



## **CWI Syllabi**

### **Managing Editors**

A.M.H. Gerards (CWI, Amsterdam)  
M. Hazewinkel (CWI, Amsterdam)  
J.W. Klop (CWI, Amsterdam)  
N.M. Temme (CWI, Amsterdam)

### **Executive Editor**

M. Bakker (CWI Amsterdam, e-mail: [Miente.Bakker@cwi.nl](mailto:Miente.Bakker@cwi.nl))

### **Editorial Board**

W. Albers (Enschede)  
K.R. Apt (Amsterdam)  
M.S. Keane (Amsterdam)  
P.W.H. Lemmens (Utrecht)  
J.K. Lenstra (Eindhoven)  
M. van der Put (Groningen)  
A.J. van der Schaft (Enschede)  
J.M. Schumacher (Tilburg)  
H.J. Sips (Delft, Amsterdam)  
M.N. Spijker (Leiden)  
H.C. Tijms (Amsterdam)

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/publications.bibl/>

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

Vakantiecursus 2001  
Experimentele wiskunde

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

ISBN 90 6196 505 5  
NUGI-code: 811

Copyright ©2001, Stichting Centrum voor Wiskunde en Informatica, Amsterdam  
Printed in the Netherlands

## Inhoud

Ten geleide J. VAN DE CRAATS	1
Experimentele getaltheorie F. BEUKERS	3
Het begrip dimensie A. DOELMAN	17
Uitputting en evenwicht J.A. VAN MAANEN	39
Snel oplossen is een experiment waard H.A. VAN DER VORST	49
Experimenten met Cabri A.J. GODDIJN	61
De wondere wereld van de Poisson-kansverdeling H.C. TIJMS	93
Netwerkplanning R.D. VAN DER MEI	111
Medewerkers aan de Vakantiecursus	125





## Ten Geleide

Jan van de Craats  
KMA Breda

e-mail: [jcr@euronet.nl](mailto:jcr@euronet.nl)

Experimentele wiskunde – nog maar kort geleden zou die term in universitaire wiskundekringen vooral fronsende wenkbrauwen opgeleverd hebben. In de wiskunde werd niet geëxperimenteerd, in de wiskunde werd de geest gescherpt door zuiver denkwerk. Experimenten bewijzen niets, en een bewijs is in de wiskunde het enige dat telt.

Inmiddels is dat beeld sterk veranderd. Natuurlijk, nog steeds gaat het in de wiskunde om denkwerk, en nog steeds vormen bewijzen de kern van de zaak, maar ook de zuiver wiskundige gaat experimenten niet meer uit de weg. Zo valt er bijvoorbeeld in de getaltheorie tegenwoordig heel wat te experimenteren, met name met behulp van de computer. Daar komen dan allerlei verrassende resultaten, maar ook weer nieuwe onderzoeksvragen uit voort. Dit voorbeeld staat niet op zichzelf, zoals uit de teksten in deze Syllabus moge blijken. Op allerlei gebieden in de wiskunde spelen experimenten een rol, zowel in het onderzoek als in het onderwijs. Daarmee is Experimentele Wiskunde een thema dat een rijk gevarieerde Vakantiecursus voor wiskundeleraren mogelijk maakt, met oog voor historische achtergronden en moderne ontwikkelingen in het onderzoek, maar ook met aandacht voor de beroepspraktijk van de wiskundige en dwarsverbanden met de schoolpraktijk.

Wiskunde is de wetenschap van structuren en patronen. Experimenten, al dan niet met de computer, helpen wiskundigen structuren te ontdekken, te verkennen en in beeld te brengen. Intelligent experimenteren in combinatie met abstract denkwerk vormen een vruchtbare voedingsbodem waarop de wiskunde in de komende eeuw een nieuwe bloeitijd tegemoet gaat.

Gaarne wil ik op deze plaats allen bedanken die in 2001 opnieuw een Vakantiecursus mogelijk hebben gemaakt. In de eerste plaats natuurlijk de sprekers, die naast hun lezing ook een tekst voor deze Syllabus hebben geleverd. Daarmee wordt opnieuw een aantrekkelijk deel toegevoegd aan een serie syllabi met voor leraren en andere belangstellenden uiterst waardevol materiaal. Het Centrum voor Wiskunde en Informatica te Amsterdam en de Technische Universiteit Eindhoven stelden zaalruimte beschikbaar, de administratieve en praktische organisatie van de cursus was in handen van mevrouw Wilmy van Ojik en dr. Miente Bakker, die ook samen met mevrouw Minnie Middelberg de inhoudelijke coördinatie van de Syllabus verzorgde.

Allen hartelijk dank!





## Experimentele Getaltheorie

Frits Beukers  
Universiteit Utrecht  
e-mail: beukers@math.uu.nl

### 1. DE VERGELIJKING VAN MORDELL

In de 17e eeuw merkte Fermat al op dat  $5^2 = 3^3 - 2$  en vroeg zich af of er meer gehele getallen  $x, y$  bestaan zó dat  $y^2 = x^3 - 2$ . Al spoedig liet Fermat zien dat er geen andere oplossingen bestaan. Fermat's vraag laat zich gemakkelijk generaliseren. Gegeven een getal  $k \neq 0$ , positief of negatief, bestaan er gehele getallen  $x, y$  zó dat

$$y^2 = x^3 + k.$$

Dit is een voorbeeld van een *diophantische vergelijking*, dat wil zeggen een vergelijking waarvan we de oplossing in gehele getallen willen weten. De vergelijking  $y^2 = x^3 + k$  staat bekend als de *vergelijking van Mordell*, naar de eerste wiskundige die aantoonde dat voor gegeven  $k \neq 0$  er hooguit eindig veel oplossingen kunnen zijn. Het bewijs hiervan, dat niet makkelijk is, is te vinden in Mordell, *Diophantine Equations*, Academic Press 1969. Mordell's vergelijking wordt gezien als een van de meest "klassieke" diophantische vergelijkingen. Hoewel we weten dat er voor gegeven  $k$  hoogstens eindig veel oplossingen zijn, is het vaak een lastige zaak om die oplossingen ook inderdaad te bepalen. In het boek van London en Finkelstein, *On Mordell's equation*, Bowling Green State University 1973, wordt voor de meeste waarden van  $k$  met  $|k| \leq 100$  de volledige oplossing gegeven. Het feit dat hieraan een heel boek besteed wordt, geeft al aan dat het geen eenvoudige klus is! Hier is een voorbeeld, de vergelijking  $y^2 = x^3 + 17$  heeft de oplossingen

$$(x, y) = (-2, 3), (-1, 4), (2, 5), (4, 9), (8, 23)$$

$$(43, 282), (52, 375), (5234, 378661)$$

waarbij we alleen de oplossingen met  $y > 0$  hebben opgeschreven. De andere krijg je gewoon door het teken van  $y$  te veranderen. Aan de andere kant zijn er voor veel waarden van  $k$  helemaal geen oplossingen. Sinds een twintigtal jaren zijn er systematische methoden ontwikkeld om Mordell's vergelijking aan te pakken. Deze ontwikkeling is gestimuleerd door de beschikbaarheid van steeds grotere hoeveelheden reken capaciteit in de vorm van steeds sneller wordende computers. In een artikel uit 1998 van Gebel, Pethö en Zimmer, *On Mordell's equation*, *Compositio Math.* 110(1998), p. 335–367 worden alle vergelijkingen met  $|k| < 10.000$  opgelost en het merendeel van de vergelijkingen met  $|k| < 100.000$ . Deze laatste lijst is afgemaakt door K.Wildanger. De lijst met  $|k| <$

10.000 is te vinden op <http://emmy.math.uni-sb.de/~simath/MORDELL/>. Uit deze tabel reproduceren we in de Appendix A de oplossingen voor alle  $k$  met  $0 < |k| \leq 100$ .

Sleutelwoorden in de gebruikte technieken zijn: lineaire vormen in elliptische logaritmen en LLL-algoritme, ingrediënten die pas de laatste vijftien jaar gevonden zijn.

## 2. HALL'S VERMOEDEN

Ondanks het feit dat er nu uitgebreide tabellen voor de Mordell-vergelijking bestaan, blijft er nog een groot aantal vragen over. Bijvoorbeeld, om de Mordell-vergelijking op te lossen zou het handig zijn om te weten hoe klein het verschil tussen een kwadraat en een derde macht kan zijn. Het bekendste vermoeden in die richting is het volgende.

**Vermoeden 2.1 (Hall).** *Er bestaat een getal  $C > 0$  zó dat voor elk tweetal  $x, y \in \mathbb{Z}_{>0}$  met  $x^3 \neq y^2$  geldt  $|x^3 - y^2| > C\sqrt{x}$ .*

Dit vermoeden werd naar voren gebracht door Marshall Hall tijdens een symposium *Computers in Number Theory* [H] in 1969. Dit symposium was een van de eerste waarin computer-experimenten op getaltheoriegebied voorop staan.

Laten we eens aannemen dat Hall's vermoeden waar is en kijken wat het gevolg is voor de Mordell vergelijking. Aangezien we nog geen enkel voorbeeld van  $x, y$  kennen met  $0 < |x^3 - y^2| < 0.01\sqrt{x}$ , nemen we even  $C = 0.01$ . Voor een oplossing van de vergelijking  $y^2 = x^3 + k$  zou dit betekenen dat  $|k| = |x^3 - y^2| > 0.01\sqrt{x}$  en dus,  $x < (100k)^2$ . Om de vergelijking van Mordell op te lossen met bijvoorbeeld  $k = 17$  zouden we voor  $x = -2, -1, 0, 1, 2, \dots, 1700^2$  moeten proberen of  $x^3 + 17$  een kwadraat is. Voor de meeste PC's is dit tegenwoordig een paar seconden werk.

Het vermoeden van Hall is, zoals de naam zegt, een vermoeden. De ondergrenzen voor  $|x^3 - y^2|$  die werkelijk bewezen zijn in de getaltheorie liggen daar nog heel ver af. Een recente ondergrens is de volgende

**Stelling 2.2 (Sprindzuk, 1982).** *Als  $x^3 \neq y^2$  en  $x > 10$ , dan geldt*

$$|x^3 - y^2| > \gamma \log x / (\log \log x)^6$$

waarin  $\gamma$  een berekenbaar positief getal is.

De techniek om deze stelling te bewijzen, is A.Baker's theorie van lineaire vormen in logaritmen uit 1970 en verbeteringen daarop in latere jaren. Dit resultaat is het beste resultaat dat we kunnen krijgen met de hedendaagse technieken uit de getaltheorie. Om echter in de buurt van Hall's ongelijkheid te komen zullen er waarschijnlijk fundamenteel nieuwe technieken in de getaltheorie ontwikkeld moeten worden.

Omdat we natuurlijk geen zin hebben om op die nieuwe technieken te wachten, gaan we aan het experimenteren. Allereerst zouden we eens kunnen proberen om getallen  $x, y$  te vinden zo dat  $0 < |x^3 - y^2| < \sqrt{x}$ . De lezer is bij deze

uitgenodigd een poging hiertoe te wagen, zonder naar de tabellen verderop te kijken. Hij of zij zal waarschijnlijk verbaasd zijn om te moeten constateren dat dit vrijwel niet lukt. De hardnekkigheid waarmee  $|x^3 - y^2| > \sqrt{x}$  voor elk doorsnee tweetal  $x, y$  is opvallend. Het lijkt erop dat we te maken hebben met een harde wetmatigheid die zich in de getallenwereld afspeelt. Een systematische manier om erachter te komen hoe hardnekkig dit verschijnsel is, is natuurlijk voor  $x = 1, 2, 3, \dots$  de waarde  $x^3$  uitrekenen en het verschil met het dichtstbijzijnde kwadraat  $y^2$  bepalen. Dit experiment is talloze malen op vele computers gedaan. Een aantal jaren geleden heb ik zelf ook een dergelijke proef genomen en alle getallen  $x < 10^{11}$  getest. De waarde  $|x^3 - y^2|/\sqrt{x}$  zullen we de *Hallwaarde* van het paar  $x, y$  (of van  $x$ ) noemen. We spreken ook af dat we  $x, y$  met Hallwaarde nul, dus  $x^3 = y^2$  voortaan buiten beschouwing laten. Hieronder volgen alle  $x < 10^{11}$  met Hallwaarde  $< 1$ .

$x$	$x^3 - y^2$	Hallwaarde
5234	-17	0.234
8158	-24	0.265
93844	-297	0.969
367806	207	0.341
421351	-618	0.952
720114	-225	0.265
939787	307	0.316
28187351	-1090	0.205
110781386	-8569	0.814
154319269	-11492	0.925
384242766	-14668	0.748
390620082	-14857	0.751
3790689201	-28024	0.455
65589428378	-117073	0.456

Een fysicus zou, aangemoedigd door deze tabel, kunnen concluderen dat  $C = 0.2$  in Hall's vermoeden. Er is echter geen enkele garantie daarvoor. Laten we dus verder gaan en grotere waarden van  $x$  testen. Nu treedt er echter een probleem op. Vroeger of later botsen we tegen de beperkingen van onze hardware op. Om een idee te geven op welk moment we met deze beperkingen te maken krijgen, stellen voor het gemak dat het testen van één  $x$ -waarde ongeveer  $10^{-7}$  seconde kost. Dit zijn ongeveer 100 klok cycli met de tegenwoordige GigaHertz processoren, amper genoeg om een getal van zo'n tien cijfers tot de derde macht te verheffen, een bijbehorende  $y$  te vinden en  $y^2 - x^3$  te bepalen. Voor het testen van  $10^{12}$   $x$ -waarden hebben we dus ongeveer  $10^5$  seconden nodig, ofwel 30 uur. Om alle  $x < 10^{15}$  te testen hebben we 30.000 uur nodig. Het moge duidelijk zijn dat we hier aan de grens van onze mogelijkheden zitten.

