Bewijs: $2^n - 1$ is deelbaar door $2^d - 1$.

We kunnen bovenstaande heuristiek aanpassen door alleen over *priemgetallen* n te sommeren. Dat geeft wegens stelling 2 nog steeds een divergente som, hetgeen

vermoeden 3 'ondersteunt'. Omdat de som $\sum_p \frac{1}{p}$ extreem langzaam naar oneindig gaat – met de priemen tot aan 10^{50} komen we nog niet boven de 5 – zouden er maar heel weinig Mersennepriemen bekend moeten zijn. Dat we er nu 37 kennen is onder meer te danken aan uitgebreid rekenwerk op computers. Er is zelfs een Great Internet Mersenne Prime Search (GIMPS), waaraan iedereen met een PC mee kan doen [2]. Wat ook helpt is dat er voor getallen van deze vorm een tamelijk snelle *primaliteitstest* is, en dat de 'correctieconstante' in de heuristiek tamelijk groot is.

Opgave 12. Zij p een priemgetal. Laat zien dat $2^p - 1$ geen delers kleiner dan p heeft.

Opgave 13. Een priemgetal heet een 'repunit' als het decimaal volledig uit cijfers 1 bestaat. Voorbeelden: 11, 1111111111111111111, 11111111111111111111111. Hoeveel van zulke priemen verwacht je tot aan X, voor X groot?
[Of oneindig veel priemen repunits zijn is – je raadt het al – onbekend.]

Na het voorafgaande vermoeden komt het volgende misschien als een verrassing.

Vermoeden 4. *Er zijn maar eindig veel priemgetallen p van de vorm $p = 2^n + 1$.*

De priemgetallen in het laatste vermoeden heten *Fermat-priemgetallen*. Fermat vond namelijk dat voor $k = 0, 1, 2, 3, 4$ de getallen

$$F_k = 2^{2^k} + 1$$

steeds priem zijn. De corresponderende waarden zijn 3, 5, 17, 257 en 65537.

Opgave 14. Bewijs dat $2^{1000} + 1$ geen priemgetal is. Laat algemener zien dat $2^n + 1$ niet priem is als n niet een macht van 2 is.

Fermatpriemgetallen zijn van speciaal belang vanwege de stelling van Gauss die zegt dat een regelmatige p-hoek in het platte vlak te construeren is met passer en lineaal dan en slechts dan als het priemgetal p een Fermat-priemgetal is. Dat een regelmatige 65537-hoek op het oog niet van een cirkel te onderscheiden is maakt dit resultaat voor de wiskundige niet minder charmant. Fermat's vermoeden dat F_k altijd priem is werd door Euler ontkracht:

$$F_5 = 2^{32} + 1 = 4294967297 = 641 \cdot 6700417.$$

Er zijn op dit moment geen waarden $k \geq 5$ bekend waarvoor F_k *wel* priem is. Voor $5 \leq k \leq 23$ weten we dat F_k *geen* priemgetal is. De dubbel-exponentiële groei van de getallen F_k maakt het echter niet gemakkelijk om te rekenen met getallen F_k voor grote k. Van F_{13} is bijvoorbeeld de priemfactorisatie nog niet bekend.

Opgave 15. Geef een heuristiek voor vermoeden 4 gebaseerd op de priemgetalstelling.

Een van de redenen dat de problemen in deze sectie zo weerbarstig zijn, is dat ze op *additieve eigenschappen* van priemgetallen betrekking hebben. Uit het feit dat p een priemgetal is – een bij uitstek *multiplicatieve* eigenschap – kan men namelijk erg weinig afleiden over multiplicatieve eigenschappen van $p - 1$, $p + 1$ of $p + 2$. Dit nu is precies waar onze 4 vermoedens betrekking op hebben!

Het bekendste puur additieve probleem betreffende priemgetallen is zonder twijfel het *Goldbach-vermoeden*, dat zegt dat ieder even getal een som van twee priemgetallen is. De numerieke evidentie is ook hier overweldigend: ieder groot even getal

heeft de neiging op vele manieren een som van twee priemgetallen te zijn. Met zeeftechnieken kon de Chinese wiskundige Chen Jing-Run bewijzen dat ieder voldoende groot even getal een som van een priemgetal en een getal met ten hoogste twee priemfactoren is. Ook hier is het gewenste eindresultaat echter nog niet in zicht.

Iedere vraag – opgelost of niet – geeft ook hier weer aanleiding tot talloze generalisaties. Zou bijvoorbeeld ook ieder even getal n het *verschil* van twee priemgetallen zijn? Het ligt voor de hand te denken dat dit zo is, en wel op oneindig veel manieren. Voor $n = 2$ is dit precies het priemgetaltweelingen-vermoeden.

5. Primitieve wortels

Een enigszins algebraïsche vraag betreffende priemgetallen die weliswaar onopgelost is, maar waarop het antwoord gegeven kan worden onder aanname van de zogenaamde *gegeneraliseerde Riemann-hypothese*, is het *Artin-vermoeden* voor primitieve wortels modulo een priemgetal. Het vermoeden heeft betrekking op de *vermenigvuldiggroep* $(\mathbf{Z}/p\mathbf{Z})^*$ van gehele getallen modulo een priemgetal p. Deze groep is een eenvoudig voorbeeld van wat in de groepentheorie een *eindige abelse groep* wordt genoemd. Hij is zeer geschikt om kennis te maken met de heel algemene uitspraken die de groepentheorie doet. Bewijzen van de stellingen in deze sectie vindt men dan ook in inleidende algebra-boeken [1].

Modulo een priemgetal $p$blijkt de verzameling $(\mathbf{Z}/p\mathbf{Z})^* = \{\overline{1}, \overline{2}, \overline{3}, \ldots, \overline{p-1}\}$ van restklassen verschillend van $\overline{0}$ altijd te bestaan uit de machten van een enkele klasse. Nemen we bijvoorbeeld $p = 7$, dan hebben we

$$(\mathbf{Z}/7\mathbf{Z})^* = \{3, \quad 3^2 = 2, \quad 3^3 = 6, \quad 3^4 = 4, \quad 3^5 = 5, \quad 3^6 = 1\} = \{1, 2, 3, 4, 5, 6\}.$$

Een dergelijke 'voortbrengende restklasse', zoals hier 3 mod 7, heet een *voortbrenger* van $(\mathbf{Z}/7\mathbf{Z})^*$. Men zegt ook wel dat 3 een *primitieve wortel* is modulo 7.

Opgave 16. Laat zien dat 5 ook een primitieve wortel is modulo 7.

Opgave 17. Laat zien dat 3 een primitieve wortel is modulo 10, en dat er geen primitieve wortels bestaan modulo 8 en modulo 12.

Dat er – anders dan modulo willekeurige getallen n – modulo ieder priemgetal p een primitieve wortel bestaat werd ontdekt door Euler. In de groepentheorie formuleert men zijn resultaat als volgt.

Stelling 5. *De vermenigvuldiggroep $(\mathbf{Z}/p\mathbf{Z})^*$ modulo een priemgetal p is cyclisch.*

Meestal zijn er meerdere primitieve wortels modulo p. Van de restklassen a modulo p die geen voortbrengers zijn kan men eveneens onderzoeken waar de verzameling $\langle\overline{a}\rangle$ van machten van $\overline{a}$ in $(\mathbf{Z}/p\mathbf{Z})^*$ uit bestaat. In $(\mathbf{Z}/7\mathbf{Z})^*$ heeft men bijvoorbeeld

$$\langle\overline{2}\rangle = \{2, \quad 4, \quad 1\} = \{3^2, \quad 3^4, \quad 3^6\}.$$

De deelverzameling $\langle\overline{a}\rangle \subset (\mathbf{Z}/p\mathbf{Z})^*$ van machten van $\overline{a}$ is gesloten onder vermenigvuldiging:

$$\overline{x}, \overline{y} \in \langle\overline{a}\rangle \quad \Longrightarrow \quad \overline{x} \cdot \overline{y} \in \langle\overline{a}\rangle.$$

Het is, in de terminologie van de groepentheorie, een *ondergroep* van $(\mathbf{Z}/p\mathbf{Z})^*$.

Het aantal elementen van $\langle\overline{a}\rangle$ heet de *orde* van a modulo p. Merk op dat de restklassen van de primitieve wortels modulo p precies de elementen van orde $p-1$ zijn. Door enig experimenteren ontdekt men spelenderwijs de volgende stelling.

Stelling 6. *De orde van een element a modulo p is een deler van $p-1$. Is a een primitieve wortel modulo p, dan is de orde van a^i modulo p gelijk aan $(p-1)/d$, met d de grootste gemeenschappelijke deler van i en $p-1$.*

In de groep $(\mathbf{Z}/7\mathbf{Z})^*$ hebben de primitieve wortels 3 en 3^5 orde 6, terwijl 3^2 en 3^4 orde 3 hebben. De orde van $3^3 = -1$ en $3^6 = 1$ zijn gelijk aan respectievelijk 2 en 1.

De cyclische structuur van de groep $(\mathbf{Z}/p\mathbf{Z})^*$ is zeer behulpzaam bij het bewijzen van allerhande resultaten over priemgetallen – bijvoorbeeld die in de 'sterretjesopgaven' 2 en 12.

Opgave 18. Bepaal de orde van elk van de elementen van $(\mathbf{Z}/19\mathbf{Z})^*$.

Opgave 19. Bepaal de kleinste primitieve wortel modulo 41.

Opgave 20. Laat zien dat de orde van a in $(\mathbf{Z}/p\mathbf{Z})^*$ een deler is van $(p-1)/d$ dan en slechts dan als er een element b bestaat met $a \equiv b^d \bmod p$.

Als we een vast priemgetal p nemen, is het niet moeilijk om alle primitieve wortels modulo p te bepalen. Nemen we omgekeerd een vast getal a, dan is het moeilijker om de verzameling van priemgetallen p te beschrijven waarvoor a een primitieve wortel is modulo p.

Artin-vermoeden voor primitieve wortels. *Zij $a > 1$ een geheel getal dat geen kwadraat is. Dan is a een primitieve wortel modulo oneindig veel priemgetallen p.*

Er is geen enkele waarde van a waarvoor we dit in 1927 door Artin geformuleerde vermoeden kunnen bewijzen. Er zijn wel zwakkere resultaten die er enigszins in de buurt komen, zie [5].

Ook voor Artin's vermoeden kan men een heuristisch resultaat formuleren, dat laat zien dat de verzameling priemen p met de eigenschap dat a een primitieve wortel modulo p is oneindig is, en zelfs een *positieve dichtheid* heeft. Hiertoe merkt men op dat het 'ongeluk' dat a geen primitieve wortel modulo p is zich dan en slechst dan vooordoet als een priemgetal ℓ bestaat zodat de orde van a een deler is van $\frac{p-1}{\ell}$. Voor vaste a en ℓ is de heuristische 'kans' dat de orde van a modulo p zo'n factor ℓ mist gelijk aan $\frac{1}{\ell(\ell-1)}$. Immers, er moeten zich gelijktijdig twee 'toevalligheden' voordoen. Allereerst moet $p-1$ deelbaar zijn door ℓ. De kans dat $p \equiv 1 \bmod \ell$ geldt is wegens de stelling van Dirichlet gelijk aan $\frac{1}{\ell-1}$. Ten tweede moet het zo zijn dat dat $\overline{a}$ in de ondergroep van $\frac{p-1}{\ell}$ elementen in $(\mathbf{Z}/p\mathbf{Z})^*$ ligt met orde een deler van $\frac{p-1}{\ell}$. Kans hierop: $\frac{1}{\ell}$.

Voor vaste ℓ concluderen we dat een 'fractie' $\frac{1}{\ell(\ell-1)}$ van de priemgetallen p de eigenschap dat de orde van $a \bmod p$ een factor ℓ van het maximum $p-1$ afzit. Gooien we al deze fracties weg, dan blijft een fractie

$$A = \prod_{\ell \text{ priem}} \left(1 - \frac{1}{\ell(\ell-1)}\right) \approx 0.3739558136$$

van de priemen over. Voor deze priemen is a een primitieve wortel. Tot aan X moeten we dus wegens de priemgetalstelling ongeveer $A \cdot \frac{X}{\log X}$ priemen p verwachten waarvoor a een primitieve wortel is modulo p. Hierin is A de zojuist gedefinieerde *Artin-constante*.

Onder aanname van de gegeneraliseerde Riemann-hypothese, die wij verder niet uit zullen leggen, kan men bewijzen dat bovenstaande heuristiek correct is voor veel waarden van a, zoals $a = 2$ en $a = 3$. Voor sommige a zijn kleine aanpassingen nodig. Een enigszins subtiele aanpassing, die men aanvankelijk over het hoofd zag, werd op grond van numeriek materiaal ontdekt [4].

Literatuur.

1. M. A. Armstrong, *Groups and symmetry*, Springer Undergraduate Text, 1988.
   [Een goed leesbare inleiding tot de groepentheorie.]
2. C. K. Caldwell, *The Prime Pages – prime number research, records and resources*, webpagina http://www.utm.edu/research/primes/.
   [Deze *website* is een goede introductie tot de overvloed aan informatie, numerieke gegevens en record-lijsten betreffende priemgetallen die op internet te vinden is.]
3. J. H. Conway, R. K. Guy, *The Book of Numbers*, Copernicus, Springer Verlag, 1996.
   [Speels, fraai geïllustreerd boek over getallen in de breedste zin des woords.]
4. D. H. Lehmer en E. Lehmer, *Heuristics, anyone?*, Studies in mathematical analysis and related topics, Stanford Univ. Press, 1962, pp. 202–210.
5. M. Ram Murty, *Artin's conjecture for primitive roots*, Math. Intelligencer **10/4** (1988), 59–67.
6. D. Zagier, *Elemente der Mathematik, Beiheft Nr. 15*, Birkhäuser Verlag, 1977.

FACULTEIT WINS
UNIVERSITEIT VAN AMSTERDAM
PLANTAGE MUIDERGRACHT 24
1018 TV AMSTERDAM
E-MAIL: psh@wins.uva.nl

Vakantiecursus 1999

pp. 31 – 44

De Riemann-Hypothese en het ABC-Vermoeden

R. Tijdeman
Mathematisch Instituut
Postbus 9512
2300 RA Leiden
e-mail:tijdeman@math.leidenuniv.nl

Deze inleiding bestaat uit drie delen. In het eerste deel wordt de Riemann-hypothese uitgelegd en het verband met de verdeling van de priemgetallen aangeduid. Als voorkennis wordt hierbij aangenomen dat de lezer weet wat priemgetallen en wat complexe getallen zijn. Het tweede deel gaat over de gegeneraliseerde Riemann-hypothese en de verde- ling van priemgetallen over rekenkundige rijen. In dit deel worden congruenties, grootste gemene delers en eenheidswortels gebruikt. Het laatste deel staat vrijwel los van de eerste twee en heeft voornamelijk betrekking op vergelijkingen in gehele getallen. Er worden enkele resultaten genoemd die verkregen zijn nadat Wiles met hulp van Taylor er in slaagde de laatste stelling van Fermat te bewijzen. Er worden enkele vermoedens geformuleerd, waaronder het abc-vermoeden, en samenhang tussen deze vermoedens wordt aangegeven. In dit deel wordt alleen het begrip priemgetal bekend verondersteld.

In de Appendix, die alleen betrekking heeft op het tweede deel, wordt het verband tussen karakters en eenheidswortels uitgewerkt. Karakters worden gebruikt in de definitie van L-functies. De gegeneraliseerde Riemann-hypothese is een uitspraak over de nulpunten van L-functies. De appendix bevat goede aanknopingspunten voor eigen onderzoek van de wonderlijke samenhang tussen gehele getallen, die de grondslag vormt van een diepgaande theorie.

Voor meer informatie over de eerste twee delen verwijs ik naar [D] en [A] en voor het derde deel naar [B2].

1. De Riemann-hypothese (RH)

Door velen wordt de Riemann-hypothese als het belangrijkste open probleem in de getaltheorie beschouwd, misschien wel in de hele wiskunde. Feit is dat de RH, zoals het vermoeden vaak wordt afgekort, sinds 1859 velen heeft gefascineerd. In dat jaar verscheen een kort artikel van Riemann [R] waarin hij de zeta-functie onderzocht en een vermoeden over haar nulpunten formuleerde. Uit Riemann's analyse bleek dat zijn hypothese belangrijke gevolgen heeft voor de verdeling van priemgetallen. Sindsdien zijn veel eigenschappen van getallen alleen bewezen onder aanname van (een generalisatie van) de RH, of soms juist

onder de aanname dat de RH niet waar is. Zo droeg Riemann's enige artikel over getaltheorie meer bij dan het levenswerk van veel andere getaltheoretici.

Euler (1707-1783) had al een verband tussen priemgetallen en de zetafunctie opgemerkt. De Hoofdstelling van de Rekenkunde zegt dat elk positief geheel getal op precies één manier te schrijven is als product van priemgetallen, als we niet op de volgorde letten. Daarbij spreken we af dat 1 het product is van nul priemgetallen, en een priemgetal het product van één priemgetal. Het betekent dat elk positief geheel getal $n > 1$ op precies één manier te schrijven is als $p_1^{k_1} p_2^{k_2} ... p_r^{k_r}$ waarbij $p_1 < p_2 < ... < p_r$ priemgetallen zijn en $k_1, k_2, ..., k_r$ positieve gehele getallen. Voor een willekeurig reëel getal x volgt dat

$$n^x = p_1^{k_1 x} p_2^{k_2 x} ... p_r^{k_r x}.$$

Wellicht herinnert u zich nog dat $\sum_{n=1}^{\infty} \frac{1}{n^x}$ convergeert als $x > 1$ en divergeert als $x \leq 1$. (De logaritmische reeks $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ...$ heeft som ∞, maar $1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + ...$ heeft de eindige som $\frac{\pi^2}{6}$). Voor $x > 1$ geldt daarom

$$\begin{aligned}(1 + \frac{1}{2^x} + \frac{1}{2^{2x}} + \frac{1}{2^{3x}} + ...)(1 + \frac{1}{3^x} + \frac{1}{3^{2x}} + ...)(1 + \frac{1}{5^x} + \frac{1}{5^{2x}} + ...)(1 + \frac{1}{7^x} + ... \\ = 1 + \frac{1}{2^x} + \frac{1}{3^x} + \frac{1}{2^{2x}} + \frac{1}{5^x} + \frac{1}{2^x 3^x} + \frac{1}{7^x} + \frac{1}{2^{3x}} + \frac{1}{3^{2x}} + \frac{1}{2^x 5^x} + ... \\ = \sum_{n=1}^{\infty} \frac{1}{n^x}.\end{aligned}$$

We schrijven hiervoor kortweg

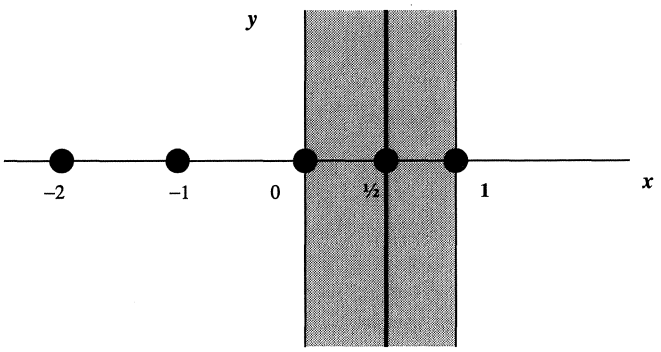
$$\prod_p (1 + \frac{1}{p^x} + \frac{1}{p^{2x}} + \frac{1}{p^{3x}} + ...) = \sum_{n=1}^{\infty} \frac{1}{n^x}. \tag{1}$$

De functie $\sum_{n=1}^{\infty} \frac{1}{n^x}$ noemen we nu de *Riemann-zeta functie* en noteren we met $\zeta(x)$. Formule (1), het zg. *Euler-product*, geeft aan dat $\zeta(x)$ nauw met de priemgetallen samenhangt. Als $x > 0$ dan is $\frac{1}{p^x} < 1$ en kunnen we nog de sommatie-formule $1 + y + y^2 + y^3 + ... = \frac{1}{1-y}$ voor $|y| < 1$ gebruiken om het Eulerproduct als volgt te schrijven:

$$\zeta(x) = \prod_p \frac{1}{1 - p^{-x}} \qquad (x > 1). \tag{2}$$

Riemann (1826-1866) beschouwde de zeta-functie niet als reële functie, maar als complexe functie. In de eerste helft van de 19e eeuw hadden met name Cauchy en Weierstraß de complexe functietheorie ontwikkeld, waarbij het domein niet reële getallen x betreft, maar complexe getallen $z = x + iy$ waarbij x en y reële getallen zijn en $i^2 = -1$. Riemann kon daarom in plaats van (2) schrijven dat

$$\zeta(z) = \prod_p \frac{1}{1 - p^{-z}} \qquad (x > 1), \tag{3}$$



FIGUUR 1. De kritieke strip en de kritieke as

waardoor de zeta-functie in het halfvlak rechts van de lijn $x = Re\ z = 1$ geheel werd vastgelegd. Uit de complexe functietheorie volgde dat deze functie analytisch is en maar op één manier analytisch voortgezet kan worden tot het gehele complexe vlak met uitzondering van het punt $z = 1$, waar zich een enkelvoudige pool met residu 1 bevindt. Eenvoudiger gezegd, $\zeta(1) = \infty$, zoals we al wisten, maar $(z-1)\zeta(z)$ is een functie die, als we voor $z = 1$ de waarde 1 lezen, voor elke waarde van z willekeurig vaak differentieerbaar is. Verder volgt uit (3) dat $\zeta(z) \neq 0$ voor $x > 1$.

Riemann's fundamentele ontdekking was dat er een eenvoudig verband is tussen de functiewaarden $\zeta(z)$ en $\zeta(1-z)$. Zo'n verband heet een *functionaalvergelijking*. We kennen dat van de cosinusfunctie: $\cos z = \cos(-z)$ en van de sinusfunctie: $\sin z = \sin(\pi - z)$. Het is niet zo dat $\zeta(z) = \zeta(1-z)$. In de functionaalvergelijking voor $\zeta(z)$ speelt ook de Gamma-functie $\Gamma(z)$ een rol, een complexe generalisatie van de faculteiten: $\Gamma(n) = (n-1)!$ voor $n = 1, 2, \ldots$. De *functionaalvergelijking van Riemann* zegt dat

$$\pi^{-\frac{1}{2}z}\Gamma(\frac{1}{2}z)\zeta(z) = \pi^{-\frac{1}{2}(1-z)}\Gamma(\frac{1}{2}(1-z))\zeta(1-z). \tag{4}$$

Hieruit zijn de nulpunten van $\zeta(z)$ voor $x < 0$ gemakkelijk te bepalen. Omdat $\Gamma(z) = \infty$ voor $z = 0, -1, -2, \ldots$ en nergens anders, volgt dat $\zeta(z) = 0$ voor $z = -2, -4, -6, \ldots$ en voor geen andere waarde van z met $x < 0$. De nulpunten $-2, -4, -6, \ldots$ heten de *triviale nulpunten* van $\zeta(z)$. Er moeten echter nog oneindig veel meer nulpunten zijn en deze moeten dus wel in de *kritieke strip* $0 \leq x \leq 1$ liggen.

Uit de functionaalvergelijking volgt dat als $\frac{1}{2} + x + iy$ een nulpunt van $\zeta(z)$ is, ook $\frac{1}{2} - x + iy, \frac{1}{2} + x - iy$ en $\frac{1}{2} - x - iy$ nulpunten van $\zeta(z)$ zijn. We hoeven dus alleen naar het bovenhalfvlak te kijken en de nulpunten komen in paren, tenzij ze op de *kritieke as* $x = \frac{1}{2}$ liggen. De RH zegt nu dat *alle* niet-trivale nulpunten van $\zeta(z)$ op deze lijn liggen. Inmiddels is dit, door van de Lune, Te Riele en Winter [LRW] van het CWI te Amsterdam, voor de eerste 1.500.000.000 nulpunten $x + iy$ met $y > 0$ geverifieerd. In 1974 bewees Levinson dat tenminste een derde deel van de niet-trivale nulpunten van $\zeta(z)$

op de kritieke as ligt. De eerste nulpunten op de kritieke as vanaf $\frac{1}{2}$ zijn bij benadering:

$$\frac{1}{2} \pm 14,13i, \ \frac{1}{2} \pm 21,02i.$$

Wat is nu het verband tussen de RH en de priemgetallen? Gauss 1777-1855) had al vermoed dat het aantal $\pi(x)$ van priemgetallen $\leq x$ groeit als $\frac{x}{\ln x}$ wanneer $x \to \infty$ en Chebyshev had al omstreeks 1859 aangetoond dat $0,92\frac{x}{\ln x} < \pi(x) < 1,11\frac{x}{\ln x}$ voor x groot genoeg en dat als $\pi(x)$ uiteindelijk voor elke $\varepsilon > 0$ tussen $(c-\varepsilon)\frac{x}{\ln x}$ en $(c+\varepsilon)\frac{x}{\ln x}$ ingeklemd zou worden, het getal c gelijk aan 1 moet zijn. Het probleem om de *priemgetalstelling*, $\pi(x) \sim \frac{x}{\ln x}$, d.w.z. dat voor elke $\varepsilon > 0$ en x groot genoeg $\pi(x)$ ingeklemd wordt tussen $(1-\varepsilon)\frac{x}{\ln x}$ en $(1+\varepsilon)\frac{x}{\ln x}$, te bewijzen was dus aan te tonen dat $\pi(x)$ voldoende regelmatig groeit. Dit nu is equivalent met aan te tonen dat $\zeta(1+iy) \neq 0$ voor $y > 0$. In 1896 werd dit laatste tegelijkertijd door Hadamard en door De la Vallée Poussin bewezen. Andere formuleringen van de priemgetalstelling zijn:

$$\begin{aligned} \pi(x) &\sim \int_2^x \frac{dt}{\ln t} =: \mathrm{li}x \qquad (x \to \infty), \\ \theta(x) &:= \sum_{p \leq x} \log p \sim x \quad (x \to \infty), \\ \psi(x) &:= \sum_{p^m \leq x} \log p \sim x \quad (x \to \infty). \end{aligned}$$

Een verband tussen $\psi(x)$ en de niet-trivale nulpunten ρ van $\zeta(z)$ met $(y > 0)$ wordt gegeven door de volgende formule van Von Mangoldt (1895):

$$\psi(x) - x = -\sum_{\rho} \frac{x^{\rho}}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\ln(1 - \frac{1}{x^2}) \quad (x \in \mathbf{R}_{>0}, x \notin \mathbf{Z}).$$

De volgende vraag was hoe regelmatig deze functies $\pi(x), \theta(x), \psi(x)$ groeien, m.a.w. hoe groot de restfuncties $\pi(x) - \mathrm{li}x, \theta(x) - x, \psi(x) - x$ kunnen worden. Uit von Mangoldt's formule blijkt dat dit nauw verbonden is met de ligging van de nulpunten van $\zeta(z)$. Laat a een getal zijn zó dat er nulpunten $z = x+iy$ van $\zeta(z)$ zijn met $x < a$ en $a - x$ willekeurig klein, maar geen met $x > a$. Dan geldt dat, voor elke $\varepsilon > 0$,

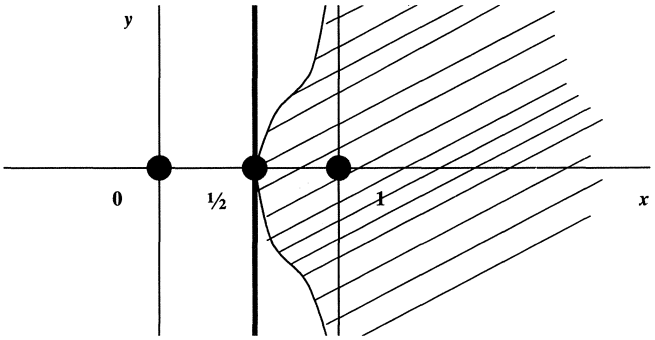
$$|\pi(x) - \mathrm{li}\ x| < x^{a+\varepsilon}$$

voor x voldoende groot, maar komt

$$|\pi(x) - \mathrm{li}\ x| > x^{a-\varepsilon}$$

voor willekeurig grote x voor. Als de RH waar is, geldt zelfs dat

$$|\pi(x) - \mathrm{li}x| > \frac{\sqrt{x}}{\ln x}\mathrm{lnlnln}x$$



FIGUUR 2. Nulpuntvrij gebied van ζ (s)

voor willekeurig grote x voorkomt. Echt regelmatig zijn de priemgetallen zeker niet verdeeld, maar als de RH geldt, gedragen de priemgetallen zich nog ongeveer zo als de wet van de grote aantallen voorschrijft. Als de RH niet geldt, is het gedrag opmerkelijk wild. Hilbert (1862-1943) dacht dat de RH eerder bewezen zou worden dan de laatste stelling van Fermat, zelfs nog tijdens zijn leven (cf. [S, blz. 84]). Het toont aan dat ook de groten der aarde niet in kunnen schatten hoe moeilijk onopgeloste problemen zijn.

Het resultaat dat het verste in de richting van de RH gaat is al weer 40 jaar oud. Richert bewees dat

$$\zeta(x+iy) \neq 0 \quad \text{voor} \quad x > 1 - \frac{c}{(\log|y|)^{\frac{2}{3}}(\log\log|y|)^{\frac{1}{3}}} \tag{5}$$

waarbij $c = \frac{1}{8757}$, als y groot genoeg is. Deze nulpuntvrije strook komt willekeurig dicht bij de lijn $x = 1$ als $y \to \infty$.

Uit (5) volgt een bovengrens voor de grootte van de afwijking van $\pi(x) - \mathrm{li}(x)$, nl.

$$|\pi(x) - \mathrm{li}(x)| < x\exp(-(\ln x)^{\frac{3}{5}}/(\ln\ln x))$$

voor x voldoende groot.

2. De gegeneraliseerde Riemann-hypothese (GRH)

Een *rekenkundige rij* is een rij gehele getallen van de vorm $a, a+d, a+2d, \ldots$. We veronderstellen dat a en d positief zijn; a heet de *beginterm*, d het *verschil*. Als a en d een deler $b > 1$ gemeen hebben, is elke term van de rij door b deelbaar. Zo'n rij bevat geen ander priemgetal dan eventueel b, nl. in geval $b = a$ een priemgetal is. Als a en d onderling ondeelbaar zijn, dan bevat de rij oneindig veel priemgetallen. Dit werd door Dirichlet (1805-1859) aangetoond. In zijn bewijs gebruikte Dirichlet een generalisatie van de zeta-functie, de zg. *Dirichlet-reeksen* of *L-reeksen* $L(z, \chi)$.

De Griekse letter chi wordt hier gebruikt om een zg. karakter modulo d aan te geven. Een *karakter* modulo d is een functie $\chi : \mathbf{Z} \to \mathbf{C}$ met de eigenschappen

(i) $\chi(1) = 1$
(ii) $\chi(n+d) = \chi(n)$ voor alle n
(iii) $\chi(n) = 0$ als ggd$(n,d) = 1$
(iv) $\chi(mn) = \chi(m)\chi(n)$ voor alle m en n.

Vanwege eigenschap (ii) hoeven we maar d opeenvolgende waarden van χ te geven om de functie vast te leggen. Nemen we $d = 2$, dan vinden we één karakter: $\chi_2(1) = 1, \chi_2(2) = 0$. Nemen we $d = 3$, dan vinden we $\chi(1) = 1, \chi(3) = 0, \chi(4) = 1$ en dus $(\chi(2))^2 = 1$. Dit levert twee karakters:

$$\begin{array}{lllll} \chi_{3,1}: & \chi(1)=1, & \chi(2)=1, & \chi(3)=0; \\ \chi_{3,2}: & \chi(1)=1, & \chi(2)=-1, & \chi(3)=0. \end{array}$$

Als $d = 4$, vinden we $\chi(1) = 1, \chi(2) = \chi(4) = 0, \chi(9) = (\chi(3))^2 = 1$ en dus weer twee karakters:

$$\begin{array}{llllll} \chi_{4,1}: & \chi(1)=1, & \chi(2)=0, & \chi(3)=1, & \chi(4)=0; \\ \chi_{4,2}: & \chi(1)=1, & \chi(2)=0, & \chi(3)=-1, & \chi(4)=0. \end{array}$$

Nemen we $d = 5$, dan vinden we $\chi(1) = 1, \chi(5) = 0, \chi(16) = 1$ dus $(\chi(2))^4 = 1$. Hieruit volgt $\chi(2) = 1$ of -1 of i of $-i$. Dit levert vier karakters op:

$$\begin{array}{llllll} \chi_{5,1}: & \chi(1)=1, & \chi(2)=1, & \chi(3)=1, & \chi(4)=1, & \chi(5)=0; \\ \chi_{5,2}: & \chi(1)=1, & \chi(2)=-1, & \chi(3),=-1, & \chi(4)=1, & \chi(5)=0; \\ \chi_{5,3}: & \chi(1)=1, & \chi(2)=i, & \chi(3)=-i, & \chi(4)=-1, & \chi(5)=0; \\ \chi_{5,4}: & \chi(1)=1; & \chi(2)=-i, & \chi(3)=i, & \chi(4)=-1, & \chi(5)=0. \end{array}$$

Een karakter $\overline{\chi}$ heet *geconjugeerd* aan χ als $\overline{\chi}(n) = \overline{\chi(n)}$ voor alle n. Zo is $\chi_{5,4}$ geconjugeerd aan $\chi_{5,3}$. De andere genoemde karakters nemen alleen reële waarden aan en zijn dus gelijk aan hun geconjugeerde.