De enige optie om verder te komen is een andere methode te bedenken die ons verder kan brengen. Dit is tevens de grootste uitdaging voor de computationele wiskunde. Goede ideeën die ons rekenbereik significant vergroten zijn zeldzaam en meestal zeer lastig te vinden. Daarom zijn ze goud waard. Een

dergelijke doorbraak op het gebied van het experimentele Hall-vermoeden werd in 1998 door Noam Elkies gedaan. Door herordening van het zoekbereik en een aantal ideeën uit de diophantische approximatie te gebruiken, slaagde hij erin alle  $x < 10^{18}$  te testen. De grens  $10^{11}$  is hiermee letterlijk verpletterd. Als we met de naïeve methode alle getallen  $x < X$  willen testen dan hebben we daarvoor ongeveer  $\gamma X$  seconden nodig, waarin  $\gamma$  afhangt van de gebruikte hardware. Een functie die begrensd wordt door een constante te maal  $X$  noemen we een *functie van orde  $X$* . Notatie:  $O(X)$ . De methode van Elkies daarentegen heeft een looptijd van de orde  $O(X^{1/2} \log X)$ . Een fors verschil met  $X$ , vooral als  $X$  groot is. Hier volgt een tabel van de nieuwe instanties met Hall-waarde  $< 1$ .

$x$	$x^3 - y^2$	Hall-waarde
12438517260105	2767769	0.784
35495694227489	5190544	0.871
53197086958290	-4401169	0.603
5853886516781223	1641843	0.021
12813608766102806	87002345	0.768
23415546067124892	105077952	0.686
38115991067861271	30032270	0.153
322001299796379844	548147655	0.965
471477085999389882	-497218657	0.724
810574762403977064	-193234265	0.214

In deze tabel komt zelfs een paar  $x, y$  met Hall-waarde 0.021 voor! Het even simpele als geniale idee van Elkies zullen we in dit stukje niet uiteenzetten. Daarvoor verwijzen we naar het artikel van Elkies in *Lecture Notes in Computer Science 1838*, Springer Verlag. Maar zelfs met Elkies' nieuwe ideeën stuiten we uiteindelijk ook weer op de beperkingen die de hardware ons oplegt. Uit de tabel krijgen we echter wel het gevoel dat er oneindig veel  $x$  met Hallwaarde  $< 1$  zijn.

De Russische wiskundige Danilov baarde in 1980 groot opzien toen hij erin slaagde oneindig veel voorbeelden met Hallwaarde  $< 1$  te construeren. Danilov's constructie is even verrassend als eenvoudig. Hij begint met de identiteit

$$((u-3)^2 - 5)^3 - (u^2 + 1)(u^2 - 9u + 19)^2 = 27(2u - 11).$$

Stel nu dat we een gehele waarde van  $u$  kunnen vinden zó dat  $u^2 + 1 = 125v^2$  voor zekere gehele  $v$ . In het bijzonder betekent dit dat  $u^2 + 1 \equiv 0 \pmod{5}$ . Hieruit volgt dat  $u \equiv \pm 3 \pmod{5}$ . Door het teken van  $u$  geschikt te kiezen kunnen we ervoor zorgen dat  $u \equiv 3 \pmod{5}$ . Met deze waarde van  $u$  volgt nu dat

$$\left(5 \left(\frac{u-3}{5}\right)^2 - 1\right)^3 - v^2(u^2 - 9u + 19)^2 = \frac{27}{125}(2u - 11).$$

Stel nu  $x = 5 \left(\frac{u-3}{5}\right)^2 - 1$  en  $y = v(u^2 - 9u + 19)$ . Dan hebben we hiermee een paar  $x, y$  met Hallwaarde

$$\frac{27|2u - 11|\sqrt{5}}{125\sqrt{(u-3)^2 - 5}}$$

en dit gaat naar  $\frac{54}{25\sqrt{5}} = 0.965\dots < 1$

De vergelijking  $u^2 + 1 = 125v^2$  kan herschreven worden als  $u^2 - 125v^2 = -1$  en staat bekend als de *vergelijking van Pell*. Het is bekend dat deze vergelijking oneindig veel gehele oplossingen heeft. Ze kunnen verkregen worden door de oneven machten van  $682 + 61\sqrt{125}$  uit te werken in de vorm  $u + v\sqrt{125}$ . Bijvoorbeeld, de eerste macht van  $682 + 61\sqrt{125}$  geeft  $u = 682, v = 61$ . We nemen nu  $u = -682$  om ervoor te zorgen dat  $u \equiv 3 \pmod{5}$ . Met deze  $u$  vinden we  $x = 5\left(\frac{u-3}{5}\right)^2 - 1 = 93844$  en  $y = v(u^2 - 9u + 19) = 28748141$ . Merk op dat  $93844^3 - 28748141^2 = -297$ . Deze komt inderdaad in de eerste tabel voor. Uitwerking van de derde macht geeft

$$(682 + 61\sqrt{125})^3 = 1268860318 + 113490317\sqrt{125}.$$

We kiezen  $u = 1268860318$ . De bijbehorende  $x$ -waarde is dan

$$x = 322001299796379844,$$

welke in de tweede tabel voorkomt.

Tenslotte nog een waarde die niet in de tabellen voorkomt.

$$(682 + 61\sqrt{125})^5 = 2360712083917682 + 211148507797805\sqrt{125}$$

We kiezen  $u = -2360712083917682$ . De resulterende  $x$ -waarde is

$$x = 1114592308630995805123571151844.$$

We kunnen ons afvragen of er misschien oneindig veel  $x$  met Hallwaarde  $< 0.5$  of zelfs  $< 0.2$  kunnen voorkomen. Een expliciete constructie van een dergelijke rij  $x$ -waarden zou bijzonder spectaculair zijn!

Bij voorlopig gebrek aan een dergelijke constructie zullen we een experiment uitvoeren waarvan de uitslag suggereert dat er inderdaad zulke rijen zijn. Sterker nog, we zullen argumenten aanvoeren waarom Hall's vermoeden niet juist kan zijn. Anders gezegd, we zullen betogen, maar niet bewijzen, dat er bij elke  $C > 0$  een paar  $x, y \in \mathbb{N}$  bestaat zó dat  $|x^3 - y^2| < C\sqrt{x}$ .

### 3. EEN EXPERIMENT

Kies een getal  $C > 0$  en beschouw de ongelijkheid

$$|y^2 - x^3| < C\sqrt{x} \tag{A}$$

waarin  $x, y$  positief gehele getallen zijn. Als  $x, y$  inderdaad aan een dergelijke ongelijkheid voldoen, dan zal, als  $x > 4C$ , het getal  $y$  het unieke gehele getal zijn dat het dichtst bij  $x^{3/2}$  ligt. We kiezen nu een interval, bijvoorbeeld  $10^6 < x < 10^7$ . Vervolgens bepalen we voor elke  $x$  daarin het gehele getal  $y$  dat het dichtst bij  $x^{3/2}$  ligt en kijken of  $x, y$  voldoen aan (A). We tellen het aantal  $x$ -waarden waarbij dat inderdaad gebeurt. In de volgende tabel vinden we de resultaten voor  $x < 10^{11}$  en  $C = 20, 10, 5, 2, 1$ .

interval	$C = 20$	$C = 10$	$C = 5$	$C = 2$	$C = 1$
$10^3 \leq x < 10^4$	43	16	7	3	2
$10^4 \leq x < 10^5$	37	19	5	2	1
$10^5 \leq x < 10^6$	43	21	12	7	4
$10^6 \leq x < 10^7$	40	18	6	1	0
$10^7 \leq x < 10^8$	40	15	6	2	1
$10^8 \leq x < 10^9$	64	38	20	8	4
$10^9 \leq x < 10^{10}$	55	25	14	3	1

Uit deze tabel wordt het volgende patroon zichtbaar. Voor gegeven  $C > 0$  bestaat er een getal  $n_C$  zó dat de gevonden aantallen oplossingen van (A) in elk interval  $10^n \leq x < 10^{n+1}$  rond deze waarde  $n_C$  fluctueren. Voor  $C = 20$  ligt  $n_{20}$  rond de veertig. In elk interval lijkt zich eenzelfde aantal "hits" af te tekenen. Er zijn natuurlijk fluctuaties, zoals in het interval  $10^8 \leq x < 10^9$  waarin de aantallen ver boven het gemiddelde uitsteken. De vraag is natuurlijk of bij uitbreiding van onze tabel de gevonden gemiddelden aan blijven houden of niet. Met veel moeite zouden we de tabel met nog een paar regels kunnen uitbreiden, maar op een gegeven moment houdt het op. Er zijn grenzen, zelfs voor de snelste hedendaagse computers.

We willen dit gedrag verklaren met behulp van een redelijk klinkende aanname. Maar eerst vertalen we ongelijkheid (A) in een iets andere. Uit deze ongelijkheid volgt namelijk dat  $|y - x^{3/2}|(y + x^{3/2}) < C\sqrt{x}$  en dus  $|y - x^{3/2}| < C\sqrt{x}/(y + x^{3/2})$ . We nemen aan dat  $x$  groot is ten aanzien van  $C$ . In dat geval is namelijk  $y$  vrijwel gelijk aan  $x^{3/2}$  en we vervangen  $y$  in de rechterzijde van onze afchatting gewoon door  $x^{3/2}$ . We vinden,

$$|y - x^{3/2}| < C\sqrt{x}/(y + x^{3/2}) \approx C/2x$$

De fout die we bij de benadering gemaakt hebben, is miniem en van geen enkele consequentie voor ons verhaal. Uit onze ongelijkheid mogen we afleiden dat  $y$  het gehele getal is dat het dichtst bij  $x^{3/2}$  ligt. Het verschil van een getal  $\xi$  met het dichtstbijgelegen gehele getal zullen we aangeven met  $\langle \xi \rangle$ . Bijvoorbeeld  $\langle \pi \rangle = 0.14159\dots$  en  $\langle e \rangle = -0.2817\dots$ . We vinden nu

$$|\langle x^{3/2} \rangle| < C/2x \tag{B}$$

Voor praktische doeleinden en voor grote  $x$  is (B) vrijwel equivalent met ongelijkheid (A).

We formuleren nu onze hypothese. Deze berust erop dat in de rij getallen  $\langle x^{3/2} \rangle$  voor  $x = 1, 2, 3, \dots$  geen enkel herkenbaar patroon zit. Er is geen enkele reden waarom sommige deelintervallen van  $[-1/2, 1/2]$  relatief meer waarden van  $\langle x^{3/2} \rangle$  zouden bevatten dan andere deelintervallen. Bij gebrek aan enig zichtbaar patroon maken we de volgende aanname:

*De rij getallen  $\langle x^{3/2} \rangle$  voor  $x = 1, 2, 3, \dots$  is verdeeld over het interval  $[-1/2, 1/2]$  via een uniforme kansverdeling op dit interval.*

Anders gezegd:

Gegeven een interval  $I \subset [-1/2, 1/2]$ , dan is de kans dat  $\langle x^{3/2} \rangle$  in  $I$  zit gelijk aan de lengte van  $I$ .

Met behulp van deze hypothese kunnen we een verklaring geven van de uitslag van ons experiment. Kies twee grote gehele getallen  $X, Y$  met  $X < Y$ . In onze toepassingen zal  $Y$  een bepaalde factor maal  $X$  zijn, bijvoorbeeld  $Y = 2X$  of  $Y = 10X$ , zoals in ons experiment. We gaan kijken hoeveel getallen  $x$  met  $X \leq x < Y$  volgens verwachting voldoen aan ongelijkheid (B). De kans dat  $x$  aan (B) voldoet is volgens onze hypothese gelijk aan de lengte van het bijbehorende interval,  $C/x$ . De kans dat  $x$  niet aan (B) voldoet is derhalve  $1 - C/x$ .

Met  $P_n$  geven we de kans aan dat precies  $n$  getallen  $x$  met  $X \leq x < Y$  aan (B) voldoen. Om te beginnen de kans dat geen enkele  $x$  voldoet. Deze is gelijk aan

$$P_0 := \prod_{X \leq x < Y} \left(1 - \frac{C}{x}\right).$$

We zullen de waarde van  $P_0$  niet uitrekenen. De kans  $P_1$  is gelijk aan de som over  $y$  van alle kansen dat  $y$  aan (B) voldoet en alle andere waarden niet. Dus

$$P_1 = \sum_{X \leq y < Y} \frac{C}{y} \left(1 - \frac{C}{y}\right)^{-1} \prod_{X \leq x < Y} \left(1 - \frac{C}{x}\right).$$

Omdat we  $X$  zeer groot nemen ten opzichte van  $C$ , kunnen we veronderstellen dat de factor  $\left(1 - \frac{C}{y}\right)^{-1}$  vrijwel gelijk aan 1 is. Hiermee krijgen we de zeer accurate benadering

$$P_1 \approx \sum_{X \leq y < Y} \frac{C}{y} P_0.$$

Uit de appendix van deze tekst blijkt dat we een fout van hooguit  $1/X$  maken als we  $\sum_{X \leq y < Y} \frac{1}{y}$  vervangen door  $\log(Y/X)$ . Dus,

$$P_1 \approx C \log(X/Y) P_0.$$

De kans dat er twee oplossingen voor (B) zijn in het interval  $X \leq x < Y$  is gelijk aan

$$P_2 \approx \sum_{X \leq y_1 < y_2 < Y} \frac{C}{y_1} \frac{C}{y_2} P_0.$$

Merk op dat

$$\sum_{X \leq y_1 < y_2 < Y} \frac{1}{y_1 y_2} = \frac{1}{2} \sum_{\substack{X \leq y_1, y_2 < Y \\ y_1 \neq y_2}} \frac{1}{y_1 y_2}.$$

Als we hier  $\sum_{X \leq y < Y} \frac{1}{y^2}$  bij optellen, dan kunnen we de eis  $y_1 \neq y_2$  in de sommatie laten vervallen. Uit de appendix blijkt dat

$$\sum_{X \leq y < Y} \frac{1}{y^2} < \frac{1}{X-1}.$$

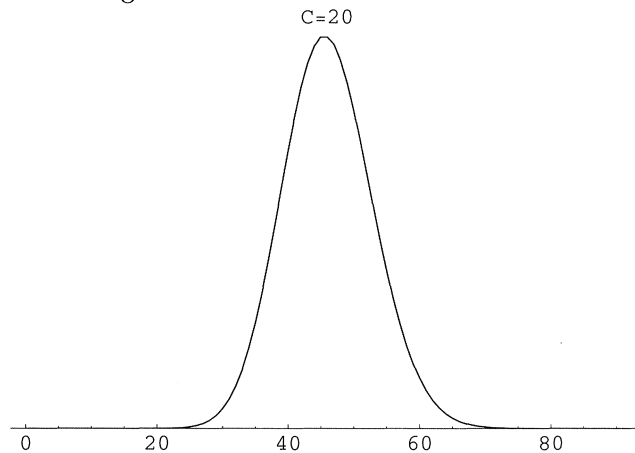
Een zeer klein getal dus. We vinden daarom de accurate benadering

$$\begin{aligned} P_2 &\approx \frac{1}{2} \sum_{X \leq y_1, y_2 < Y} \frac{C}{y_1} \frac{C}{y_2} P_0 \\ &\approx \frac{1}{2} \left( \sum_{X \leq y < Y} \frac{C}{y} \right)^2 P_0 \\ &\approx \frac{1}{2} (C \log(Y/X))^2 P_0 \end{aligned}$$

Op analoge wijze vinden we dat

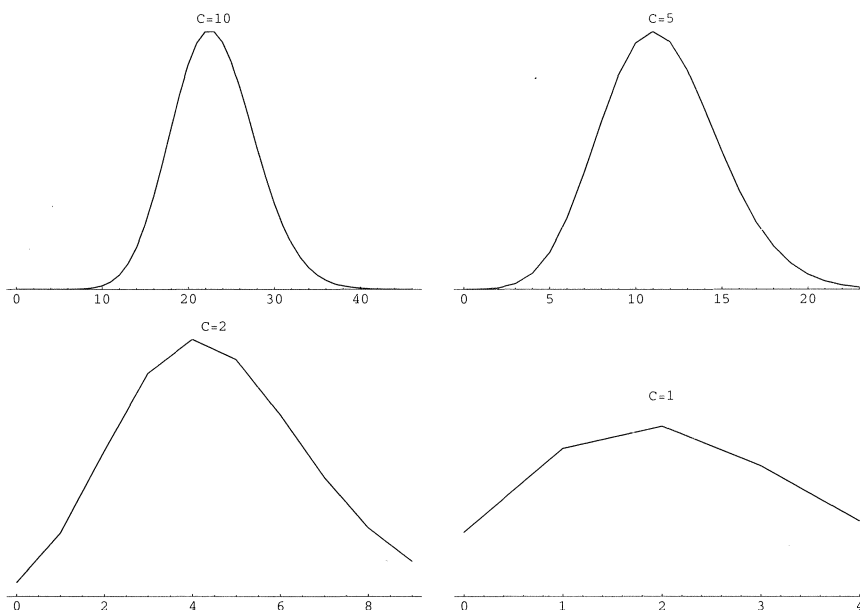
$$P_n \approx \frac{1}{n!} (C \log(Y/X))^n P_0.$$

Met andere woorden, de aantallen oplossingen van (B) zijn verdeeld volgens een Poisson verdeling met verwachting  $C \log(Y/X)$ . We willen hier niet ingaan op de details van het begrip Poissonverdeling. Daarvoor refereren we naar de voordracht van professor Tijms, of naar zijn boek *Spelen met Kansen* in de Epsilonreeks. Wel geven we hier een aantal plots van de grafiek van  $P_n$ . In ons experiment hadden we  $Y = 10X$  genomen. De grafiek van  $\frac{1}{n!} (C \log 10)^n$  met  $C = 20$  ziet er als volgt uit.



De waarden die we in ons experiment vonden, passen goed in dit plaatje. Enige uitzondering is misschien de waarde 64 die we in het interval  $10^8 \leq x < 10^9$  vonden. Deze valt nog net in het bereik waar de kans significant positief is. Hieronder volgen de grafieken van de kansverdelingen bij de overige waarden van  $C$ .





Het blijkt dat onze experimentele resultaten redelijk binnen het bereik van bovenstaande grafieken vallen. De spreiding van de grafieken wordt groter naarmate de waarde van  $C$  kleiner wordt. Dit hangt samen met het feit dat de standaarddeviatie van een Poissonverdeling met gemiddelde  $\Lambda$  gelijk is aan  $\sqrt{\Lambda}$ . De verhouding standaarddeviatie/gemiddelde is dus  $1/\sqrt{\Lambda}$ , een dalende functie in  $\Lambda$ .

#### 4. CONCLUSIES

Het feit dat de voorspellingen in de vorige paragraaf in redelijke overeenstemming zijn met ons experiment, suggereert dat de voorspellingen ook opgaan voor andere waarden van  $C, X, Y$ . Neem bijvoorbeeld  $C = 0.1$  en als interval  $X, Y = 10^{30}X$ . Dan geldt  $C \log(Y/X) = 6.90\dots$  en standaarddeviatie  $2.6\dots$ . Dit suggereert dat we in elk interval van de vorm  $X, 10^{30}X$  een zestal oplossingen van (B), en dus ook (A), mogen verwachten. Nu is de waarde  $C = 0.1$  willekeurige gekozen en hadden we ook iedere andere positieve waarde kunnen nemen. Onze conclusie luidt dus dat we verwachten dat het vermoeden van Hall niet waar is. Iets anders gezegd, we verwachten dat er een oneindige rij paren  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), \dots$  bestaat waarvan de bijbehorende Hall-waarde naar nul gaat als  $n \rightarrow \infty$ .

Tegenwoordig is men voorzichtiger en formuleert het Hallvermoeden als volgt.

**Vermoeden 4.1 (Gewijzigd Hall-vermoeden).** *Bij elke  $\epsilon > 0$  is er een  $C(\epsilon) > 0$  zó dat*

$$|y^2 - x^3| > C(\epsilon)x^{1/2-\epsilon}$$

*voor elk tweetal  $x, y \in \mathbb{N}$  met  $y^2 \neq x^3$ .*

We willen dit verhaal besluiten met een kleine waarschuwing die illustreert dat de formulering van aannemelijke hypothesen binnen de getaltheorie een uiterst precare bezigheid is. Vaak laat men op het Hall-vermoeden een heuristische redenatie toe die maakt gebruik van het feit dat in de buurt van een groot geheel getal  $X$  het verschil van twee opeenvolgende kwadraten ongeveer  $2\sqrt{X}$  is. Stel namelijk  $n^2 \leq X < (n+1)^2$ . Dan is  $n$  ongeveer  $\sqrt{X}$  en het verschil tussen  $n^2$  en  $(n+1)^2$  is  $2n+1 \approx 2\sqrt{X}$ . Kies  $x$  willekeurig. In de buurt van  $x^3$  hebben kwadraten een onderling verschil van ongeveer  $2x^{3/2}$ . De kans dat het interval  $[x^3 - C\sqrt{x}, x^3 + C\sqrt{x}]$  een kwadraat bevat is dus de lengte  $2C\sqrt{x}$  van het interval gedeeld door  $2x^{3/2}$ . Met andere woorden,  $C/x$ . De verwachting van het aantal  $x$  tussen twee grenzen  $X$  en  $Y$  waarvan de Hallwaarde kleiner dan  $C$  is, is dus  $\sum_{x=X}^Y \frac{C}{x} \approx C \log(Y/X)$ . Dit komt goed overeen met onze beschouwingen in de voorgaande paragrafen, en is dus bemoedigend.

Het blijkt echter dat in veel gevallen soortgelijke redeneringen volkomen foutieve antwoorden geven. Bijvoorbeeld, de vergelijking  $x^3 + y^3 = z^3$  met  $\text{ggd}(x, y, z) = 1$  heeft alleen maar de triviale oplossingen met  $xyz = 0$ , terwijl de vergelijking  $x^3 + y^3 + 1 = z^3$  er oneindig veel heeft. Kies bijvoorbeeld  $x = 9t^4 + 3t, y = 9t^3 - 1, z = 9t^4$  met  $t$  willekeurig. Vanuit statistisch oogpunt zouden we echter verwachten dat de functies  $\sqrt[3]{x^3 + y^3}$  en  $\sqrt[3]{x^3 + y^3 + 1}$  elkaar niet zoveel ontlopen. Dit geeft al een indicatie dat als er zoiets als zinvolle experimentele getaltheorie bestaat, dan zullen we zeer voorzichtig moeten zijn in het formuleren van goede uitgangspunten.

## 5. DIVERSEN

Behalve het Hall-vermoeden bestaan er nog vele andere gebieden waarbij men met de computer probeert het antwoord te zoeken. Een bekend voorbeeld is dat van generalisaties van het vermoeden van Fermat. Gegeven een getal  $n > 2$ , dan weten we sinds 1994 dat de vergelijking van Fermat:  $x^n + y^n = z^n$  geen oplossing in positieve gehele getallen  $x, y, z$  heeft. Generalisaties van de Fermatvergelijking zijn er ook altijd geweest. Bijvoorbeeld  $x^4 + y^4 + z^4 = u^4$ . Sinds de 18e eeuw heeft men zich afgevraagd of er behalve de voor de hand liggende oplossing  $0^4 + 0^4 + 0^4 + 1^4 = 1^4$  nog andere oplossingen zijn. Ook in het computertijdperk is door veel mensen naar zo'n oplossing gezocht. Totdat Noam Elkies in 1988 ontdekte dat er oneindig veel oplossingen zijn, met als "kleinste":

$$95800^4 + 217519^4 + 414560^4 = 422481^4.$$

Deze oplossingen werden op puur theoretische gronden ontdekt, er is dus geen computer aan te pas gekomen. De enig bekende methode om via de computer oplossingen van  $x^4 + y^4 + z^4 = u^4$  te bepalen is alle mogelijke drietallen  $x, y, z$  met  $x \leq y \leq z$  aflopen en testen of de som van hun vierde machten weer een vierde macht is. Nemen we als bovengrens  $x, y, z < 500.000$  dan moeten we  $\frac{1}{3!}(500.000)^3 \approx 21 \times 10^{15}$  mogelijkheden testen. Veel te veel voor zelfs de snelste PC's. Hoewel er voor 1988 vele pogingen zijn gedaan om oplossingen voor het vierde graads probleem te vinden, is het nu duidelijk waarom dit niet gelukt is.

Dit betekent niet dat computers nutteloos zijn bij het vinden van oplos-

singen. Er is namelijk een generalisatie van Fermat's vergelijking waarvoor in 1966 door Lander en Parkin wel een oplossing per computer werd gevonden, namelijk:

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5.$$

Ook hier geldt weer dat naïeve zoekacties al gauw stranden op het feit dat looptijden voor het computerprogramma uit de hand lopen. Ondanks de vele pogingen daartoe zijn er nog steeds geen andere oplossingen dan bovenstaande voor het vijfde graads geval gevonden. Overigens zijn er wel oneindig veel gevallen van vijf vijfde machten waarvan de som een vijfde macht is. Deze worden bijvoorbeeld gegeven door de identiteit

$$\begin{aligned} & (75y^5 - x^5)^5 + (x^5 + 25y^5)^5 + (x^5 - 25y^5)^5 + (10x^3y^2)^5 + (50xy^4)^5 \\ & = (x^5 + 75y^5)^5. \end{aligned}$$

Op de website <http://euler.free.fr> wordt een verzameling aangelegd van gevallen waarin sommen van  $k$ -de machten gelijk zijn aan sommen van andere  $k$ -de machten. Iedereen die zin en klokcykels op de PC over heeft, kan zich aanmelden en meedoen met de zoektocht naar deze rariteiten. Een andere beroemd internetproject is het GIMPS-project, waarin gezocht wordt naar de grootste bekende priemgetallen. De laatste vier records zijn afkomstig uit dit project. Zie <http://www.mersenne.org>, waar ook vele links staan naar andere projecten waar volop aan gerekend wordt.

## 6. APPENDIX A

Oplossingen van de Mordell-vergelijking  $y^2 = x^3 + k$  voor alle  $k$  met  $0 < |k| \leq 100$ . Als een waarde  $k$  niet in de tabel voorkomt dan betekent dit dat er geen oplossingen zijn. We schrijven alleen de oplossingen met  $y \geq 0$  op en beginnen met  $k > 0$ .

$k$	Oplossingen	$k$	Oplossingen
1	(-1, 0), (2, 3), (0, 1)	44	(-2, 6), (5, 13)
2	(-1, 1)	48	(1, 7)
3	(1, 2)	49	(0, 7)
4	(0, 2)	50	(-1, 7)
5	(-1, 2)	52	(-3, 5)
8	(-2, 0), (1, 3), (2, 4), (46, 312)	54	(3, 9)
9	(0, 3), (-2, 1), (3, 6), (6, 15), (40, 253)	55	(9, 28)
10	(-1, 3)	56	(2, 8)
12	(-2, 2), (13, 47)	57	(-2, 7), (7, 20), (4, 11)
15	(1, 4), (109, 1138)	63	(-3, 6), (1, 8)
16	(0, 4)	64	(-4, 0), (8, 24), (0, 8)
17	(-2, 3), (-1, 4), (2, 5), (8, 23), (4, 9), (43, 282), (52, 375), (5234, 378661)	65	(-4, 1), (-1, 8), (14, 53), (584, 14113)
18	(7, 19)	68	(-4, 2), (152, 1874)
19	(5, 12)	71	(5, 14)
22	(3, 7)	72	(-2, 8)
24	(-2, 4), (1, 5), (10, 32), (8158, 736844)	73	(-4, 3), (2, 9), (3, 10), (6, 17), (72, 611), (356, 6717)
25	(0, 5)	76	(-3, 7)
26	(-1, 5)	79	(45, 302)
27	(-3, 0)	80	(-4, 4), (1, 9), (4, 12), (44, 292)
28	(-3, 1), (2, 6)	81	(0, 9)
30	(19, 83)	82	(-1, 9)
31	(-3, 2)	89	(-4, 5), (-2, 9), (10, 33), (55, 408)
33	(-2, 5)	91	(-3, 8)
35	(1, 6)	92	(2, 10)
36	(-3, 3), (0, 6), (4, 10), (12, 42)	94	(3, 11)
37	(-1, 6), (3, 8), (243, 3788)	97	(18, 77)
38	(11, 37)	98	(7, 21)
40	(6, 16)	99	(1, 10)
41	(2, 7)	100	(-4, 6), (0, 10), (5, 15), (20, 90), (24, 118), (2660, 137190)
43	(-3, 4)		

In de volgende tabel staan de oplossingen van  $y^2 = x^3 + k, y \geq 0$  met  $k < 0$ .