Een karakter χ modulo d leidt tot een karakter χ^* modulo dm voor elk geheel getal $m > 1$ volgens

$$\chi^*(n) = \begin{cases} \chi(n) & \text{als ggd } (n, md) = 1 \\ 0 & \text{als ggd } (n, md) > 1 \end{cases}$$

Zulke karakters χ^* noemen we niet-primitief. Zo leidt $\chi_{2,1}$ tot $\chi_{3,1}, \chi_{4,1}$ en $\chi_{5,1}$. Karakters die alleen waarden 0 en 1 aannemen noemen we *hoofdkarakters*. Verder leidt $\chi_{3,2}$ bijvoorbeeld tot het volgende karakter modulo 6.

$$\chi_{6,2}: \chi(1) = 1, \chi(2) = 0, \chi(3) = 0, \chi(4) = 0, \chi(5) = -1, \chi(6) = 0.$$

Een karakter dat niet op deze manier uit een karakter met kleinere modulus gedefinieerd kan worden heet *primitief*. De karakters $\chi_{3,2}, \chi_{4,2}, \chi_{5,2}, \chi_{5,3}$ en $\chi_{5,4}$ zijn primitief. Bij de moduli 2 en 6 bestaan geen primitieve karakters.

De L-functie die bij het karakter χ hoort wordt gegeven door

$$L(z, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

Omdat volgens (iv) geldt dat $\frac{\chi(m)}{m^s} \cdot \frac{\chi(n)}{n^s} = \frac{\chi(mn)}{(mn)^s}$, hebben L-functies veel eigenschappen met de ζ-functie gemeen. In plaats van (3) krijgen we de Euler-product-formule

$$L(s,\chi) = \prod_p \frac{1}{1-(\chi(p)p^z)^{-1}} \quad (x>1).$$

Als χ niet een hoofdkarakter is, geldt de convergentie zelfs voor $x > 0$. De functionaalvergelijking voor een primitieve L-functie met $\chi(-1) = 1$ is (vgl. (4))

$$\pi^{-\frac{1}{2}z} d^{\frac{1}{2}z} \Gamma(\frac{1}{2}z) L(z,\overline{\chi}) = \tau\pi^{-\frac{1}{2}(1-z)} d^{\frac{1}{2}(1-z)} \Gamma(\frac{1}{2}(1-z)) L(1-z,\chi)$$

waarbij $\overline{\chi}$ het geconjugeerde karakter van χ is en τ een eenheidswortel. Als $\chi(-1) = -1$ is de vergelijking een beetje anders. Omdat $(\chi(-1))^2 = \chi(1) = 1$, geldt altijd $\chi(-1) = 1$ of $\chi(-1) = -1$. Ook $L(z,\chi)$ heeft geen nulpunten met $x > 1$. Ook de nulpunten van $L(z,\chi)$ met $x \leq 0$ liggen op de reële as en deze zg. triviale nulpunten worden gegeven door $0, -2, -4, -6, ..$ als $\chi(-1) = 1$ en door $-1, -3, -5, -7, ..$ als $\chi(-1) = -1$. Een verrassend onderscheid met $\zeta(z)$ is dat het bestaan van reële nulpunten van $L(z,\chi)$ met $0 < x < 1$ niet uitgesloten is. Het zal niet bij veel karakters voorkomen, maar er zouden enkele karakters χ kunnen bestaan waarvoor $L(z,\chi)$ een reëel nulpunt tussen 0 en 1 heeft, dat dan vlak bij 0 of vlak bij 1 ligt, een zg. *Siegel nulpunt*. Het bestaan van Siegel-nulpunten is nog een groot mysterie. De verdere theorie lijkt wel veel op die van de zeta-functie. In het bijzonder zegt de *gegeneraliseerde Riemann-hypothese* (GRH) dat alle niet-triviale nulpunten van alle L-functies op de kritieke as $x = \frac{1}{2}$ liggen. Veel resultaten uit de analytische getaltheorie zijn onder aanname van GRH of een variant bewezen. Zo bewees Hooley in 1967 onder aanname van een uitbreiding van de RH voor Dedekind zeta-functies dat het vermoeden van Artin over primitieve wortels waar is. De beste resultaten voor L-functies in de richting van GRH lijken op (5).

Uit de theorie van de L-functies corresponderend met karakters modulo d volgt dat niet alleen elke rekenkundige rij $a, a+d, a+2d, ...$ met ggd $(a,d) = 1$ oneindig veel priemgetallen bevat, maar ook dat de priemgetallen in gelijke verhoudingen over de rekenkundige rijen modulo d verdeeld zijn. Ongeveer de helft van de priemgetallen is dus van de vorm $4n+1$, de andere helft is van de vorm $4n+3$ (met als enige uitzondering het priemgetal 2). Van alle priemgetallen is een kwart van de vorm $8n+1$, een kwart van de vorm $8n+3$, een kwart van de vorm $8n+5$ en een kwart van de vorm $8n+7$. Eén achtste van alle priemgetallen is van de vorm $24n+5$, enz. Voor de resttermen gelden hetzelfde soort resultaten als voor priemgetallen zelf. Als de GRH geldt, is de verdeling over de verschillende rekenkundige rijen betrekkelijk regelmatig, anders onregelmatiger.

Wanneer komt het eerste priemgetal in een rekenkundige rij $a, a+d, a+2d, ...$, met $0 < a < d$ en a en d onderling ondeelbaar voor? Ju. V. Linnik

bewees in 1947 dat er constanten c_1 en c_2 zijn zó dat het eerste priemgetal in de rij kleiner is dan $c_1 d^{c_2}$. Het bepalen van de beste constante c_2 is een open probleem. Deze constante wordt wel *Linnik's constante* genoemd.

Een rekenkundige rij kun je opvatten als de lineaire functie $dx+a$. Voor hogeregraadspolynomen bestaan eigenlijk alleen vermoedens. Zo wordt vermoed dat er oneindig veel priemgetallen van de vorm x^2+1 met $x \in \mathbf{Z}$ bestaan.

3. Gegeneraliseerde Fermat-vergelijkingen en het *abc*-vermoeden

Sinds de laatste stelling van Fermat in 1670 bekend werd, is fervent naar een bewijs ervan gezocht. Uiteindelijk wist Wiles met hulp van Taylor in 1995 hiervoor een sluitende redenering op te stellen. Hij deed dit door het semi-stabiele geval van een centraal vermoeden uit de algebraïsche meetkunde te bewijzen, het Taniyama-Shimura-Weil vermoeden. (Dit vermoeden wordt ook vaak aangegeven met permutaties van de namen of weglating van één der namen.) Het is nu dus zeker dat er geen positieve gehele x, y, z, n met $n \geq 3$ bestaan zó dat

$$x^n + y^n = z^n.$$

Het is niet duidelijk waarom de exponenten aan elkaar gelijk moeten zijn. In de vakliteratuur komt men vaak vergelijkingen

$$ax^k + by^l = cz^m \tag{6}$$

tegen waarbij a, b, c gegeven coëfficiënten zijn en $k > 1, l > 1, m > 1, x \geq 1, y \geq 1, z \geq 1$ met ggd $(x, y, z) = 1$ onbekenden.

Zo vermoedde Catalan in 1842 dat $8 = 2^3$ en $9 = 3^2$ het enige paar van opeenvolgende natuurlijke getallen is die beide machten zijn. Dit correspondeert met de vergelijking

$$x^k + 1 = z^m \quad (x \geq 1, z \geq 1, k > 1, m > 1),$$

dus $a = b = c = y = 1$ in (6).

Er zijn oneindig veel rekenkundige rijen van drie kwadraten, bijv. $1, 25, 49$; $1, 1641$; $1, 28561, 57121$. Euler bewees een bewering van Fermat dat vier kwadraten geen niet-constante rekenkundige rij kunnen vormen.

De vraag of drie n-de machten met $n > 2$ een rekenkundige rij (met positief verschil) kunnen vormen is pas heel onlangs door Darmon en Merel [DM] beantwoord. Als x^n, z^n, y^n zo'n rij vormen, geldt $z^n - x^n = y^n - z^n$, ofwel

$$x^n + y^n = 2z^n.$$

Door een ingewikkelde uitbreiding van de methode van Wiles slaagden zij er in aan te tonen dat deze vergelijking geen oplossingen heeft. Ook voor sommige andere waarden van a is immiddels aangetoond dat

$$x^n + y^n = az^n \quad (x > 0, y > 0, z > 0, n > 2)$$

geen oplossingen heeft.

Ik zal me hier verder beperken tot de vergelijking

$$(7) \quad x^k + y^l = z^m \ (k > 1, l > 1, m > 1, x > 1, y > 1, z > 1, \text{ggd}(x, y, z) = 1).$$

Ik veronderstel nu verder dat $\frac{1}{k} + \frac{1}{l} + \frac{1}{m} < 1$. Zoals ik tijdens de Fermatdag in Utrecht in november 1993 onthulde, vonden Beukers en Zagier naast de reeds bekende kleine oplossingen nog vijf opmerkelijk grote oplossingen (zie [T]):

$$2^5 + 7^2 = 3^4, 7^3 + 13^2 = 2^9,$$
$$2^7 + 17^3 = 71^2, 3^5 + 11^4 = 122^2,$$
$$17^7 + 76271^3 = 21063928^2,$$
$$1414^3 + 2213459^2 = 65^7$$
$$9262^3 + 15312283^2 = 113^7$$
$$43^8 + 96222^3 = 30042907^2$$
$$33^8 + 1549034^2 = 15613^3$$

Toch wordt vermoed dat (7) maar eindig veel oplossingen met $\frac{1}{k} + \frac{1}{l} + \frac{1}{m} < 1$ heeft.

Ik liet ook zien (wat algemeen bekend was) dat het zonder de beperking ggd $(x, y, z) = 1$ heel gemakkelijk is om oplossingen te construeren. Tijdens mijn voordracht in Utrecht merkte ik op dat in alle negen oplossingen van (7) een kwadraat voorkomt en sprak ik het vermoeden uit dat er geen getallen $k \geq 3, l \geq 3, m \geq 3, x > 0, y > 0, z > 0$ met ggd $(x, y, z) = 1$ zijn zó dat

$$x^k + y^l = z^m.$$

Immiddels heeft de Amerikaanse bankier Andrew Beal ook dit vermoeden uitgesproken en aan het bewijs van dit vermoeden een geldprijs van tenminste \$10.000 verbonden [M].

Enige vooruitgang is inmiddels geboekt. In Leiden heeft Nils Bruin zg. Chabauty-technieken gebruikt om voor bepaalde drietallen (k, l, m) aan te tonen dat we nu alle oplossingen kennen, nl. alle permutaties van $(2, 4, 6)$, $(2, 4, 5)$ en $(2, 3, 8)$. Het ziet er naar uit dat zijn methode ook werkt voor $(2, 3, 9)$, maar zeker niet voor $(2, 3, 7)$. Nils hoopt in oktober op dit werk te promoveren.

Darmon en Merel [DM] en Poonen [P] hebben met een uitbreiding van de methode van Wiles aangetoond dat de vergelijking $x^n + y^n = z^2$ geen oplossingen heeft in positieve gehele getallen $n \geq 4, x, y, z$ met ggd $(x, y, z) = 1$.

Ook wat betreft het vermoeden van Catalan vinden gestaag vorderingen plaats. Nadat ik in 1976 bewees dat er maar eindig veel oplossingen van $x^m - y^n = 1$ kunnen zijn, heeft met name Mignotte de mogelijkheden flink ingeperkt. Als er nog een andere oplossing is dan 8 en 9, dienen de exponenten m, n volgens de meeste recente resultaten van Mignotte en Roy in te liggen tussen ongeveer 10^5 en 10^{20}.

Om een goed gevoel voor dit soort vergelijkingen te krijgen, is het nuttig het *abc*-vermoeden te kennen. Dit zegt dat als er drie natuurlijke getallen a, b en c zijn, onderling ondeelbaar, zó dat

$$a + b = c,$$

dat dan c kleiner is dan $c_\varepsilon N^{1+\varepsilon}$ waarbij N het product van de priemdelers van abc is, $\varepsilon > 0$ willekeurig en c_ε een geschikt getal dat alleen van ε afhangt. Ik vermoed dat bij $\varepsilon = 1$ de constante $c_\varepsilon = 1$ genomen kan worden zodat altijd

$$c < N^2.$$

Zo geldt bijvoorbeeld voor de oplossingen van (7):

$$\begin{array}{rcl}
2^5 + 7^2 = 3^4 & N = 2 \cdot 7 \cdot 3 = 42, & c = N^{1,176} \\
7^3 + 13^2 = 2^9 & N = 7 \cdot 13 \cdot 2, & c = N^{1,199} \\
2^7 + 17^3 = 71^2 & N = 2 \cdot 17 \cdot 71, & c = N^{1,095} \\
3^5 + 11^4 = 122^2 & N = 3 \cdot 11 \cdot 2 \cdot 61, & c = N^{1,158} \\
17^7 + \ldots & N \leq 17 \cdot 76271 \cdot 5265982, & c = N^{1,090} \\
1414^3 + \ldots & N \leq 1414 \cdot 2213459 \cdot 65, & c = N^{1,122} \\
9262^3 + \ldots & N \leq 9262 \cdot 15312283 \cdot 113, & c = N^{1,088} \\
43^8 + \ldots & N \leq 43 \cdot 96222 \cdot 30042907, & c = N^{0,927} \\
33^8 + \ldots & N \leq 33 \cdot 1549034 \cdot 15613, & c = N^{1,057}
\end{array}$$

De drie grootste exponenten die tot nog toe gevonden zijn, worden gegeven door:

Reyssat: $3^{10} \cdot 109 + 2 = 23^5$, exponent $= 1,62991$
De Weger: $3^2 \cdot 5^6 \cdot 7^3 + 11^2 = 2^{21} \cdot 23$, exponent $= 1,62599$
Browkin & Brzezinski:
$19 \cdot 1307 + 7 \cdot 29^2 \cdot 31^8 = 2^8 \cdot 3^{22} \cdot 5^4$, exponent $= 1,62349$.

Ondanks veel computerrekenwerk zijn de laatste vijf jaar geen verbeteringen gevonden en het vermoeden dat altijd $c < N^2$ is dus niet uit de lucht gegrepen. Als we deze variant van het abc-vermoeden voor waar aannemen, is het bewijzen van de laatste stelling van Fermat een fluitje van een cent. We mogen aannemen dat x, y en z onderling ondeelbaar zijn en dat $x^n + y^n = z^n$. Nemen we $a = x^n, b = y^n, c = z^n$, dan vinden we $a + b = c$ en $N \leq xyz$ en dus

$$z^n < N^2 \leq (xyz)^2 < z^6.$$

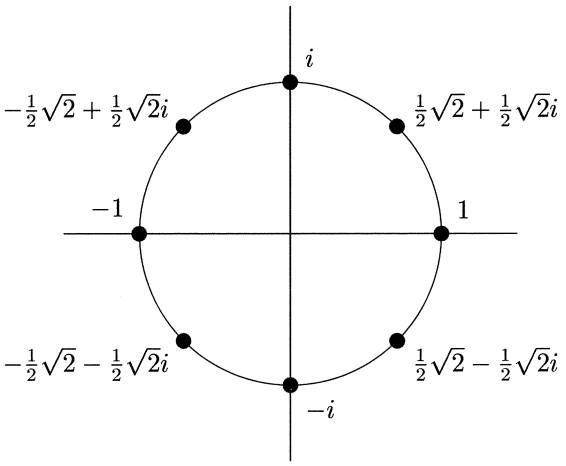
Hieruit volgt dat $n < 6$. Het is al meer dan 150 jaar bekend dat er geen oplossingen van $x^n + y^n = z^n$ zijn met $n = 3, 4$ of 5.

Het echte abc-vermoeden leidt op eenzelfde wijze tot de bewering dat (3.2) maar eindig veel oplossingen heeft. Immers, we nemen $a = x^k, b = y^l, c = z^m$ zodat $N < xyz$ en vinden met $\varepsilon = 1/83$ dat er een $c > 0$ is met

$$z^m < cN^{1+\frac{1}{83}} < c(xyz)^{\frac{84}{83}} < cz^{\frac{84}{83}(\frac{1}{k}+\frac{1}{l}+\frac{1}{m})m}.$$

Merk nu op dat $\frac{1}{k} + \frac{1}{l} + \frac{1}{m}$ voor k, l, m altijd ≥ 1 of $\leq \frac{41}{42}$ is, waarbij de waarde $\frac{41}{42}$ wordt aangenomen voor $k = 2, l = 3, m = 7$. Als $\frac{1}{k} + \frac{1}{l} + \frac{1}{m} < 1$, volgt dus $z^m < cz^{\frac{84}{83} \cdot \frac{41}{42} m}$, ofwel $z^{\frac{m}{83}} < c$. Hieruit volgt dat $z^m < c^{83}$ zodat er maar eindig veel oplossingen x^k, y^l, z^m kunnen zijn.

Inmiddels is aangetoond dat veel vermoedens uit de getaltheorie en algebraïsche meetkunde tot het abc-vermoeden herleid kunnen worden. Ondanks



FIGUUR 3.

pogingen daartoe is het nog niet gelukt aan te tonen dat de RH een gevolg is van het *abc*-vermoeden. Wel is onlangs door Granville en Stark aangetoond dat uit een generalisatie van een *abc*-vermoeden voor algebraïsche getallen volgt dat er geen Siegel-nulpunten bestaan. Het is verbijsterend om te zien hoe een pas in 1985 voor het eerst geformuleerd simpel vermoeden over de som van twee onderling ondeelbare getallen zulke verstrekkende gevolgen heeft.

4. APPENDIX: KARAKTERS.

Voordat we de structuur van karakters beschrijven, onderzoeken we eerst de structuur van m-de eenheidswortels. Ik herinner eraan dat optellen van complexe getallen de vectoroptelling in het x-y-vlak is. Vermenigvuldigen van complexe getallen komt in het x-y-vlak neer op het vermenigvuldigen van de lengtes van de vectoren om de lengte van het product te berekenen en het optellen van de hoeken met de positieve x-as om de richting van de productvector aan te geven. Als $z^m = 1$ met z complex en m geheel, dan moet de vector (x, y) met $z = x + iy$ dus lengte 1 hebben. Verder moet m maal de hoek die de vector (x, y) met de positieve x-as maakt een veelvoud $2k\pi$ van 2π radialen opleveren. De hoek is dus van de vorm $\frac{2k\pi}{m}$ en we vinden $z = \cos\frac{2k\pi}{m} + i\sin\frac{2k\pi}{m}$ voor $k = 0, 1, ..., m-1$. Het komt erop neer dat de complexe getallen z met $z^m = 1$ de punten van een regelmatige m-hoek om de oorsprong vormen met 1 als één van de hoekpunten. Voor $d = 4$ vinden we de vierde-eenheidswortels $\pm 1, \pm i$ corresponderend met $(x, y) = (1, 0), (-1, 0), (0, 1), (0, -1)$. De zesde-eenheidswortels zijn $\pm 1, \pm\frac{1}{2} \pm \frac{1}{2}i\sqrt{3}$ en de achtste-eenheidswortels $\pm 1, \pm i, \pm\frac{1}{2}\sqrt{2} \pm \frac{1}{2}i\sqrt{2}$.

Een *primitieve m-de eenheidswortel* is een eenheidswortel van de vorm $z = \cos\frac{2k\pi}{m} + i\sin\frac{2k\pi}{m}$ met ggd $(k, m) = 1$. Zo'n getal is niet de eenheidswortel voor een kleinere m en de machten $z, z^2, ..., z^m$ zijn precies alle m-de eenheidswortels. Zo'n primitieve eenheidswortel is een *voortbrenger* van de

m-de eenheidswortels. Het aantal primitieve eenheidswortels noteren we met $\varphi(m)$. Omdat we $k = 1$ en $k = m$ kunnen kiezen, geldt $0 < \varphi(m) < m$.

Opgave 1: Bereken $\varphi(m)$ voor $m = 7, 8, 9, 10, 11, 12$.

Elke functiewaarde van een karakter die niet 0 is, is een eenheidswortel. Dit is een gevolg van de stelling van Fermat-Euler die zegt dat $a^{\varphi(d)} \equiv 1 (\text{mod } d)$ als ggd $(a, d) = 1$. Uit (iv) volgt dat $(\chi(a))^{\varphi(d)} = \chi(a^{\varphi(d)}) = \chi(1) = 1$. Dus is $\chi(a)$ een $\varphi(d)$-de eenheidswortel. Omdat $\varphi(d) = d \prod_{p|d}(1 - \frac{1}{p})$, waarbij het product zich uitstrekt over alle priemdelers p van d, is $m = \varphi(d)$ gemakkelijk te bepalen als de priemontbinding van d bekend is.

Voor de structuur van karakters is het verder belangrijk te weten dat bij elke priemmacht $d = p^l (p$ priem, $l \in \mathbf{Z}_{>0})$ een getal g bestaat zó dat $g, g^2, ..., g^{\varphi(d)}$ modulo d precies alle getallen onderling ondeelbaar met d representeren, met als enige uitzonderingen de getallen $d = 2^l$ met $l \geq 3$. Zo'n g heet een *primitieve wortel* van $d = p^l$ en is vergelijkbaar met een primitieve $\varphi(d)$-de eenheidswortel. Bijv. voor $p = 5$ is 2 een primitieve wortel:

$$2^1 = 2, 2^2 = 4, 2^3 \equiv 3, 2^4 \equiv 1 (\text{mod } 5), \quad \varphi(5) = 4.$$

Als χ een karakter mod 5 is, is $\chi(2)$ dus een 4-de eenheidswortel en is het voldoende om $\chi(2)$ te weten, want de andere functiewaarden worden hierdoor vastgelegd: $\chi(4) = (\chi(2))^2, \chi(3) = (\chi(2))^3, \chi(1) = (\chi(2))^4 = 1$. Zo vinden we de vier karakters $\chi_{5,1}, \chi_{5,2}, \chi_{5,3}, \chi_{5,4}$.

Bijv. voor $p = 7$ is 3 een primitieve wortel (2 niet!), want

$$3^1 = 3, 3^2 \equiv 2, 3^3 \equiv 6, 3^4 \equiv 4, 3^5 \equiv 5, 3^6 \equiv 1 (\text{mod } 7), \varphi(7) = 6.$$

Een karakter χ modulo 7 is dus volledig bepaald door de zesde-eenheidswortel $\chi(3)$. Zo vinden we zes karakters mod 7. De enige twee primitieve karakters (mod 7), vinden we door voor $\chi(3)$ een primitieve zesde-eenheidswortel te nemen, $\frac{1}{2} \pm \frac{1}{2} i\sqrt{3}$.

Voor $p = 9$ is 2 een primitieve wortel (3, 4 en 7 niet). Er geldt $2^1 = 2, 2^2 = 4, 2^3 = 8, 2^4 \equiv 7, 2^5 \equiv 5, 2^6 \equiv 1 (\text{mod } 9)$ en $\varphi(9) = 6$. Een karakter χ modulo 9 is dus volledig bepaald door de zesde-eenheidswortel $\chi(2)$. Net als in het vorige voorbeeld zijn er zes karakters waarvan twee primitief.

Opgave 2 a): Zoek primitieve wortels van $4, 11, 13, 25$ en 27.
b) Beschrijf de karakters die bij deze moduli horen.

Voor machten 2^l van 2 is het belangrijk dat 5 alle resten $\equiv 1 (\text{mod } 4)$ voortbrengt. Voor resten $n \equiv 3 (\text{mod} 4)$ gebruiken we

$$\chi(n) = \chi(-1)\chi(-n)$$

met $-n \equiv 1 (\text{mod } 4)$. Omdat $(\chi(-1))^2 = \chi(1) = 1$, kunnen we voor $\chi(-1)$ alleen 1 en -1 kiezen. Voor $\chi(5)$ kunnen we een willekeurige 2^{l-2}-de eenheidswortel kiezen en alle andere functiewaarden zijn daardoor vastgelegd.

Bijv. $d = 2^3$. Er geldt $5^1 = 5, 5^2 \equiv 1 (\text{mod } 8), \varphi(8) = 4$,
De waarde van $\chi(5)$ moet dus 1 of -1 zijn. Ook de waarde van $\chi(-1)$ moet 1 of -1 zijn. Dit levert de volgende vier karakters modulo 8. (We laten functiewaarden 0 weg.):

$$\begin{array}{llllll}
\chi_{8,1} : & \chi(1) = 1, & \chi(3) = 1, & \chi(5) = 1, & \chi(7) = 1 & \text{(hoofdkarakter)} \\
\chi_{8,2} : & \chi(1) = 1, & \chi(3) = -1, & \chi(5) = -1, & \chi(7) = 1 & \text{(primitief)} \\
\chi_{8,3} : & \chi(1) = 1, & \chi(3) = -1, & \chi(5) = 1 & \chi(7) = -1 & \text{(karakter mod.4)} \\
\chi_{8,4} : & \chi(1) = 1, & \chi(3) = 1, & \chi(5) = -1, & \chi(7) = -1 & \text{(primitief)}
\end{array}$$

Opgave 3: Bepaal de acht karakters modulo 16.

Als $d = p_1^{k_1} p_2^{k_2} ... p_r^{k_r}$ de priemontbinding van d met $p_1 < p_2 < ... < p_r$ is, dan worden de $\varphi(d)$ karakters modulo d verkregen door karakters χ_1 mod $p_1^{k_1}, \chi_2$ mod $p_2^{k_2}, ..., \chi_r$ mod $p_r^{k_r}$ te vermenigvuldigen en vervolgens waar nodig 0 te stellen:

$$\chi(n) = \begin{cases} \chi_1(n)\chi_2(n)...\chi_r(n) & \text{als ggd } (n,d) = 1 \\ 0 & \text{anders.} \end{cases}$$

Zo vinden we de volgende acht karakters mod 24:

	$\chi(1)$	$\chi(5)$	$\chi(7)$	$\chi(11)$	$\chi(13)$	$\chi(17)$	$\chi(19)$	$\chi(23)$
$\chi_{3,1}\chi_{8,1}$:	1	1	1	1	1	1	1	1
$\chi_{3,1}\chi_{8,2}$:	1	−1	1	−1	−1	1	−1	1
$\chi_{3,1}\chi_{8,3}$:	1	1	−1	−1	1	1	−1	−1
$\chi_{3,1}\chi_{8,4}$:	1	−1	−1	1	−1	1	1	−1
$\chi_{3,2}\chi_{8,1}$:	1	−1	1	−1	1	−1	1	−1
$\chi_{3,2}\chi_{8,2}$:	1	1	1	1	−1	−1	−1	−1
$\chi_{3,2}\chi_{8,3}$:	1	−1	−1	1	1	−1	−1	1
$\chi_{3,2}\chi_{8,4}$:	1	1	−1	−1	−1	−1	1	1

Het eerste karakter is het hoofdkarakter, de volgende drie komen van karakters mod 8 (de derde zelfs mod 4), het vijfde karakter komt van $\chi_{3,2}$. Het zevende karkater komt van een karakter mod 12. De andere twee karakters zijn primitief modulo 24.

Opgave 4: a) Bepaal alle karakters modulo 12, modulo 18 en modulo 20.
b) Bepaal de primitieve karakters modulo 12, modulo 18 en modulo 20.

5. Literatuur

[A] T.M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, New York etc., 1976.

[B1] F. Beukers, *The diophantine equations* $Ax^p + By^q = Cz^r$, Duke Math. J. **91** (1998), 61-68.

[B2] F. Beukers, *Getaltheorie voor Beginners*, Epsilon Uitgaven, Utrecht, 1999.

[D] H. DAVENPORT, *Multiplicative Number Theory*, Springer-Verlag, New York etc., 2e druk, 1980.

[DM] H. DARMON AND L. MEREL, *Winding quotiens and some variants of Fermat's Last Theorem*, J. reine angew. Math. **490** (1997), 81-100.

[H] C. HOOLEY, *On Artin's conjecture*, J. reine angew. Math. **225** (1967), 209-220.

[LRW] J. VAN DE LUNE, H.J.J. TE RIELE AND D.T. WINTER, *On the zeros of the Riemann zeta function in the critical strip IV*, Math. Comput. **46** (1986), 667-681.

[M] R.D. MAULDIN, *A generalization of Fermat's Last Theorem: The Beal conjecture and prize problem*, Notices Amer. Math. Soc. **44** (1997), 1436-1437.

[P] B. POONEN, *Some diophantine equations of the form* $x^n+y^n=z^m$, Acta Arith. **86** (1998), 193-205.

[R] B. RIEMANN, *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsber. Berliner Akademie, 1859, blz. 671-680.

[S] C.L. SIEGEL, *Transcendental Numbers*, Annals of Mathematics Studies 16, Princeton University Press, 1949.

[T] R. TIJDEMAN, *De Laatste Stelling van Fermat*, Symposium op 6 november 1993 in Utrecht, blz. 15-24.

[W] A. WILES, *Modular elliptic curves and Fermat's Last Theorem*, Ann. Math. **141** (1995), 443-551.

Vakantiecursus 1999

pp. 45 – 61

Het Poincaré Vermoeden

P.W.H. Lemmens
Mathematisch Instituut
Universiteit Utrecht

1. Inleiding

Henri Poincaré werd geboren in 1854 te Nancy en overleed in 1912 te Parijs. Hij was een neef van Raymond Poincaré, die president van Frankrijk was in de Eerste Wereldoorlog. Hij zag slecht en was buitengewoon onhandig. In *Men of Mathematics* noemt E.T. Bell hem "The Last Universalist", en verder schrijft hij: "Had Poincaré been as strong in practical science as he was in theoretical he might have made a fourth with the incomparable three, Archimedes, Newton, and Gauss."

Het POINCARÉ VERMOEDEN luidt:

> Een gesloten, samenhangende, en enkelvoudig samenhangende 3-dimensionale variëteit M^3 is homeomorf met de 3-dimensionale sfeer S^3.

Hierin komen nogal wat termen voor die om uitleg vragen.

Definities

Een n-dimensionale *variëteit* is een topologische ruimte die *lokaal* homeomorf is met $\mathbf{R}^n$.

Een variëteit heet *gesloten* als hij *compact* is en *geen rand* heeft.

Een n-dimensionale *variëteit met rand* is in een randpunt lokaal homeomorf met een halfruimte van $\mathbf{R}^n$ in een randpunt.

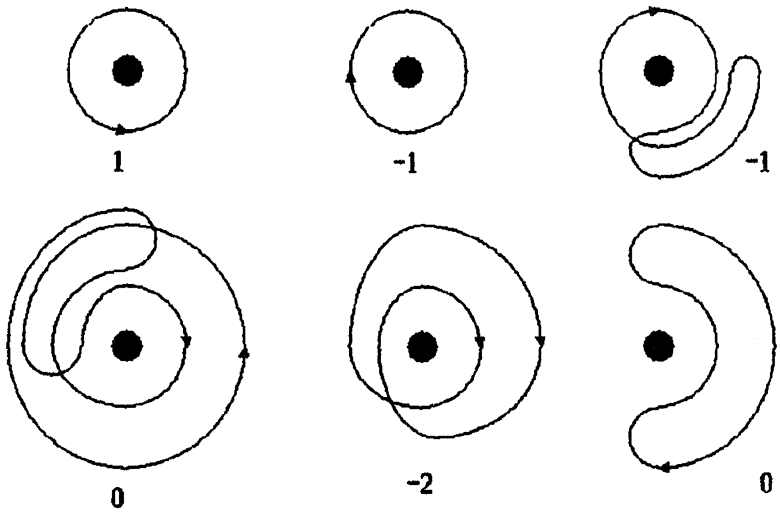
Een variëteit heet *samenhangend* als elk tweetal punten ervan binnen de variëteit verbonden kan worden door een continue kromme.

Een variëteit X heet *enkelvoudig samenhangend* als iedere gesloten kromme (continue afbeelding van de cirkel S^1 naar X) continu deformeerbaar is (*binnen* X) tot een *constante* afbeelding (we zeggen dat de kromme *samentrekbaar* is).

Twee ruimten X en Y heten *homeomorf* als er een eeneenduidige continue afbeelding X op Y bestaat waarvan de inverse ook continu is. Populair heeft men het in dit verband wel eens over 'rubbermeetkunde', maar daarmee is toch enige voorzichtigheid geboden: zie bijvoorbeeld verderop de tekst bij figuur 4.

De sfeer S^n is de deelverzameling van de punten in $\mathbf{R}^{n+1}$ met afstand 1 tot de oorsprong. Zo is bijvoorbeeld
$S^3 = \{(u, v, x, y) \in \mathbf{R}^4 \mid u^2 + v^2 + x^2 + y^2 = 1\}$.

Om het Poincaré vermoeden beter te begrijpen, zullen we eerst aandacht besteden aan de situatie in lagere dimensies.



FIGUUR 1. Verschillende omloopsgetallen

2. DIMENSIE 1

De letter "H" is als topologische ruimte niet lokaal homeomorf met $\mathbf{R}^1$, onder andere omdat er punten in zitten waar "H" er uitziet als een driesprong. Voor lokale homeomorfie van een ruimte X met $\mathbf{R}^1$ moet X er in de buurt van ieder punt uitzien zoals $\mathbf{R}^1$ er in de buurt van een punt uitziet.
Op een gesloten 1-dimensionale variëteit kunnen we ergens beginnen en in een van de twee richtingen gaan lopen. Omdat er geen rand is, lijkt het wegens de compactheid onvermijdelijk dat we moeten terugkomen op het uitgangspunt. Aldus is aannemelijk te maken dat een gesloten, samenhangende 1-dimensionale variëteit er uitziet als een cirkel, als S^1. Maar dat betekent geenszins dat het ook een deelverzameling moet zijn van $\mathbf{R}^2$. Het kan bijvoorbeeld best een knoop in $\mathbf{R}^3$ zijn.
Merk echter op dat S^1 niet enkelvoudig samenhangend is, immers de identiteitsafbeelding $S^1 \to S^1$ is niet binnen S^1 te deformeren tot een afbeelding die S^1 op één punt afbeeldt. Een belangrijke *invariant* voor *homotopie* (= deformatie) is het *omloopsgetal.* Het omloopsgetal meet hoe vaak een gesloten kromme ergens omheen loopt, en in welke richting. Bij het omsluiten van een andere gesloten kromme heeft men een soortgelijke invariant, het *schakelgetal.*
Een afbeelding van S^1 naar S^1 kan men zich voorstellen als een rondwandeling in de beeldruimte S^1. Daarbij loopt men een *geheel* aantal malen over de cirkel om het middelpunt daarvan in $\mathbf{R}^2$ (tegen de klok in wordt gewoonlijk positief geteld). In figuur 1 is een aantal voorbeelden getekend van omloopsgetallen van gesloten krommen om een punt.

Omdat een deformatie een *continu* proces is, moeten twee homotope rondwandelingen op S^1 hetzelfde omloopsgetal hebben. Voor rondwandelingen op S^1 geldt ook het omgekeerde: twee rondwandelingen met hetzelfde omloops-

getal zijn homotoop. Dit is niet triviaal, hetgeen zich uit in het feit dat de analoge generalisatie niet waar is voor de meeste andere ruimten dan S^1.
Ook rondwandelingen in $\mathbf{R}^2$ die niet door $(0,0)$ gaan, hebben een omloopsgetal om $(0,0)$, maar voor homotopieën in $\mathbf{R}^2$ is dat geen homotopie-invariant, omdat men bij zo'n homotopie door $(0,0)$ zou kunnen gaan. Op een dergelijk 'moment' is het omloopsgetal niet gedefinieerd, en na de 'passage' kan het versprongen zijn.

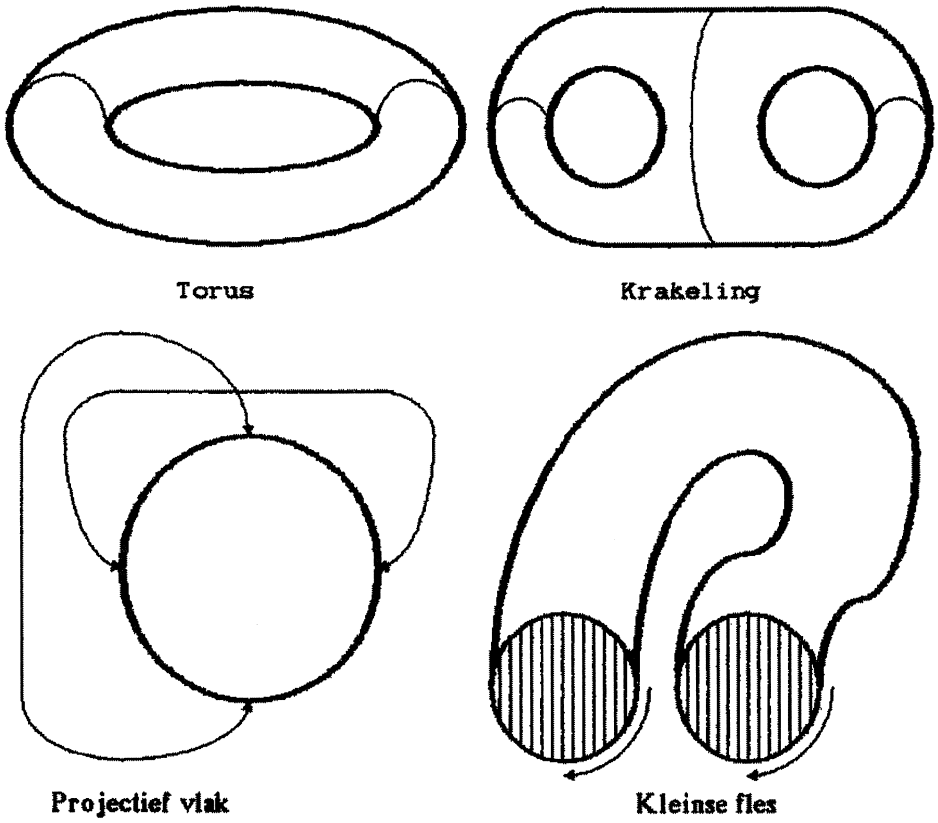
3. Dimensie 2

Voorbeelden van gesloten 2-dimensionale variëteiten zijn:
- het *boloppervlak* S^2.
- het *torusoppervlak* T^2 (= fietsband zonder ventiel).
- het oppervlak van een *krakeling.*
- het *projectieve vlak* P^2.
- de *Kleinse fles* K^2.

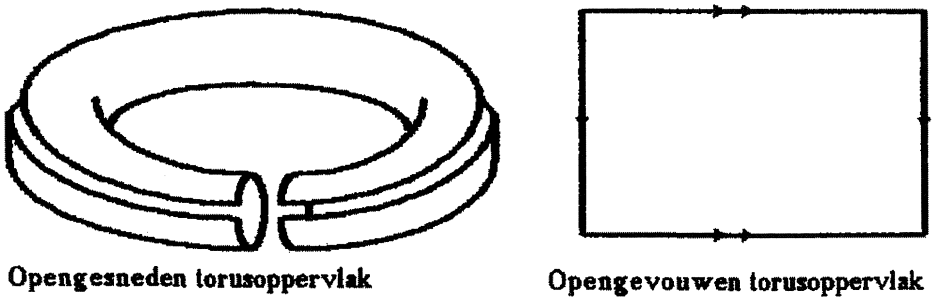
Een 2-dimensionale variëteit wordt ook wel een *oppervlak* genoemd.

Het projectieve vlak en de Kleinse fles zijn niet goed te tekenen, omdat ze niet in de 3-dimensionale ruimte kunnen worden ingebed zonder zelfdoorsnijdingen.
Daarom zijn ze zoals in figuur 2 op een speciale manier voorgesteld. Het projectieve vlak kan men krijgen uit een schijf waarvan ieder punt op de rand wordt wordt vastgemaakt aan het diametraal daar tegenover liggende punt.
De Kleinse fles kan worden verkregen uit een cilinder waarvan de randcirkels aan elkaar worden gezet, rekening houdend met de in figuur 2 aangegeven richting. Bij andersom plakken ontstaat het torusoppervlak.
Het is zeer instructief en verhelderend om een zakdoek te nemen en daarvan stukken van de rand op de voorgeschreven wijze (projectief vlak resp. Kleinse fles) aan elkaar te rijgen.
Elk gesloten, samenhangend oppervlak *ongelijk aan* S^2 kan langs een of meer gesloten krommen (beelden van S^1) worden *opengeknipt zonder dat het oppervlak uiteenvalt* (onsamenhangend wordt). Het eindresultaat is een *schijf* (althans homeomorf met een schijf) met een *rand* die is ontstaan uit de krommen waarlangs het oppervlak is opengeknipt. Iedere knipkromme geeft aanleiding tot *twee* krommen op de rand van de schijf. In figuur 3 is het torusoppervlak op een dergelijke manier opengesneden, en vervolgens opengevouwen tot een rechthoek.

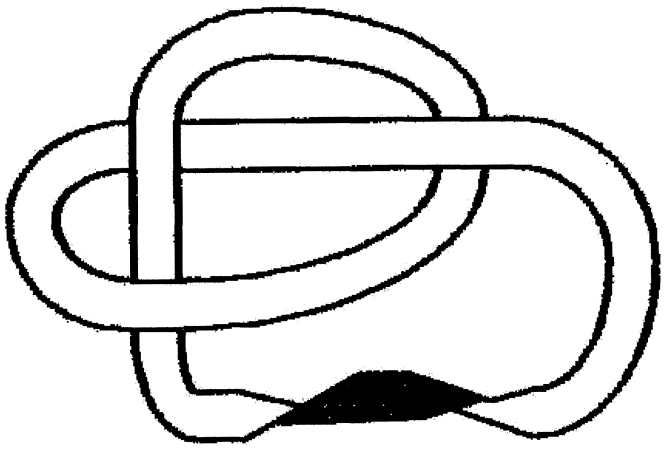
Identificeert men elk bij elkaar behorend paar randstukken op de geëigende manier, dan ontstaat het oorspronkelijke oppervlak weer. Daarbij is het mogelijk dat men het zo ontstane oppervlak niet direct visueel kan herkennen als het oorspronkelijke oppervlak.
Een voorbeeld is een smalle *cilinder* die eerst wordt opengeknipt langs een meridiaan. Daarbij resulteert een rechthoek waarvan twee overstaande zijden ontstaan zijn uit de meridiaan waarlangs de cilinder is opengeknipt. Men kan de cilinder terugkrijgen door deze twee zijden weer aan elkaar te plakken op de manier waarop ze oorspronkelijk aan elkaar hebben gezeten. Alvorens ze aan elkaar te plakken, kan men een aantal malen een complete draaiing van



FIGUUR 2. Een aantal oppervlakken



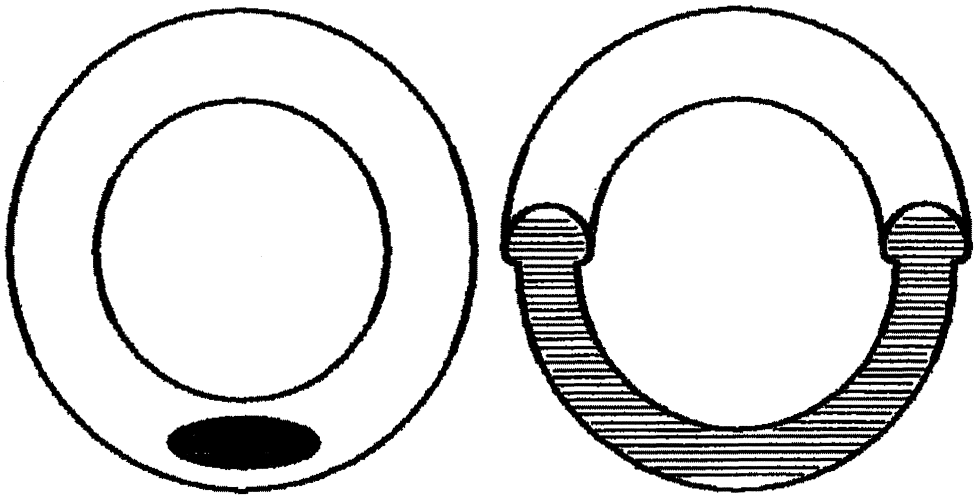
FIGUUR 3. Opengesneden torusoppervlak



FIGUUR 4. Geknoopte en gedraaide cilinder

360 graden uitvoeren. Bovendien is het mogelijk om de rechthoek eerst in een knoop te leggen. Een voorbeeld hiervan is getekend in figuur 4.
Merk op dat een zo ontstane gedraaide en geknoopte cilinder homeomorf is met de oorspronkelijk cilinder, maar dat hij binnen $\mathbf{R}^3$ niet als een brede elastieken band zonder scheuren kan worden gedeformeerd tot de oorspronkelijke cilinder.
Naast gesloten krommen die het oppervlak niet in twee stukken verdelen, zijn soms ook gesloten krommen bruikbaar die het oppervlak wel in twee delen splitsen. Vaak kan men van een stuk de "wond" groter maken en daardoor een beter inzicht krijgen in het topologisch karakter ervan.
Een voorbeeld is het uitknippen van een schijfje uit het torusoppervlak. Het ene deel kan dan gedeformeerd worden tot een variëteit met rand, die er uitziet als een vierkant waaraan twee stroken zijn vastgeplakt. Het eerste stadium van deze deformatie is te zien in figuur 5. Op de stroken kan men allerlei operaties toepassen, bijvoorbeeld het doorknippen van een van de stroken, deze om de andere strook winden en weer langs de knip herstellen. Na afloop van deze experimenten wordt het torusoppervlak hersteld door de rand van de uitgesneden schijf op de voorgeschreven wijze weer vast te maken aan de rand van het andere stuk.

De experimenten met de stroken kan men zich nog voorstellen in $\mathbf{R}^3$, maar het vastmaken van de schijf kan vaak niet in $\mathbf{R}^3$ geschieden, omdat er dan onvermijdelijk ongewenste doorsnijdingen ontstaan. Net zomin als bij het projectieve vlak en de Kleinse fles is dit een probleem, omdat bij de definitie van een variëteit niet wordt geëist dat hij een deelverzameling moet zijn van een Euclidische ruimte. Er is overigens wel een bekende stelling, die garandeert dat iedere gesloten n-dimensionale variëteit ingebed kan worden in $\mathbf{R}^{2n+1}$.
Er is voor gesloten, samenhangende oppervlakken een mooie stelling die het verband aangeeft tussen het karakter van het oppervlak en het gedrag van gesloten krommen daarop:



FIGUUR 5. Deformatie van torusoppervlak met gat

STELLING
Twee gesloten, samenhangende oppervlakken U en V zijn homeomorf dan en slechts dan als hun eerste homologiegroepen $H_1(U)$ en $H_1(V)$ isomorf zijn.

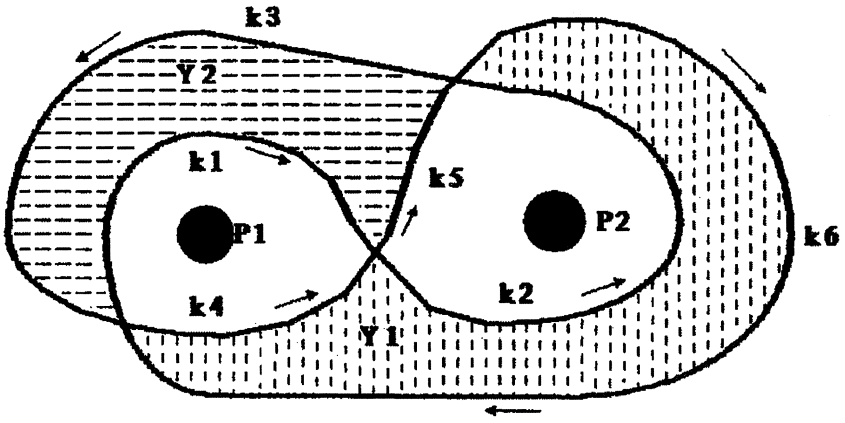
VOORBEELDEN
$H_1(S^2) = 0$, $H_1(T^2) = \mathbf{Z} \oplus \mathbf{Z}$, $H_1(P^2) = \mathbf{Z}_2$, $H_1(K^2) = \mathbf{Z} \oplus \mathbf{Z}_2$.

Met name geldt dus de "Poincaré stelling in dimensie 2":

> Een gesloten, samenhangend oppervlak X is homeomorf met S^2 dan en slechts dan als $H_1(X) = 0$.

$H_1(X) = 0$ zegt ons dat iedere 'fatsoenlijke' gesloten kromme k in X de *rand* is van een *georiënteerd* 2-dimensionaal complex Y in X.
Er zijn twee fundamenteel belangrijke zaken waarvan men zich hierbij bewust moet zijn:
(1) Y hoeft geen schijf of zelfs maar homeomorf met een schijf te zijn,
(2) k wordt hierbij niet geïnterpreteerd als een afbeelding van S^1 naar X, maar als een 'som' van afbeeldingen van segmenten naar X die in een andere volgorde aan elkaar mogen worden vastgemaakt (waarbij wel de eindpunten en beginpunten moeten passen).

Gezien de vele verschillende vormen waarin eenzelfde oppervlak zich blijkens het voorgaande in $\mathbf{R}^3$ kan voordoen, is deze Poincaré stelling erg belangrijk. Hij geeft een mogelijkheid om intrinsiek (zonder buiten het oppervlak te treden) vast te stellen om welk oppervlak het gaat. Bovendien wordt een topologisch begrip (homeomorfie) omgezet in een algebraïsch begrip (groepen).

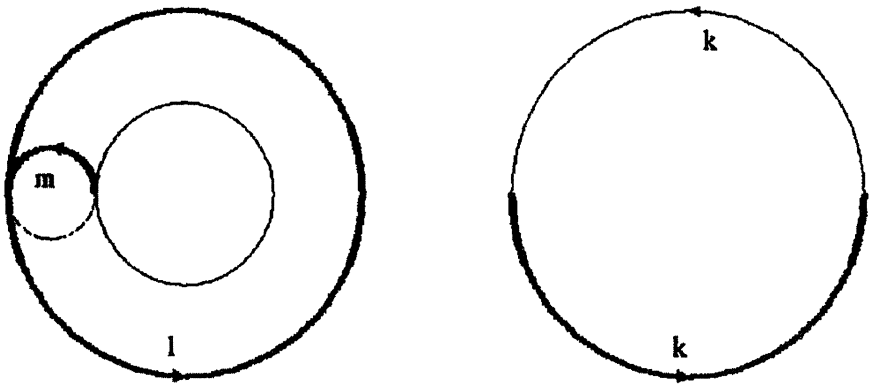


FIGUUR 6. Kromme in vlak zonder twee punten

VOORBEELD
Zie figuur 2. De kromme die het oppervlak van een krakeling in twee helften (links, rechts) deelt, is de rand van elk van de helften, maar geen van beide zijn homeomorf met een schijf.

VOORBEELD
Voor X nemen we het vlak met daaruit twee punten P_1 en P_2 verwijderd. Voor k nemen we een kromme die eerst rechts om P_1 loopt, dan links om P_2, dan links om P_1, en tenslotte rechts om P_2, waarna hij weer bij zijn beginpunt komt. De omloopsgetallen van k om P_1 en P_2 zijn dus 0.
Zoals in figuur 6 aangegeven, kan men de kromme opsplitsen in 6 beelden van segmenten, k_1 t/m k_6, en k wordt als afbeelding beschreven door $k = k_1 + k_2 + \cdots + k_6$, terwijl we k als rand zien door hem te splitsen in de vorm $(k_4 + k_2 + k_6) + (k_1 + k_5 + k_3)$. Het complex Y bestaat uit twee schijven Y_1 en Y_2 die elkaar raken in drie punten. De rand van Y_1 is $(k_4 + k_2 + k_6)$, en de rand van Y_2 is $(k_1 + k_5 + k_3)$.

Normaliter schrijven we het aan elkaar zetten van afbeeldingen van segmenten met het vermenigvuldigingsteken "$*$", maar in homologie-beschouwingen gebruiken we de "$+$" om het commutatieve karakter te benadrukken.
Merk ook op dat de oriëntering van Y_2 niet past bij de oriëntering van Y_1 ten opzichte van de oriëntering van X.
Uit het bovenstaande blijkt dat de kromme k in X homoloog nul is.
Dat k niet binnen X te deformeren is tot een constante afbeelding, kan men illustreren door in een plankje twee spijkers te slaan, en dan een tot een gesloten lus gemaakt koord op de door k aangegeven manier op het plankje te leggen. Hoe we ook proberen dit koord te verschuiven, zonder het over de spijkers te tillen of achter het plankje om te gaan, het lukt niet om het koord vrij van de spijkers te krijgen. Door de fysieke beperkingen van het koord (een deel kan

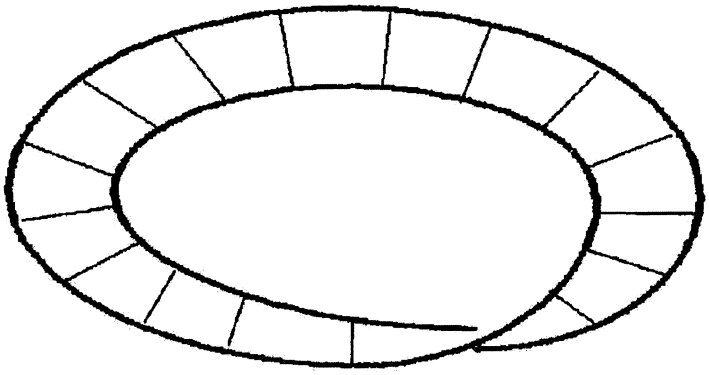


FIGUUR 7. Gesloten krommen op torusoppervlak en projectieve vlak

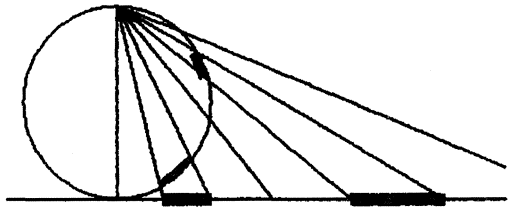
bijvoorbeeld niet door een ander deel bewegen, en het is niet uitrekbaar) is dit experiment niet helemaal betrouwbaar, maar het geeft een idee.
Voor een echt bewijs van het feit dat k in X niet samentrekbaar is, hebben we meer techniek nodig.

$H_1(T^2) = \mathbf{Z} \oplus \mathbf{Z}$ omdat er op het torusoppervlak T^2 twee gesloten krommen zijn, een *longitudinale* l en een *meridiane* m (zie figuur 7) die geen van beide een rand zijn, waarvan geen gehele combinatie (behalve de 0-combinatie) een rand is, terwijl elke "fatsoenlijke" gesloten kromme op T^2 *homoloog* (d.w.z. op een rand na gelijk) is met een of andere gehele combinatie van deze twee. De meridiaan heeft een schakelgetal ± 1 met de centrale cirkel die binnen de torus loopt, en heeft omloopsgetal 0 met de rotatieas van de torus. De longitudinaal heeft daarentegen schakelgetal 0 met de centrale cirkel en omloopsgetal ± 1 met de rotatieas. Hieruit zien we dat meridiaan en longitudinaal op het torusoppervlak niet samentrekbaar en niet in elkaar deformeerbaar kunnen zijn.
$H_1(P^2) = \mathbf{Z}_2$ omdat er op het projectieve vlak P^2 een gesloten kromme k bestaat, die geen rand is en de eigenschap heeft dat iedere fatsoenlijke kromme op P^2 homoloog is met een geheel veelvoud van k, maar bovendien is $2k$ wel een rand! In figuur 7 zien we k twee keer wegens de identificaties op de rand van de schijf (vergelijk figuur 2). Samen zijn ze de rand van de schijf. Dus rekenen we met $2k = 0$ in de homologie. Dit impliceert dat voor elke gesloten kromme h op P^2 homologisch geldt $h = 0$ of $2h = 0$.

Het verschijnsel van gesloten krommen k die zelf niet homoloog 0 zijn, maar waarvoor $2k$ wel homoloog 0 is, doet zich alleen voor bij *niet-oriënteerbare* oppervlakken. Maakt men over het oppervlak een rondreis die zo'n kromme eenmaal snijdt, dan komt men ondersteboven op het uitgangspunt terug: men heeft een *oriëntering-omkerende kromme* doorlopen. Een smalle omgeving van zo'n kromme ziet eruit als een *Möbiusband.*
Een Möbiusband ontstaat uit een smalle cilinder door deze langs een meridiaan open te knippen en de twee einden met een halve slag van 180 graden weer aan



FIGUUR 8. Möbiusband



FIGUUR 9. Stereografische projectie

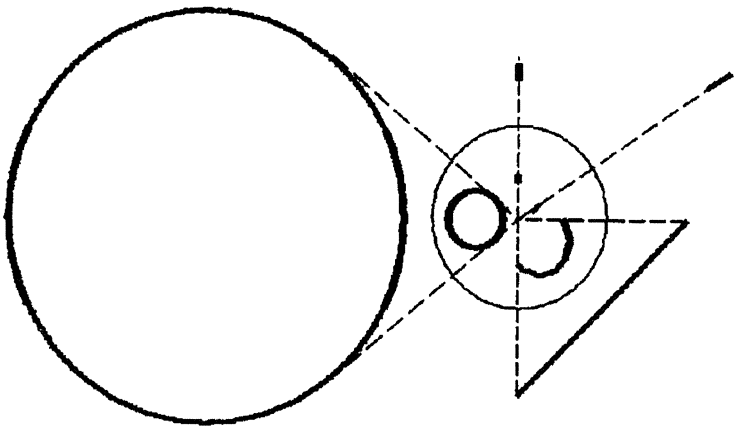
elkaar te plakken, zoals in figuur 8 aangegeven.

De Möbiusband is niet homeomorf met de oorpronkelijke cilinder.
Bij knippen langs een gesloten kromme valt de cilinder uiteen in minstens twee losse stukken, terwijl de 'centrale cirkel' van de Möbiusband die eigenschap niet heeft. Kunt u inzien dat knippen langs de centrale cirkel van de Möbiusband resulteert in één (weliswaar gedraaide) cilinder?

4. DIMENSIE 3

We kunnen ons heel moeilijk een gesloten 3-dimensionale variëteit voorstellen omdat we zelf in een *open* 3-dimensionale variëteit denken te leven, en daar alleen compacte 3-dimensionale variëteiten met rand kunnen zien.
Kijken we voor analogie naar S^2, en snijden we die door over de equator, dan houden we twee halfronden over die beide homeomorf zijn met een cirkelschijf in $\mathbf{R}^2$.
In feite is geheel S^2 op één punt na homeomorf af te beelden op $\mathbf{R}^2$. Dat kan heel fraai via *stereografische projectie*, die S^2 vanuit de noordpool perspectivisch projecteert op het raakvlak aan de zuidpool. Het zuidelijk halfrond wordt dan keurig afgebeeld op een cirkelschijf.

Het noordelijk halfrond komt daarbij terecht op het complement van deze



FIGUUR 10. Inversie in een cirkel

cirkelschijf, *met uitzondering van de noordpool*, die zich niet laat afbeelden. Nabij de noordpool worden stukken van S^2 sterk uitgerekt afgebeeld, maar meridianen door de noordpool (zonder de noordpool zelf) worden wel als rechte lijnen door het beeld van de zuidpool afgebeeld. In figuur 9 is deze situatie aangegeven. Wist u trouwens dat cirkels op S^2 bij deze stereografische projectie in echte cirkels of in rechte lijnen worden afgebeeld? Het is alsof we $\mathbf{R}^2$ rondbuigen tot een boloppervlak met een gat en de "rand" van dat gat in één punt laten uitmonden.
Eigenlijk is stereografische projectie ook lastig voor te stellen om S^3 in verband te brengen met $\mathbf{R}^3$, want dat speelt zich af in de $\mathbf{R}^4$.
Er is een mooie methode om het complement van een ronde schijf in $\mathbf{R}^2$ af te beelden naar een schijf, min of meer zoals de stereografische projectie dat doet, namelijk via *inversie* in de rand(cirkel) van de schijf.
Inversie is een proces dat we ons ook in $\mathbf{R}^3$ kunnen voorstellen.
Het complement van een bol met straal R in $\mathbf{R}^3$ wordt als volgt afgebeeld binnen die bol. Van een punt (z_1, z_2, z_3) in $\mathbf{R}^3$ met $r^2 = z_1^2 + z_2^2 + z_3^2 \geq R^2$ is het beeldpunt $(z_1 \cdot R^2/r^2, z_2 \cdot R^2/r^2, z_3 \cdot R^2/r^2)$. Een punt en zijn beeldpunt liggen dus op dezelfde halfrechte door de oorsprong. Punten met $r^2 = R^2$ blijven hierbij op hun plaats. Naarmate een punt verder van de oorsprong ligt, komt het beeldpunt dichter bij de oorsprong.
Het blijkt dat boloppervlakken worden afgebeeld op boloppervlakken, en cirkels op cirkels. Vlakken worden afgebeeld in boloppervlakken die door de oorsprong gaan, en rechten in cirkels die door de oorsprong gaan. De oorsprong is geen echt beeldpunt, deze is als het ware het beeld van het oneindige.
In figuur 10 is een inversie in het vlak voorgesteld.

Zo komen we op een model van S^3, dat bestaat uit twee massieve bollen B_1^3 met rand S_1^2 en B_2^3 met rand S_2^2, die met hun randen aan elkaar gehecht zijn. Dat aanhechten beschrijven we gewoonlijk met een afbeelding $f : S_2^2 \to S_1^2$.

In dit geval is f te interpreteren als de identieke afbeelding. Het moeilijke zit er nu in, zich steeds voor ogen te houden dat B_1^3 en B_2^3 *alleen met hun randen* aan elkaar zitten.

Niet elke gesloten 3-dimensionale variëteit M^3 laat zich schrijven als een vereniging van bollen die op een of andere manier met hun randen aan elkaar zitten. Er is nog een andere fundamentele manier om S^3 open te snijden langs een gesloten oppervlak waarbij ook twee voorstelbare 3-dimensionale variëteiten met rand (elk van die randen is dus eigenlijk dat oppervlak waarlangs geknipt is) ontstaan, en waaruit men S^3 weer terug krijgt door de twee randen aan elkaar te plakken.
De methode die we bedoelen is, te snijden langs een torusoppervlak T^2. Nemen we een gewone torus in $\mathbf{R}^3$, niet geknoopt, bijvoorbeeld zoals getekend in figuur 2, en snijden we $\mathbf{R}^3$ open langs het oppervlak van deze torus, dan valt $\mathbf{R}^3$ uiteen in een torus en het complement ervan. Uit het beeld dat we ons gevormd hebben van S^3, volgt dat ook S^3 onder deze chirurgische ingreep in twee stukken uiteenvalt. Door een kleine bol binnen de torus te nemen en inversie toe te passen, kunnen we ons ervan overtuigen dat ook het complement in S^3 van de torus weer een torus is.
Dus S^3 is de vereniging van twee torussen, die met hun oppervlakken aan elkaar vast zitten. Een longitudinaal respectievelijk meridiaan van het ene oppervlak correspondeert met een meridiaan respectievelijk longitudinaal van het andere oppervlak (dus precies andersom).
Enigszins abstract kunnen we deze splitsing van S^3 ook inzien door te schrijven:

$$S^3 = \partial(I^4) = \partial(I^2 \times I^2) = (\partial(I^2) \times I^2) \cup (I^2 \times \partial(I^2)) = (S^1 \times I^2) \cup (I^2 \times S^1).$$

Hierbij stellen we S^3 voor door de daarmee homeomorfe rand van de 4-dimensionale kubus I^4, waarbij I staat voor het segment $[0, 1]$.

We hebben nu een erg simpele operatie uitgevoerd, namelijk eerst S^3 opensnijden langs een torusoppervlak en dan de twee oppervlakken van de ontstane delen weer terugzetten "zoals ze aan elkaar zaten". Maar we kunnen ook wat creatiever zijn en meer plastische chirurgie uitvoeren, door bijvoorbeeld longitudinaal op longitudinaal en meridiaan op meridiaan te hechten. Dat kan zonder veel bezwaar, alleen krijgen we dan S^3 niet terug, maar een andere gesloten variëteit M^3. Dat het echt iets anders wordt dan S^3, zien we aan de gemeenschappelijke longitudinaal, die nu niet de rand is van een 2-dimensionaal complex in M^3, terwijl in S^3 elke fatsoenlijke gesloten kromme een rand is.