$k$	Oplossingen	$k$	Oplossingen
-1	(1, 0)	-47	(6, 13), (12, 41), (63, 500)
-2	(3, 5)	-48	(4, 4), (28, 148)
-4	(2, 2), (5, 11)	-49	(65, 524)
-7	(2, 10), (32, 181)	-53	(9, 26), (29, 156)
-8	(2, 0)	-54	(7, 17)
-11	(3, 4), (15, 58)	-55	(4, 3), (56, 419)
-13	(17, 70)	-56	(18, 76)
-15	(4, 7)	-60	(4, 2), (136, 1586)
-18	(3, 3)	-61	(5, 8)
-19	(7, 18)	-63	(4, 1), (568, 13537)
-20	(6, 14)	-64	(4, 0)
-23	(3, 2)	-67	(23, 110)
-25	(5, 10)	-71	(8, 21)
-26	(3, 10), (35, 207)	-72	(6, 12)
-27	(3, 0)	-74	(99, 985)
-28	(4, 6), (8, 22), (37, 225)	-76	(5, 7), (101, 1015)
-35	(11, 36)	-79	(20, 89)
-39	(4, 5), (10, 31), (22, 103)	-81	(13, 46)
-40	(14, 52)	-83	(27, 140)
-44	(5, 9)	-87	(7, 16)
-45	(21, 96)	-89	(5, 6)
		-95	(6, 11)
		-100	(5, 5), (10, 30), (34, 198)

## 7. APPENDIX B

In deze appendix geven we afschattingen voor de sommaties

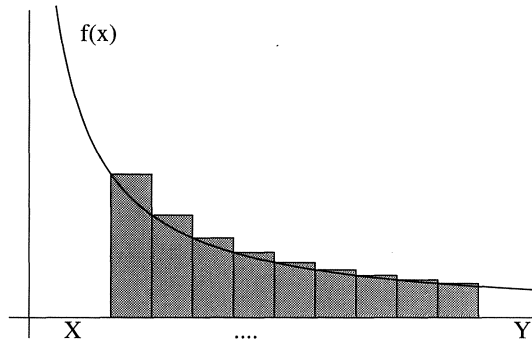
$$S = \sum_{X \leq n < Y} \frac{1}{n^k}$$

voor  $k = 1, 2, 3, \dots$ . We gebruiken hiervoor het zogenaamde *integraalcriterium* voor de sommatie van reeksen.

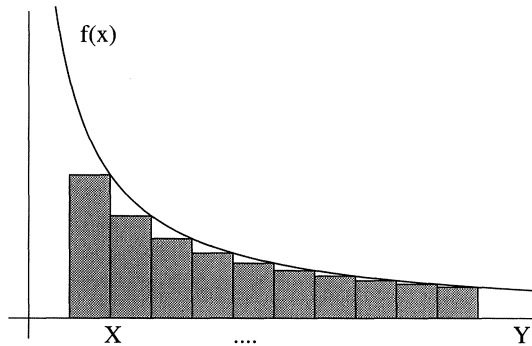
**Stelling 7.1.** *Zij  $f(x)$  een monotoon dalende functie op de positieve reële getallen. Dan geldt voor elk tweetal positief gehele getallen  $X < Y$  dat*

$$\int_X^Y f(x) dx \leq \sum_{X \leq n < Y} f(n) \leq f(X) + \int_X^Y f(x) dx.$$

Het bewijs van deze stelling is het makkelijkst aan de hand van een plaatje uit te leggen. De sommatie  $\sum_{X \leq n < Y} f(n)$  is gelijk aan het oppervlak van de gearceerde figuur in onderstaand plaatje.



Hieruit volgt direct dat onze sommatie groter is dan de integraal  $\int_X^Y f(x)dx$ . De bovengrens volgt uit het volgende plaatje.



Het gearceerde gebied bestaat nu uit de eerste kolom met hoogte  $f(X)$  en verder een gebied dat onder de grafiek van  $f(x)$  ligt. De bovengrens voor onze sommatie volgt nu direct.

Passen we deze Stelling toe op  $f(x) = 1/x$  dan vinden we,

$$\sum_{X \leq n < Y} \frac{1}{n} \geq \int_X^Y \frac{1}{x} dx = \log(Y/X)$$

en

$$\sum_{X \leq n < Y} \frac{1}{n} \leq \frac{1}{X} + \int_X^Y \frac{1}{x} dx = \frac{1}{X} + \log(Y/X).$$

Het integraal criterium toegepast op  $f(x) = 1/x^2$  geeft

$$\sum_{X \leq n < Y} \frac{1}{n} \leq \frac{1}{X^2} + \int_X^Y \frac{1}{x^2} dx = \frac{1}{X^2} + \frac{1}{X} - \frac{1}{Y}.$$

Het laatste getal is weer kleiner dan  $\frac{1}{X^2} + \frac{1}{X}$ .

## Het Begrip Dimensie: Theorie en Experiment

Arjen Doelman  
Universiteit van Amsterdam  
e-mail: [doelman@science.uva.nl](mailto:doelman@science.uva.nl)

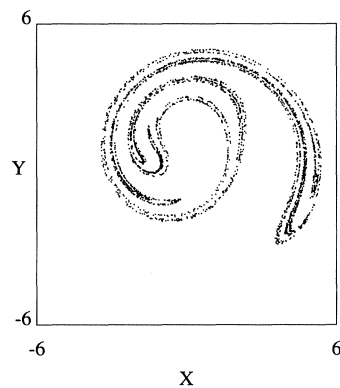
### 1. INLEIDING

Wij leven in een 3-dimensionale wereld (voor zover wij weten). Wat betekent dat eigenlijk? In een 3-dimensionale ruimte zijn er 3 onafhankelijke ‘vrijheidsgraden’ waarmee je een positie in die ruimte kan vastleggen. In de wiskunde maak je die 3 vrijheidsgraden vaak expliciet, bijvoorbeeld met Cartesische coördinaten: de positie van een punt leg je vast met 3 coördinaten ten opzichte van de 3 ( $x$ -,  $y$ - en  $z$ -) assen van het Cartesische coördinatensysteem. Op eenzelfde manier heeft een 3-dimensionaal voorwerp ook 3 ‘onafhankelijke richtingen’: een lengte, een breedte en een hoogte. Zo’n voorwerp heeft een volume, dat wordt gemeten in kubieke meters,  $m^3$ , en dat bij eenvoudige, blok-vormige voorwerpen verkregen wordt uit het product van de lengte, breedte en hoogte.

In een 2-dimensionale ruimte, een oppervlak, zijn er maar 2 van die vrijheidsgraden, die in Cartesische coördinaten worden vastgelegd met de  $x$ -as en de  $y$ -as. Op eenzelfde manier heeft een 2-dimensionaal object ook bijvoorbeeld wel een lengte en een breedte, maar geen hoogte. Dit heeft natuurlijk niets te maken met het al dan niet ‘plat’ zijn van het oppervlak: ook op een gekromd oppervlak zijn er volgens onze intuïtie 2, en niet meer dan 2, coördinaten nodig om een positie vast te leggen. Een 2-dimensionaal voorwerp heeft een oppervlak dat wordt gemeten in vierkante meters,  $m^2$ . Voor de meest eenvoudige 2-dimensionale objecten wordt dit bepaald door het product lengte  $\times$  breedte.

Neem je nog zo’n vrijheid weg, dan hou je een 1-dimensionale ruimte, ofwel een kromme, over. Een kromme heeft alleen een lengte, en die wordt gemeten in meters,  $m = m^1$ . Wiskundig kan je ook die laatste vrijheid wegnemen, wat overblijft is een 0-dimensionaal object, een mathematisch punt, dat geen lengte, breedte of hoogte heeft. Het enige wat je met zo’n 0-dimensionaal object kan doen is het tellen, ofwel bepalen uit hoeveel (losse) punten het bestaat.

Tot zover lijkt ons intuïtieve idee van het begrip dimensie goed te werken. Onze intuïtie kan ons echter in de problemen brengen: als één punt dimensie 0 heeft, dan heeft een verzameling bestaande uit twee punten natuurlijk ook



FIGUUR 1. Een zogenaamde 'chaotische attractor' die optreedt in een model voor de eb en vloed-cyclus (zie paragraaf 6). Dit plaatje is verkregen met behulp van een computer-simulatie van het bijbehorende dynamische systeem.

dimensie 0. Hetzelfde geldt voor iedere verzameling van eindig veel punten. Maar zijn lijnen en vlakken niet verzamelingen van oneindig veel punten? Hoe en wanneer maak je dan de stap van 0 naar 1 of naar 2? Moeten de overgangen per se discontinu zijn? Kunnen verzamelingen misschien ook een dimensie hebben die tussen 0 en 1 of 1 en 2 inzit? Wat zijn dat dan voor verzamelingen? Zulke in eerste instantie abstracte vragen spelen ook een belangrijke rol buiten de zuiver wiskundige context. Met de opkomst van de computer is het duidelijk geworden dat zeer veel natuurlijke processen gerelateerd kunnen worden aan verzamelingen waarvan het niet a priori duidelijk is wat de dimensie zou moeten zijn. Deze verzamelingen ontstaan uit computer-experimenten met de dynamische systemen die deze processen wiskundig beschrijven. In dit verhaal nemen we als voorbeeld een representatie van de benadering van een zogenaamde chaotische attractor, zie figuur 1. Deze attractor speelt een rol bij het bestuderen van getijde-systemen en geeft aan hoe de op het eerste gezicht zeer regelmatige eb en vloed-cyclus chaotisch gemoduleerd kan worden (zie paragraaf 6). De computer geeft ons vanzelfsprekend alleen een eindige hoeveelheid punten, maar de attractor bestaat 'natuurlijk' uit oneindig veel punten. De vraag is nu: wat is de dimensie van deze attractor? De puntenwolk heeft op meerdere plaatsen al een duidelijke lijnenstructuur, dus het lijkt erop dat de dimensie (minstens) 1 is. Op sommige plaatsen liggen deze lijnen echter zo dicht bij elkaar dat ze een tweedimensionaal gebiedje lijken te vullen en je dus eerder zou denken dat de dimensie van de attractor 2 moet zijn. Of is het zo dat deze '2-dimensionale gebiedjes' toch weer blijken te bestaan uit vele



aparte, maar zeer dichtbij elkaar liggende lijnen, zoals gebruikelijk is bij ‘fractale objecten’ (zie [4])? Maar als dat nu eens oneindig veel lijnen blijken te zijn?

Wat is eigenlijk de dimensie van deze puntenwolk? Hoe zou je de dimensie van dit wellicht ‘fractale’ object moeten ‘meten’?

We zullen hier een methode presenteren die ons in staat stelt een soort (!) dimensie, ofwel de *limiet-capaciteit*, van verzameling, of een puntenwolk te bepalen, of te meten (vaak met behulp van de computer). Deze methode is in essentie op niets anders dan een nauwkeurige wiskundige (her)formulering van het intuïtieve dimensie begrip gebaseerd (paragraaf 4). We zullen laten zien dat de methode voor mooie ‘gladde’ verzamelingen ook het antwoord geeft waarop we rekenen: een ‘gewone’ lijn heeft gewoon dimensie 1, etc. Echter, als we de methode gebruiken om de dimensie van een fractale verzameling, zoals een Cantor verzameling (paragraaf 2), een Koch kromme (paragraaf 3) of een chaotische attractor (figuur 1, paragraaf 6), te bepalen, dan vinden we uitkomsten die niet per se geheeltalig zijn. Met andere woorden, zulke verzamelingen kunnen een zogenaamde ‘fractale dimensie’ hebben.

Hoewel het wiskundige fundament van de methode op een limiet-procedure berust, kan de methode ook op een eenvoudige manier worden toegepast op ‘experimentele’ en dus eindige ‘data-sets’, zoals de output van een computer-experiment of waarnemingen van natuurlijke structuren. Omdat de verzameling in zo’n geval maar uit eindig veel elementen bestaat, of omdat de meetnauwkeurigheid begrensd is, kan nu de wiskundige limiet niet worden genomen. Echter, ook zonder de limiet te nemen kan de methode een benadering of een indicatie van de dimensie van een data-set geven. Wij zullen bijvoorbeeld in sectie 6 de dimensie van de attractor in figuur 1 schatten en vinden dat deze (ongeveer) gelijk aan 1.6 is.

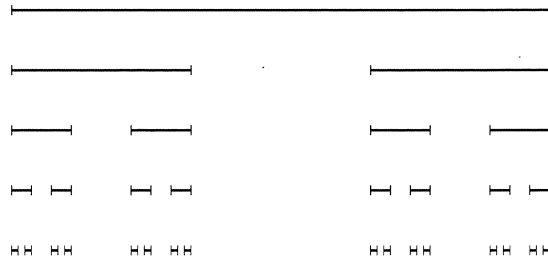
## 2. VERZAMELINGEN EN LIMIET-PROCEDURES

In deze, en de komende paragraaf, geven we voorbeelden van verzamelingen die aan de hand van een limiet-procedure geconstrueerd worden. Mede als gevolg daarvan is het intuïtief niet duidelijk wat de dimensie van deze verzamelingen zou kunnen/moeten zijn. Met andere woorden: de verzamelingen die we in deze en de komende paragraaf construeren tonen aan hoe gebrekkig ons intuïtieve dimensie begrip eigenlijk is.

Allereerst construeren we de (standaard) Cantor-verzameling  $\mathcal{C}$ ;  $\mathcal{C}$  is een deelverzameling van het gesloten interval  $[0, 1]$ . De Cantor-verzameling  $\mathcal{C}$  wordt stapsgewijs opgebouwd. Voor de eerste stap definiëren we:

$$V_0 = [0, 1].$$

Merk op dat de lengte van  $V_0$ ,  $L_0$ , gelijk is aan 1. Vervolgens definiëren we  $V_1$  door  $V_0$  in 3 gelijke stukken te verdelen en het middelste stuk, het interval



FIGUUR 2. De eerste 5 stappen in de constructie van de Cantor-verzameling  $\mathcal{C}$ : van boven naar onderen zien we  $V_0$ ,  $V_1$ ,  $V_2$ ,  $V_3$  en  $V_4$ .