De aanhechting van de twee torusoppervlakken aan elkaar kan nog wilder dan hierboven beschreven. Zo kan men de meridiaan van het ene torusoppervlak vasthechten aan het andere torusoppervlak volgens een kromme die op het andere oppervlak homoloog is met p keer de longitudinaal + q keer de meridiaan. Hierin moeten p en q gehele getallen zijn met $\text{ggd}(p, q) = 1$, anders is de aanhechting niet af te maken tot een homeomorfisme tussen de oppervlakken. Stellen we ons het torusoppervlak T^2 voor als een vierkant met zijden van lengte 1, waarvan elke twee overstaanden aan elkaar worden geplakt, dan kunnen we

zo'n afbeelding definiëren door

$$(x, y) \to (rx + py - [rx + py], sx + qy - [sx + qy]) \text{ voor } 0 \leq x \leq 1 \text{ en } 0 \leq y \leq 1,$$

waarbij r en s gehele getallen zijn zo dat $rq - sp = 1$, en $[z]$ de entier van z aanduidt. Deze afbeelding is dus lokaal lineair met determinant 1.
De na het aanhechten resulterende M^3 staat bekend als de *lensruimte* $L(p, q)$.

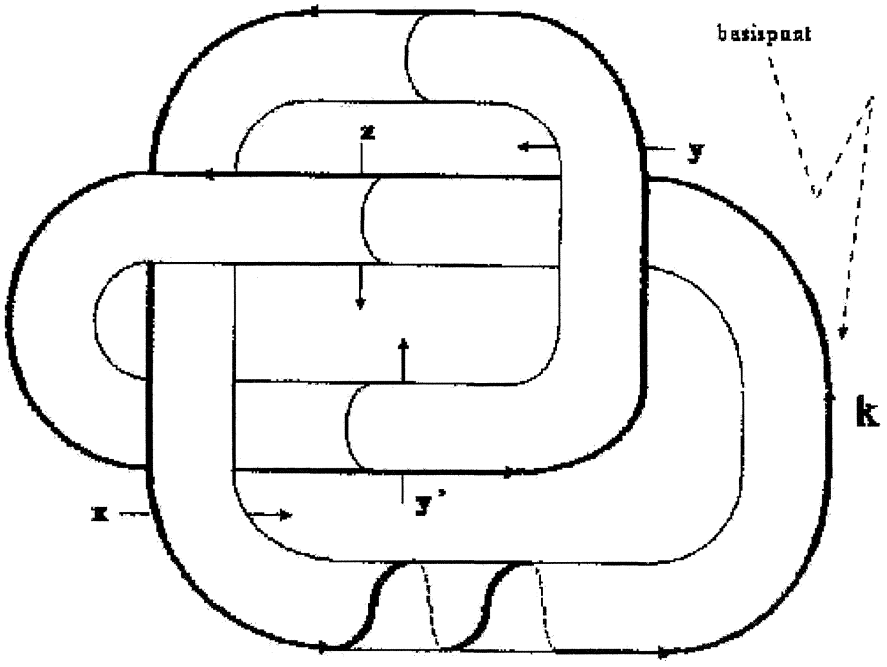
VOORBEELDEN
$p = 1$, q willekeurig: $L(1, q) = S^3$,
$L(0, 1) = S^2 \times S^1$, $L(2, 1) = P^3$.

Over het verband tussen twee lensruimten vermelden we de volgende stelling.

STELLING
$L(p, q)$ is homeomorf met $L(p', q')$ dan en slechts dan als $p = p'$ en $\pm q' \equiv q^{\pm 1}$ mod p.
Om het belang van deze constructie te illustreren, vermelden we de volgende stelling.

STELLING
Elke gesloten, samenhangende, oriënteerbare 3-dimensionale variëteit M^3 is te verkrijgen door twee gegeneraliseerde krakelingen met hun randen (oppervlakken) aan elkaar te hechten.
In deze stelling zit een extra veronderstelling: oriënteerbaarheid.
In het kader van onze probleemstelling, namelijk enkelvoudig samenhangende 3-dimensionale variëteiten, is dat geen probleem, want een oriëntering-omkerende gesloten kromme kan niet samentrekbaar zijn. Een enkelvoudig samenhangende gesloten variëteit is dus automatisch oriënteerbaar.

Het eerste vermoeden dat Poincaré publiceerde [Proc. London Math. Soc. **32** (1900), 277–308] was dat een gesloten, samenhangende 3-dimensionale variëteit M^3 homeomorf zou zijn met S^3 dan en slechts dan als elke gesloten kromme k in S^3 de rand is van een 2-dimensionaal complex. Dit zou dus een directe generalisering zijn van de corresponderende stelling voor oppervlakken. Al spoedig vond Poincaré zelf een tegenvoorbeeld van dit vermoeden [Rend. Circ. Mat. Palermo **18** (1904), 45–110]. Dit tegenvoorbeeld staat bekend als *The Poincaré Manifold.*
Daarna zijn vele andere voorbeelden geconstrueerd van niet enkelvoudig samenhangende 3-dimensionale variëteiten M^3 met $H_1(M^3) = 0$. Een dergelijke constructie van Max Dehn uit 1910 is homeomorf met de Poincaré variëteit. Dehn's methode gaat uit van een torus die enkelvoudig geknoopt in S^3 ligt. Deze torus wordt langs zijn oppervlak uit S^3 gesneden, waarna een standaard torus wordt teruggeplaatst door een meridiaan van het oppervlak daarvan af te beelden op de in figuur 11 aangegeven kromme k.



FIGUUR 11. Dehn's constructie

Met N^3 duiden we de 3-dimensionale variëteit met rand aan die resteert na het uitsnijden van de geknoopte torus.

Om de homotopie van gesloten krommen op N^3 te bestuderen, rekent men gewoonlijk de *fundamentaalgroep* $\pi_1(N^3)$ uit. Daarbij bestudeert men het gedrag van continue afbeeldingen van het segment $[0,1]$ naar N^3, die 0 en 1 afbeelden op een vast gekozen *basispunt* in N^3.
Visueel kan men zo'n afbeelding zien als een rondwandeling in N^3, beginnend en eindigend in het basispunt.
De elementen van de groep zijn de klassen van onderling homotope afbeeldingen van dit type (tijdens de deformatie van de ene afbeelding naar de andere moeten steeds 0 en 1 in het basispunt blijven afgebeeld).
Het groepskarakter ontstaat door twee wandelingen na elkaar uit te voeren (hier valt heel wat te verifiëren ten aanzien van de homotopieklassen waarover het eigenlijk gaat!). Een wandeling w_1 gevolgd door een wandeling w_2 noteren we met $w_1 * w_2$, en men dient zich ervan bewust te zijn dat $w_2 * w_1$ in het algemeen niet homotoop is met $w_1 * w_2$.
Stil blijven staan in het basispunt hoort bij het eenheidselement 1 van de groep, en een wandeling w in omgekeerde richting uitvoeren geeft de inverse w^{-1}.

Omdat we uiteindelijk willen bestuderen wat er met N^3 gebeurt na het daaraan vastplakken van een standaardtorus, is het verstandig om het basispunt

van N^3 te kiezen op het oppervlak van de geknoopte torus. Zo kan het immers als beeldpunt optreden van het basispunt van de standaardtorus onder de plakafbeelding. Voor het gemak kiezen we het basispunt van N^3 op de kromme k zoals aangegeven in figuur 11.
Vervolgens maken we bij de in figuur 11 aangegeven pijl x een *elementaire* rondwandeling door vanuit het basispunt naar het beginpunt van x te gaan *zonder ergens achter de uitgesneden torus te komen*, vervolgens langs x naar zijn eindpunt, en vandaar weer *zonder ergens achter de uitgesneden torus te komen* terug naar het basispunt. Deze rondwandeling duiden we eenvoudigheidshalve ook aan met x.
We komen daarbij dus nergens achter de uitgesneden torus, behalve eventueel in het gedeelte waar de pijl x doorlopen wordt Verder is het belangrijk dat de gehele rondwandeling in N^3 plaatsvindt: we mogen niet in het inwendige van de geknoopte torus komen.
Geheel analoog definiëren we de elementaire rondwandelingen y en z.
Dan is het min of meer duidelijk dat ieder element van $\pi_1(N^3)$ gerepresenteerd kan worden als een product van de rondwandelingen x, y en z, in een bepaalde volgorde en eventueel met herhalingen en inverses. Hierbij hoeven we er eigenlijk alleen maar op te letten waar de rondwandeling achter de uitgesneden torus verdwijnt en in welke richting.

De lezer wordt uitgenodigd om eens te kijken welke rondwandeling hoort bij $x * y^{-1} * z$. Kunt u ook inzien dat $y * z * y^{-1} * x^{-1}$ homotoop is met stil blijven staan in het basispunt? Eerst kunnen we de hierbij horende wandeling deformeren tot de wandeling die hoort bij $y * z * (y')^{-1} * x^{-1}$ (zie figuur 11) door y^{-1} naar $(y')^{-1}$ te deformeren. Vervolgens kunnen we $y * z * (y')^{-1} * x^{-1}$ deformeren tot een wandeling waarbij van het eindpunt van pijl y *rechtstreeks* naar het beginpunt van z wordt gegaan in plaats van via het basispunt te gaan (bedenk dat bij een deformatie alleen wordt geëist dat begin- en eindpunt van de *totale* wandeling in het basispunt blijven). Deze "truuk" kan ook worden gebruikt om rechtstreeks van het eindpunt van pijl z naar het eindpunt van pijl y' en rechtstreeks van het beginpunt van pijl y' naar het eindpunt van pijl x. We zien zo dat de wandeling $y * z * y^{-1} * x^{-1}$ gedeformeerd is tot een wandeling die vanuit het basispunt naar het beginpunt van pijl y gaat en verder *geheel achter de torus om* naar het beginpunt van pijl x en dan terug naar het basispunt gaat. Vervolgens kan deze wandeling gemakkelijk worden deformeerd tot een wandeling die in het basispunt blijft.
Dit laatste betekent in de fundamentaalgroep dat $yzy^{-1}x^{-1} = 1$, en dus dat $z = y^{-1}xy$. We kunnen z dus missen als voortbrenger van de fundamentaalgroep, en alleen x en y gebruiken. Verder geldt de relatie $y^{-1}xy = xyx^{-1}$. De lezer kan dit zelf nagaan door aan de hand van figuur 11 te beredeneren dat de rondwandeling $y * z * x^{-1} * z^{-1}$ samentrekbaar is in N^3.
Gebruiken we alleen x en y als voortbrengers, dan blijkt $y^{-1}xy = xyx^{-1}$ alle relaties in $\pi_1(N^3)$ te genereren. Bovendien is $y^{-1}xy = xyx^{-1}$ equivalent met de relatie $xyx = yxy$, waarin geen negatieve exponenten voorkomen. Daarom wordt $\pi_1(N^3)$ als groep gepresenteerd door de voortbrengers en relaties

$$\pi_1(N^3) = (x, y\,;\, xyx = yxy).$$

In N^3 is k volgens figuur 11 homotoop met $y * x * z * x^{-1} * x^{-1}$, dus met $y * x * (y^{-1} * x * y) * x^{-1} * x^{-1}$. In $\pi_1(N^3)$ geldt dus

$$k = yxy^{-1}xyx^{-1}x^{-1} = yx^2yx^{-3}.$$

Omdat k bij de aanhechting het beeld is van een meridiaan m van de standaardtorus, en m daarin samentrekbaar is, krijgen we de relatie $k = 1$, dus $yx^2y = x^3$ in de fundamentaalgroep van de resulterende ruimte M^3. De stelling van E.R. van Kampen (zie bijv. het in de literatuurlijst genoemde boek van Crowell en Fox) zegt dat we deze relatie slechts hoeven toe te voegen aan de relaties van $\pi_1(N^3)$ om daaruit de presentatie van $\pi_1(M^3)$ te verkrijgen:

$$\pi_1(M^3) = (x, y\,;\, xyx = yxy, yx^2y = x^3).$$

Door nog de extra relatie $xy = yx$ toe te voegen, wordt de fundamentaalgroep *commutatief* gemaakt, en het is bekend dat de commutatief gemaakte (geabelianiseerde) fundamentaalgroep precies de eerste homologiegroep is. Het toevoegen van die extra relatie stelt ons in staat om het een en ander te vereenvoudigen, immers dan geldt bijvoorbeeld $y^{-1}xy = x$ en $xyx^{-1} = y$, dus $xyx = yxy$ impliceert dan $x = y$. In $H_1(M^3)$ kunnen we y dus schrappen als voortbrenger, en in de relaties de letter y vervangen door x. Hierdoor gaat de relatie $yx^2y = x^3$ over in $x = 1$, met andere woorden: de eerste homologiegroep bestaat uit slechts één element.
Commutatief schrijven we dat gewoonlijk als het nul-element, dus $H_1(M^3) = 0$.

Dat de fundamentaalgroep niet uit slechts 1 bestaat, is in principe veel moeilijker in te zien. Het volgt bijvoorbeeld uit het feit dat er in de groep S_5 van alle permutaties van $\{1, 2, 3, 4, 5\}$ twee elementen zijn die binnen die groep aan dezelfde relaties voldoen als x en y in $\pi_1(M^3)$. Zo ziet men dat $\pi_1(M^3)$ zich surjectief laat afbeelden op een niet-triviale subgroep van S_5.
Voor de bedoelde elementen van S_5 kan men bijvoorbeeld respectievelijk kiezen (in cykel-notatie): (1 2 3 4 5) en (1 5 3 2 4). De lezer kan dit zelf nagaan.
Hieruit blijkt dan dat M^3 niet homeomorf met S^3 kan zijn, want daarin is iedere gesloten kromme samentrekbaar.
Het bovenstaande komt in essentie uit het in de literatuurlijst genoemde boek van Rolfsen.

5. Dimensie > 3

In dimensies > 3 is het Poincaré vermoeden opgelost, dat wil zeggen dat daar de volgende stelling geldt:

> Voor $n > 3$ is een gesloten, samenhangende en enkelvoudig samenhangende n-dimensionale variëteit M^n homeomorf met de n-dimensionale sfeer S^n als alle homologiegroepen van M^n overeenkomen met die van S^n.

Voordat deze stelling bewezen was, heette hij *het gegeneraliseerde Poincaré vermoeden.*

Dat we in dimensie 3 alleen iets hoeven te eisen over de fundamentaalgroep van M^3 (namelijk enkelvoudige samenhang), komt omdat daar wegens de lage dimensie de hele homologiestructuur wordt bepaald door de fundamentaalgroep.
Het eerste bewijs van het gegeneraliseerde Poincaré vermoeden dateert uit 1960 en is afkomstig van Steve Smale. Hij bewees het voor $n \geq 5$ en met een extra conditie, namelijk dat M^n een differentieerbare variëteit is. Zijn resultaat was daarmee dan ook dat M^n diffeomorf is met S^n, hetgeen veel sterker is dan homeomorf. Er zijn voorbeelden van onderling niet diffeomorfe ruimten die homeomorf zijn met S^n.
John Milnor was de eerste die dergelijke ruimten construeerde met behulp van z.g. Morsetheorie.
Voor zijn bewijs gebruikte Smale nieuwe ontwikkelingen in de theorie van dynamische systemen. Daardoor kon hij, steunend op resultaten van René Thom, aantonen dat iedere differentieerbare gesloten variëteit te verkrijgen is als een gegeneraliseerd n-dimensionaal oppervlak van een krakeling (een z.g. handlebody). De extra veronderstellingen stelden hem in staat om aan te tonen dat de handles twee aan twee tegen elkaar uitgewisseld kunnen worden.
In het begin van de zestiger jaren zijn er stormachtige ontwikkelingen geweest ten aanzien van het Poincaré vermoeden. Aangespoord door de resultaten van Smale, waren ook anderen in staat om het in verschillende variaties te bewijzen. Bekende namen in dit verband zijn J.R. Stallings, E.C. Zeeman, C.T.C. Wall, A.H. Wallace, M. Hirsch, A. Haefliger. Wie er meer over wil weten kan terecht in de in de literatuurlijst genoemde artikelen van Smale uit 1963, en (voor wat meer pikante details) uit 1990. In 1983 heeft M. Freedman het vermoeden voor $n = 4$ bewezen.

6. Slotopmerkingen

Het Poincaré vermoeden is tot nu toe een niet aflatende bron van wetenschappelijk onderzoek. Het is zeker eenvoudiger om samenhang en enkelvoudige samenhang vast te stellen dan om rechtstreeks aan te tonen dat een variëteit homeomorf is met S^3. Er zijn ook andere interessante problemen in de 3- en 4-dimensionale topologie die opgelost zouden zijn, of die men diepgaander zou kunnen bestuderen als het Poincaré vermoeden waar zou zijn.
Een recentere tak van onderzoek betreft de *metrische eigenschappen* van 3-dimensionale variëteiten, met consequenties voor de ruimte waarin we zelf leven. Hiervoor verwijzen we naar de literatuurlijst, met name naar het artikel van Thurston en Weeks en/of naar het boek van Weeks. Ook het zeer recente artikel van Luminet, Starkman en Weeks is in dit opzicht interessant.

Graag dank ik Prof. Dr. D. Siersma voor zijn hulp bij het vinden en begrijpen van de literatuur, en voor zijn commentaar op een eerdere versie.

7. Literatuur

1. E.T. Bell, *Men of Mathematics*, Simon & Schuster, New York, 1965.
2. Richard H. Crowell, Ralph H. Fox, *Introduction to Knot Theory*, Springer Verlag, New York, 1977 (Graduate texts in mathematics 57).
3. Jean-Pierre Luminet, Glenn D. Starkman and Jeffrey R. Weeks, *Is space finite?*, Scientific American **280** (4) (1999), 68–75. (April 1999)
4. Dale Rolfsen, *Knots and links*, Publish or Perish, Wilmington (USA), 1976 (Mathematical Lecture Series 7).
5. S. Smale, *A survey of some recent developments in differential toplogy*, Bulletin American Math. Soc., **69** (1963), 131–145.
6. Steve Smale, *The story of the higher dimensional Poincaré Conjecture (What actually happened on the beaches of Rio)*, The Mathematical Intelligencer **12** (2) (1990), 44–51.
7. Steve Smale, *Mathematical problems for the next century*, The Mathematical Intelligencer **20** (2) (1998), 7–15.
8. William P. Thurston and Jeffrey R. Weeks, *The mathematics of three-dimensional manifolds*, Scientific American **251** (1) (1984), 94–106.
9. Jeffrey R. Weeks, *The shape of space*, Marcel Dekker, New York, 1985.

Vakantiecursus 1999

pp. 63 – 73

Pakkende Problemen

Hans Melissen

In dit verhaal laten we een aantal vermoedens en problemen de revue passeren die te maken hebben met het pakken (stapelen) en overdekken van objecten. We zullen ons daarbij beperken tot cirkels en bollen van een gelijke grootte. We zullen zien dat zelfs met deze restricties er genoeg interessante problemen over blijven. Even voor de duidelijkheid: Met pakken of stapelen bedoelen we het plaatsen van objecten (bijvoorbeeld bollen of cirkelschijven) zodanig dat ze elkaar nier overlappen. Met overdekken bedoelen we het plaatsen van objecten zodanig dat elk punt van de ruimte (of een afgesproken verzameling) in minstens één van de objecten ligt. Een voor de hand liggende vraag bij het pakken is om zoveel mogelijk objecten te pakken, ze zo dicht mogelijk bij elkaar te pakken, de ruimte optimaal te benutten, terwijl het bij overdekken gaat om een zo zuinig mogelijke overdekking, met zo weinig mogelijk of zo klein mogelijke objecten.

1. Eéndimensionaal pakken: Het vermoeden van Palásti

Een goede aanpak om inzicht in een probleem te krijgen is, om het eerst maar eens zo eenvoudig mogelijk te maken. Helaas gaat daarmee soms ook het wezen of de moeilijkheid van het probleem verloren. Als we het cirkel- of bolpakkingsprobleem in één dimensie bekijken, blijkt het bijzonder simpel te zijn. Een ééndimensionale bol is een segment, met lengte tweemaal de straal van de "bol". Een stapeling van zulke segmenten die zo dicht mogelijk is is zeer eenvoudig te bedenken. Je legt gewoon alle segmenten aansluitend achter elkaar en je krijgt zo een maximale pakking. Elk punt van de lijn wordt namelijk zelfs overdekt, dus er valt geen speld meer tussen te krijgen. Beter kan niet.

Nu is het gelukkig zo dat wiskundigen zelfs triviaal ogende problemen weer moeilijk weten te maken, en dat is ook hier het geval. Een variatie op het ééndimensionale pakkingsprobleem is namelijk het *Random Parking Problem.* Hierbij willen auto's met een vaste lengte (zeg: vijf meter) parkeren langs een oneindig lange weg. De eerste bestuurder komt aanrijden, kiest een willekeurige plek en zet daar zijn auto neer. Elke volgende auto doet hetzelfde. Er kan een probleem optreden als de uitgekozen plek niet groot genoeg is voor de auto. In dat geval kiest de bestuurder willekeurig een nieuwe plek, eventueel net zolang tot er een passend gat is gevonden. Dit zal heel lang goed gaan, maar op een gegeven moment kan de situatie ontstaan dat alle gaten die overblijven te klein zijn voor een nieuwe auto. De parkeerruimte is dan verzadigd. Een praktische vraag die meteen bij je opkomt is: Hoe economisch zijn de auto's nu gemiddeld

geparkeerd? Hoeveel ruimte neemt een auto gemiddeld gesproken in? Het lijkt een kwestie van wat elementaire kansrekening. Eén mogelijk argument gaat als volgt: Een auto neemt minimaal een plek van vijf meter lang in. Op z'n ergst is er vóór de auto een ruimte van (bijna) vijf meter, waar nét geen nieuwe auto meer in past. Dat betekent dat elke auto een plek inneemt tussen 5 en 10 meter, gemiddeld is dat 7,5 meter. De gemiddelde bezettingsgraad is dus 75%.

De grote fout in het bovenstaande argument is natuurlijk de aanname dat de kansverdeling van de open ruimte voor elke auto uniform is. Als je het precies gaat bekijken blijkt het een lastig probleem te zijn. Als je het probleem eerst op een eindige weg met lengte x bekijkt, dan kun je het verwachte aantal auto's dat past $M(x)$ noemen. Deze $M(x)$ voldoet aan een ingewikkelde vergelijking, een differentie-differentiaalvergelijking:

$$xM'(x+1) + M(x+1) = 2M(x) + 1.$$

Het moeilijke aan deze vergelijking is dat er zowel afgeleiden (M') in voorkomen, zoals in een gewone differentiaalvergelijking, als ook verschuivingen in de argumenten (x en $x+1$). De vergelijking is niet analytisch op te lossen, maar merkwaardig genoeg is wel de limiet voor grote x te bepalen, en dat is alles wat we nodig hebben:

$$c = \lim_{x\to\infty} \frac{M(x)}{x} = \int_0^\infty e^{-2\int_0^t \frac{1-e^{-u}}{u}\mathrm{d}u}\mathrm{d}t = 0,7475\ldots$$

Dit antwoord lijkt heel erg op de 0,75 die we eerder vonden, maar dat is meer geluk dan wijsheid. De berekening werd voor het eerst in 1958 gedaan door de Hongaar Rényi.

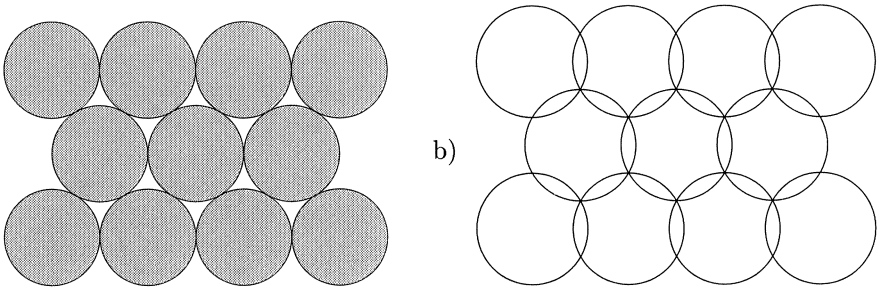
De oplossing werd in 1960 opgepakt door Palásti. Hij uitte het vermoeden, dat als je random gaat parkeren met d-dimensionale kubussen in d dimensies, dat dan de gemiddelde efficiëntie gelijk is aan c^d. Hierin is c het getal dat we in één dimensie hebben gezien.

In 1991 werd door Brosilow en medewerkers met behulp van zeer uitvoerige computersimulaties aangetoond dat de gemiddelde bedekkingsgraad gelijk is aan $0,562009 \pm 0,00004$. Dit klopt niet met de voorspelling van Palásti, want $c^2 = 0,558903\ldots$ Het vermoeden is dus niet waar, maar er is niet bekend wat het antwoord dan wel is.

2. Tweedimensionaal pakken: Het vermoeden van Fejes Tóth

De beste pakking van even grote munten op een grote tafel is eenvoudig te vinden. Je legt één munt neer. Vervolgens passen er precies zes omheen. Zo kun je doorgaan. Je krijgt een hexagonaal rooster met munten, waarbij elke munt aan precies zes andere raakt. De pakkingsdichtheid is $\pi/\sqrt{12} = 0,906899\ldots$, de bedekkingsgraad is dus ruimt 90%. De vermaarde meetkundige Coxeter merkte over deze pakking al op: *'The problem op packing, as densely as possible, an unlimited number of equal non-overlapping circles in a plane was solved millions of years ago by the bees.'* Ondanks het feit dat deze pakking er zo simpel uit ziet, liet het bewijs ervan door mensen nog geruime tijd op zich wachten. Het

was pas honderd jaar geleden, in 1892, dat de Noorse getaltheoreticus Thue een bewijs hiervoor gaf.

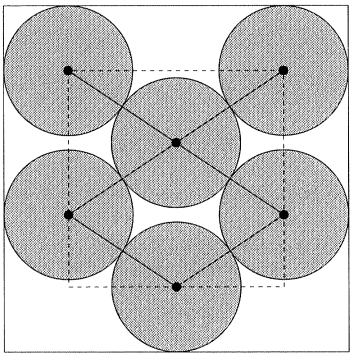


FIGUUR 1. De dichtste pakking en dunste overdekking van het vlak met congruente cirkels leveren allebei het hexagonale rooster.

Gerelateerd aan deze cirkelpakking is een vermoeden van de Hongaarse wiskundige László Fejes Tóth: Stel, je gaat uit van de beste cirkelpakking in het vlak, de hexagonale dus. Je pakt er n munten uit, en legt weer $n-1$ munten terug. Het vermoeden is dat de munten altijd op oude posities moeten komen te liggen, met andere woorden, er ontstaat weer een hexagonale pakking, alleen is er één gat ontstaan. Dit vermoeden is tot op heden niet bewezen.

3. STAPELEN IN EEN VIERKANT

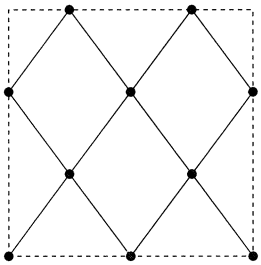
Als je even nadenkt, dan zie je snel dat het vinden van de dichtste cirkelpakking in een vierkant equivalent is met het neerleggen van evenveel punten zodanig dat de minimale afstand tussen deze punten maximaal is. De best mogelijke configuraties zijn ondertussen gevonden voor tot en met 20 cirkels. Daarboven zijn er nog veel vermoedens.



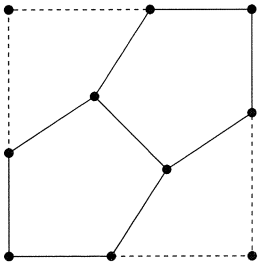
FIGUUR 2. Het vinden van de dichtste cirkelpakking is equivalent met het vinden van punten met een grootst mogelijke minimale afstand.

Uit de volgende figuur is duidelijk dat het vinden van de beste configuratie niet altijd eenvoudig was. Zelfs nadat de beste configuratie was gevonden bleven

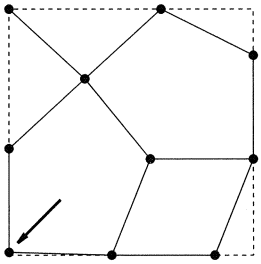
inferieure oplossingen opduiken.



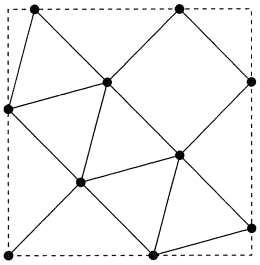
a) Goldberg, 1970
0.416666...



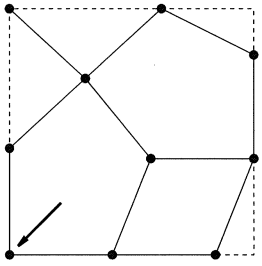
b) Schaer, 1971
0.419542...



c) Schlüter, 1979
0.421279...



d) Milano, 1987
0.420143...



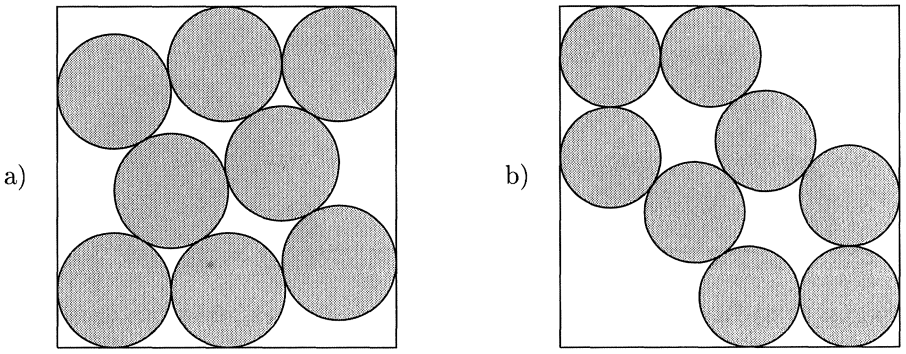
e) Valette, 1989
0.421190...

FIGUUR 3. De zoektocht naar de beste configuratie van tien punten in een vierkant.

4. STABIELE PAKKINGEN: HET SIGARETTENPROBLEEM

Het moet een vervent roker zijn geweest die het volgende probleem bedacht: Neem een nieuw pakje sigaretten. De sigaretten zitten daar stevig ingepakt. Als je één sigaret verwijdert dan blijft dat zo. Hoeveel sigaretten kun je nu weglaten zodanig dat de overgebleven sigaretten nog een stabiele structuur vormen? Wat is in het algemeen de minst dichte structuur van cirkels van

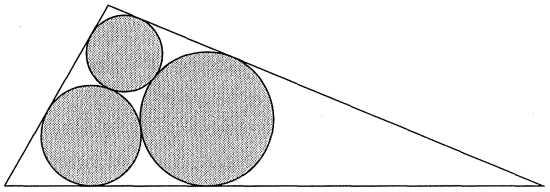
dezelfde grootte in een container (rechthoek, cirkel, etc.) die nog steeds stabiel is, in de zin dat geen enkele cirkel kan bewegen? Voor dit probleem zijn nog weinig vermoedens en geen bewijzen bekend.



FIGUUR 4. Dunne stabiele pakkingen van acht gelijke cirkels in een vierkant.

5. DRIE CIRKELS IN EEN DRIEHOEK: HET MALFATTI PROBLEEM

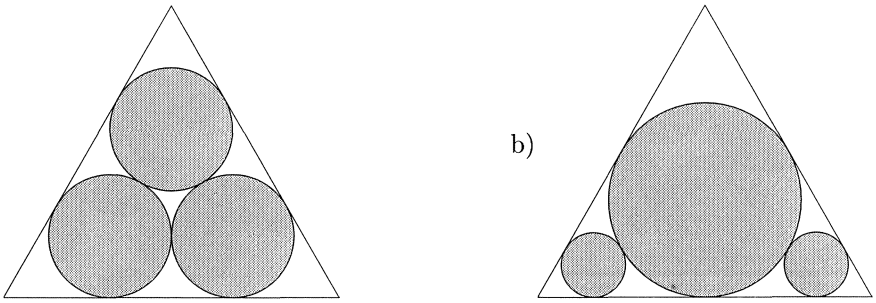
In 1803 stelde zich de Italiaan Malfatti de volgende vraag: Stel je hebt een stuk marmer in de vorm van een prisma en je wilt hier drie cilinders uit hakken, zodanig dat de cilinders samen een zo groot mogelijk volume hebben. Het is duidelijk hoe je dit moet doen, zei Malfatti: "*... cosicchè ciascun de circolo toccasse gli alti due ed insieme due lati del triangolo*". In de driehoek moet je drie cirkels tekenen die twee aan twee aan elkaar raken, zoals in de figuur. Het



FIGUUR 5. De drie Malfatticirkels in een driehoek.

probleem om bij een gegeven driehoek met passer en liniaal de stralen van deze drie cirkels te bepalen heette sindsdien het Malfattiprobleem. In 1930 begon er wat twijfel te rijzen. Lob en Richmond vonden namelijk een configuratie in een gelijkzijdige driehoek, die qua oppervlakte net iets beter was dan de configuratie die Malfatti voorstelde. Zij plaatsen namelijk eerst één grote cirkel, en vulden vervolgens twee van de gaten met een zo groot mogelijke cirkel. In 1967 kwam Goldberg met een verrassende constatering: Malfatti heeft nooit gelijk. Het is altijd beter om de zogenaamde "greedy" aanpak te volgen, waarbij telkens een zo groot mogelijke cirkel wordt geplaatst in de overblijvende ruimte. Goldberg maakte zijn bewering plausibel met numerieke resultaten. Een hard bewijs is

tot nu toe echter nog niet geleverd. Een voor de hand liggende generalisatie, als je de driehoek vervangt door een algemenere figuur, is niet juist.



FIGUUR 6. a) De Malfatti cirkels in een gelijkzijdige driehoek. De totale oppervlakte van de cirkels is $(6-3\sqrt{3})/8 = 0.100480\ldots$ b) De "greedy" configuratie met een totaal oppervlakte van $11/108 = 0.101851\ldots$ is beter!

6. STAPELEN IN DRIE DIMENSIES: HET PROBLEEM VAN KEPLER

Hoe stapel je bollen in drie dimensies, of, zeg maar: sinasappels of kanonskogels? Vooral het laatste bleek van groot praktisch belang. Aan het eind van de zestiende eeuw vroeg Sir Walter Raleigh, de kaper-admiraal en gunsteling van koningin Elisabeth I, zich af hoeveel kanonskogels er in zo'n nette piramide gestapeld konden worden. Hij droeg zijn wetenschappelijk adviseur, de astronoom en wiskundige Thomas Harriot, op om dit uit te zoeken. Deze had het probleem snel genoeg opgelost, maar het belangwekkende was dat hij hierover in correspondentie raakte met Johannes Kepler. Kepler kreeg het idee dat hij de zestallige structuur van sneeuwkristallen kon verklaren door uit te gaan van atomen die op een regelmatige manier gestapeld waren. In een geschrift uit 1609 beschreef hij voor het eerst de dichtste bolstapeling. Deze stapeling krijg je door eerst in het vlak een dichtste stapeling te maken van één laag: een hexagonaal patroon, waarbij iedere bol raakt aan zes buren. Vervolgens maak je een zelfde laag e n die leg je er iets verschoven op, zodanig dat ze net in de kuiltjes van de laag eronder rusten. Bij elke nieuwe laag zijn er overigens twee keuzes te maken die niet equivalent zijn. Als we de positie van de eerste laag met A aangeven en van de tweede met B, dan kan de derde laag precies boven de eerste laag komen te liggen (A) of of een andere plek: C. De volgorde $ABCABCABCABC\ldots$ heet wel de *Face-Centered Cubic* stapeling, de keuze $ABABABABABAB\ldots$ wordt *Hexagonal Close Packing* genoemd. Door deze keuzemogelijkheid kun je ook heel eenvoudig een niet-periodieke stapeling maken: $ABCBABABCBABCBCBCBABABC\ldots$. De dichtheid van de stapeling wordt niet beïnvloed door deze keuzes, de dichtheid is $\pi/\sqrt{18}$. Kepler ging er eigenlijk zondermeer vanuit dat deze voor de hand liggende stapeling inderdaad niet verbeterd kon worden. Hij gaf geen bewijs, maar sindsdien staat het probleem van het bepalen van de dichtste bolstapeling bekend als het probleem van Kepler.

Het bleek inderdaad een formidabel probleem. Honderden jaren lang lukte het wiskundigen niet om een bewijs te leveren. Gauss was de eerste die iets in de richting bewees. In een recensie van een boek bewees hij in 1840 terloops even dat de beste bolstapeling die een regelmatig rooster vormt inderdaad degene is die Kepler al voor ogen had. Veel wetenschappers hebben geprobeerd om bovengrenzen af te leiden, in de hoop om zo dichter bij de waarheid te komen. De meest recente bovengrens voor de dichtheid is die van Muder (1993): $0,773055\ldots$. Deze is nog ver verwijderd van de echte waarde: $0,7405\ldots$. Veel experimentatoren hebben ook geprobeerd om de maximale stapelingsdichtheid te bepalen door bussen met metalen kogeltjes, hagelkorrels, en zelfs erwten te schudden en te trillen. Helaas blijkt dat het onmogelijk is om zo de buurt te komen van de beste pakkingsdichtheid. Als je een heleboel kogeltjes in een potje doet krijg je een stapeling die de *random loose packing* heet, de dichtheid hiervan is ongeveer zo'n 0,6. Als je vervolgens goed gaat schudden en stampen (denk aan een pot suiker die je zo kunt inklinken) win je nog wat ruimte en krijg je een *random dense packing* met een dichtheid van ongeveer 0,64. Dit is nog ver verwijderd van de optimale waarde van $0,7405\ldots$. Dit heeft te maken met het feit dat er zich lokaal kristalstructuren vormen die zich verzetten tegen lokale verstoringen. Er zijn heel veel van deze lokaal optimale structuren.

Er zijn een aantal redenen te noemen waarom het probleem van Kepler zo moeilijk is. Zo is er niet één optimale oplossing, maar oneindig veel verschillende, zoals we al eerder hebben gezien. De locale structuur in een bolstapeling ligt niet vast, zoals in de tweedimensionale situatie. Daar wordt elke cirkel omgeven door precies zes cirkels, en voor de omringende cirkels is geen speelruimte meer. Ze kunnen hoogstens tegelijk draaien om de centrale cirkel. Deze situatie is compleet anders in de ruimte. De eerste vraag is al: hoeveel bollen passen er om een bol? Hoeveel bollen kunnen tegelijkertijd aan één bol raken (In vaktaal heet dit het *kissing number*)? Twaalf kan in ieder geval, want dat is zo in de FCC pakking, maar passen er misschien ook dertien omheen? Deze vraag is het onderwerp geweest van een controverse in 1694 tussen de Schotse astronoom David Gregory en Isaac Newton. Gregory beweerde dat dertien mogelijk was, maar kon geen configuratie aangeven, terwijl Newton het gevoel had dat tw aalf maximaal is. Het probleem is dat met 12 rakende buren de situatie nog niet vast ligt. Als je de buitenste bollen volgens een icosaëder rangschikt zou je de buitenste bollen nog zo'n 5% groter kunnen maken zonder dat ze contact met de binnenste bol verliezen. Deze vrijheid is voldoende om de buitenste bollen allemaal van plaats te kunnen laten verwisselen, zonder dat ze contact met de binnenste bol verliezen. Verder is de ruimtehoek die een rakende bol maakt, gezien vanuit het middelpunt van de middelste bol minder dan $1/13$ van de totale ruimtehoek, dus wat dat betreft zouden er 13 kunnen passen. Het verlossende woord werd pas in 1874 gesproken door de Duitser Hoppe die het gelijk van Newton bewees.

Een andere dissonant is dat, anders dan in het vlak, de 'locale dichtheid' wel groter kan zijn dan de maximale dichtheid van $\pi/\sqrt{18}$. Een stapeling kan dus ergens wat dichter zijn, maar kennelijk wordt dat altijd ergens anders weer opgeheven door een minder dicht stuk in de stapeling. Dit verschijnsel werd

in de vijftiger jaren opgemerkt door onze landgenoot Boerdijk. Een vervelende consequentie is dat je geen bewijs kunt vinden door het afschatten van die locale dichtheid.

Een lastig, zeer oud probleem met een voor de hand liggende oplossing. De pakkingsdeskundige C. A. Rogers verwoordde het in 1958 als volgt: *'Many mathematicians believe, and all physicists know that* 0.7405... *is the highest possible packing density.'* De oplossing ligt zo voor de hand, dat veel mensen niet eens wisten dat het nog niet is bewezen. Milnor schreef twintig jaar geleden: *'However, the corresponding situation in 3 dimensions remains unsolved. This is a scandalous situation, since the (presumably) correct answer has been known since the time of Gauss.'* In 1900 nam David Hilbert het probleem tijdens zijn fameuze toespraak tot het Internationale Wiskunde Congres in Parijs op als onderdeel van zijn achttiende probleem. Dit was een impliciete erkenning voor de moeilijkheidsgraad.

De Chinees-Amerikaanse differentiaalmeetkundige Hsiang publiceerde in 1993 een bewijs van het vermoeden van Kepler. Het 90 pagina's tellende artikel in het *International Journal of Mathematics* ontlokten aanvankelijk euforische reacties aan populair wetenschappelijke tijdschriften zoals *Nature* en *New Scientist* en de internationale pers, maar deze geluiden verstomden al snel toen bleek dat de wetenschappelijke wereld na kritische lezing een andere mening was toegedaan. Hsiang maakte in zijn bewijs een aantal foutjes die wel te corrigeren waren, en een aantal erg grove *easy to see* stappen, die nog heel wat details zouden vergen. In de discussie die volgde bleek Hsiang niet te overtuigen te zijn van de noodzaak om de nodige details in te vullen. Zijn bewijs wordt dan in de wetenschappelijke wereld niet als bewijs geaccepteerd.

Op 9 augustus 1998 kondigde Thomas Hales van de universiteit van Michigan per e-mail aan dat het hem gelukt was om het bewijs af te ronden: '*I have started to distribute copies of a series of papers giving a solution to the Kepler conjecture, the oldest problem in discrete geometry*'. Hales heeft jaren gewerkt aan een reeks van artikelen die tot doel hadden het complete bewijs te leveren. Het bewijs bestaat nu uit zeven artikelen en een proefschrift van één van zijn studenten, in totaal 250 pagina's, samen met 3 gigabytes aan computercode. Het bewijs van Hales is gebaseerd op een idee van de Hongaar László Fejes Tóth, die al in 1953 opmerkte dat het probleem kon worden gereduceerd tot een aantal optimaliseringsproblemen in een groot aantal variabelen, in die tijd nog een onoverkomelijk probleem. In het bewijs van Hales worden een 5000 mogelijke configuraties doorgerekend en afgeschat. Elke situatie levert een optimaliseringsprobleem met zo'n 150 variabelen. Hales heeft hiervoor onder andere int ervalaritmetiek toegepast, waarmee aan te tonen is dat het numeriek gevonden optimum ook echt is.

Het bewijs van Hales betekent nog niet dat dit hoofdstuk helemaal is afgesloten. Hoe zit het bijvoorbeeld met de optimale stapeling van vierdimensionale bollen? Het is ook nog niet bekend of een vierdimensionale bol tegelijk aan maximaal 24 niet-overlappende bollen kan raken, of aan 25.

7. Het probleem van Tammes

De Nederlandse botanicus Tammes heeft de wereld een probleem nagelaten dat nu zijn naam draagt. Tijdens onderzoekingen in de jaren dertig aan pollenkorrels merkte hij dat de uittree-openingen op deze korrels allemaal ongeveer even ver van elkaar afzitten, ze zijn ongeveer uniform verdeeld. Dat riep bij hem de vraag op hoe je eigenlijk n punten uniform op een bol verdeelt. Een criterium daarbij zou kunnen zijn dat je de punten zo ver mogelijk uit elkaar wilt plaatsen. Dat betekent dat de minimale afstand tussen alle punten op de bol maximaal moet zijn. Op een cirkel zou dat flauw zijn, ze komen op gelijke afstand van elkaar te liggen, maar op een bol ligt het even niet voor de hand hoe je dat moet oplossen. Een equivalente formulering van het probleem is om een aantal niet-overlappende cirkels op de bol te plaatsen, zodanig dat de cirkels even groot en zo groot mogelijk zijn, met andere woorden: de dichtst mogelijke cirkelpakking op de bol. Het probleem is opgelost voor $2 \leq n \leq 14$ en voor $n = 24$, onde r andere door Fejes Tóth en de Nederlander Van der Waerden.

8. Het worstvermoeden van Fejes Tóth

Bij een stapeling van een eindig aantal cirkels of bollen kun je ook spreken van een dichtst mogelijke stapeling. Hierbij kun je echter verschillende criteria hanteren. Zo kun je bijvoorbeeld bij een verzameling bollen de gemiddelde afstand van de middelpunten minimaliseren, of de inhoud van het convexe omhulsel (de kleinste convexe verzameling die de bollen omvat), of de oppervlakte van het convex omhulsel. Als je bijvoorbeeld in het vlak de oppervlakte van het convex omhulsel van een aantal cirkels minimaliseert krijg je als beste configuratie een cirkelvormige kluit met cirkels die een deel vormen van hexagonale pakking die globaal het beste is. Als je hetzelfde in de ruimte doet en de inhoud van het convexe omhulsel minimaliseert gebeurt er iets heel anders. Voor kleine aantallen bollen bestaat de beste oplossing uit een aantal bollen die op één rij achter elkaar liggen. Het convex omhulsel is dan een worst. Dit geldt tenminste als je niet meer dan 56 bollen hebt, daarboven treden er andere oplossing en op. In vier dimensies gebeurt er iets soortgelijks. Tot een bepaald aantal (dit aantal is niet bekend, maar maximaal 367,300) is de rechte sliert het beste. Vanaf dimensie vijf en hoger heeft echter László Fejes Tóth het vermoeden uitgesproken (de *Sausage Conjecture*, of ook *Wurstvermutung*) dat de rechte worst altijd het beste is. Een bewijs hiervan is in 1996 geleverd door Betke en Henk voor dimensie 42 en hoger. Er blijft nog een interessant gat over om te bewijzen!

9. Overdekkingen van een cirkel

Rond de vorige eeuwwisseling bestond er op Engelse kermissen een bekend spel "Spot-the-spot". Op een tafel lag een kleed met daarop een rode cirkel geschilderd. Een speler kreeg vijf blikken bordjes en moest daarmee de rode cirkel compleet bedekken. De bordjes mochten na neerleggen niet meer worden aangeraakt. Het zal duidelijk zijn dat de moeilijkheid van het spel was gelegen de kleine afmeting van de kleine bordjes, waardoor de rode schijf moeilijk te overdekken was. Dit werpt de vraag op naar de minimale straal σ_n van de n

TABEL 1. Zuinigste overdekkingsstraal σ_n voor een eenheidscirkel.

n	Vermoeden	Bewijs	σ_n	Waarde
1,2		elementair	1	=1.000000...
3		elementair	$\sqrt{3}/2$	=0.866025...
4		elementair	$1/\sqrt{2}$	=0.707106...
5	Neville 1915	K. Bezdek 1983		0.609382...
6		K. Bezdek 1979		0.555905...
7		elementair	1/2	=0.500000...
8	Nagy 1974	G. Fejes Tóth '96	$(1+\cos(2\pi/7))^{-1}$	=0.445041...
9	Nagy 1974	G. Fejes Tóth '96	$\sqrt{2}-1$	=0.414213...
10	Nagy 1974		$(1+\cos(2\pi/9))^{-1}$	=0.394930...
11	Melissen/Schuur '96			0.380006...
12	Melissen/Schuur '96		$(1+s-8/s)/3$, $s=\sqrt[3]{1+3\sqrt{57}}$	=0.361103...

kleine cirkels, zodanig dat overdekking van een eenheidscirkel nog net mogelijk is. Dit probleem is opgelost tot en met negen cirkels: Zie Figuur 7 op pagina 73.

10. OPGAVEN

1. Kanonskogels kun je op twee manieren als een piramide stapelen. Je kunt namelijk beginnen met een driehoek, of met een vierkant.

1a. Vind een formule voor d_n, het aantal kanonskogels in een driehoekige piramide die uit n lagen bestaat.

1b. Doe hetzelfde voor v_n, het aantal kogels in een vierkante piramide.

1c. Laat zien dat je met de kogels uit twee opeenvolgende driehoekige stapels precies genoeg hebt om een vierkante piramide te maken: $d_{n-1}+d_n=v_n$.

1d. Is er een driehoekige piramide die precies genoeg kogels bevat om er een vierkante piramide van te stapelen?

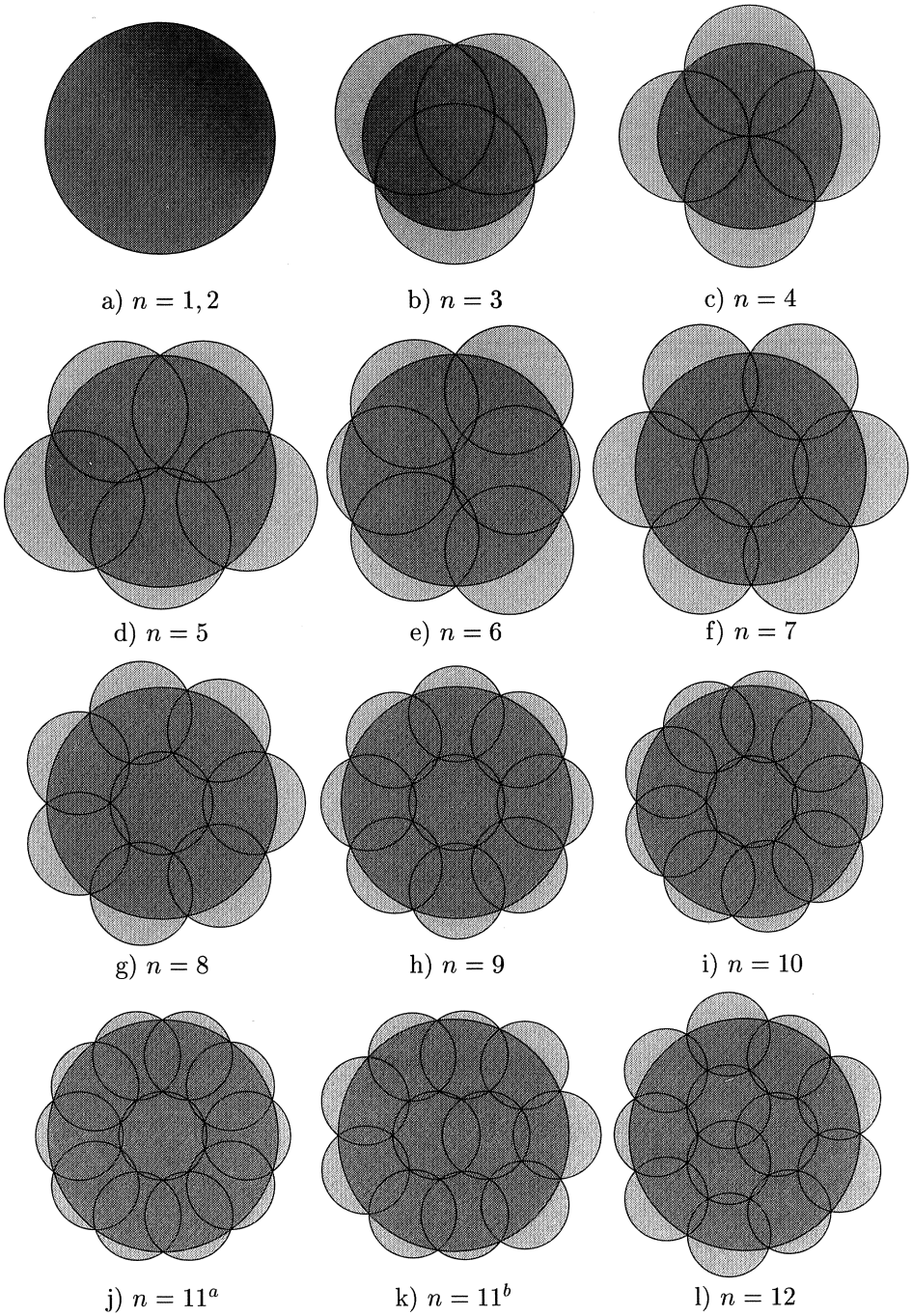
1e. Hoe hoog is een stapel kanonskogels van n lagen hoog? De kogels hebben een straal van 10 cm.

2. Er zijn twee voor de hand liggende manieren om drie cirkels of drie bollen zo dicht mogelijk tegen elkaar te stapelen. De eerste manier is tegen elkaar aan, met de middelpunten op een rechte lijn. De tweede is allemaal rakend aan elkaar, met de middelpunten volgens een gelijkzijdige driehoek. Reken in beide gevallen zowel voor cirkels als bollen de omtrek (oppervlakte) en de oppervlakte (inhoud) van het convex omhulsel uit, en trek daar eventueel conclusies uit.

LITERATUUR

J.B.M. Melissen, *Packing and Covering with Circles*, Proefschrift, Universiteit Utrecht, 1997 (bevat vrijwel alle relevante verwijzingen).

http://www.stetson.edu/~efriedma/packing.html



FIGUUR 7. Optimale ($n \leq 9$) en vermoedelijk zuinigste overdekking van een cirkel met cirkels.

Knopen

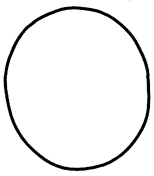
GERARD VAN DER GEER

Iedereen weet hoe je een knoop in een touwtje kunt leggen. Die knoop kun je er daarna ook weer uithalen. Dat lukt echter niet altijd meer als je de uiteinden van het touwtje met elkaar verbindt. Daardoor ontstaat dan een cirkelvormige figuur die om zichzelf verstrengeld is: een knoop. Daarom neemt de wiskundige als definitie van een knoop: *een cirkel die in de 3-dimensionale ruimte* $\mathbb{R}^3$ *is ingebed.* De eerste vraag die dan opkomt is of deze knoop te ontwarren is of dat we te maken hebben met een echte knoop. Een knoop ontwarren betekent hier: de knoop continu deformeren zodat de triviale knoop ontstaat:

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, z = 0\}. \tag{1}$$

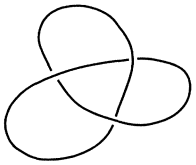
Hierbij moeten de spelregels wel duidelijk vastgelegd worden. Zo is het niet toegestaan de knoop oneindig klein te maken en daarmee te laten verdwijnen, alhoewel onze draad oneindig dun is. De deformatie wordt gegeven door een continue familie $\{h_t : 0 \leq t \leq 1\}$ van deformaties $h_t : \mathbb{R}^3 \to \mathbb{R}^3$ van de omringende ruimte, die de knoop in een cirkel overvoeren terwijl de tijd van 0 tot 1 loopt. Hierbij is h_0 de identiteit, en iedere h_t is een homeomorfisme, d.w.z., een continue bijectieve afbeelding waarvoor de inverse afbeelding ook continu is. Terwijl de tijd loopt blijft het touwtje dus intact; het mag niet 'breken'. Je kunt de deformatie van de knoop vergelijken met de beweging van het touwtje in een vloeistof; de omringende vloeistof wordt meebewogen.

Laat ik eerst wat voorbeelden van knopen geven: de triviale knoop gegeven door de vergelijking (1)



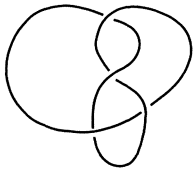
triviale knoop

en bekende knopen als



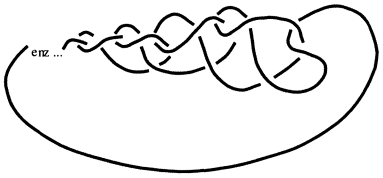
klaverbladknoop

en



cijfer 8-knoop

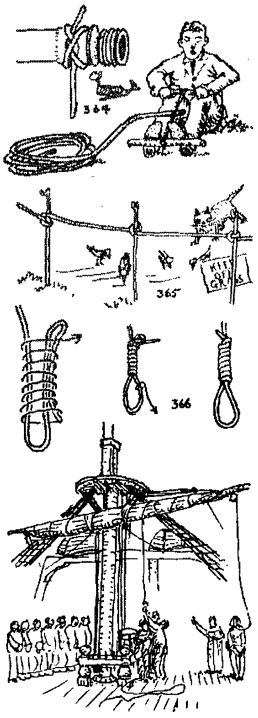
of een wat vervaarlijker type



Een wilde knoop

Laatstgenoemde knoop laat zich ontwarren door rechts aan de knoop te trekken; maar daarvoor hebben we wel oneindig veel tijd nodig. Dit soort knopen zijn ook veel wiskundigen te wild en wij zullen ons alleen met *tamme* knopen bezighouden.

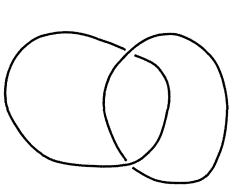
Knopen vinden we overal in het dagelijks leven, bij het zeilen, bij de kleermaker, in onze veters en zelfs de beul heeft zijn eigen knoop:



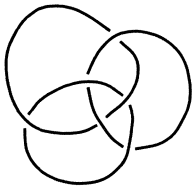
Knoop van de Beul; uit: The Ashley Book of Knots

In de voorbeelden werden de knopen met een figuur of diagram in het vlak aangegeven. Zo een figuur heet een *knopendiagram*. Je kunt zo een figuur opvatten als een schaduw van de knoop, waarbij we boven- en onderkruisingen hebben gemarkeerd. Een probleem daarbij is dat een en dezelfde knoop veel verschillende knopendiagrammen kan hebben.

Er is geen reden om ons te beperken tot één cirkelfiguur; vaak komen we er meer tegen. Dus krijgen we een variatie op ons thema door meerdere cirkels in de drie-dimensionale ruimte in te bedden. Het resultaat heet dan een *schakel* of *verstrengeling*. Hier zijn een paar voorbeelden:



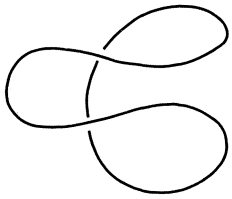
Hopf-schakel



Borromeïsche ringen

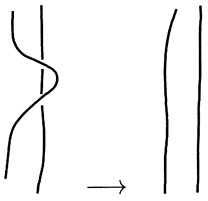
De laatste schakel heeft de bijzondere eigenschap dat het verwijderen van één cirkel ervoor zorgt dat de beide anderen niet meer geschakeld zijn.

Knopen die we door een deformatie als boven in elkaar kunnen overvoeren heten equivalent of 'knopen van hetzelfde type'. Zo beschouwen we bijvoorbeeld de volgende knoop als gelijkwaardig met de triviale knoop



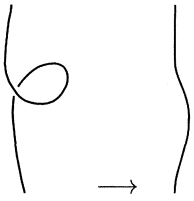
Triviale knoop incognito

Dat betekent dat we in het diagram de volgende beweging of 'zet' mogen uitvoeren



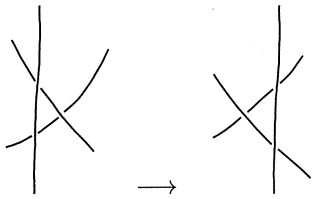
Reidemeisterzet II

Ook de volgende bewegingen of zetten zijn toegestaan



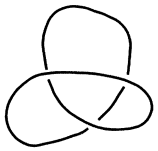
Reidemeisterzet I

en



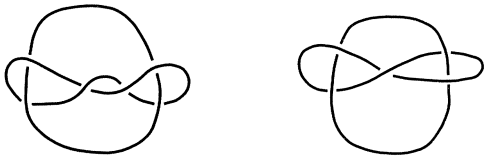
Reidemeisterzet III

zoals men ziet door de projectie (schaduw) een beetje te deformeren. Ik geef nu een aantal equivalente diagrammen. De triviale knoop komt in vele vermommingen voor:

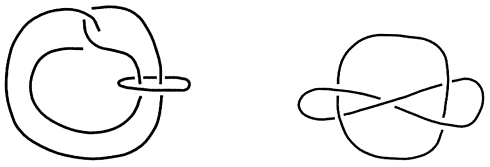


Triviale knoop vermomd

Het is niet moelijk de benodigde Reidemeisterzetten te vinden, zoals ook bij het volgende paar



of bij het paar



Maar vaak is het niet zo simpel.

De Duitse wiskundige Kurt Reidemeister heeft laten zien dat twee verschillende diagrammen precies dan hetzelfde knopentype in de ruimte voorstellen, als de beide diagrammen door herhaalde toepassing van de Reidemeisterzetten in elkaar kunnen worden overgevoerd. Helaas is er geen goed algoritme om dit in de praktijk uit te voeren. Dat leidt tot het hoofdprobleem van de knopentheorie: *Hoe kan men vaststellen of twee knopendiagrammen (schakeldiagrammen) dezelfde knoop (schakel) voorstellen?* Dat het ons na lang proberen niet gelukt is het ene diagram door Reidemeisterzetten in het andere over te voeren zegt misschien maar weinig over de knoop, maar meer over ons eigen onbegrip.

Soms is het echter wel mogelijk vast te stellen dat twee knopen of schakels niet van hetzelfde type zijn. Daarvoor worden *invarianten* gebruikt. Ik probeer dit met een voorbeeld duidelijk te maken. Neem een diagram van een schakel die uit twee al dan niet verstrengelde cirkels bestaat. We bekijken daarin uitsluitend de kruisingen van de twee cirkels, dus niet de kruisingen waar een

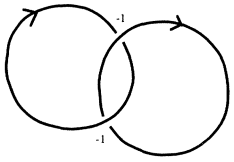
cirkel zichzelf kruist. We voorzien de twee cirkels elk van een oriëntatie, dat wil zeggen een looprichting. Voor zo'n kruising zijn er dan twee mogelijkheden (eventueel na draaien van het diagram):

$+1$ of -1

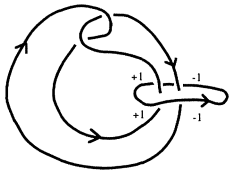
(dwz. het is $+1$ als je op de bovenkruising aankomt en het verkeer beneden komt van rechts). Op deze manier kunnen we iedere kruising van beide cirkels een voorteken ϵ voorzien; we tellen al deze voortekens op en noemen de helft van de totale sum "het schakelgetal":

$$\ell(K) = \frac{1}{2}\sum \epsilon.$$

Als we op een van de twee cirkels de orientatie omkeren dan keert het teken van $\ell(K)$ ook om: $\ell(K) \mapsto -\ell(K)$. Hier volgen eerst twee voorbeelden:



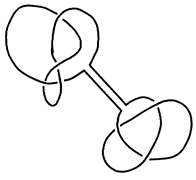
Hopf-schakel met schakelgetal: $\frac{-1-1}{2} = -1$



Whitehead-schakel met schakelgetal: 0

Men gaat nu na dat het schakelgetal van zo een schakeldiagram niet verandert onder toepassing van Reidemeisterzetten. (Omdat we alleen naar kruisingen van de twee cirkels en niet die van een cirkel met zichzelf kijken, spelen alleen zetten van type II en III een rol; daarvoor is het niet moeilijk.) Daarmee hebben we nu een invariant van schakels met twee cirkels gevonden. Zijn twee zulke schakels van hetzelfde type, dan zijn hun schakelgetallen gelijk. We zien nu bijvoorbeeld dat de Hopfschakel en de Whiteheadschakel niet van hetzelfde type zijn.

Knopen kan men net als getallen met elkaar vermenigvuldigen: men verbindt beide knopen als in de figuur.



De knopen die zo ontstaan heten samengestelde knopen. Knopen die men niet ontbinden kan heten *priemknopen*. Er zijn zeer veel priemknopen en daarmee ook vele knopen. Het aantal priemknopen met k kruisingen is een exponentiële functie van k; voor lage waarden van k is dit aantal in de volgende tabel aangegeven.

#kruisingen	3	4	5	6	7	8	9	10	11	12	13	14
#priemknopen	1	1	2	3	7	21	49	165	552	2176	9988	?

Het aantal priemknopen met 14 kruisingen is nog niet bekend.

Geschiedenis van de Knopentheorie

Het is een goede gewoonte de geschiedenis van welk thema dan ook in de klassieke oudheid te laten aanvangen, en ik sluit me graag aan bij deze traditie. Zo dus: het eerst bekende probleem uit de knopentheorie is de *Gordiaanse Knoop*. In Gordion, een stad in Frygië (in het huidige Turkije), stond bij de tempel voor Zeus een ossenkar waarmee volgens een legende Midas de stad in gekomen was voor hij tot koning gekroond werd. Volgens overlevering zou diegene koning van de wereld worden, die de reusachtige knoop los zou maken waarmee de kar aan de tempel vastgemaakt was. Toen Alexander de Grote, koning van Macedonië, in het jaar –334 Gordion veroverde ging hij de beroemde knoop bekijken en hij ruimde het probleem resoluut uit de weg zoals we van hem mogen verwachten. (Hoe hij de knoop precies losmaakte, daarover verschillen de bronnen van mening; bij Plutarchos kan men de verschillende versies nalezen.)