$(\frac{1}{3}, \frac{2}{3})$ , ‘weg te snijden’ (zie figuur 2):

$$V_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1];$$

$V_1$  heeft (totale) lengte  $L_1 = \frac{2}{3}$ . We definiëren  $V_2$  door hetzelfde te doen met elk van de twee componenten van  $V_1$ : deel het in 3 gelijke stukken in en snij het middelste stuk weg. Dus,

$$V_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1],$$

met totale lengte  $L_2 = \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}$ . In het algemeen wordt  $V_{n+1}$  uit  $V_n$  geconstrueerd door ieder component van  $V_n$  in 3 gelijke stukken te delen en de middelste daarvan weg te snijden (figuur 2). Het gevolg hiervan is dat  $L_{n+1} = \frac{2}{3}L_n$ . De Cantor-verzameling  $\mathcal{C}$  is nu gedefinieerd door

$$\mathcal{C} = \lim_{n \rightarrow \infty} V_n.$$

Hoewel  $\mathcal{C}$  is geconstrueerd uit elementen die stuk voor stuk ‘duidelijk’ dimensie 1 hebben, is het nu niet meer zo duidelijk of  $\mathcal{C}$  zelf ook nog wel dimensie 1 heeft: de lengte van  $\mathcal{C}$  is bijvoorbeeld niet gedefinieerd, want  $L_n \rightarrow 0$  als  $n \rightarrow \infty$ , zodat  $\mathcal{C}$  dus wel ‘lengte 0’ moet hebben. Toch is  $\mathcal{C}$  nog een behoorlijk ‘volle’ verzameling, die veel meer punten bevat dan alleen de randpunten  $\frac{m}{3^n}$  van de intervallen  $V_n$  (met  $0 \leq m \leq 3^n$ ). De Cantor-verzameling  $\mathcal{C}$  bevat zelfs *overaftelbaar* veel punten. Dit betekent dat  $\mathcal{C}$  ‘net zoveel’ punten bevat als het interval  $[0, 1]$ , hetgeen natuurlijk merkwaardig is omdat we  $\mathcal{C}$  juist geconstrueerd hebben door oneindig veel stukken uit  $[0, 1]$  weg te snijden (zie Opmerking 1).

Een andere belangrijke eigenschap van  $\mathcal{C}$  is dat de verzameling *zelfgelijkvormig* is, dat wil zeggen dat je door op de goede manier in te zoomen op  $\mathcal{C}$ , steeds weer in essentie dezelfde verzameling  $\mathcal{C}$  terugvindt. Bekijk hiervoor bijvoorbeeld de linkerhelft van  $\mathcal{C}$ , dat wil zeggen de verzameling  $\mathcal{C}_1 = \mathcal{C} \cap [0, \frac{1}{3}] \subset \mathcal{C}$ . Na vermenigvuldiging van ieder element met een factor 3 is deze verzameling



FIGUUR 3. Een benadering van het Cantor-gordijn;  $\alpha$  loopt van boven naar beneden van 0 tot 1. Merk op dat in deze figuur de verzameling  $V_5^\alpha \supset C^\alpha$  is getekend (voor alle  $\alpha \in [0, 1]$ ).

identiek aan  $C$  zelf! Hetzelfde geldt voor iedere doorsnede van  $C$  met een van de componenten van een verzameling  $V_n$ : na verschuiving en vermenigvuldiging met de factor  $3^n$  is deze identiek aan de oorspronkelijke verzameling  $C$ .

Natuurlijk kan je op een eenvoudige manier andere, maar soortgelijke, Cantor-verzamelingen  $C^\alpha$  construeren. Hiervoor kijken we nog eens naar de constructie van  $C$ . Bij iedere stap in de limiet-procedure wordt ieder deelinterval van  $V_n$ , dat dus een lengte heeft van  $3^{-n}$ , in 3 stukken gedeeld. Ieder deel heeft telkens dezelfde lengte, te weten  $3^{-(n+1)}$ . We kunnen dit natuurlijk ook anders doen: we beginnen weer bij  $V_0$  en geven het middelste deelinterval een lengte  $\alpha$ , waarbij  $\alpha$  een willekeurig getal in  $[0, 1]$  is. Vervolgens maken we de twee deelintervallen aan de uiteinden precies even lang, hetgeen betekent dat ze een lengte van  $\frac{1}{2}(1 - \alpha)$  hebben. Net als bij de constructie van  $C$  snijden we nu het middelste interval weg, zodat

$$V_1^\alpha = [0, \frac{1}{2}(1 - \alpha)] \cup [\frac{1}{2}(1 + \alpha), 1]$$

overblijft. Vervolgens snijden we uit beide deelintervallen van  $V_1^\alpha$  het ‘middelste  $\alpha$ -gedeelte’, etc. De *middelste- $\alpha$  Cantor-verzameling*  $C^\alpha$  wordt nu weer verkregen door de limiet te nemen. Net als in het oorspronkelijke geval geldt dat de totale lengte van  $V_n^\alpha$ ,  $L_n^\alpha$ , naar 0 gaat als  $n \rightarrow \infty$ , tenminste, als  $\alpha > 1$ :

$$L_n^\alpha = (1 - \alpha)L_{n-1}^\alpha = (1 - \alpha)^n \rightarrow 0 \text{ als } n \rightarrow \infty,$$

dus, de ‘totale lengte’ van  $C^\alpha$  is, net als in het oorspronkelijke geval, 0 (als  $\alpha > 0$ ). Merk op dat  $C^0 = [0, 1]$ ,  $C^1 = \{0, 1\}$  (ofwel: 2 losse punten) en dat  $C = C^{\frac{1}{3}}$ . In figuur 3 is (een benadering van) het zogenaamde *Cantor-gordijn*

geschetst:  $\alpha$  loopt van 0 tot 1 van boven naar beneden, bovenaan staat  $\mathcal{C}^0$ , onderaan  $\mathcal{C}^1$ . Voor kleine  $\alpha$  liggen de elementen van  $\mathcal{C}^\alpha$  zo'n beetje homogeen verdeeld in  $[0, 1]$  en lijkt  $\mathcal{C}^\alpha$  wel eens een verzameling van dimensie 1 te kunnen zijn (hetgeen het natuurlijk duidelijk is voor  $\alpha = 0$ ). Voor  $\alpha$  dicht bij 1 bestaat  $\mathcal{C}^\alpha$  alleen uit ijle puntenverzamelingen dicht bij de randpunten 0 en 1:  $\mathcal{C}^\alpha$  lijkt nu meer op een verzameling met dimensie 0 (de dimensie van  $\mathcal{C}^1$  is natuurlijk 0).

De verzamelingen  $\mathcal{C}^\alpha$  lijken dus een mooie continue 'verbinding' te geven tussen een overduidelijk 0-dimensionale verzameling ( $\alpha = 1$ ) en een 1-dimensionale verzameling ( $\alpha = 0$ ). Dit maakt het wel erg natuurlijk om het bestaan van verzamelingen met een niet-gehele dimensie te accepteren: indien de dimensie van  $\mathcal{C}^\alpha$  altijd 0 of 1 moet zijn, bij welke  $\alpha \in [0, 1]$  zou dan de overgang van 0 naar 1 moeten zijn?

OPMERKING 1. Om in te zien dat  $\mathcal{C} = \mathcal{C}^{\frac{1}{3}}$  overaftelbaar is, dat wil zeggen, 'evenveel' punten bevat als het interval  $[0, 1]$ , merken we op dat er een grote overeenkomst bestaat tussen de bovenstaande meetkundige constructie van  $\mathcal{C}$  en de 3-tallige representatie van getallen in het interval  $(0, 1)$ . De vraag is dan eerst: hoe bepaal je de 3-tallige representatie van een getal  $p \in [0, 1]$ ? Dat gaat als volgt:

- Kies  $p \in [0, 1]$  en bereken  $q_1 = 3p$ .
- Definieer  $p_1 \in \{0, 1, 2\}$  door  $p_1 \leq q_1 = 3p < p_1 + 1$  ( $p_1$  heet de 'entier' van  $q_1$ ).
- Bereken  $q_2 = 3(q_1 - p_1)$ .
- Definieer  $p_2 \in \{0, 1, 2\}$  door  $p_2 \leq q_2 < p_2 + 1$ .
- Bereken  $q_3 = 3(q_2 - p_2)$ .
- Definieer  $p_3 \in \{0, 1, 2\}$  door  $p_3 \leq q_3 < p_3 + 1$ .
- Etcetera.

Merk op dat nu per constructie geldt dat

$$p = \frac{p_1}{3} + \frac{p_2}{3^2} + \frac{p_3}{3^3} + \dots = \sum_{j=1}^{\infty} \frac{p_j}{3^j}.$$

Het getal  $p \in (0, 1)$  ligt dus eenduidig vast als we het rijtje  $\{p_1, p_2, \dots\}$  – met  $p_j \in \{0, 1, 2\}$  – kennen en kan nu worden gerepresenteerd door  $p =_{(3)} 0.p_1p_2p_3\dots$ .  
Bijvoorbeeld

$$\frac{1}{2} =_{(3)} 0.11111\dots, \quad \frac{1}{3} =_{(3)} 0.1, \quad \frac{1}{4} =_{(3)} 0.020202\dots$$

Echter, verschillende rijtjes  $\{p_1, p_2, \dots\}$  kunnen wel hetzelfde getal  $p \in [0, 1]$  representeren:

$$\begin{aligned} \frac{1}{3} =_{(3)} 0.1 =_{(3)} 0.02222\dots, & \quad \frac{2}{3} =_{(3)} 0.2 =_{(3)} 0.12222\dots, \\ 1 =_{(3)} 1.0 =_{(3)} 0.22222\dots, & \quad \text{etc.} \end{aligned}$$

Dit fenomeen kennen we natuurlijk al van de gewone 10-tallige representatie van getallen:  $1 =_{(10)} 0.9999\dots$  etc. We kunnen nu terugkeren naar de constructie van de Cantor-verzameling  $\mathcal{C}$  en constateren dat

$$p \in V_1 \Leftrightarrow p_1 \neq 1$$

met andere woorden:  $p \in V_1$  impliceert dat  $p_1 = 0$  of  $2$ , en  $p_1 = 0$  of  $2$  betekent dat  $p \in V_1$ . Analooq zien we dat

$$p \in V_n \Leftrightarrow p_1, p_2, \dots, p_n \neq 1 \text{ ofwel } p_1, p_2, \dots, p_n \in \{0, 2\}.$$

Hieruit volgt dat  $\mathcal{C}$  als volgt kan worden gekarakteriseerd:

$$\mathcal{C} = \{p =_{(3)} 0.p_1p_2\dots, \text{ met } p_j \in \{0, 2\}\}.$$

Merk op dat we op deze manier veel ‘concrete elementen’ van  $\mathcal{C}$  – die geen randpunt van een van de basis-intervallen  $V_m$  zijn – kunnen vinden, bijvoorbeeld:  $\frac{1}{4} =_{(3)} 0.020202\dots \in \mathcal{C}$ . Aan de hand van de zojuist geconstateerde niet-uniciteit van de representatie zien we tevens dat deze randpunten  $\frac{m}{3^n}$ , die natuurlijk ook allemaal elementen van  $\mathcal{C}$  zijn, ook op deze manier gekarakteriseerd kunnen worden. Als we ons nu realiseren dat ten opzichte van het 2-tallig stelsel het interval  $[0, 1]$  als volgt kan worden gerepresenteerd:

$$[0, 1] = \{p =_{(2)} 0.p_1p_2\dots, \text{ met } p_j \in \{0, 1\}\},$$

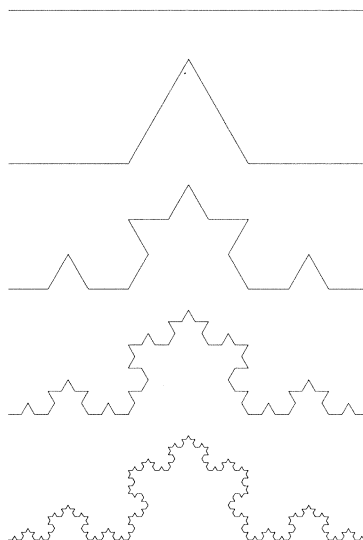
dan zien we hoe we aan ieder punt van  $\mathcal{C}$  een uniek punt van  $[0, 1]$  kunnen koppelen (en omgekeerd): verander elke ‘2’ in de 3-tallige representatie van een element  $p$  van  $\mathcal{C}$  in een ‘1’ (of andersom). Op deze manier wordt bijvoorbeeld  $\frac{1}{3} \in [0, 1]$  gekoppeld aan  $\frac{1}{4} \in \mathcal{C}$ . Op deze manier heeft  $\mathcal{C}$  dus ‘precies evenveel’ elementen als  $[0, 1]$ !

OPMERKING 2. Op eenzelfde manier als in Opmerking 1 kan worden aangetoond dat alle in deze paragraaf geconstrueerde Cantor-verzamelingen  $\mathcal{C}^\alpha$  – natuurlijk met  $0 \leq \alpha < 1$  – overaftelbaar zijn, dat wil zeggen in essentie net zoveel punten hebben als het interval  $[0, 1]$ .

### 3. EEN KROMME MET ONBEGRENSDE LENGTE

De Cantor-verzamelingen uit de vorige paragraaf kunnen natuurlijk nooit meer dan 1-dimensionaal zijn, ze zijn tenslotte geconstrueerd als deelverzamelingen van  $[0, 1]$ . In deze paragraaf construeren we, wederom met behulp van een limiet-procedure, een ‘kromme’ waarvan de lengte ‘oneindig’ moet zijn. Dit zou kunnen betekenen dat deze kromme een dimensie heeft die groter is dan 1.

We starten weer met het (gesloten) eenheids-interval,  $W_0 = [0, 1]$ , delen het weer in 3 gelijke stukken op en snijden het middelste interval,  $(\frac{1}{3}, \frac{2}{3})$ , weg. Echter, we vervangen dit interval nu door een tent- of driehoek-constructie, bestaande uit 2 lijnstukken met dezelfde lengte als het oorspronkelijke stuk,



FIGUUR 4. De eerste 5 stappen (de verzamelingen  $W_0, W_1, W_2, W_3, W_4$ ) in de constructie van de Koch kromme  $\mathcal{K}$ .