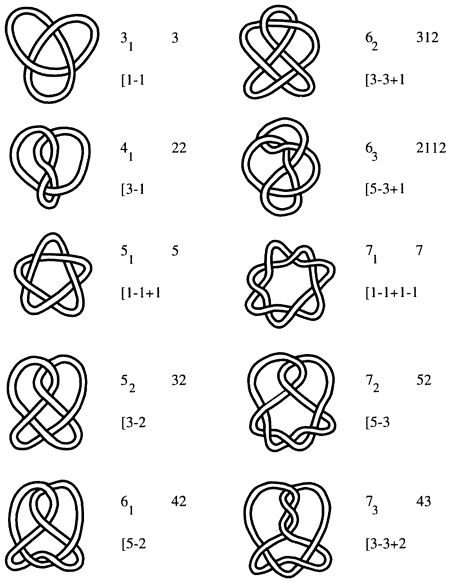
Het heeft lang geduurd voordat de knopen de aandacht van de geleerden trokken. C.F. Gauss was misschien de eerste die knopen als wiskundige objecten beschouwde, maar hij heeft er nauwelijks over gepubliceerd. Zijn leerling J.B. Listing wijdde een gedeelte van zijn boek *Vorstudien zur Topologie* aan knopen. Maar men zou met recht kunnen beweren dat de knopentheorie pas tot ontwikkeling is gekomen door de pogingen van Lord Kelvin (Willam Thompson, 1824–1907, Edinburgh) om de aard van atomen te begrijpen. Hij had het idee dat atomen knopen in de ether waren. De ether was hierbij een hypothetische substantie, waarvan de hele ruimte doortrokken was. Hij uitte zijn idee in het jaar 1869 (On vortex motion, *Trans. Royal Soc. Edinburgh* **25**, 217–260). Verschillende knopen zouden tot verschillende atomen, dus ook tot verschillende chemische elementen aanleiding geven. Dat elementen stabiel zijn, zou een extra voorwaarde in dit model zijn. Hij hoopte op deze manier niet alleeen de elementen, maar ook de spectraallijnen in de stralingsspectra te kunnen verklaren. Ik moet toegeven, dat ik dit nog steeds een zeer aantrekkelijk idee vind. Het zou de veelheid van de elementen kunnen verklaren en ook hoe elementen door transmutaties in elkaar overgaan.

De eerste classificatie van knopen stamt van de dominee Thomas P. Kirkman (The enumeration, description and construction of knots with fewer than 10 crossings. *Trans. Royal Soc. Edinburgh* **32**, (1884), p. 281–309.) Helaas zijn

zijn geschriften niet erg toegankelijk. De Schotse natuurkundige Peter Guthrie Tait (1831–1901) heeft de methoden van Kirkman gebruikt en geprobeerd knopen te classificeren om zo het idee van Lord Kelvin te verwezenlijken en de verschillende elementen (in de zin van de scheikunde) te bepalen. Twintig jaar lang werd het idee van Lord Kelvin serieus bestudeerd. Maxwell merkte op dat dit atoommodel aan meer voorwaarden voldeed dan alle andere. Maar helaas (?!) is de ether verdwenen en daarmee ook dit atoommodel.

Terwijl Tait knopen klassificeerde, werkte tegelijkertijd ook de Amerikaanse wiskundige C.N. Little daaraan; ook hij maakte gebruik van de lijsten van Kirkman. In 1899 publiceerde hij een lijst van (niet-alternerende) knopen met 10 kruisingen.* Bij deze klassificatie werden heuristische principes, onbewezen aannames, gebruikt. Veel daarvan golden tot voor kort nog als vermoedens.

Ik geef hieronder de eerste bladzijde van een moderne klassificatie (uit D. Rolfsen: *Knots and Links*).



Classificatie

Langzamerhand begonnen steeds meer wiskundigen zich voor knopen te interesseren, maar het systematische onderzoek naar knopen ontwikkelde zich pas in de 20ste eeuw. In dit verband moeten de namen van M. Dehn, J.W. Alexander en natuurlijk K. Reidemeister, die we al tegenkwamen, genoemd worden; daarbij hoort ook de naam van Emil Artin (Theorie der Zöpfe, 1925).

De eerste mijlpaal was het Alexanderpolynoom (1928); daar kom ik later op terug. Het boek van Reidemeister (Knopentheorie, 1932) vat de resultaten van dit eerste tijdvak goed samen.

---

* Een fout in de lijst van Little werd pas in 1974 door de jurist Perko uit New York ontdekt.

De knopentheorie ontwikkelde zich dan gedurende meer dan vijftig jaar gestaag verder. Van de hoger-dimensionale topologie en meetkunde, die zojuist was ontwikkeld, werd dankbaar gebruik gemaakt om knopen beter te begrijpen. Maar dan vindt in 1984 plotseling een ware revolutie plaats, die deze hele tak van wiskunde dramatisch verandert. Terugblikkend op de eerdere periode moet men constateren dat de het tijdvak 1930–1984 weliswaar vooruitgang, maar geen grote doorbraken heeft gebracht.

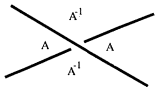
In 1984 ontdekt Vaughan Jones, een wiskundige uit Nieuw-Zeeland, een nieuwe invariant van knopen, het *Jones-polynoom*. Deze invariant is niet afkomstig uit de hoger-dimensionale meetkunde, maar via een omweg (zie later) direct uit de knopendiagrammen. Jones heeft hiervoor de Fieldsmedaille, de belangrijkste onderscheiding voor wiskundig werk, gekregen. Ons voert het nu, na een lijst met de belangrijkste namen uit de knopentheorie, tot het volgende thema.

C.F. Gauss (1777-1855), J.B. Listing (1808-1882), W. Thompson (Lord Kelvin; 1824-1907), J.C. Maxwell (1831-1879), Th. P. Kirkman (1806-1895), P.G. Tait (1831-1901), C.N. Little, J.W. Alexander (1888-1971), M. Dehn (1878-1952), W. Burau, H. Seifert (1907-), E. Artin (1898-1962), K. Reidemeister (1893-1971), J.H. Conway (1937-), V. Jones (1952-), V. Vassiliev (19??).

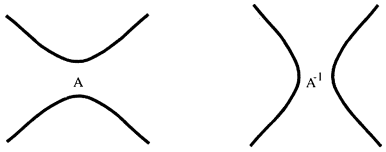
Polynomen

We spelen nu een spel met een knoopdiagram en ik hoop dat u meespeelt. De regels zijn eenvoudiger dan die van het schaakspel. We illustreren ze aan de hand van het voorbeeld van de Hopfschakel.

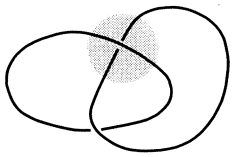
Eerste stap. Neem een kruising uit het diagram van de schakel. In de vier gebieden noteren we afwisselend de uitdrukkingen A en A^{-1}. Het wordt zo gedaan, dat als je op de bovenste weg van het kruispunt rijdt je voor aankomst bij de kruising het A-gebied aan de rechterhand hebt.



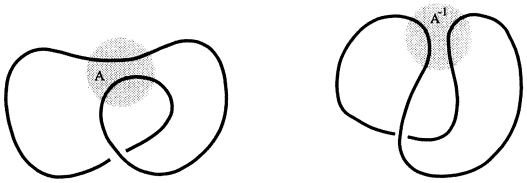
We kunnen nu op twee verschillende manieren de kruising opheffen om zo een eenvoudiger figuur te krijgen; de een wordt genoteerd met A, de andere met A^{-1} (afhankelijk van welke gebieden worden verbonden):



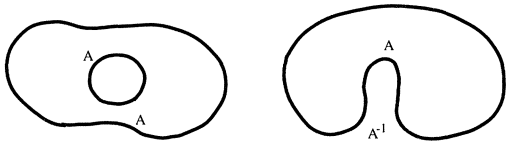
We krijgen op deze manier uit het voorbeeld van de Hopfschakel



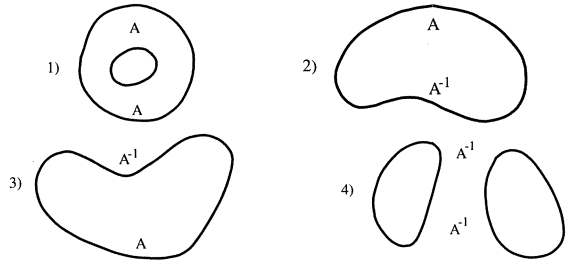
twee nieuwe schakeldiagrammen die zijn voorzien van een symbool (A of A^{-1}).



Tweede stap. We nemen nu op elk van de diagrammen een willekeurige kruising en voeren de eerste stap weer uit. Uit het eerste diagam ontstaan twee nieuwe



en uit het andere ook, zodat we in totaal nu vier diagrammen vinden die met symbolen A of A^{-1} versierd zijn:



de vier diagrammen

Voorlaatste stap. Nu er geen kruising meer voorhanden is om het proces te herhalen doen we de boekhouding voor het totaal: we nemen het product van de symbolen in iedere figuur en vermenigvuldingen het met de uitdrukking

$$(-A^2 - A^{-2})^{\ell-1},$$

waarbij ℓ het aantal cirkels in de figuur is. Voor de vier figuren in het bovenstaande voorbeeld levert dat de producten

$$\begin{aligned} A^2 \cdot (-A^2 - A^{-2})^1 &= -A^4 - 1, \\ 1 \cdot (-A^2 - A^{-2})^0 &= 1, \\ 1 \cdot (-A^2 - A^{-2})^0 &= 1, \\ A^{-2} \cdot (-A^2 - A^{-2})^1 &= -1 - A^{-4}, \end{aligned}$$

Optellen levert in totaal:

$$\mathcal{K} = -A^{-4} - A^4.$$

Deze uitdrukking $\mathcal{K} = \mathcal{K}(K)$ heet *Kauffman-haakje* van K, en wordt ook wel genoteerd als $\langle K \rangle$, vanwaar de naam. We kunnen deze stappen, misschien vaak herhaald, voor ieder diagram uitvoeren. Een diagram met k kruisingen levert na dit proces 2^k diagrammen, waarvan men de uitdrukkingen moeten uitrekenen en optellen. Deze 2^k diagrammen noemen we de '*toestanden*' van de knoop.

Wat gebeurt er als we voor dezelfde knoop of schakel een ander diagram kiezen? Het aardige is nu dat de uitdrukking $\mathcal{K}$ niet verandert onder toepassing van Reidemeisterzetten van type II en type III ! Rest ons nog na te gaan hoe het zit met die van type I. We kiezen nu een oriëntatie op (iedere component van) de knoop of schakel. Iedere kruising kan dan van een voorteken worden voorzien:

voorteken $+1$ en voorteken -1

Laatste stap. Laat nu $w(K)$ de som van de voortekens van het diagram K zijn. Men kan dan nagaan dat de uitdrukking

$$P(K) = (-A^3)^{-w(K)} \times \mathcal{K}(K)$$

niet verandert onder *alle* Reidemeisterzetten. Dat wil zeggen: we hebben een invariant van de knoop of schakel gevonden; deze hangt er niet vanaf hoe we de knoop of schakel met een diagram presenteren, alleen maar van het type van de schakel.

Voorbeeld. Voor de Hopf-schakel (met de eerder gegeven oriëntatie) vinden we:

$$P(K) = (-A^3)^2 \times (-A^{-4} - A^4) = -A^{-2} - A^{10},$$

en voor de klaverbladknoop

$$\begin{aligned} P(K) &= (-A^3)^{-3} \times (-A^5 - A^{-3} + A^{-7}) \\ &= A^{-4} + A^{-12} - A^{-16}. \end{aligned}$$

Wat het nut hiervan is zien we direct:

Stelling. *De klaverbladknoop is een echte knoop. De klaverbladknoop en zijn spiegelbeeld zijn verschillende knopen.*

Bewijs. Als de klaverbladknoop triviaal was, dan zou hij hetzelfde polynoom P hebben als de triviale knoop, d.w.z. 1. Dat is echter niet het geval. De klaverbladknoop is niet van hetzelfde type als zijn spiegelbeeld, omdat onder het spiegelen het symbool A in A^{-1} overgaat, terwijl het polynoom $A^{-4} + A^{-12} - A^{-16}$ verschillend is van $A^4 + A^{12} - A^{16}$. □

Het *Jones-polynoom* wordt nu uit P verkregen door A te vervangen door $t^{1/4} = \sqrt[4]{t}$. Het is een Laurent-polynoom, d.w.z. een veelterm in $t^{1/4}$ plus een veelterm in $t^{-1/4}$. We kunnen deze invariant nu ook met axioma's definiëren. Maar merk op dat het a priori helemaal niet duidelijk is of er wel een invariant bestaat die aan deze axioma's voldoet en of die eenduidig vastligt.

Axioma's. Het Jones-polynoom $V_K(t)$ van een geöriënteerde knoop K is een Laurent-polynoom in $\sqrt[4]{t}$ met de volgende eigenschappen:

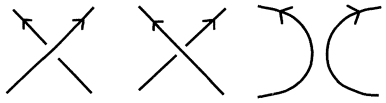
i) Als K_1 en K_2 equivalente geöriënteerde knopen zijn dan $V_{K_1}(t) = V_{K_2}(t)$;

ii) $V_K(t) = 1$ voor de triviale knoop K;

iii) Er geldt de schematische relatie:

$$\frac{1}{t}V_{K_+} - tV_{K_-} = (\sqrt{t} - \frac{1}{\sqrt{t}})V_{K_0},$$

waarbij de knoop K_- (respectievelijk de knoop K_0) uit de knoop K_+ ontstaat door één kruising van type + te vervangen door een van type − (respectievelijk van type 0), zie de figuur.



K_+ K_- K_0

Het Jones-polynoom voldoet aan de regels i) en ii). Met een berekening volgt dat het Jones-polynoom ook aan iii) voldoet. (Probeer dit zelf!) Dat polynoom is door de regels i), ii) en iii) ook volledig bepaald. Daarmee kunnen we nu snel rekenen, Neem bijvoorbeeld de knoop K_+ uit het rijtje van drie knopen hieronder

K_+ K_- K_0

Het is niet moeilijk in te zien dit dit een vermomde triviale knoop is, evenals K_- en dus dat regel iii) zegt

$$\frac{1}{t} - t = (\sqrt{t} - \frac{1}{\sqrt{t}})V_{K_0},$$

waaruit volgt dat $V_{K_0} = -(\sqrt{t} + 1/\sqrt{t})$.

Het nut van dit polynoom is nu dat je ermee kunt laten zien dat vele knopen of schakels echt verchillend zijn door gewoon het polynoom uit te rekenen. Maar we kunnen er meer mee doen zoals in de volgende paragraaf zal blijken.

Het Jones-polynoom is een wezenlijke verfijning van een eerdere invariant, het *Alexanderpolynoom*, die al in 1928 was gevonden door Alexander, met behulp van de topologie. Het Alexander-polynoom $A_K(t)$ kan ook door axioma's als voor het Jonespolynoom worden gegeven. Het verschil met het Jonespolynoom zit in regel iii), die nu luidt

iii)' Er geldt de schematische relatie:

$$A_{K_+}(t) - A_{K_-}(t) = (t^{1/2} - t^{-1/2})A_{K_0},$$

Zo vinden we nu voor K = twee disjuncte cirkels dat $A_K = 0$; voor de Hopfschakel vinden we $A_K = t^{-1/2} - t^{1/2}$ en voor de klaverbladknoop $A_K = t^{-1} - 1 + t$. Conway definieerde een variant $C_K(t)$ daarvan door naast regels i) en ii) als derde regel te nemen

$$C_{K_+}(t) - C_{K_-}(t) = tC_{K_0}.$$

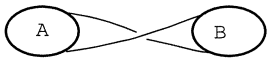
Dit is dan een polynoom $C_K = c_0 + c_1 t + c_2 t^2 + \ldots + c_m t^m$ in t. Het blijkt dat $c_0 = 1$ als het aantal componenten van de schakel K gelijk aan 1 is, en $c_0 = 0$ anders, en dat c_1 het schakelgetal is als K uit precies 2 componenten bestaat en $c_1 = 0$ anders. Deze coëfficiënten c_i zijn voorbeelden van geheel nieuwe invarianten die in de afgelopen jaren door Vassiliev zijn gedefinieerd en die alle eerder bekende invarianten omvatten. Het is niet bekend of deze Vassiliev-invarianten alle knopen geheel karakteriseren. Het is een nog openstaand vermoeden dat dit inderdaad het geval is.

Jones heeft zijn polynoom niet gevonden door op boven aangegeven manier met knopendiagramen te spelen. Hij werkte op een heel ander gebied van de wiskunde, de functionaalanalyse, en stootte bij zijn onderzoek aan oneindig-dimensionale ruimten op een algebraïsche structuur die je ook bij knopen vindt. Dit bracht hem op het idee voor zijn invariant. Later heeft men de eenvoudiger benadering met diagrammen gevonden. Zo lijkt het telkens weer te gaan in de knopentheorie: via grote omwegen komt men uiteindelijk uit bij begrippen die men direct met de diagrammen kan definiëren. Het lijkt erop dat de omweg nodig is om bij de definitie uit te komen.

Een oud vermoeden van Tait

In deze paragraaf bekijken we *alternerende* knopendiagrammen: dat zijn diagrammen waar je bij het doorlopen van de knoop afwisselend een onderkruising en een bovenkruising tegenkomt. De cijfer-8-knoop bijvoorbeeld, zoals in de inleiding gegeven, heeft een alternerend diagram

We noemen zo een alternerend diagram *gereduceerd* als het niet de volgende gedaante heeft



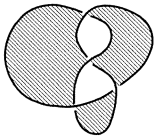
Zo een diagram kunnen we immers vereenvoudigen door B op te pakken en om te draaien zodat de kruising verdwijnt:

Intuïtief lijkt het alsof er aan een alternerend diagram dat gereduceerd is niet veel meer te vereenvoudigen is. Inderdaad formuleerden Tait en de andere knopenklassificeerders van het eerste uur het volgende heuristische principe.

Vermoeden. *Het aantal kruisingen in een samenhangend gereduceerd alternerend diagram van een knoop K hangt niet van de keuze van dat diagram af.*

Dit vermoeden werd in 1987 bewezen door Kauffman en onafhankelijk van hem ook door de Japanse wiskundige Murasagi. We schetsen het bewijs van Kauffman, dat het Jonespolynoom gebruikt.

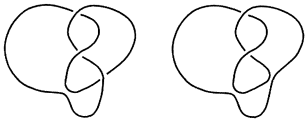
Laat K een samenhangend alternerend knoop- of schakeldiagram zijn. We voorzien het diagram eerst van een dambordpatroon (wit/zwart). Hierbij wordt het buitenste gebied wit gelaten en twee gebieden die aan elkaar grenzen langs een stuk van de knoop hebben verschillende kleuren, zoals in onderstaand voorbeeld.



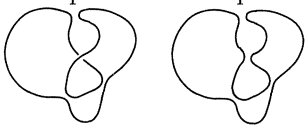
Het is niet moeilijk in te zien dat dit altijd kan. Namelijk, hef een voor een alle kruisingen op terwijl je ervoor zorgt het diagram samenhangend blijft. Gaat dat op een gegeven moment mis, d.w.z. zou het diagram niet meer samenhangend zijn, dan ziet het er zo uit

maar dan zou de andere splitsing

samenhangend zijn. We vinden uiteindelijk een samenhangende figuur zonder kruisingen, dus een cirkel. Dan zegt een stelling van Jordan dat (een figuur homeomorf met) een cirkel een buiten- en een binnengebied heeft. Toepassing hiervan leidt ertoe dat we altijd zo een dambordpatroon kunnen vinden. Hieronder staat het procédé voor de cijfer-8 knoop.



stap 1 en stap 2



stap 3 en stap 4

We gaan nu weer uit van een gegeven samenhangend gereduceerd schakeldiagram en voorzien dit diagram van een dambordpatroon. Met dit dambordpatroon kunnen we de kruisingen in dit diagram opheffen en wel zó dat de zwarte gebieden met elkaar verbonden worden. Het resultaat T is een van de 2^k 'toestanden' die we vonden bij de berekening van het Kauffman-haakje. Als $k(K)$ het aantal kruisingen in K is, dan is volgens het daar gedefinieerde procédé de bijdrage van deze 'toestand' T aan het Kauffman-haakje de uitdrukking

$$(-A^2 - A^{-2})^{\ell(T)-1} A^{k(K)},$$

omdat we bij iedere kruising een A gezet hebben. Dus de term van de hoogste graad die deze toestand T levert is

$$(-1)^{\ell(T)-1} A^{k(K)+2\ell(T)-2}.$$

Wat leveren de andere toestanden op? Merk hier op dat de uitdrukking $\ell(T)$ het aantal witte gebieden is dat overblijft na opheffing van de kruisingen.

De bewering is nu dat alle andere toestanden alleen termen van lagere graad leveren in hun bijdrage aan het Kauffmanhaakje. Neem een andere toestand T'. Stel die kan uit T worden verkregen door in één kruising niet de twee zwarte, maar de twee witte gebieden met elkaar te verbinden. Maar dat betekent dat er in T' een wit gebied minder is doordat twee witte gebieden met elkaar zijn verbonden; waren die twee witte gebieden al verbonden, dan was het diagram niet gereduceerd. Verder is er een A vervangen door een A^{-1}. Dus de 'toestand' T' levert een uitdrukking

$$A^{k(K)-2}(-A^2 - A^{-2})^{\ell(T)-2}.$$

De graad hiervan is 4 minder dan de bijdrage van T. Een precieze analyse leert dat de hoogste en laagste graad van de termen in het Kauffmanhaakje van K zijn

$$\text{max. graad} = k(K) + 2W - 2, \qquad \text{min. graad} = -k(K) - 2Z + 2, \qquad (2)$$

waar W en Z het aantal witte en zwarte gebieden aanduiden.

De *spanwijdte* van een knoopdiagram is nu per definitie het verschil tussen hoogste en laagste graad in het Kauffmanhaakje. Maar we weten dat het polynoom $P_K = (-A^3)^{-w(K)}\mathcal{K}(K)$ een invariant van de knoop is. Het verschil tussen hoogste en laagste graad in P_K is duidelijk gelijk aan dat in $\mathcal{K}(K)$. Dus de spanwijdte van een knoop is een invariant van de knoop.

Volgens het resultaat (2) is dit verschil gelijk aan

$$\begin{aligned} \text{spanwijdte}(K) &= 2k(K) + 2(W + Z) - 4 \\ &= 2k(K) + 2G - 4, \end{aligned}$$

met G het aantal gebieden in ons diagram. Het verband tussen het aantal gebieden en het aantal kruisingen is $G = k(K) + 2$. Dit zie je in door een

knopendiagram te tekenen en te kijken wat er gebeurt als je een kruising toevoegt. Dus als we dit substitueren vinden we

$$\begin{aligned} \text{spanwijdte}(K) &= 2k(K) + 2G - 4 \\ &= 4k(K). \end{aligned}$$

Dus is het aantal kruisingen $k(K)$ van een alternerend gereduceerd diagram een invariant van de knoop. Q.E.D.

Opmerking. Het bewijs leert ons ook dat de term van de hoogste en laagste graad in P_K als coefficient ± 1 hebben. Verder is het aantal kruisingen dus precies de spanwijdte van het Jonespolynoom als Laurentpolynoom in t.
Voorbeeld. De cijfer-8-knoop heeft Jonespolynoom $V_K = t^{-2} - t^{-1} + 1 - t + t^2$ met spanwijdte 4. Inderdaad zijn er in ons diagram 4 kruisingen.

Met wat meer werk kan men ook bewijzen dat het aantal kruisingen van een gereduceerde alternerende projectie van een knoop minimaal is voor alle projecties van deze knoop. Ook dit was een vermoeden van Tait.

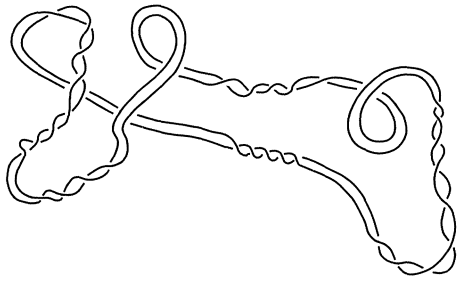
Deze vermoedens van Tait konden worden opgeruimd, maar vele andere vermoedens stammend uit de tijd van de eerste classificatiepogingen zijn gebleven. Zoals bijvoorbeeld het volgende vermoeden van Tait uit 1890. Onder het *kruisingsgetal* van een knoop wordt het minimale aantal kruisingen in een projectie van de knoop verstaan.

Vermoeden. (Tait) *Als het kruisingsgetal van een knoop oneven is, dan is de knoop niet van hetzelfde type als zijn spiegelbeeld.*

Onze klaverbadknoop is een voorbeeld; het aantal kruisingen is 3 en we weten al met behulp van de invariant P_K dat het type van deze knoop verandert als we spiegelen, dwz. als we alle kruisingen in het diagram vervangen door de tegengestelde kruisingen.

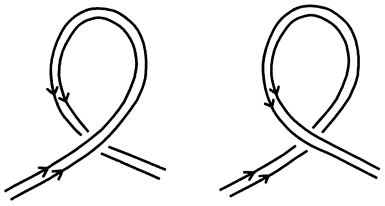
DNA-verstrengelingen

We nemen weer een knopendiagram en vervangen de draad door een gummi-band. We noemen het resultaat een DNA-verstrengeling of DNA-schakel. De beide randen leveren dan een schakel van twee cirkelvormige figuren die we C en C' noemen. Dat ziet er dan bijvoorbeeld zo uit:



DNA-verstrengeling

Ik hoop dat U zich de definitie van het schakelgetal nog herinnert. De twee cirkelfiguren C en C' hebben een schakelgetal. Maar in deze situatie kunnen we nog twee andere getallen definiëren, de *torsie* en de *draaiing*. Om de torsie te definiëren nemen we aan dat de beide zijden van onze gummiband een verschillende kleur hebben, zeg rood en groen. De torsie geeft aan hoe vaak we een kleurverandering zien in het diagram (rood/groen of groen/rood). Voor de draaiing tellen we hoe vaak (met voorteken) de gummiband zichzelf snijdt:

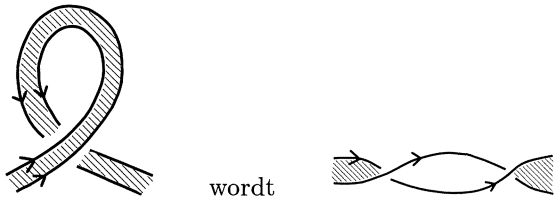


draaiing -1 en draaiing $+1$

Er is dan een formule die deze drie getallen met elkaar verbindt:

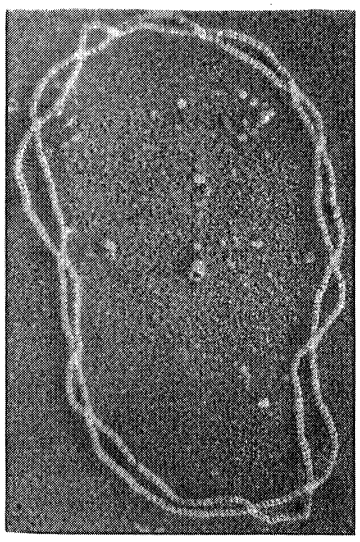
$$\text{schakelgetal} = \text{torsie} + \text{draaiing}, \qquad \ell = t + d$$

Om deze formule in te zien kijk je gewoon naar de verandering die optreedt als de verstrengeling lokaal strakgetrokken wordt:



wordt

Deze twee-dimensionale formule heeft een generalisatie voor DNA-verstrengelingen in de driedimensionale ruimte. Deze generalisatie (Formule van White) ziet er precies zo uit, maar de definitie van t en d zijn wat moeilijker. Voor de skeptici kan men deze formule met een riem illustreren; de telefoondraad tussen hoorn en toestel vertoont een analoog gedrag. Maar deze formule heeft een veel diepere toepassing in de natuur. Het volgende plaatje toont een DNA-verstrengeling (uit [S2]).



Een DNA-knoop

Ik ga ervan uit, dat U weet dat iedere cel van een levend wezen DNA bevat waarin de genetische informatie opgeslagen is. Om een idee van de orde van grootte te geven: als we een model van de cel zo groot als een voetbal maken dan moet dit model ongeveer 200 km aan vissnoer (hengeldraad) als DNA bevatten. Het is moeilijk voorstelbaar dat een dergelijke hoeveelheid niet in de knoop raakt. De cel heeft dus kennis van de knopentheorie nodig om te kunnen functioneren, en zij maakt zeer subtiel gebruik van deze kennis. Zo schijnen er enzymen te bestaan die het schakelgetal van een DNA-verstrengeling veranderen, maar de torsie onveranderd laten. Volgens bovenstaande formule of de formule van White moet dan de draaiing veranderen. Daarmee kan het DNA kompakter of juist veel minder kompact gemaakt worden. Mijn kennis van de biologie is te gering om hier verder over uit te weiden. Biologen verzekeren mij dat de formule van White en andere resultaten uit de knopentheorie hun weg in de moleculaire biologie hebben gevonden.

Zoals zo vaak is een heel deelgebied van de wiskunde ontstaan uit de wens een onderdeel van de natuur beter te begrijpen; dit deelgebied heeft zich dan zelfstandig als zuivere wiskunde verder ontwikkeld om dan tenslotte weer te leiden tot volledig onverwachte en diepe toepassingen op een volslagen ander gebied.

Literatuur

[A] C.C. Adams: *The Knot Book.* Freeman 1994.

[As] C. Ashley: *The Ashley Book of Knots.* Doubleday and Co. New York, 1944.

[BCW] W.R. Bauer, F.H.C. Crick, J.H. White: Supercoiled DNA. *Scientific American* **243**, 1980, p. 118-133.

[K] L.H. Kauffman: *Knots and Physics.* Second edition. World Scientific Publishing Co., Inc., River Edge, NJ, 1993. xiv+723 pp.

[J] V.F.R. Jones: A polynomial invariant for knots and links via Van Neumann algebras. *Bull. Am. Math. Soc.* **12**, (1985), p. 103–111.

[L] J.B. Listing: *Vorstudien zur Topologie.* Göttingen, 1848.

[R] L. Reidemeister: *Knotentheorie.* Julius Springer, Berlin, 1932.

[Ro] D. Rolfsen: *Knots and Links.* Math. Lect. Series 7. Publish or Perish Press (1976).

[S1] D.W. Sumners: Untangling DNA. *The Mathematical Intelligencer* **12**, (1990), p. 71–80.

[S2] D.W. Sumners: Lifting the curtain: using topology to probe the hidden action of enzymes. *Notices Amer. Math. Soc.* **42** (1995), no. 5, p. 528–537.

[T] William P. Thurston: Three-dimensional geometry and topology. Vol. 1. Edited by Silvio Levy. Princeton Mathematical Series, 35. Princeton University Press, Princeton, NJ, 1997.

Toelichting. Als eerste inleiding in de knopentheorie kan het boek van Adams worden aanbevolen. Het bespreekt op toegankelijke wijze vele aspecten van de knopentheorie. De tabellen achterin zijn niet betrouwbaar omdat er twee notaties van knopen (de klassieke en die van Thistletwaite) worden verwisseld. Het boek van Ashley is een omvangrijk niet-wiskundig curiosum uit 1944 met een indrukwekkende verzameling van knopen uit het dagelijkse leven. Het boek van Thurston behandelt delen van de meetkunde/topologie die ook voor de knopentheorie belangrijk zijn en kan ook worden aanbevolen voor verdere studie.

Vakantiecursus 1999

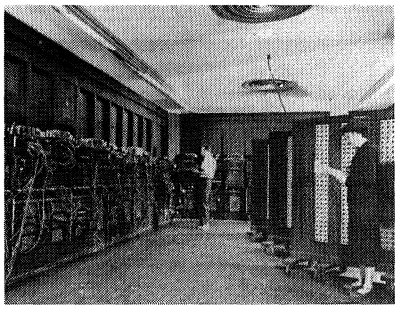
pp. 93 – 108

$\mathcal{P} = \mathcal{NP}$?

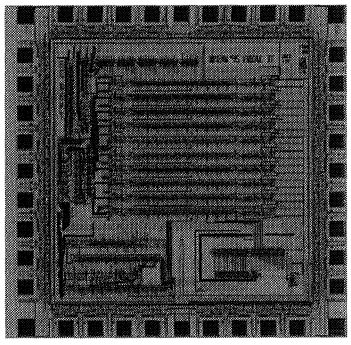
Frits Beukers

1. Inleiding

De eerste electronische computer, de ENIAC(="Electronic Numeric Integrator and Calculator")) werd in 1946 in gebruik genomen en luidde het tijdperk van electronisch rekenen in. Deze kolos bevatte maar liefst 18.800 radiobuizen en was in staat om zo'n 5000 elementaire bewerkingen per seconde uit te voeren. Voor die tijd een onvoorstelbare snelheid. Maar de tijd heeft niet stilgestaan.



ENIAC, 1946



ENIAC-ON-A-CHIP, 1996

Tegenwoordig bevat een gemiddelde processor in een PC een miljoen of meer logische componenten en is in staat tot enkele miljoenen berekeningen per seconde. Hiermee vergeleken was ENIAC slechts kinderspel. De enorme explosie van rekenkracht die we de afgelopen jaren om ons heen gezien hebben, heeft een onvoorstel effect op onze maatschappij gehad. Hoezeer computers in onze maatschappij verweven zijn moge blijken uit onze zorgen over het millenniumprobleem. Ook in de wereld van het wetenschappelijk rekenen hebben we zo'n omwenteling meegemaakt. Met computers kunnen we tegenwoordig weersvoorspellingen doen, windtunnels simuleren, stromingen van de oceaan berekenen, sterrenstelsels laten botsen, optimale strategiën voor grote organisaties bepalen, het dienstrooster van de NS samenstellen, medische scans zichtbaar maken, en ga zo maar door. Dit alles vestigt de indruk dat als er zaken zijn die we nu niet kunnen berekenen, dat we dit over een aantal jaren wel zouden moeten kunnen.

Merkwaardig genoeg blijken bij onze pogingen om de grenzen van onze mogelijkheden te verleggen er nieuwe onverwachtte barrières op te duiken. In de natuurkunde stuiten we regelmatig op dat soort muren. Materie blijkt ondeelbaar op atomair- of quarkniveau, niets kan sneller gaan dan het licht, en we

kunnen snelheid en positie van een deeltje niet beiden precies kennen (Heisenberg's onzekerheidsrelatie). Dergelijke begrenzingen kunnen heel onaangenaam zijn, maar ook bijzonder boeiend. Op het gebied van berekenbaarheid blijken er ook zulke nieuwe barrières op te doemen. Een heel bekend voorbeeld is dat van de *chaostheorie* met als belangrijkste voorbeeld de weersvoorspelling. De differentiaalvergelijkingen die het gedrag van de atmosfeer bepalen laten zien dat onnauwkeurigheden in de beschrijving van de beginsituatie zich in de loop der dagen exponentieel vermenigvuldigen, totdat na een week deze uitvergrote onnauwkeurigheden de werkelijke beschrijving totaal overheersen. Om deze reden gaan huidige weersvoorspellingen niet veel verder dan een week. Willen we de weersvoorspelling een dag verder uitstrekken, dan moet de nauwkeurigheid van de beginsituatie met een factor, zeg 2, verbeterd worden. Voor iedere volgende dag moeten we de begintoestand een factor 2 nauwkeuriger kennen. De toestand van de atmosfeer met een dergelijke nauwkeurigheid opmeten wordt al snel onmogelijk, afgezien van het probleem de vereiste berekeningen uit te moeten voeren. We hebben te maken met een chaotisch systeem. Reeds aan het begin van deze eeuw werd een dergelijk gedrag door de beroemde wiskundige H.Poincaré voorspeld. Nu we computers tot onze beschikking hebben, worden we met deze harde werkelijkheid geconfronteerd. Wie belangstelling heeft voor dit onderwerp, raad ik aan het boekje [BCV] (zie referenties achterin) te lezen. In dit stuk willen we echter de berekenbaarheid bespreken van een heel ander soort problemen dan chaotische dynamische systemen. Wij zullen het hier hebben over gehele getallen, lijsten van symbolen en andere discrete zaken. Hier zijn een paar voorbeelden.

Sorteerproblemen Gegeven een lijst woorden, bijvoorbeeld een miljoen stuks. Sorteer ze op alfabet.

Deelsomproblemen Gegeven een rij X van n natuurlijke getallen en een natuurlijk getal S. Ga na of er een deelrij van X bestaat waarvan de som van de elementen gelijk aan S is.

Ontbinden in factoren Gegegeven een natuurlijk getal. Bepaal zijn ontbinding in priemfactoren.

Handelsreizigerprobleem Gegegeven een aantal steden waarvan de onderlinge afstanden bekend zijn. Bepaal het kortste gesloten pad dat elk van deze steden precies éénmaal bezoekt.

Schaakprobleem Gegeven een stelling in een schaakpartij. Bepaal of wit in gewonnen positie staat.

Bovenstaande problemen hebben allemaal als gemeenschappelijk kenmerk dat er een oplosmethode is die bestaat uit het uitproberen van alle mogelijkheden. Zo zouden we het handelsreizigerprobleem kunnen oplossen door alle gesloten paden op te schrijven, hun lengte te bepalen, en de kortste daarvan als oplossing te nemen. Een getal van 200 cijfers kunnen we ontbinden door gewoon alle getallen van 2 tot 10^{100} als deler te proberen. Of deze oplosmethoden efficiënt

zijn is weer een heel andere kwestie. Voor een oplosmethode is het namelijk niet alleen belangrijk dat hij een antwoord oplevert, maar ook dat dit gebeurt binnen een tijd waarin het antwoord nog voor ons van nut is. Aan een oplosmethode die eeuwen in beslag neemt hebben we niets. De vraag naar het bestaan van efficiënte methoden zal de leidraad van deze voordracht zijn. We zullen echter eerst beginnen met het vastleggen van wat begrippen.

2. Computers en algoritmen

Het eerste onderzoek naar berekenbaarheid werd al in de dertiger jaren gedaan door de wiskundige Alan Turing. In die tijd bestonden er nog geen electronische computers. Men kende wel mechanische rekenmachines en men had ook een primitief begrip van de mogelijkheid zo'n rekenmachine een bepaald rekenschema of programma te laten volgen. Voor zijn onderzoekingen maakte Turing gebruik van een denkbeeldige machine die later bekend is geworden onder de naam (universele) *Turing machine*. We laten voortaan het woordje "universeel" weg. Het getuigt van de visie van Turing dat onze tegenwoordige computers in principe realisaties zijn van een Turing machine, met als enig verschil dat een Turing machine over onbeperkt veel geheugen kan beschikken. Door deze overeenkomst zijn Turing's beschouwingen ook van toepassing op onze tegenwoordige computers. Sterker nog, een Turing machine, een PC of een supercomputer zijn equivalent in die zin dat ze zijn in staat tot het uitvoeren van dezelfde soort berekeningen. Het enige verschil is natuurlijk dat de ene machine langer over een bepaalde berekening zal doen dan de ander. Echter, met betrekking tot de vraag wèlke berekeningen uitgevoerd kunnen worden zijn alle drie machines gelijkwaardig. Zelfs een ijverige klerk die trouw alle instructies opvolgt zou ook aan het Turing model voldoen. Door deze uitwisselbaarheid van machinetypes is het nu niet meer nodig Turing's definitie van zijn machine uit te leggen. We kunnen gewoon denken aan een computer met onbeperkt veel geheugen.

Om computers problemen te laten oplossen hebben we *algoritmen* nodig. Een algoritme is een computerprogramma dat een invoer leest en vervolgens uitvoer produceert die het antwoord op ons probleem geeft. Iedereen heeft wel eens een computerprogramma gezien. Het is altijd een eindige lijst met instructies. Voor ons zijn die instructies niet altijd even helder, voor de computer moeten ze ondubbelzinnige opdrachten voorstellen die één voor één uitgevoerd moeten worden door de processor.

Het soort problemen dat we bekijken heeft trouwens altijd *parameters* in zich. Kijk maar naar de lijst van voorbeeldproblemen uit de vorige paragraaf. In het ontbindingsprobleem gaat het om een probleem om een getal n te ontbinden. Het getal n is de parameter. Bij het sorteeralgoritme is de parameter de te sorteren lijst van woorden. In al deze problemen vormen de parameters de invoer voor het eventuele algoritme dat het probleem op moet lossen. De uitvoer, die op een beeldscherm verschijnt of op papier wordt uitgeprint, kan ook verschillende vormen hebben. Bij het sorteerprobleem is dat gesorteerde lijst, bij het schaakprobleem of deelsomprobleem is dat een simpel "ja" of "nee", al naar gelang de positieve of negatieve uitkomst. Problemen met alleen "ja" of "nee"

als uitvoer zullen voor ons van bijzondere interesse zijn. We noemen ze *beslissingsproblemen.* Het sorteerprobleem is duidelijk geen beslissingsprobleem. Het aardige is echter dat ontbinding in factoren daarentegen wel geformuleerd kan worden als een beslissingsprobleem.

Ontbinding als beslissingsprobleem Gegeven een tweetal positief gehele getallen n, m met $m \leq n$. Ga na of n een echte (dwz $\neq 1, n$) deler $< m$ heeft.

Als we een algoritme hebben om het oorspronkelijke ontbindingsprobleem op te lossen, dan kunnen we bovenstaand probleem uiteraard ook oplossen. Stel nu omgekeerd dat we een algoritme hebben om bovenstaand beslissingsprobleem op te lossen. Dan kunnen we het oorspronkelijke ontbindingsprobleem ook oplossen. We zoeken namelijk de kleinste echte deler van n door middel van een binaire zoekactie, bij velen beter bekend als het "hoger-lager" spel. Dit spel bestaat eruit dat we, gegeven n, iemand een getal $x < n$ laten raden door alleen met "ja" of "nee" op zijn vragen te antwoorden. De vragen mogen alleen van het type 'is x kleiner dan y?' zijn waarin y een willekeurig door de vraagsteller te benoemen getal is. De handigste manier is ervoor te zorgen dat we het interval waarin x moet liggen bij elke vraag in lengte halveren. Te beginnen bij het interval $[1, n]$. Op deze manier kan de vraagsteller met $\log_2(n) + 1$ vragen klaar zijn.
We kunnen dit spelletje spelen met de verzameling echte delers van een getal n. Geef deze verzameling aan met S. Het algoritme om ons beslissingsprobleem op te lossen speelt de rol van een orakel dat "ja" of "nee" antwoordt op de vraag of S een element $< m$ bevat. Neem als voorbeeld $n = 989$. Als er een echte deler is dan is deze kleiner of gelijk $\sqrt{989} < 32$. De eerste vraag is 'is er een echte deler van n kleiner dan 32?'. Antwoord 'ja'. We zijn op zoek naar de kleinste niet-triviale deler. Dus stellen we de volgende vraag: 'is er een echte deler van n kleiner dan 16?'. Antwoord 'nee'. We weten dus dat $16 \leq d < 32$ waarin d de kleinste echte deler van n is. Hier volgt een kleine tabel van de rest van het vraag-en-antwoord spel.

vraag	antwoord
$d < 24$	ja
$d < 20$	nee
$d < 22$	nee
$d < 23$	nee

Uit de laatste regel concluderen we dat $d = 23$.
Met enige moeite kan op soortgelijke wijze ook het handelsreizigerprobleem geherformuleerd worden als beslissingsprobleem. Voor de details ervan verwijzen we naar [LLRS, p. 46]
We leggen hier de nadruk op het begrip beslissingsprobleem omdat het een grote rol speelt in de bespreking van het probleem $\mathcal{P} = \mathcal{NP}$? Voordat het zover is besteden we eerst nog wat aandacht aan een merkwaardige klasse van problemen die geen oplossing hebben.

3. Onoplosbare problemen

De problemen die we in de de inleiding noemden hadden allen als eigenschap dat er een algoritme voor hun oplossing bestaat. Ontbinding van een getal n in factoren zou in principe kunnen gebeuren door alle gehele getallen $< \sqrt{n}$ als deler te testen. Het handelsreiziger probleem kunnen we oplossen door alle gesloten paden op te schrijven, hun lengte uit te rekenen, en vervolgens het pad met minimale lengte uit te kiezen. Dit zijn allemaal procedures die in eindige tijd uitgevoerd kunnen worden. Of er ook efficiëntere oplosmethoden bestaan is een zaak die in de volgende paragrafen aan de orde komt.
In deze paragraaf kunnen we het niet nalaten om een aantal problemen te noemen waarvan het helemaal niet voor de hand ligt dat er een algoritme voor hun oplossing bestaat. Hier zijn twee beroemde voorbeelden. Merk op dat het beide beslissingsproblemen zijn.

Halting probleem (Turing). Bestaat er een algoritme dat voor elk computerprogramma en voor elke gegeven invoer, kan beslissen of het oneindig lang blijft doorgaan, of dat het daadwerkelijk stopt (termineert). Zoals we weten kan een computerprogramma oneindig lang door blijven gaan als het programma in een zogenaamde loop terechtkomt. Een eenvoudig voorbeeld hiervan is

```
Domprogramma()
begin
(a) Ga naar (a)
(b) STOP
end
```

Hoewel er een STOP-instructie in dit programma staat, blijft het bij uitvoering in de eerste regel "hangen". Uiteraard is dit een zeer naïef voorbeeld. Veel loops komen op een veel ingewikkelder manier tot stand en voor software testers zou het ideaal zijn om een algoritme te hebben dat het bestaan van deze verborgen loops aan het licht brengt.

Hilbert's tiende probleem Bestaat er een algoritme dat voor elke diophantische vergelijking beslist of het een geheeltallige oplossing heeft? Een diophantische vergelijking is een vergelijking van het type

$$F(x_1, \ldots, x_n) = 0$$

waarin F een polynoom met gehele coëfficienten is in de n variabelen $x_1, \ldots, x_n$ is. Het woord diophantisch heeft betrekking op het feit dat we voor deze vergelijkingen alleen de geheeltallige oplossingen zoeken. De bekendste illustratie van het feit dat diophantische vergelijkingen moeilijk zijn is het laatste probleem van Fermat waarover elders in deze syllabus meer wordt gezegd (zie de bijdrage van R.Tijdeman). Voor de aardigheid zou de

lezer kunnen stoeien met het probleem om aan te tonen dat $y^2 = x^3 - 1$ geen geheeltallige oplossingen x, y heeft.

Een ander saillant detail is dat veel bekende problemen uit de wiskunde kunnen worden geherformuleerd als een diophantische vergelijking. Het bekendste voorbeeld hiervan is het Riemann-vermoeden (zie de bijdrage van R.Tijdeman in dit boekje). De juistheid van dit vermoeden is equivalent met het bestaan van een oplossing van een contrueerbare, maar zeer ingewikkelde, diophantische vergelijking (voor deze constructie zie [DMR, P. 334–337]). Het bestaan van een positief antwoord op Hilbert's tiende probleem zou in het bijzonder impliceren dat er een mechanische oplossing voor het Riemann-vermoeden bestaat.

In de dertiger jaren maakte Turing de volgende schokkende ontdekking.

STELLING 3.1 *Er bestaat geen algoritme om het Halting Probleem op te lossen.*

Deze ontdekking was schokkend omdat dit de eerste stelling in de wiskunde was die het niet-bestaan van een algoritme aantoonde. In de tweede plaats was het een grote verrassing dat men überhaupt een dergelijke non-existentie kon aantonen. Turing's bewijs is echter geniaal, en bovendien makkelijk te presenteren. We geven het hier. Om te beginnen beperkte Turing zich tot algoritmen die positieve gehele getallen als invoer accepteren. Bij veel problemen hebben we al gehele getallen als invoergegeven, maar ook als dat niet zo is, dan geeft dit geen beperking. Dit berust op de volgende stelling.

STELLING 3.2 *Zij S een eindige verzameling van symbolen. Zij $\mathcal{S}$ de verzameling van eindige rijtjes elementen van S. Dan is $\mathcal{S}$ aftelbaar. Dat wil zeggen, we kunnen de elementen van $\mathcal{S}$ op een rij $S_1, S_2, S_3, \ldots, S_n \ldots$ zetten zó dat elk element van $\mathcal{S}$ in deze rij voorkomt.*

BEWIJS We kunnen de verzameling S zien als een alfabet en $\mathcal{S}$ als de verzameling woorden gemaakt met dat alfabet. Het is nu heel makkelijk om alle elementen uit $\mathcal{S}$ op een rij te zetten. Omdat het alfabet eindig is zijn er van elke gegeven lengte L een eindig aantal woorden. We schrijven nu eerst alle woorden van lengte 1 op, vervolgens alle woorden van lengte 2, dan alle woorden van lengte 3, etcetera. Op deze manier weten we zeker dat elk woord op den duur ook inderdaad aan bod komt. Om de rangschikking nog wat ondubbelzinniger te maken zouden we nog een alfabetische volgorde voor de elementen van S kunnen afspreken en binnen elke lengte de woorden alfabetisch rangschikken. □

Elke invoer bestaat uit (hoofd)letters, cijfers, spaties en andere leestekens. Een eindige collectie van tekens dus. Volgens bovenstaande Stelling kunnen we nu alle mogelijke invoeren op een rij zetten. Vervolgens kan elke invoer met zijn rangnummer worden aangegeven. We gebruiken dit rangnummer in ons bewijs als invoer van onze algoritmen.
Wat voor invoer geldt, geldt ook voor algoritmen. Ook deze kunnen we keurig netjes op een rij zetten die we aangeven met $A_1, A_2, A_3, \ldots$. Een algoritme

A_k waar we input l aan geven noteren we met $A_k(l)$ en met deze twee data (algoritme plus invoer) kan de computer aan de slag.
Stel nu dat er een algoritme $\mathcal{A}$ is dat van een gegeven algoritme met gegeven invoer kan beslissen of het termineert of niet. Beschouw nu het volgende algoritme,

`catch`$(j \in \mathbb{N})$

begin

(a) Als $A_j(j)$ termineert volgens $\mathcal{A}$, ga dan in een ONEINDIGE LOOP.

(b) Als $A_j(j)$ niet termineert volgens $\mathcal{A}$, dan STOP.

end

Omdat `catch` een algoritme is, heeft het een rangnummer. Zeg `catch` $= A_r$. Stel dat $A_r(r) =$ `catch`(r) termineert. Uit de code van het algoritme `catch` zien we dat we blijkbaar de STOP-instructie bereikt hebben, met andere woorden, $\mathcal{A}$ heeft besloten dat $A_r(r)$ niet termineert. Dit is in tegenspraak met onze aanname dat $A_r(r)$ termineert. Stel nu dat $A_r(r) =$ `catch`(r) niet termineert. In ons algoritme `catch` zien we dat `catch` in regel (a) blijft hangen en dit komt doordat $\mathcal{A}$ besloten heeft dat $A_r(r)$ termineert. Wederom een tegenspraak! Wat de uitslag van $A_r(r)$ ook is, altijd komen we op een tegenspraak uit. Blijkbaar is het bestaan van $\mathcal{A}$ een onmogelijkheid en daarmee is ook Turing's stelling bewezen.
Het tiende probleem van Hilbert is er één uit D.Hilbert's lijst van 23. Hilbert was rond 1900 één van de bekendste wiskundigen, en in 1900 gaf hij tijdens het wereldcongres voor wiskundigen een voordracht over de wiskundeproblemen die een uitdaging vormden voor de komende eeuw. Een groot aantal problemen van de lijst is in deze eeuw daadwerkelijk opgelost, waaronder ook het tiende probleem. In 1970 toonde Matijasevich, na veel voorbereidend werk van Davis en Robinson, de volgende stelling aan.

STELLING 3.3 (MATIJASEVICH) *Er bestaat geen algoritme dat van een willekeurige diophantische vergelijking beslist of er gehele oplossingen zijn of niet.*

De technieken die hiervoor gebruik zijn vele malen ingewikkelder dan die in het bewijs bij Turing's Halting Probleem. De laatste stap van het bewijs bevat echter weer Turing's idee. Voor wiskundigen zou Matijasevich's resultaat heel geruststellend moeten zijn. Blijkbaar kan het oplossen van diophantische vergelijkingen, en in het bijzonder het Riemann-vermoeden, niet aan computers worden overgelaten. Voor elk nieuw type vergelijking zijn er weer nieuwe ideeën nodig, en dit is wat veel wiskunde zo boeiend maakt.
Tenslotte nog een opmerking om misverstanden te vermijden. Het feit dat er geen algemeen algoritme voor het Halting probleem en Hilbert's tiende probleem bestaat, betekent niet dat alle instanties van deze problemen onoplosbaar zijn. In het geval van `domprogramma` zagen we immers meteen dat het niet termineert. Ook van de diophantische vergelijking $x^2 + y^2 + z^2 = -1$ zal

meteen duidelijk zijn dat het geen oplossingen heeft. Zo zijn er ook talloze andere, wellicht lastiger, gevallen waarvoor we een oplossing kunnen vinden. Het enige wat in ons bovenstaande verhaal beweerd hebben is dat er geen algoritme bestaat dat àlle gevallen aankan. Eem zelfde soort opmerking geldt ook als we het later over moeilijke problemen gaan hebben. Niet alle instanties van een moeilijk probleem hoeven perse moeilijk te zijn.

4. Makkelijke problemen

Gelukkig bestaan er ook problemen die gemakkelijk oplosbaar zijn. Het succes van de computer is voor een groot deel aan dit soort problemen te danken. Neem bijvoorbeeld de sorteer algoritmen die in elk database programma gebruikt worden. Stel we hebben een lijst van L woorden, waarbij L in grote databases kan oplopen tot een miljoen of meer. Gevraagd wordt deze lijst op alfabet te sorteren. Als het woord v in de alfabetische rangschikking aan w voorafgaat noteren we dit als $v < w$. We kunnen ook zeggen dat v vroeger is dan w. Een heel voor de hand liggend algoritme om onze lijst te sorteren zou het volgende zijn. Kies uit de lijst het vroegste woord, d.w.z. het woord dat het eerst in de alfabetische volgorde aan bod komt. Haal dit woord uit onze lijst en zet dit in de uitvoer, bijvoorbeeld naar een printer. Voor het vinden van het vroegste woord zijn $L-1$ tests van het type $a < b$? nodig. Uit de overgebleven lijst kiezen we weer het vroegste woord en herhalen het proces. Voor deze stap zijn $L-2$ tests nodig. Zo doorgaand zal na hooguit $(L-1)L+(L-2)+\cdots+2+1 = L(L-1)/2 < L^2$ tests de lijst op volgorde gezet zijn. Nu is L^2 nog een onacceptabel groot getal als $L = 10^6$, maar gelukkig zijn er ook slimmere algoritmen die het klusje in hooguit $c \cdot L\ln(L)$ stappen klaren, waarin $c > 0$ een constante is. We noteren dit vaak door te zeggen dat er sorteeralgoritmen van de orde $O(L\ln(L))$ zijn. Het symbool O noemen we het *orde symbool.*

Een ander simpel probleem is het optellen of vermenigvuldigen van twee positieve gehele getallen a, b. Normaal gesproken is dit een fluitje van een cent, maar als a, b uit enkele honderden tot duizenden cijfers gaan bestaan, dan wordt het wat anders. Om te zien hoeveel tijd een optelling van twee getallen in beslag kan nemen moeten we afspreken wat een elementaire bewerking is. Dat wil zeggen, een basisbewerking die een vaste maximale tijd in beslag neemt. We nemen hiervoor de optelling van twee cijfers van elk van de getallen. Stel dat a, b uit maximaal L cijfers bestaan. Op de lagere school hebben we geleerd dat we voor de optelling van a, b dan ook L maal twee cijfers moeten optellen. De tijd die zo'n optelling duurt is dus hooguit $c \cdot L$ waarin c een positief getal is, die aangeeft hoe lang een computer over één elementaire bewerking doet. Omdat de waarde van c ons voor dit verhaal niet interesseert gebruiken we het ordesymbool O en zeggen we dat de looptijd van het optelalgoritme van de orde $O(L)$ is. Evenzo hebben we voor de vermenigvuldiging van a, b maximaal L^2 elementaire vermenigvuldigen van cijfers nodig, en daarna nog eens L^2 optellingen. Het vermenigvuldigingsalgoritme van de lagere school is dus van de orde $O(L^2)$. Overigens is het mogelijk door slimme technieken uit de Fourieranalyse (Schönhage-Strassen) deze looptijd te verbeteren tot $O(L\ln(L))$, maar dit

algoritme valt ver buiten het bestek van dit verhaal. Het zal verder hopelijk duidelijk zijn dat als $N = \max(a, b)$, het verband met L gegeven wordt door $\log_{10} N < L \leq \log_{10} N + 1$. De looptijd van het naïeve vermenigvuldigingsalgoritme is dus $O(\log_{10}(N)^2) = O(\ln(N)^2)$.
Algoritmen van het bovenstaande type hebben de eigenschap dat hun looptijd *polynomiaal* in de lengte van de invoer L is. Dat wil zeggen, hun looptijd is van de orde $O(L^a)$ voor zekere $a > 0$. Bij het naïeve optelalgoritme is $a = 1$, bij het naïeve sorteren en vermenigvuldigen is $a = 2$.
Het is lang niet bij alle problemen duidelijk of er een oplossing bestaat met een polynomiale looptijd. Neem bijvoorbeeld het probleem om te testen of een getal N priem is. Een methode zou zijn om voor $d = 2, 3, \ldots, [\sqrt{N}]$ te testen of d een deler van N is. Als dat niet zo is dan is N priem. Een dergelijke methode neemt maximaal $\sqrt{N}$ stappen in beslag en is dus van orde $O(\sqrt{N})$. Dit is echter geen polynomiale methode. De lengte van de invoer is immers het aantal cijfers L van N en dat is ongeveer $\log_{10}(N)$. Dus de looptijd van onze naïeve priemtest is $O(10^{L/2})$. De looptijd hangt nu exponentieel af van de invoerlengte, en dat is heel vervelend. Exponentiële functies hebben de neiging razend snel te groeien. Dit kunnen we mooi in de volgende tabel zien, waarin we een aantal polynomiale functies en exponentiële functies met elkaar vergelijken. In deze, tabel ontleend aan [GG], nemen we $c = 10^{-6}$ seconde als tijd nodig voor één basisbewerking.

L	10	20	30	40	50
L^2	0.0001 sec	0.0004 sec	0.0009 sec	0.0016 sec	0.0025 sec
L^3	0.001 sec	0.008 sec	0.027 sec	0.064 sec	0.125 sec
L^5	0.1 sec	3.2 sec	24.3 sec	1.7 min	5.2 min
2^L	0.001 sec	1 sec	17.9 min	12.7 dag	35.7 jaar
3^L	0.059 sec	58 min	6.5 jaar	3855 eeuw	2×10^8 eeuw

Uit deze tabel zien we dat bij exponentiële looptijden een kleine toename in de invoerlengte een gigantische, bijna explosieve toename van de maximale looptijd inhoudt. In het bijzonder zien we dat het naïeve priemtest algoritme voor getallen van 50 cijfers tot in de eeuwigheid kan duren!
Hopelijk is het nu begrijpelijk waarom we graag algoritmen met polynomiale looptijd willen hebben. We maken hier meteen de aantekening bij dat een polynomiaal algoritme nog niet alles zaligmakend hoeft te zijn. Een beroemd voorbeeld hiervan is het *lineaire programmeringsprobleem*, dat in polynomiale tijd oplosbaar is (Khachian, 1978). De theorie achter lineaire programmering werd vlak na de tweede wereldoorlog door G.B.Dantzig ontwikkeld en had tot doel de ingewikkelde logistiek van het Amerikaanse leger te optimaliseren. Later volgden er ook toepassingen in de economie en de pioniers op dit gebied, Koopmans en Kantorowitz, hebben voor hun werk de nobelprijs in de economie ontvangen. Kort gezegd komt het lineaire programmeringsprobleem erop neer dat we $x_1, x_2, \ldots, x_n$ moeten bepalen zó dat een lineaire functie van de vorm

$$c(x_1, \ldots, x_n) = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

een maximale waarde heeft onder een aantal beperkende condities van de vorm

$$\begin{aligned} x_1, x_2, \ldots, x_n &\geq 0 \\ a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &\leq b_2 \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &\leq b_n \end{aligned}$$

De functie c zou bijvoorbeeld een winstfunctie kunnen zijn en $x_1, x_2, \ldots, x_n$ investeringen in diverse activiteiten. De ongelijkheden staan model voor de beperkingen waar we in het dagelijks leven allemaal bloot aan staan. Op de middelbare school worden lineaire programmeringsproblemen voor $n = 2$ nog wel eens geformuleerd en met grafieken opgelost. In de praktijk werkt men vaak met tientallen of honderden variabelen en wordt het zogenaamde *simplexalgoritme* gebruikt. Dit algoritme is tamelijk eenvoudig te begrijpen en op een computer te programmeren. Vandaar dat het simplexalgoritme één van de vaakst gebruikte algoritmen in de praktijk is. Het is echter heel lastig om een goede afschatting van de looptijd van het algoritme te geven. In de praktijk, op doorsnee problemen, lijkt het zich polynomiaal te gedragen t.a.v. het aantal variabelen n. Echter, men is er in geslaagd om gekunstelde voorbeelden te maken waarin de looptijd van het simplexalgoritme exponentiëel met n toeneemt. Het aardige is dat deze gevallen in de praktijk zo zelden voorkomen, dat we er in de regel geen last van hebben. In 1978 slaagde de rus Khachian erin een nieuw algoritme (de zogenaamde *ellipsoïde methode*) te geven dat het lineaire programmeringsprobleem aantoonbaar in polynomiale tijd oplost. Dit was een opzienbarende vondst die aantoonde dat lineaire programmering een polynomiaal, dus makkelijk op te lossen, probleem is. Ironisch genoeg gebruikt men echter nog steeds het simplex-algoritme in plaats van het polynomiale ellipsoïde algoritme. Reden hiervoor is de grote eenvoud van het simplex-algoritme. De voordelen van dit makkelijk te onderhouden en meestal snelle algoritme wegen op tegen de paar lastig op te lossen gevallen. Die neemt men graag voor lief.
Een andere kanttekening die we bij polynomialiteit van een algoritme kunnen plaatsen is dat een algoritme met looptijd $O(L^{10})$, dus met een hoge exponent, ook niet echt prettig is. Praktisch gezien zal een dergelijk algoritme ook veel te veel tijd gaan kosten voor grotere invoerlengtes L.
Daar staat tegenover dat de klasse van polynomiale algoritmen een zeer belangrijke rol speelt in theoretische beschouwingen over efficiëntie van algoritmen. Eén mooie eigenschap is dat polynomialiteit van een algoritme tamelijk ongevoelig is voor het soort computer of computermodel dat gebruikt wordt. Voor theoretische beschouwingen zullen we trouwens alleen naar beslissingsproblemen kijken en we komen tot het volgende belangrijke begrip,

DEFINITIE 4.1 *De verzameling van alle beslissingsproblemen waarvoor een algoritme met polynomiale looptijd bestaat geven we aan met $\mathcal{P}$.*

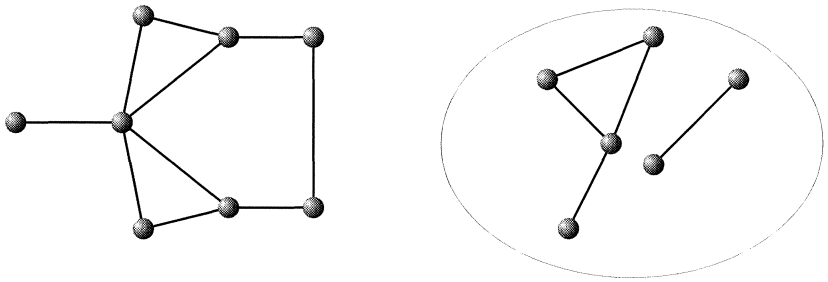
5. Moeilijke problemen

In de vorige paragraaf zagen we dat we het liefst een polynomiaal algoritme voor de oplossing van een probleem willen hebben. Helaas kan dat niet altijd. We hebben immers al gezien dat er onoplosbare problemen zijn. Maar ook binnen de klasse van oplosbare problemen kunnen we voorbeelden geven die geen polynomiale oplossing toelaten. Bekende voorbeelden hiervan zijn het probleem van de *Presburger rekenkunde* (zie [GG,SC]), *Busy beaver problem* en het *Wegblokkade spel* (zie [SC]).
Daarnaast is er ook een groot aantal problemen waarvoor geen polynomiale oplossing bekend is, maar waarvan evenmin aangetoond kan worden dat er geen polynomiaal algoritme voor bestaat. Het bekendste voorbeeld hiervan is het probleem van de ontbinding in factoren. Ondanks de boeiende ontwikkelingen van de afgelopen vijftien jaar op dit gebied is het ontbinden van een doorsnee getal van zo'n 150 cijfers nog steeds een huzarenstukje waar honderden of duizenden computers bij betrokken zijn. Het beste algoritme, de *General Number Theory Sieve*, heeft een looptijd van de orde $O(\exp(\sqrt[3]{\ln(N)\ln\ln(N)}))$.
Om te benadrukken waarom dit probleem zo belangrijk is, het zogenaamde RSA-protocol in de cryptografie is gebaseerd op onze onkunde om grote getallen in factoren te ontbinden. Het is echter wel zo dat voor dit protocol grote toepassingen gezien worden in datacommunicatie, waaronder ook electronische financiële transacties.
Veel beslissingsproblemen waarvoor noch een polynomiaal algoritme, noch het niet-bestaan ervan kan worden aangetoond hebben een interessant kenmerk gemeen. Dit verwoorden we in de volgende definitie.