$\frac{1}{3}$  (zie figuur 4). Merk op dat de zo geconstrueerde ‘tent’ met de hoekpunten  $(\frac{1}{3}, 0)$ ,  $T_1^1 = (\frac{1}{2}, \frac{1}{6}\sqrt{3})$ ,  $(\frac{2}{3}, 0)$  een gelijkzijdige driehoek is.

De verzameling  $W_1$  bestaat nu dus uit 4 lijnstukken met lengte  $\frac{1}{3}$ : de verbindingen tussen de punten  $(0, 0)$  en  $(\frac{1}{3}, 0)$ ,  $(\frac{1}{3}, 0)$  en  $T_1^1$ ,  $T_1^1$  en  $(\frac{2}{3}, 0)$ , en tenslotte  $(\frac{2}{3}, 0)$  en  $(1, 0)$ . Dit betekent dat de lengte  $M_1$  van  $W_1$  gelijk is aan  $\frac{4}{3}$ .

Vervolgens passen we dezelfde ‘tent-constructie procedure’ toe op ieder van de 4 deel-lijnstukken van  $W_1$ . Hieruit ontstaat de verzameling  $W_2$ , die uit 16 lijnstukjes met lengte  $\frac{1}{9}$  bestaat; deze lijnstukjes verbinden 17 punten met elkaar, waaronder de 4 nieuwe ‘tent-punten’

$$T_2^1 = (\frac{1}{6}, \frac{1}{18}\sqrt{3}), T_2^2 = (\frac{1}{3}, \frac{1}{9}\sqrt{3}), T_2^3 = (\frac{2}{3}, \frac{1}{9}\sqrt{3}), T_2^4 = (\frac{5}{6}, \frac{1}{18}\sqrt{3}).$$

De totale lengte  $M_2$  van  $W_2$  is dus  $\frac{16}{9} = \frac{4}{3} \times \frac{4}{3}$ . In de volgende stap wordt natuurlijk op ieder van de 16 lijnstukjes weer een gelijkzijdig tentje ‘opgezet’, zodat  $W_3$  uit 64 lijnstukjes van lengte  $\frac{1}{27}$  bestaat:  $M_3 = 64 \times \frac{1}{27} = (\frac{4}{3})^3$ . Etcetera (zie figuur 4).

De *Koch-kromme*  $\mathcal{K}$  ontstaat nu door het nemen van de limiet  $n \rightarrow \infty$  in  $W_n$ . Vanzelfsprekend geldt dat

$$\text{lengte}(\mathcal{K}) = \lim_{n \rightarrow \infty} \text{lengte}(W_n) = \lim_{n \rightarrow \infty} (\frac{4}{3})^n = \infty.$$

Echter, ondanks het feit dat  $\mathcal{K}$  ‘oneindig lang’ is, is  $\mathcal{K}$  wel keurig begrensd:  $\mathcal{K}$  past bijvoorbeeld in z’n geheel in de rechthoek met hoekpunten  $(0, 0)$ ,  $(0, \frac{1}{6}\sqrt{3})$ ,  $(1, \frac{1}{6}\sqrt{3})$ ,  $(1, 0)$ . Maar als  $\mathcal{K}$  begrensd is en toch oneindig lang is, is  $\mathcal{K}$  dan wel een kromme, dat wil zeggen, is  $\mathcal{K}$  dan wel 1-dimensionaal? Intuïtief was het hebben van een (eindige) lengte toch de bepalende eigenschap van 1-dimensionale objecten. De ‘kromme’  $\mathcal{K}$  lijkt zeker ook niet 2-dimensionaal te zijn, al was het maar omdat de Koch kromme ‘duidelijk’ geen oppervlak lijkt te hebben.

Net als de Cantor verzamelingen uit de vorige paragraaf is ook  $\mathcal{K}$  weer zelfgelykvormig. Ieder stukje van  $\mathcal{K}$  dat opgebouwd wordt uit een van de deelstukjes van een  $W_n$  kan door middel van schuiven, draaien en ‘opblazen’ geïdentificeerd worden met de oorspronkelijke ‘hele’  $\mathcal{K}$ .

OPMERKING 3. Natuurlijk kan je ook hier weer algemenere structuren dan  $\mathcal{K}$  definiëren, net als in de vorige paragraaf waarin  $C^\alpha$  de Cantor verzameling  $\mathcal{C}$  generaliseerde. In het geval van een gegeneraliseerde Koch kromme ontstaat er wel het ‘probleem’ – dat helemaal geen probleem hoeft te zijn – dat deze ‘kromme’ zichzelf gaat doorsnijden.

#### 4. DE LIMIET-CAPACITEIT VAN EEN VERZAMELING

In deze paragraaf, en in de rest van dit verhaal, beperken we ons tot deelverzamelingen van het (platte) vlak. De verzamelingen die we beschouwen kunnen dus maximaal dimensie 2 hebben. Alles wat we hier bespreken kan worden gegeneraliseerd naar  $N$  dimensies.

Het basis-idee achter het begrip ‘*limiet-capaciteit*’ is dat de dimensie van een begrensde verzameling  $U \subset \mathbf{R}^2$  bepaald hoeveel ‘kleine vierkantjes’ er nodig zijn om  $U$  te overdekken. Hierbij bedoelen we met ‘overdekken’ dat ieder element van  $U$  in minstens een van de overdekkende vierkantjes zit.

Het lijnstuk  $[0, 1]$  kan natuurlijk (precies) worden overdekt met 2 vierkantjes  $S(\frac{1}{2})$ , dat zijn vierkantjes waarvan beide zijden lengte  $\frac{1}{2}$  hebben. In het algemeen definiëren we  $S(\varepsilon)$  als een vierkantje met zijden van lengte  $\varepsilon$ . Er zijn 4 vierkantjes  $S(\frac{1}{4})$  en, in het algemeen,  $n = n^1$  vierkantjes  $S(\frac{1}{n})$  nodig om  $[0, 1]$  te bedekken. Natuurlijk zijn er  $n^2$  vierkantjes  $S(\frac{1}{n})$  nodig om het ‘grote’ vierkant  $S(1)$  te overdekken. Aan de andere kant is er altijd maar  $1 = n^0$   $S(\frac{1}{n})$  nodig om het punt  $(0, 0)$  te overdekken.

Als we nu  $\varepsilon$  schrijven in plaats van  $\frac{1}{n}$  en  $\mathcal{N}(\varepsilon) = \mathcal{N}(\varepsilon; U)$  introduceren als het getal dat aangeeft hoeveel vierkantjes  $S(\varepsilon)$  er nodig zijn om een gegeven verzameling  $U \subset \mathbf{R}^2$  te overdekken, dan kunnen we concluderen dat  $\mathcal{N}(\varepsilon) = \varepsilon^{-2}$  voor het duidelijk 2-dimensionale vierkant  $S(1)$ , dat  $\mathcal{N}(\varepsilon) = \varepsilon^{-1}$  voor het 1-dimensionale interval  $[0, 1]$  en dat  $\mathcal{N}(\varepsilon) = 1 = \varepsilon^0$  voor een 0-dimensionaal punt. We zullen verderop zien dat er nog een aantal haken en ogen aan deze conclusies zitten, maar, het lijkt duidelijk dat de dimensie van een verzame-

ling  $U$  iets te maken heeft met de (absolute waarde van de) exponent van  $\varepsilon$  in de uitdrukking voor  $\mathcal{N}(\varepsilon; U)$ . Nu hebben de zojuist gegeven drie voorbeelden natuurlijk wel een zeer simpele, en bovendien regelmatige, structuur. Als gevolg daarvan zijn deze voor relatief grote waarden van  $\varepsilon$  al zeer ‘netjes’ te overdekken. Dit geldt natuurlijk niet voor alle waarden van  $\varepsilon$ :  $S(1)$  wordt bijvoorbeeld ‘sluitend’ overdekt met 4  $S(\frac{1}{2})$ ’s, maar voor waarden van  $\varepsilon$  die ook maar iets kleiner zijn dan  $\frac{1}{2}$  geldt meteen dat  $\mathcal{N}(\varepsilon; S(1)) \geq 7$  (maar niet automatisch 9, zie hieronder), met als gevolg dat er een aantal vierkantjes  $S(\varepsilon)$  zullen zijn die flink ‘uitsteken’ – we komen hier verderop nog uitgebreid op terug. Met name voor verzamelingen met een wat meer grillig karakter zal gelden dat de overdekking passender/nauwkeuriger wordt naarmate de waarde van  $\varepsilon$  afneemt. Gebaseerd op deze constatering definieren we de *limiet-capaciteit*  $\text{limcap}(U)$  van een verzameling  $U \subset \mathbf{R}^2$  als volgt:

$$\text{limcap}(U) = - \lim_{\varepsilon \rightarrow 0} \frac{\log \mathcal{N}(\varepsilon; U)}{\log \varepsilon}. \quad (1)$$

Merk op dat in ieder geval geldt dat  $0 \leq \text{limcap}(U) \leq 2$ , voor iedere (begrensd)  $U \subset \mathbf{R}^2$  (omdat zeker geldt dat  $0 \leq \mathcal{N}(\varepsilon; U) \leq C\varepsilon^{-2}$ , voor zekere positieve  $C$ ). Opvallend is dat het niet uitmaakt welk grondtal we gebruiken voor de logaritmen in (1), de extra omrekeningsfactor (bijvoorbeeld van  $e$  naar 10) treedt in teller en noemer op en valt dus weg. Omdat  $\mathcal{N}(\varepsilon; S(1)) = \varepsilon^{-2}$  geldt nu natuurlijk dat

$$\text{limcap}(S(1)) = - \lim_{\varepsilon \rightarrow 0} \frac{\log \varepsilon^{-2}}{\log \varepsilon} = 2 = \dim(S(1))$$

en, natuurlijk,

$$\text{limcap}([0, 1]) = 1 = \dim([0, 1]) \quad \text{en} \quad \text{limcap}(P = (p_1, p_2)) = 0 = \dim(P).$$

Dus, voor dit soort eenvoudige verzamelingen lijkt de formule (1) een correcte, berekenbare uitdrukking voor de dimensie van een verzameling te geven.

Echter, tot nu toe hebben we de situatie iets eenvoudiger voorgesteld dan hij in werkelijkheid is. We moeten eigenlijk het getal  $\mathcal{N}(\varepsilon; U)$  precieser definiëren: het is het *minimale* aantal vierkantjes  $S(\varepsilon)$  waarmee de verzameling  $U \subset \mathbf{R}^2$  in z’n geheel overdekt kan worden. Deze voorwaarde dat  $\mathcal{N}(\varepsilon)$  minimaal is kan a priori aanleiding geven tot een flink gepuzzel.

Laten we daarom nog eens terugkeren naar het vierkant  $S(1)$ . Zoals we al opmerkten geldt namelijk in het algemeen niet dat  $\mathcal{N}(\varepsilon; S(1)) = \varepsilon^{-2}$ , dat is alleen het geval voor  $\varepsilon = \frac{1}{n}$  voor zekere  $n \in \mathbf{N}$ . Als we  $\varepsilon$  gelijk aan  $\frac{1}{2} - \sigma$  kiezen, en aannemen dat  $\sigma$  ‘klein genoeg’ is dan zien we meteen dat  $4 < \mathcal{N}(\varepsilon) \leq 9$ . Echter,  $\mathcal{N}(\varepsilon)$  is kleiner dan 9. Om dit in te zien maken we eerst van 4 vierkantjes  $S(\varepsilon)$  te samen het vierkant  $S(2\varepsilon)$ , en geven dat de hoekpunten  $(0, 0)$ ,  $(0, 2\varepsilon)$ ,  $(2\varepsilon, 2\varepsilon)$ ,  $(2\varepsilon, 0)$ . Als we deze  $S(2\varepsilon)$  over  $S(1)$  met de hoekpunten  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 1)$ ,  $(1, 0)$  leggen dan blijft er nog een strookje ter breedte van  $\sigma$



over langs de zijdes van  $S(1)$  die de punten  $(1,0)$ ,  $(1,1)$  en  $(0,1)$  met elkaar verbinden. De ‘zuinigste’ methode om dit (smalle) strookje te overdekken is om 3 vierkantjes  $S(\varepsilon)$  te nemen: eentje leg je met het middelpunt op  $(1,1)$  zo dat de diagonalen samenvallen met de twee onbedekte zijdes van  $S(1)$ , de andere 2 kunnen nu de resterende overgebleven stukjes van de nog open strookjes overdekken door ook deze een kwartslag te draaien (en dus de diagonalen weer parallel aan de zijdes van  $S(1)$  te kiezen). Dus, we hebben 7  $S(\varepsilon)$ ’s nodig gehad om  $S(1)$  te overdekken. Merk op dat dit alleen kan voor  $\varepsilon$  behoorlijk dicht bij  $\frac{1}{2}$ : zodra  $\varepsilon$  te klein wordt (of  $\sigma$  te groot) zal het niet meer lukken om het overgebleven strookje op deze manier te overdekken (het volgt vrij eenvoudig dat er zeker ruimte zal overblijven als  $\varepsilon \leq \frac{1}{3}\sqrt{2} \approx 0.471\dots$ ). Verder hebben we nog niet bewezen dat er nu ook werkelijk geldt dat  $\mathcal{N}(\varepsilon) = 7$  voor  $\varepsilon$  dicht genoeg bij  $\frac{1}{2}$ : wie weet lukt het toch met 6 vierkantjes  $S(\varepsilon)$  als we de overdekking geheel anders aanpakken.

Nu willen we een ‘werkbaar’ – en voor een computer berekenbaar – dimensie begrip hanteren, en dan willen we bovenstaand combinatorisch gepuzzel natuurlijk vermijden. En dat kan ook met het in (1) geïntroduceerde begrip limietcapaciteit. We zijn namelijk niet in de exacte waarde van  $\mathcal{N}(\varepsilon)$  geïnteresseerd, maar alleen in het (exponentiële) groeigedrag van  $\mathcal{N}(\varepsilon)$  als functie van  $\varepsilon$  als  $\varepsilon \rightarrow 0$ . We kijken daarom nog eens naar  $S(1)$ . Als  $\varepsilon$  voor zekere  $n$  tussen  $\frac{1}{n+1}$  en  $\frac{1}{n}$  in zit, dan zal in ieder geval gelden dat  $n^2 \leq \mathcal{N}(\varepsilon) < (n+1)^2$ . Ofwel, voor ieder  $\varepsilon (\leq 1)$  geldt dat

$$\left(\frac{1}{\varepsilon} - 1\right)^2 \leq \mathcal{N}(\varepsilon; S(1)) \leq \left(\frac{1}{\varepsilon} + 1\right)^2.$$

Dus

$$-\lim_{\varepsilon \rightarrow 0} \frac{\log\left(\frac{1}{\varepsilon} - 1\right)^2}{\log \varepsilon} \leq \text{limcap}(S(1)) \leq -\lim_{\varepsilon \rightarrow 0} \frac{\log\left(\frac{1}{\varepsilon} + 1\right)^2}{\log \varepsilon}.$$

Beide limieten zijn gelijk aan 2: het maakt dus helemaal niet uit of we de preciese waarde van  $\mathcal{N}(\varepsilon; S(1))$  kennen, het vinden van een geschikte boven- en ondergrens is al voldoende. Met andere woorden, het nemen van de limiet maakt, in combinatie met de logaritmies, het exact bepalen van  $\mathcal{N}(\varepsilon)$  overbodig! De combinatorische subtiliteiten worden er als het ware uitgefilterd.

De vraag is natuurlijk of dit geldt voor alle mogelijke verzamelingen  $U \subset \mathbf{R}^2$ . Allereerst beschouwen we verzamelingen  $U$  waarvan het ‘intuïtief duidelijk’ is dat  $\dim(U) = 2$ . Wat betekent dat? In ieder geval dat  $U$  zo is dat er een  $\delta > 0$  te vinden is zodat er een (wellicht zeer klein vierkantje)  $S(\delta)$  bestaat dat in z’n geheel in  $U$  past. Lukt het niet om zo’n  $S(\delta)$  te vinden, dan is het namelijk meteen de vraag of  $U$  wel 2-dimensionaal is. Echter, alleen al om  $S(\delta)$  te overdekken zijn er minstens een zekere constante  $\times \varepsilon^{-2}$  vierkantjes  $S(\varepsilon)$  nodig (als  $\varepsilon$  klein genoeg is, en zeker kleiner dan  $\delta$ ), zoals we zojuist gezien hebben (voor  $\delta = 1$ , maar dat maakt natuurlijk niet echt uit). Er geldt dus dat  $\mathcal{N}(\varepsilon; U) \geq C\varepsilon^{-2}$  (voor zekere  $C$ ), zodat  $\text{limcap}(U) = 2$  (omdat  $\text{limcap}(U) \leq 2$

voor alle begrensde  $U \subset \mathbf{R}^2$ ). We concluderen dat  $\dim(U) = \limcap(U) = 2$  voor elk van deze verzamelingen  $U$ .

Hetzelfde geldt voor verzamelingen die ‘duidelijk’ 1-dimensionaal zijn. Laat  $\gamma \subset \mathbf{R}^2$  zo’n duidelijk 1-dimensionale (begrensde) verzameling zijn: laten we aannemen dat  $\gamma$  een zogenaamde gladde kromme is. In dat geval heeft  $\gamma$  een welgedefinieerde lengte  $L_\gamma$  (zie Opmerking 4). Kies nu  $\varepsilon$  klein genoeg en plaats een vierkantje  $S(\varepsilon)$  over  $\gamma$ . De doorsnede  $\gamma \cap S(\varepsilon)$  zal er ‘bijna’ uitzien als een recht lijnstukje (door steeds dieper in te zoomen – ofwel: door  $\varepsilon$  steeds kleiner te kiezen – wordt het net alsof  $\gamma$  lokaal niet langer gekromd is – dit komt natuurlijk omdat we hebben aangenomen dat  $\gamma$  ‘glad’ is). We kunnen  $S(\varepsilon)$  dus zo neerleggen dat  $\gamma \cap S(\varepsilon)$  ingaat en uitgaat via twee diagonaal tegenover elkaar liggende hoekpunten van  $S(\varepsilon)$ ;  $\gamma \cap S(\varepsilon)$  valt dan dus bijna samen met een diagonaal van  $S(\varepsilon)$ , hetgeen betekent dat de lengte van  $\gamma \cap S(\varepsilon)$  ‘bijna gelijk’ is aan  $\varepsilon\sqrt{2}$ . Door alle  $S(\varepsilon)$  op deze manier neer te leggen, en ze netjes op elkaar te laten aansluiten zien we dat

$$\mathcal{N}(\varepsilon; \gamma) \approx \frac{L_\gamma}{\sqrt{2}\varepsilon} \text{ als } \varepsilon \rightarrow 0,$$

waarbij het nu wel duidelijk is dat deze  $\mathcal{N}(\varepsilon; \gamma)$  daadwerkelijk het minimale aantal vierkantjes  $S(\varepsilon)$  is (hetgeen ook betekent dat de hierboven gevonden  $\mathcal{N}(\varepsilon) = \varepsilon^{-1}$  voor  $\gamma = [0, 1]$  veel te grof was!). We zien nu dus ook dat voor iedere gladde kromme  $\gamma$  geldt dat  $\dim(\gamma) = \limcap(\gamma) = 1$ .

Ook voor verzamelingen  $P_N$  bestaande uit eindig veel, ofwel  $N$ , losse punten geldt natuurlijk dat  $\dim(P) = \limcap(P) = 0$  (eenvoudigweg omdat  $\mathcal{N}(\varepsilon; P_N) = N$ , onafhankelijk van  $\varepsilon$ , als  $\varepsilon$  klein genoeg is). Dus, we mogen concluderen dat het begrip limiet-capaciteit inderdaad overeenkomt met ons intuïtieve dimensie begrip, tenminste, voor verzamelingen  $U$  waarvoor er geen twijfel kan bestaan over de dimensie van  $U$  (zie opmerking 5). In de volgende paragrafen zullen we zien dat de interpretatie van het begrip dimensie als limiet-capaciteit (1) van het dimensie begrip ook een zeer berekenbare (en benaderbare) grootheid maakt.

OPMERKING 4. Een gladde kromme  $\gamma$  wordt vastgelegd door een afbeelding  $F = (f_1, f_2) : [0, 1] \rightarrow \mathbf{R}^2$ , zodat  $\gamma = \{(f_1(t), f_2(t)) : t \in [0, 1]\}$ . Voor beide componenten  $f_i$  moet nu gelden dat de afgeleide bestaat en continu is als functie van  $t$ , voor iedere  $t \in [0, 1]$ . Dit betekent dat  $L_\gamma$  bestaat en eindig is (zie bijvoorbeeld [8]). Merk ook op dat dit impliceert dat de Koch kromme  $\mathcal{K}$  onmogelijk door zo’n gladde afbeelding  $F$  gegeven kan worden.

OPMERKING 5. De limiet-capaciteit is zeker niet de ‘beste’ generalisatie van het dimensie begrip. Het is bijvoorbeeld zeer eenvoudig om te laten zien dat de limiet-capaciteit van de verzameling van rationale getallen tussen 0 en 1,  $\mathbf{Q} \cap (0, 1)$ , gelijk is aan die van het hele interval  $[0, 1]$ :  $\limcap(\mathbf{Q} \cap (0, 1)) = 1$ . Echter,  $\mathbf{Q} \cap (0, 1)$  bevat slechts aftelbaar oneindig veel punten (zie bijvoor-

beeld [8]) en is dus zeer ‘ijl’. In het bijzonder geldt dat  $\mathbf{Q} \cap (0, 1)$  dus veel en veel minder punten dan de Cantor verzameling  $\mathcal{C}^\alpha$  bevat, voor iedere  $\alpha < 1$  – zie Opmerkingen 1 en 2 en figuur 3 (!). In de volgende paragraaf zullen we laten zien dat  $0 < \text{limcap}(\mathcal{C}^\alpha) < 1$  voor iedere  $\alpha \in (0, 1)$  en het feit dat  $\text{limcap}(\mathbf{Q} \cap (0, 1)) = 1$  is dan dus niet erg ‘logisch’. Dit probleem kan worden weggevangen door het begrip Hausdorff-dimensie in te voeren, dat gezien kan worden als een generalisatie van de limiet-capaciteit en waarvan de waarde samenvalt met die van de limiet-capaciteit voor alle andere in dit artikel genoemde verzamelingen  $U \subset \mathbf{R}^2$ , zie [1, 7, 9].

OPMERKING 6. We hebben de wiskundige vraag of de in de definitie van de limiet-capaciteit genomen limiet voor iedere mogelijke  $U \subset \mathbf{R}^2$  bestaat niet gesteld. Het is inderdaad niet uitgesloten dat deze limiet niet bestaat, en dat het quotiënt in (1) tussen twee uitersten blijft op en neer schommelen als  $\varepsilon \rightarrow 0$ . In dat geval is de limiet-capaciteit gedefinieerd als de maximale waarde die het quotiënt aanneemt in de limiet  $\varepsilon \rightarrow 0$ .

#### 5. VERZAMELINGEN MET FRACTIONELE DIMENSIE/LIMIET-CAPACITEIT

De limiet-capaciteit van de Cantor-verzameling  $\mathcal{C}$  kan op een eenvoudige manier expliciet berekend worden.

Allereerst constateren we dat  $\mathcal{N}(\varepsilon_n; \mathcal{C})$  direct te bepalen is als  $\varepsilon_n = \frac{1}{2}\sqrt{2} \times (\frac{1}{3})^n$  voor zekere  $n \geq 1$ . Voor iedere waarde van  $n$  kunnen we de verzameling  $V_n \supset \mathcal{C}$  volledig overdekken met  $2^n$  vierkantjes  $S(\varepsilon_n)$  door ieder van de  $2^n$  deel-lijnstukjes van  $V_n$  precies passend te bedekken met de diagonaal van een vierkantje  $S(\varepsilon_n)$  – deze heeft voor deze waarde van  $\varepsilon$  namelijk exact de lengte  $\frac{1}{3^n}$  (zie figuur 2). Het is duidelijk dat deze waarde van  $\mathcal{N}(\varepsilon, \mathcal{C})$  voor deze keuzes van  $\varepsilon$  inderdaad minimaal moet zijn. Als we ons nu even niets aantrekken van alle andere mogelijke waarden van  $\varepsilon$  dan vinden we:

$$\begin{aligned} \text{limcap}(\mathcal{C}) &= - \lim_{\varepsilon_n \rightarrow 0} \frac{\log \mathcal{N}(\varepsilon_n; \mathcal{C})}{\log \varepsilon_n} = \\ &= \lim_{n \rightarrow \infty} \frac{\log 2^n}{\log 3^n \sqrt{2}} = \lim_{n \rightarrow \infty} \frac{n \log 2}{n \log 3 + \log \sqrt{2}} = \frac{\log 2}{\log 3} \approx 0.6309... \end{aligned}$$

Het preciese bewijs dat inderdaad  $\text{limcap}(\mathcal{C}) = \log 2 / \log 3 \in (0, 1)$  (!) heeft nog wat meer voeten in de aarde, omdat we vooralsnog alleen maar naar a priori heel speciale waarden van  $\varepsilon$  gekeken hebben. Maar, net als in het geval van het vierkant  $S(1)$ , volgt er dat deze andere waarden van  $\varepsilon$  geen leidende orde invloed hebben op de exponent van  $\varepsilon$  in de uitdrukking voor  $\mathcal{N}(\varepsilon; \mathcal{C})$  in de limiet  $\varepsilon \rightarrow 0$ . We zullen hier niet verder op ingaan.

Dus, de Cantor-verzameling  $\mathcal{C}$  heeft een dimensie/limiet-capaciteit die niet gelijk is aan een geheel getal! Mede om deze reden wordt  $\mathcal{C}$  wel een *fractale*

*verzameling* genoemd. De verzameling  $\mathcal{C}$  is dus duidelijk meer dan een verzameling losse punten (die zou dimensie 0 hebben), maar, zoals verwacht, ook minder dan een ‘kromme’, want dan zou de dimensie 1 hebben moeten zijn.

Op eenzelfde manier kunnen we de dimensie/limiet-capaciteit van de Koch kromme  $\mathcal{K}$  bepalen. Ook hier trekken we ons weinig meer aan van de vraag of we wel voor iedere waarde van  $\varepsilon$  de optimale – en dus ‘echte’ – waarde van  $\mathcal{N}(\varepsilon; \mathcal{K})$  kennen. We zullen ons verder ook niet meer bezig houden met deze kwestie, uiteindelijk kan met wat meer moeite worden laten zien dat dit – alweer – geen invloed heeft op de dimensie van  $\mathcal{K}$ . Net als bij het bepalen van  $\text{limcap}(\mathcal{C})$  maken we ook hier weer gebruik van het feit dat  $\mathcal{K}$  aan de hand van een limiet-procedure geconstrueerd is.

We kiezen wederom  $\varepsilon_1 = \frac{1}{2}\sqrt{2} \times \frac{1}{3}$  en zien dat we de verzameling  $W_1$  netjes kunnen overdekken met 4 vierkantjes  $S(\varepsilon_1)$  (zie figuur 4): ieder van de 4 lijnstukjes waaruit  $W_1$  is opgebouwd laten we precies samenvallen met een diagonaal van een overdekkende vierkantje  $S(\varepsilon_1)$ , die per constructie exact de goede lengte ( $= \frac{1}{3}$ ) heeft (dit is natuurlijk volledig analoog aan de overdekking van  $V_n$  hierboven). Anders dan bij  $\mathcal{C}$  moeten we nu nog wel expliciet nagaan of deze overdekking van  $W_1$  ook een overdekking van  $\mathcal{K}$  geeft. Dat was in het geval van  $\mathcal{C}$  automatisch duidelijk, omdat  $\mathcal{C} \subset V_n$ , voor iedere  $n$ . Dit is natuurlijk niet het geval voor  $\mathcal{K}$  en de  $W_n$ 's (figuur 4) en heeft daarom nog enige voeten in de aarde.