Definitie 5.1 *De verzameling $\mathcal{NP}$ is precies die verzameling van beslissingsprobleem waarbij de juistheid van een 'ja'-antwoord in polynomiale tijd geverifieerd kan worden.*

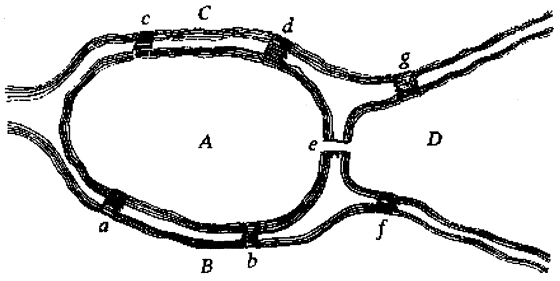
Allereerst moeten we opmerken dat $\mathcal{P} \subset \mathcal{NP}$. Het antwoord op ons beslissingsprobleem is immers in polynomiale tijd gegeven. Het interessante aan $\mathcal{NP}$ is dat er ook veel problemen toe behoren waarvan we niet weten of ze in $\mathcal{P}$ bevat zijn.
Neem als voorbeeld het deelsomprobleem. Stel we hebben 100 getallen van drie cijfers. Gevraagd wordt na te gaan of er een deelverzameling bestaat waarvan de som van de elementen gelijk is aan 50000. Ondanks alle wiskundekennis van tegenwoordig is er voor de oplossing van dit probleem weinig beters bekend dan dat we alle mogelijke deelverzamelingen uitproberen. Dit zijn dus $2^{100} \sim 10^{30}$ pogingen die we moeten uitvoeren. Als iemand er na vele dagen of jaren achter komt dat er inderdaad zo'n deelverzameling bestaat, dan is er een heel eenvoudige manier om anderen daarvan te overtuigen. Geef die ander gewoon de getallen uit de bewuste verzameling en laat hem verifiëren dat hun som inderdaad 50000 is. Dit bewijs van de juistheid van het 'ja'-antwoord kan in zeer korte tijd gevoerd worden, maar het daadwerkelijk vinden van dit bewijs is daarentegen een heel ander verhaal. In de definitie van $\mathcal{NP}$ gaat het ons echter alleen om het bestaan van een polynomiaal bewijs, er wordt niets gezegd over de moeite die men nodig had om eraan te komen.

Een ander voorbeeld uit $\mathcal{NP}$ is dat van de ontbinding van getallen als beslissingsprobleem. Het vinden van het antwoord op de vraag of er een deler van n bestaat die kleiner dan m is, is een notoir moeilijk probleem, vooral als m, n uit honderd of meer cijfers bestaan. Als het antwoord 'ja' is, dan bestaat er een heel eenvoudig bewijs voor de juistheid hiervan. We nemen gewoon de bewuste deler d, controleren of $d < m$ en controleren of d inderdaad n deelt.
Een derde voorbeeld uit de klasse $\mathcal{NP}$ is het Hamilton probleem. Dit gaat over grafen. Een graaf is een collectie van punten en lijnsegmenten zodat de uiteinden van elk lijnsegment bestaan uit een punt. Hier zijn wat voorbeelden.



Samenhangend *Niet samenhangend*

Een graaf heet samenhangend als we van elk punt via de segmenten naar elk ander punt kunnen wandelen. Een bekend probleem is de vraag of er een *Eulerpad* in zo'n graaf bestaat. Dat is gesloten pad zó dat elk segment precies éénmaal doorlopen wordt. Het bekendste voorbeeld hiervan is het klassieke Königsberger bruggen probleem, waar bij een wandelaar tijdens een rondwandeling elke brug van de zeven bruggen in het plaatje precies éénmaal wil oversteken.



De zeven bruggen van Königsberg

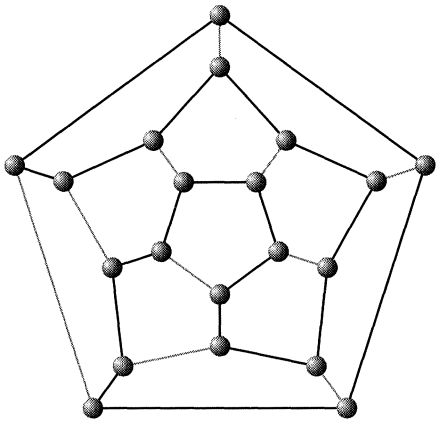
Bron: NORMAN BIGGS et al., *Graph Theory 1736–1936,* Clarendon Press, Oxford, 1976

Euler gaf hiervoor de volgende oplossing,

STELLING 5.2 (EULER) *Een samenhangende graaf bevat een Eulerpad precies dan als in elk punt van de graaf een even aantal segmenten bijeenkomen.*

Het is een leuke uitdaging hier een bewijs voor trachten te vinden. In het bijzonder zien we met deze steling dat het bruggenprobleem geen oplossing heeft.

Een variant hierop is het *Hamilton probleem* waarin gevraagd wordt of een graaf een zogenaamd *Hamilton pad* bevat. Dat is een gesloten pad dat elk punt precies éénmaal bezoekt. De naam Hamilton komt van de bekende wiskundige Rowan Hamilton, die ook verantwoordelijk is voor de Hamiltonse mechanica, de geometrische optica en de quaternionen. Hier zien we een voorbeeld van een graaf met een Hamiltonpad.



Hoeveel Hamiltonpaden zijn er in deze graaf?

Wonderlijk genoeg is er géén simpel criterium om te beslissen of een graaf een Hamiltonpad bevat. In een graaf met zo'n honderd punten kan het vinden van een Hamiltonpad een enorme zoektocht zijn. Maar ook hier geldt weer, als iemand mij een graaf geeft dat een Hamiltonpad bevat, dan kan hij mij in korte tijd van dit feit overtuigen door gewoon dat Hamiltonpad aan te geven.
Het aardige is dat bij $\mathcal{NP}$-problemen de 'nee'-antwoorden niet in polynomiale tijd bewijsbaar hoeven zijn. Dit zien we het makkelijkst bij het Hamilton probleem. Stel dat iemand mij een graaf voorlegt en beweert dat het geen Hamiltonpad bevat. Hoe zou iemand mij in korte (d.w.z. polynomiale) tijd daarvan kunnen overtuigen? Er is nu niets om ons te laten zien, er immers geen Hamiltonpad. Het is momenteel niet bekend of de afwezigheid van een Hamiltonpad in polynomiale tijd verifieerbaar is. Hoogstwaarschijnlijk niet.
Een aardig detail is dat het Hamiltonpad probleem gezien kan worden als speciaal geval van het handelsreizigerprobleem. De punten van de graaf kunnen gezien worden als steden. De afstand tussem twee van deze steden zetten we op nul als ze verbonden worden door een lijnstruk van de graaf en 1 indien ze niet verbonden worden. Het zal duidelijk zijn dat de graaf een Hamiltonpad bevat precies dan als de kortste rondweg langs de steden lengte nul heeft.
Hoewel de klasse $\mathcal{NP}$ slechts een deelverzameling van alle "moeilijke problemen" vormt, zijn wel veel voor de praktijk interessante problemen erin vertegenwoordigd. De naam $\mathcal{NP}$ betekent trouwens niet "niet polynomiaal" zoals veel mensen denken. Het staat voor *niet-deterministisch polynomiaal.* Deze benaming heeft betrekking op het niet-deterministisch computermodel. Grof gezegd is dit een computer die kan beschikken over een onbeperkte hoeveelheid naast elkaar werkende processoren. Dit is niet een erg realistisch computermodel, maar voor theoretische beschouwingen heeft het wel z'n waarde. Het verschil met de conventionele computer is dat we nu een instructie tot onze beschikking hebben waarmee we twee nieuwe processoren parallel aan het werk kunnen zetten. We noemen deze instructie BRANCH. Om een idee te geven hoe je zo'n computer zou laten werken geven we hier het voorbeeld van een

oplossing van het deelsomprobleem zoals gegeven in [L].
We hebben en rij getallen $X = (x_1, \ldots, x_n)$ en een getal S en we nemen aan dat alle processoren over deze getallen kunnen beschikken. We willen weten of er een deelrij van X bestaat zó dat de som van de elementen S is. We gebruiken hiervoor het programma `testsom`.

`testsom`(index i, tussentijdse deelsom T)

begin

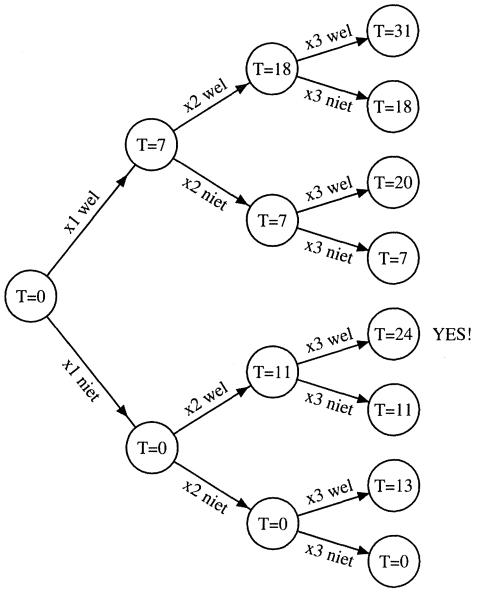
als $i > n$ en T is gelijk aan S roep "YES!"

als $i > n$ STOP

BRANCH(`testsom`$(i+1, T)$, `testsom`$(i+1, T+x_i)$)

end

Vervolgens starten we ons programma op met `testsom`$(1, 0)$. Merk op dat bij elke BRANCH-instructie een splitsing wordt gemaakt aan de hand van de beslissing of we x_i wel of niet in onze deelsom meenemen. Het totale aantal tijdstappen voor dit algoritme bedraagt n en uiteraard is dit lineair in de invoer van onze n getallen. De winst zit hem in de parallelle structuur van onze denkbeeldige machine.
In het volgende plaatje zien we hoe achtereenvolgende processoren aan het werk worden gezet in het geval dat $x_1 = 7, x_2 = 11, x_3 = 13, S = 24$.



Ook voor het ontbindingsprobleem als beslissingsprobleem kan men zich voorstellen dat we een bij elke stap exponentieel toenemend aantal processoren aan het werk zetten die elk een getal als deler proberen.
Het zal de lezer niet ontgaan zijn dat we de klasse $\mathcal{NP}$ hebben ingevoerd als beslissingsproblemen met polynomiale verificatie voor de 'ja'-instanties, maar dat de benaming van deze klasse betrekking heeft op niet-deterministische algoritmen. Dat deze twee zaken desondanks op hetzelfde neerkomen wordt ons verteld door de volgende boeiende stelling.

STELLING 5.3 (COOK,1971) *Zij A een beslissingsprobleem. Dan zijn de volgende beweringen equivalent,*

1. *De 'ja'-instanties van A zij polynomiaal verifieerbaar.*
2. *$A \in \mathcal{NP}$*
3. *A is in polynomiale tijd herleidbaar tot het Hamilton-pad probleem.*

Deze stelling bevat een derde ingredient, namelijk dat elk probleem behorend tot $\mathcal{NP}$ polynomiaal herleidbaar is tot het Hamilton-pad probleem. Een voorbeeld van een polynomiale reductie van één probleem tot een ander zagen we al bij de formulering van het ontbindingsprobleem als beslissingsprobleem. Daar hadden we enige overhead nodig in de vorm van een binaire zoekactie. Deze overhead draagt echter polynomiaal bij aan de looptijd van het algoritme. Zolang dit het geval blijft spreken van een polynomiale reductie. Het opmerkelijke van de Stelling van Cook is dat elk probleem uit de klasse $\mathcal{NP}$ polynomiaal herleidbaar is tot het Hamilton-pad probleem. Dit betekent in het bijzonder dat als we ooit een polynomiaal algoritme voor het Hamilton-pad probleem zouden vinden, in één klap alle problemen uit $\mathcal{NP}$ polynomiaal oplosbaar zijn, inclusief ontbinding in factoren.
In het oorspronkelijk werk van Cook wordt niet het Hamiltonpad-probleem genoemd maar het *Satisfiability probleem* (zie [L]) voor logische uitdrukkingen. Vlak daarna ontdekte Karp dat het satisfiability probleem polynomiaal kan worden teruggevoerd diverse andere problemen in de klasse $\mathcal{NP}$. Voorbeelden hiervan zijn het Hamiltonpad-probleem en het deelsom probleem. Problemen van deze soort noemen we $\mathcal{NP}$-complete problemen. In het algemeen noemen we een beslissingsprobleem A *$\mathcal{NP}$-compleet* als $A \in \mathcal{NP}$ en als elk $\mathcal{NP}$-probleem polynomiaal tot probleem A kan worden herleid. In de paar jaar volgend op Cook's ontdekking breidde de lijst van $\mathcal{NP}$-complete problemen zich uit tot enkele honderden. Elk van deze problemen heeft dus de eigenschap dat als we een polynomiaal oplossingsalgoritme zouden vinden, elk $\mathcal{NP}$-probleem automatisch een polynomiale oplossing heeft. De omkering van deze bewering geldt uiteraard ook. Met andere woorden,

$$A\ \mathcal{NP}\text{-compleet en } A \in \mathcal{P} \iff \mathcal{NP} = \mathcal{P}$$

Om zich te oriënteren zou de lezer zelf eens moeten proberen een oplossing te vinden voor het deelsomprobleem, een computerprogramma hiervoor schrijven en hiermee problemen met $n = 40$ (zeg) aanpakken. Velen hebben dit al gedaan en zijn hierdoor gesterkt in de mening dat het deelsomprobleem waarschijnlijk geen polynomiaal probleem is. Dit impliceert automatisch het volgende vermoeden,

VERMOEDEN 5.4

$$\mathcal{NP} \neq \mathcal{P}$$

Het is absoluut niet duidelijk hoe men een dergelijke uitspraak zou moeten bewijzen, ondanks de vele pogingen daartoe. Het hardnekkige voortbestaan van dit vermoeden en het grote belang ervan voor computationele toepassingen hebben ervoor gezorgd dat dit vermoeden een plaats heeft gekregen onder de bekendste problemen in de wiskunde.

6. LITERATUUR

Bij de voorbereiding van deze tekst heb ik dankbaar gebruik gemaakt van de artikelen en boeken die hieronder vermeld staan. Vooral de algemene en meer populaire

artikelen zijn een grote steun geweest bij de voorbereiding. Mijn dank gaat uit naar Jan Karel Lenstra die mij deze artikelen te leen gaf.

Boeken

[BCF] H.W.Broer, J.van de Craats, F.Verhulst, *Het einde der voorspelbaarheid? Chaos: ideeën en toepassingen*, Epsilon Uitgaven, Utrecht 1995.

[LLRS] E.L.Lawler, J.K.Lenstra, A.H.G.Rinnooy Kan, D.B.Shmoys, editors, *The travelling salesman problem*, Wiley and Sons, 1985.

[P] C.H.Papadimitriou, *Computational Complexity*, Addison-Wesley 1994.

[RS] G.Rozenberg, A.Salomaa, *Cornerstones od undecidability*, Prentice, Hall, 1994.

Artikelen

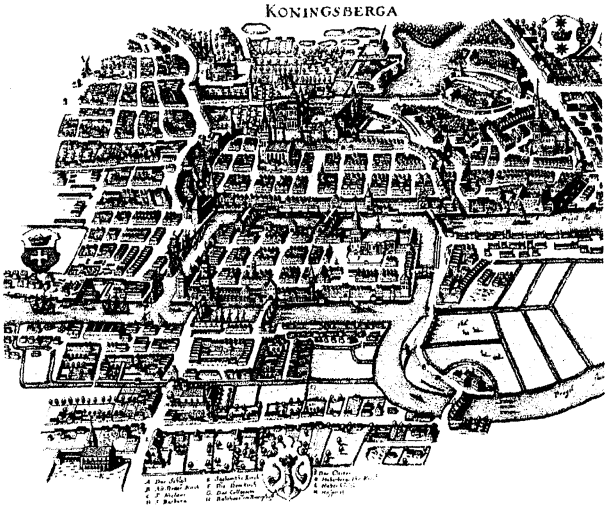
[DMR] M. Davis, Y. Matijaseirch, J. Robinson, Hilber's tenth problem, *Proc, Symp. Pure Mathematics* **28**, AMS, 1976.

[GG] R.L. Graham, M.R. Garey, *The Limits to Computation.*

[L] E.A. Lamagaa, Infeasible Computations, *Abacus* **4** (1987), p. 18-33.

[LP] H.R. Lewis, C.H. Papadimitriou, The Efficiency of Algorithms, Scientific American **238**(1), p. 96–109.

[SC] L.J. Stockmeyer, A.K. Chandra, Intrinsically Difficult Problems, Scientific American, May 1979, p. 124–133.



Königsberg in de 18^e *eeuw.*

Medewerkers aan de Vakantiecursus 1999

DOCENTEN

dr. F. Beukers
Mathematisch Instituut, Universiteit Utrecht
Postbus 80010, 3508 TA Utrecht
tel. 030-2531419, email: beukers@math.uu.nl

prof.dr. J. van de Craats
KMA, Postbus 90154, 4800 RG Breda
tel. 076-5273816, email: jcr@euronet.nl

prof.dr. G.B.M. van der Geer
Korteweg - de Vries Instituut voor Wiskunde, Universiteit van Amsterdam
Plantage Muidergracht 24, 1018 TV Amsterdam
tel. 020-5255247, email: geer@wins.uva.nl

dr. P.W.H. Lemmens
Mathematisch Instituut, Universiteit Utrecht
Postbus 80010, 3508 TA Utrecht
tel. 030-2531426, email: lemmens@math.uu.nl

dr. J.B.M. Melissen
Hogeschool 's Hertogenbosch
Postbus 732, 5201 AS 's Hertogenbosch
email: j.melissen@hsbos.nl

dr. P. Stevenhagen
Korteweg - de Vries Instituut voor Wiskunde, Universiteit van Amsterdam
Plantage Muidergracht 24, 1018 TV Amsterdam
tel. 020-5255202, email: psh@wins.uva.nl

prof.dr. R. Tijdeman
Mathematisch Instituut, Universiteit Leiden
Postbus 9512, 2300 RA Leiden
tel. 071-5277138, email: tijdeman@wi.leidenuniv.nl

CONTACTPERSONEN CWI

- dr. Miente Bakker, tel. 020-5924172, email: Miente.Bakker@cwi.nl
- Simone Panka, tel. 020-5924009, email: Simone.Panka@cwi.nl

CWI SYLLABI

1 Vacantiecursus 1984: *Hewet - plus wiskunde.* 1984.

2 E.M. de Jager, H.G.J. Pijls (eds.). *Proceedings Seminar 1981–1982. Mathematical structures in field theories.* 1984.

3 W.C.M. Kallenberg, et al. *Testing statistical hypotheses: worked solutions.* 1984.

4 J.G. Verwer (ed.). *Colloquium topics in applied numerical analysis, volume 1.* 1984.

5 J.G. Verwer (ed.). *Colloquium topics in applied numerical analysis, volume 2.* 1984.

6 P.J.M. Bongaarts, J.N. Buur, E.A. de Kerf, R. Martini, H.G.J. Pijls, J.W. de Roever. *Proceedings Seminar 1982–1983. Mathematical structures in field theories.* 1985.

7 Vacantiecurus 1985: *Variatierekening.* 1985.

8 G.M. Tuynman. *Proceedings Seminar 1983–1985. Mathematical structures in field theories, Vol.1 Geometric quantization.* 1985.

9 J. van Leeuwen, J.K. Lenstra (eds.). *Parallel computers and computations.* 1985.

10 Vacantiecursus 1986: *Matrices.* 1986.

11 P.W.H. Lemmens. *Discrete wiskunde: tellen, grafen, spelen en codes.* 1986.

12 J. van de Lune. *An introduction to Tauberian theory: from Tauber to Wiener.* 1986.

13 G.M. Tuynman, M.J. Bergvelt, A.P.E. ten Kroode. *Proceedings Seminar 1983–1985. Mathematical structures in field theories, Vol.2.* 1987.

14 Vacantiecursus 1987: *De personal computer en de wiskunde op school.* 1987.

15 Vacantiecursus 1983: *Complexe getallen.* 1987.

16 P.J.M. Bongaarts, E.A. de Kerf, P.H.M. Kersten. *Proceedings Seminar 1984–1986. Mathematical structures in field theories, Vol.1.* 1988.

17 F. den Hollander, H. Maassen (eds.). *Mark Kac seminar on probability and physics. Syllabus 1985–1987.* 1988.

18 Vacantiecursus 1988. *Differentierekening.* 1988.

19 R. de Bruin, C.G. van der Laan, J. Luyten, H.F. Vogt. *Publiceren met LATEX.* 1988.

20 R. van der Horst, R.D. Gill (eds.). *STATAL: statistical procedures in Algol 60, part 1.* 1988.

21 R. van der Horst, R.D. Gill (eds.). *STATAL: statistical procedures in Algol 60, part 2.* 1988.

22 R. van der Horst, R.D. Gill (eds.). *STATAL: statistical procedures in Algol 60, part 3.* 1988.

23 J. van Mill, G.Y. Nieuwland (eds.). *Proceedings van het symposium wiskunde en de computer.* 1989.

24 P.W.H. Lemmens (red.). *Bewijzen in de wiskunde.* 1989.

25 Vacantiecursus 1989: *Wiskunde in de Gouden Eeuw.* 1989.

26 G.G.A. Bäuerle et al. *Proceedings Seminar 1986–1987. Mathematical structures in field theories.* 1990.

27 Vacantiecursus 1990: *Getallentheorie en haar toepassingen.* 1990.

28 Vacantiecursus 1991: *Meetkundige structuren.* 1991.

29 A.G. van Asch, F. van der Blij. *Hoeken en hun Maat.* 1992.

30 M.J. Bergvelt, A.P.E. ten Kroode. *Proceedings seminar 1986–1987. Lectures on Kac-Moody algebras.* 1992.

31 Vacantiecursus 1992: *Systeemtheorie.* 1992.

32 F. den Hollander, H. Maassen (eds.). *Mark Kac seminar on probability and physics. Syllabus 1987–1992.* 1992.

33 P.W.H. Lemmens (ed.). *Meetkunde van kunst tot kunde, vroeger en nu.* 1993.

34 J.H. Kruizinga. *Toegepaste wiskunde op een PC.* 1992.

35 Vacantiecursus 1993: *Het reële getal.* 1993.

36 Vacantiecursus 1994: *Computeralgebra.* 1994.

37 G. Alberts. *Wiskunde en praktijk in historisch perspectief. Syllabus.* 1994.

38 G. Alberts, J. Schut (eds.). *Wiskunde en praktijk in historisch perspectief. Reader.* 1994.

39 E.A. de Kerf, H.G.J. Pijls (eds.). *Proceedings Seminar 1989–1990. Mathematical structures in field theory.* 1996.

40 Vacantiecursus 1995: *Kegelsneden en kwadratische vormen.* 1995.

41 Vacantiecursus 1996: *Chaos.* 1996.

42 H.C. Doets. *Wijzer in Wiskunde.* 1996.

43 Vacantiecursus 1997: *Rekenen op het Toeval.* 1997.

44 Vacantiecursus 1998: *Meetkunde, Oud en Nieuw.* 1998.

45 Vacantiecursus 1999: *Onbewezen Vermoedens.* 1999.

MC SYLLABI

1.1 F. Göbel, J. van de Lune. *Leergang besliskunde, deel 1: wiskundige basiskennis.* 1965.

1.2 J. Hemelrijk, J. Kriens. *Leergang besliskunde, deel 2: kansberekening.* 1965.

1.3 J. Hemelrijk, J. Kriens. *Leergang besliskunde, deel 3: statistiek.* 1966.

1.4 G. de Leve, W. Molenaar. *Leergang besliskunde, deel 4: Markovketens en wachttijden.* 1966.

1.5 J. Kriens, G. de Leve. *Leergang besliskunde, deel 5: inleiding tot de mathematische besliskunde.* 1966.

1.6a B. Dorhout, J. Kriens. *Leergang besliskunde, deel 6a: wiskundige programmering 1.* 1968.

1.6b B. Dorhout, J. Kriens, J.Th. van Lieshout. *Leergang besliskunde, deel 6b: wiskundige programmering 2.* 1977.

1.7a G. de Leve. *Leergang besliskunde, deel 7a: dynamische programmering 1.* 1968.

1.7b G. de Leve, H.C. Tijms. *Leergang besliskunde, deel 7b: dynamische programmering 2.* 1970.

1.7c G. de Leve, H.C. Tijms. *Leergang besliskunde, deel 7c: dynamische programmering 3.* 1971.

1.8 J. Kriens, F. Göbel, W. Molenaar. *Leergang besliskunde, deel 8: minimaxmethode, netwerkplanning, simulatie.* 1968.

2.1 G.J.R. Förch, P.J. van der Houwen, R.P. van de Riet. *Colloquium stabiliteit van differentieschema's, deel 1.* 1967.

2.2 L. Dekker, T.J. Dekker, P.J. van der Houwen, M.N. Spijker. *Colloquium stabiliteit van differentieschema's, deel 2.* 1968.

3.1 H.A. Lauwerier. *Randwaardeproblemen, deel 1.* 1967.

3.2 H.A. Lauwerier. *Randwaardeproblemen, deel 2.* 1968.

3.3 H.A. Lauwerier. *Randwaardeproblemen, deel 3.* 1968.

4 H.A. Lauwerier. *Representaties van groepen.* 1968.

5 J.H. van Lint, J.J. Seidel, P.C. Baayen. *Colloquium discrete wiskunde.* 1968.

6 K.K. Koksma. *Cursus ALGOL 60.* 1969.

7.1 *Colloquium moderne rekenmachines, deel 1.* 1969.

7.2 *Colloquium moderne rekenmachines, deel 2.* 1969.

8 H. Bavinck, J. Grasman. *Relaxatietrillingen.* 1969.

9.1 T.M.T. Coolen, G.J.R. Förch, E.M. de Jager, H.G.J. Pijls. *Colloquium elliptische differentiaalvergelijkingen, deel 1.* 1970.

9.2 W.P. van den Brink, T.M.T. Coolen, B. Dijkhuis, P.P.N. de Groen, P.J. van der Houwen, E.M. de Jager, N.M. Temme, R.J. de Vogelaere. *Colloquium elliptische differentiaalvergelijkingen, deel 2.* 1970.

10 J. Fabius, W.R. van Zwet. *Grondbegrippen van de waarschijnlijkheidsrekening.* 1970.

11 H. Bart, M.A. Kaashoek. H.G.J. Pijls, W.J. de Schipper, J. de Vries. *Colloquium halfalgebra's en positieve operatoren.* 1971.

12 T.J. Dekker. *Numerieke algebra.* 1971.

13 F.E.J. Kruseman Aretz. *Programmeren voor rekenautomaten; de MC ALGOL 60 vertaler voor de EL X8.* 1971.

14 H. Bavinck, W. Gautschi, G.M. Willems. *Colloquium approximatietheorie.* 1971.

15.1 T.J. Dekker, P.W. Hemker, P.J. van der Houwen. *Colloquium stijve differentiaalvergelijkingen, deel 1.* 1972.

15.2 P.A. Beentjes, K. Dekker, H.C. Hemker, S.P.N. van Kampen, G.M. Willems. *Colloquium stijve differentiaalvergelijkingen, deel 2.* 1973.

15.3 P.A. Beentjes, K. Dekker, P.W. Hemker, M. van Veldhuizen. *Colloquium stijve differentiaalvergelijkingen, deel 3.* 1975.

16.1 L. Geurts. *Cursus programmeren, deel 1: de elementen van het programmeren.* 1973.

16.2 L. Geurts. *Cursus programmeren, deel 2: de programmeertaal ALGOL 60.* 1973.

17.1 P.S. Stobbe. *Lineaire algebra, deel 1.* 1973.

17.2 P.S. Stobbe. *Lineaire algebra, deel 2.* 1973.

17.3 N.M. Temme. *Lineaire algebra, deel 3.* 1976.

18 F. van der Blij, H. Freudenthal, J.J. de Iongh, J.J. Seidel, A. van Wijngaarden. *Een kwart eeuw wiskunde 1946-1971, syllabus van de vakantiecursus 1971.* 1973.

19 A. Hordijk, R. Potharst, J.Th. Runnenburg. *Optimaal stoppen van Markovketens.* 1973.

20 T.M.T. Coolen, P.W. Hemker, P.J. van der Houwen, E. Slagt. *ALGOL 60 procedures voor begin- en randwaardeproblemen.* 1976.

21 J.W. de Bakker (red.). *Colloquium programmacorrectheid.* 1975.

22 R. Helmers, J. Oosterhoff, F.H. Ruymgaart, M.C.A. van Zuylen. *Asymptotische methoden in de toetsingstheorie; toepassingen van naburigheid.* 1976.

23.1 J.W. de Roever (red.). *Colloquium onderwerpen uit de biomathematica, deel 1.* 1976.

23.2 J.W. de Roever (red.). *Colloquium onderwerpen uit de biomathematica, deel 2.* 1977.

24.1 P.J. van der Houwen. *Numerieke integratie van differentiaalvergelijkingen, deel 1: eenstapsmethoden.* 1974.

25 *Colloquium structuur van programmeertalen.* 1976.

26.1 N.M. Temme (ed.). *Nonlinear analysis, volume 1.* 1976.

26.2 N.M. Temme (ed.). *Nonlinear analysis, volume 2.* 1976.

27 M. Bakker, P.W. Hemker, P.J. van der Houwen, S.J. Polak, M. van Veldhuizen. *Colloquium discretiseringsmethoden.* 1976.

28 O. Diekmann, N.M. Temme (eds.). *Nonlinear diffusion problems.* 1976.

29.1 J.C.P. Bus (red.). *Colloquium numerieke programmatuur, deel 1A, deel 1B.* 1976.

29.2 H.J.J. te Riele (red.). *Colloquium numerieke programmatuur, deel 2.* 1977.

30 J. Heering, P. Klint (red.). *Colloquium programmeeromgevingen.* 1983.

31 J.H. van Lint (red.). *Inleiding in de coderingstheorie.* 1976.

32 L. Geurts (red.). *Colloquium bedrijfsystemen.* 1976.

33 P.J. van der Houwen. *Berekening van waterstanden in zeeën en rivieren.* 1977.

34 J. Hemelrijk. *Oriënterende cursus mathematische statistiek.* 1977.

35 P.J.W. ten Hagen (red.). *Colloquium computer graphics.* 1978.

36 J.M. Aarts, J. de Vries. *Colloquium topologische dynamische systemen.* 1977.

37 J.C. van Vliet (red.). *Colloquium capita datastructuren.* 1978.

38.1 T.H. Koornwinder (ed.). *Representations of locally compact groups with applications, part 1.* 1979.

38.2 T.H. Koornwinder (ed.). *Representations of locally compact groups with applications, part II.* 1979.

39 O.J. Vrieze, G.L. Wanrooy. *Colloquium stochastische spelen.* 1978.

40 J. van Tiel. *Convexe analyse.* 1979.

41 H.J.J. te Riele (ed.). *Colloquium numerical treatment of integral equations.* 1979.

42 J.C. van Vliet (red.). *Colloquium capita implementatie van programmeertalen.* 1980.

43 A.M. Cohen, H.A. Wilbrink. *Eindige groepen (een inleidende cursus).* 1980.

44 J.G. Verwer (ed.). *Colloquium numerical solution of partial differential equations.* 1980.

45 P. Klint (red.). *Colloquium hogere programmeertalen en computerarchitectuur.* 1980.

46.1 P.M.G. Apers (red.). *Colloquium databankorganisatie, deel 1.* 1981.

46.2 P.G.M. Apers (red.). *Colloquium databankorganisatie, deel 2.* 1981.

47.1 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60: general information and indices.* 1981.

47.2 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 1: elementary procedures; vol. 2: algebraic evaluations.* 1981.

47.3 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 3A: linear algebra, part I.* 1981.

47.4 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 3B: linear algebra, part II.* 1981.

47.5 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 4: analytical evaluations; vol. 5A: analytical problems, part I.* 1981.

47.6 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 5B: analytical problems, part II.* 1981.

47.7 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 6: special functions and constants; vol. 7: interpolation and approximation.* 1981.

48.1 P.M.B. Vitányi, J. van Leeuwen, P. van Emde Boas (red.). *Colloquium complexiteit en algoritmen, deel 1.* 1982.

48.2 P.M.B. Vitányi, J. van Leeuwen, P. van Emde Boas (red.). *Colloquium complexiteit en algoritmen, deel 2.* 1982.

49 T.H. Koornwinder (ed.). *The structure of real semisimple Lie groups.* 1982.

50 H. Nijmeijer. *Inleiding systeemtheorie.* 1982.

51 P.J. Hoogendoorn (red.). *Cursus cryptografie.* 1983.