De overdekking van  $W_1$  is zeker ook een overdekking van  $W_2$ , want de 4 nieuwe ‘tentjes’ die in de volgende iteratie-stap op ieder van de deel-lijnstukjes van  $W_1$  worden ‘opgezet’ liggen keurig binnen het vierkantje  $S(\varepsilon_1)$  dat het deel-lijnstukje overdekt. Dit is eenvoudig na te rekenen, bijvoorbeeld, de ‘top’  $T_2^1 = (\frac{1}{6}, \frac{1}{18}\sqrt{3})$  van het ‘tentje’ op het eerste deel-lijnstukje  $[0, \frac{1}{3}]$  van  $W_1$  (zie paragraaf 3) ligt duidelijk onder het hoekpunt van de overdekkende  $S(\varepsilon_1)$  (dat natuurlijk op  $(\frac{1}{6}, \frac{1}{6})$  ligt). Dus ligt ook het hele ‘tentje’ binnen de overdekkende  $S(\varepsilon_1)$ . Dit geldt dan automatisch ook voor de eerste iteraties van de 3 andere deel-lijnstukjes van  $W_1$  (deze zijn namelijk op verschuiving en draaiing na identiek aan de iteratie van  $[0, \frac{1}{3}]$ ). De vraag is nu: overdekken de  $S(\varepsilon_1)$ 's ook de volgende iteraties  $W_3, W_4$ , etc. en uiteindelijk  $\mathcal{K}$ ? Dit is a priori niet duidelijk: er worden meer en meer ‘tentjes’ dichter en dichter bij de randen van de  $S(\varepsilon_1)$ 's ‘opgezet’ (figuur 4).

Laten we eerst eens naar de volgende, fijnere, overdekking kijken. Door de zelfgelijkvormigheid van  $\mathcal{K}$  kunnen we de constructie van zojuist een stapje dieper in het iteratie-proces herhalen: we kiezen  $\varepsilon_2 = \frac{1}{2}\sqrt{2} \times (\frac{1}{3})^2$  en overdekken ieder van de 16 deel-lijnstukjes van  $W_2$  met een  $S(\varepsilon_2)$  (weer zodat de diagonaal van  $S(\varepsilon_2)$  samenvalt met het deel-lijnstukje). Door de gelijkvormigheid in de constructie van  $\mathcal{K}$  geldt nu automatisch dat deze overdekking ook de volgende iteratie  $W_3$  van  $W_2$  overdekt. We kunnen nu terugkeren naar de vraag of de overdekking van  $W_1$  door de 4  $S(\varepsilon_1)$ 's ook  $W_3, W_4$ , etc. overdekt. We be-

schouwen hiervoor wederom de overdekking van het eerste deel-lijnstukje  $[0, \frac{1}{3}]$  door  $S(\varepsilon_1)$ . Het is direct expliciet na te gaan dat alle 4 de vierkantjes  $S(\varepsilon_2)$  die de  $W_2$ -iteratie van  $[0, \frac{1}{3}]$  overdekken binnen  $S(\varepsilon_1)$  liggen. Dit betekent dat de overdekking van  $W_1$  door de 4  $S(\varepsilon_1)$ 's ook een overdekking geeft van  $W_3$  (omdat de  $S(\varepsilon_2)$  overdekking van  $W_2$  hierbinnen ligt en we weten dat deze  $W_3$  overdekt). Analoog zien we een iteratie-stap verder dat de  $S(\varepsilon_2)$  overdekking van  $W_2$  ook  $W_4$  overdekt (dit is weer het gelijkvormigheids-argument). Dit betekent dan weer dat de  $S(\varepsilon_1)$  overdekking van  $W_1$  de  $W_4$  overdekt. Etcetera. Deze constructie van 'geneste overdekkingen' kunnen we dus in z'n geheel itereren: voor alle  $\varepsilon_n = \frac{1}{2}\sqrt{2} \times (\frac{1}{3})^n$  kunnen we de bijbehorende  $W_n$ , en alle daar op volgende  $W_{n+m}$ 's, overdekken met  $4^n S(\varepsilon_n)$ 's. Dit betekent dat al deze overdekkingen tevens  $\mathcal{K}$  overdekken, zodat we kunnen concluderen dat

$$\text{limcap}(\mathcal{K}) = - \lim_{\varepsilon_n \rightarrow 0} \frac{\log \mathcal{N}(\varepsilon_n; \mathcal{K})}{\log \varepsilon_n} = \lim_{n \rightarrow \infty} \frac{\log 4^n}{\log 3^n \sqrt{2}} = \frac{2 \log 2}{\log 3} \approx 1.2618\dots$$

De Koch kromme  $\mathcal{K}$  heeft dus inderdaad een dimensie (of beter: limiet-capaciteit) die niet geheel-tallig is en in ligt tussen 1 en 2!

De dimensie van de algemene Cantor verzameling  $\mathcal{C}^\alpha$  zou ook weer op eenzelfde manier als die van  $\mathcal{C}$  en  $\mathcal{K}$  kunnen worden bepaald. Echter, er is een makkelijker manier, tenminste, als we de formule (1) nog iets 'ruimer' interpreteren. Deze relatie lijkt namelijk te impliceren dat voor  $\varepsilon$  'klein genoeg'

$$\mathcal{N}(\varepsilon; U) \approx C\varepsilon^{-\text{limcap}(U)}, \text{ voor zekere } C > 0. \quad (2)$$

Dat dit niet per se zo hoeft te zijn was al in Opmerking 6 geconstateerd, echter, voor de meeste verzamelingen  $U$  is dit een zeer bruikbare benadering, die we ook in de rest van dit verhaal zullen gebruiken. Met deze formule, en het feit dat  $\mathcal{C}^\alpha$  zelfgelijkvormig is, is  $\text{limcap}(\mathcal{C}^\alpha)$  zonder at teveel expliciet rekenwerk te bepalen.

De linkerkant van de verzameling  $\mathcal{C}^\alpha$ , de verzameling  $\mathcal{C}_1^\alpha = \mathcal{C}^\alpha \cap V_1^\alpha$ , vormt natuurlijk de helft van de gehele  $\mathcal{C}^\alpha$  en kan dus ook met half zoveel vierkantjes  $S(\varepsilon)$  worden overdekt:

$$\mathcal{N}(\varepsilon, \mathcal{C}_1^\alpha) = \frac{1}{2} \mathcal{N}(\varepsilon, \mathcal{C}^\alpha).$$

Aan de andere kant kunnen we  $\mathcal{C}_1^\alpha$  ook 'opblazen', ofwel vermenigvuldigen met een factor, zodat  $\mathcal{C}_1^\alpha$  precies samenvalt met de oorspronkelijke  $\mathcal{C}^\alpha$  ( $\mathcal{C}^\alpha$  is tenslotte zelfgelijkvormig). Omdat de lengte van  $V_1^\alpha$  gelijk is aan  $\frac{1}{2}(1 - \alpha)$  (zie paragraaf 2), zal de opblaasfactor  $\frac{2}{1-\alpha}$  zijn. Door dit opblazen verandert ook de grootte van de overdekkende vierkantjes  $S(\varepsilon)$ :  $S(\varepsilon)$  blaast op tot  $S(\frac{2}{1-\alpha}\varepsilon)$ . Dus, een overdekking van  $\mathcal{C}_1^\alpha$  met  $S(\varepsilon)$ 's geeft door 'opblazing' een overdekking van  $\mathcal{C}^\alpha$  met  $S(\frac{2}{1-\alpha}\varepsilon)$ 's, hetgeen betekent dat

$$\mathcal{N}(\varepsilon, \mathcal{C}_1^\alpha) = \mathcal{N}(\frac{2}{1-\alpha}\varepsilon, \mathcal{C}^\alpha)$$

Deze laatste twee formules te samen geven de relatie:

$$\frac{1}{2}\mathcal{N}(\varepsilon, \mathcal{C}^\alpha) = \mathcal{N}\left(\frac{2}{1-\alpha}\varepsilon, \mathcal{C}^\alpha\right)$$

Als we nu aannemen dat (2) exact geldt, dan betekent dit dat

$$\frac{1}{2}C\varepsilon^{-d} = C\left(\frac{2}{1-\alpha}\varepsilon\right)^{-d} = C\left(\frac{2}{1-\alpha}\right)^{-d}\varepsilon^{-d} \quad \text{met } d = \text{limcap}(\mathcal{C}^\alpha).$$

Hieruit kunnen  $C$  en  $\varepsilon^{-d}$  worden weggedeeld, zodat volgt:

$$d = \text{limcap}(\mathcal{C}^\alpha) = \frac{\log 2}{\log 2 - \log(1-\alpha)}.$$

Deze enigszins formele afleiding van  $\text{limcap}(\mathcal{C}^\alpha)$  kan ook wiskundig hard worden gemaakt, maar daar gaan we hier verder niet op in (zie ook [1, 7, 9]).

We kennen al drie expliciete waarden van  $\text{limcap}(\mathcal{C}^\alpha)$ :  $\text{limcap}(\mathcal{C}^1) = 0$  (want  $\mathcal{C}^1 = \{0, 1\}$ ),  $\text{limcap}(\mathcal{C}^0) = 1$  (want  $\mathcal{C}^0 = [0, 1]$ ) en natuurlijk  $\text{limcap}(\mathcal{C}^{\frac{1}{3}}) = \log 2 / \log 3$ . Deze 3 waardes kloppen keurig met de algemene uitdrukking (al moet voor  $\alpha = 1$  wel een limiet genomen worden). Verder valt op dat  $\text{limcap}(\mathcal{C}^\alpha)$  monotoon afneemt als  $\alpha$  van 1 naar 0 loopt (met precies tussenin  $\text{limcap}(\mathcal{C}^{\frac{1}{2}}) = \frac{1}{2}!$ ):  $\mathcal{C}^\alpha$  geeft ons dus de mogelijkheid om voor ieder getal  $d$  tussen 0 en 1 een verzameling te construeren waarvan de dimensie/limiet-capaciteit precies gelijk is aan  $d$ !

## 6. HET BENADEREN VAN DE DIMENSIE VAN EEN ATTRACTOR/PUNTENWOLK

We keren nu terug naar de oorspronkelijke vraag uit de Inleiding: hoe moet je de dimensie (of nu: limiet-capaciteit) van een verzameling, of attractor, of puntenwolk,  $\mathcal{A}$  ‘meten’? Het grootste probleem hierbij is dat deze puntenwolken ‘in de praktijk’ door een computer gegenereerd worden, en dat je dus altijd maar eindig veel van deze punten kent (zie figuur 1). Natuurlijk gaan we er hierbij vanuit dat deze verzameling, of attractor in het geval van figuur 1, weldegelijk uit oneindig veel punten zal bestaan, het is alleen praktisch onmogelijk om al deze punten te bepalen.

De ‘officiële definitie’ van de limiet-capaciteit, (1), kan hiervoor niet gebruikt worden: er wordt een limiet  $\varepsilon \rightarrow 0$  genomen, en voor een verzameling van eindig veel punten zal dat altijd betekenen dat de limiet-capaciteit  $\equiv 0$  is. Echter, we kunnen wel weer gebruik maken van de benadering als in (2): voor  $\varepsilon$  klein genoeg, maar niet te klein, zal gelden dat

$$\text{limcap}(\mathcal{A}) \approx -\frac{\log \tilde{\mathcal{N}}(\varepsilon; \mathcal{A})}{\log \varepsilon}. \quad (3)$$

En hierbij hebben we het ons zelfs nog makkelijker gemaakt. We hebben  $\mathcal{N}(\varepsilon; \mathcal{A})$  namelijk vervangen door  $\tilde{\mathcal{N}}(\varepsilon; \mathcal{A})$ . Dit betekent dat we niet langer

‘gedwongen’ willen zijn om te gaan ‘spelen’ met vierkantjes  $S(\varepsilon)$  om zo te proberen de verzameling  $\mathcal{A}$  zo ‘zuinig mogelijk’ te overdekken (hetgeen nodig is om  $\mathcal{N}(\varepsilon; \mathcal{A})$  te bepalen). In plaats daarvan fixeren we gewoon de vierkantjes  $S(\varepsilon)$  in een rooster (als op een ruitjespapier, zie figuur 5) en tellen de ‘ruitjes’ die een stukje van  $\mathcal{A}$  overdekken. De uitkomst noemen  $\tilde{\mathcal{N}}(\varepsilon; \mathcal{A})$ . Dit getal geeft natuurlijk alleen een bovengrens op  $\mathcal{N}(\varepsilon; \mathcal{A})$ , maar we hebben in de vorige paragraaf toch steeds al gezien dat het hebben van de preciese waarde van  $\mathcal{N}(\varepsilon; \mathcal{A})$  er niet echt toe doet bij het bepalen van  $\text{limcap}(\mathcal{A})$  (zie Opmerking 7).

Het ‘berekenen’ van  $\tilde{\mathcal{N}}(\varepsilon; \mathcal{A})$  is natuurlijk zeer eenvoudig te programmeren, hetgeen duidelijk maakt waarom deze, a priori grotendeels formele benadering zo populair is. De op deze manier bepaalde dimensie van een object heeft zelfs een eigen naam gekregen: de *box-counting dimension*, een naam die precies zegt wat het is (Opmerking 7). De box-counting aanpak geeft ons (en vele met ons) een ideaal handvat om een schatting te maken van attractors als in figuur 1 en maakt van het begrip dimensie een experimenteel meetbare grootheid.

Voor we hier aan gaan beginnen moeten we nog even stilstaan bij een gevaarlijk aspect van deze methode. We beschouwen hiertoe (alweer) het eenheidsvierkant  $S(1)$ , op een rooster dat is opgebouwd uit vierkantjes  $S(0.1)$ . Als we de zijdes van  $S(1)$  precies samen laten vallen met de assen van het rooster dan is  $\tilde{\mathcal{N}}(0.1; S(1)) = 100$  en geeft de box-counting aanpak ons met behulp van (3) de volgende ‘benadering’

$$\text{limcap}(S(1)) \approx -\frac{\log 100}{\log 0.1} = 2.0$$

Echter, de dimensie van  $S(1)$  mag natuurlijk niet veranderen als we  $S(1)$  met de een of ander factor ‘opblazen’. Maar, als we  $S(1)$  twee keer zo groot maken, dan vinden we, op hetzelfde rooster (dat we dus niet mee opblazen):

$$\text{limcap}(S(2)) \approx -\frac{\log 400}{\log 0.1} = 2.602\dots,$$

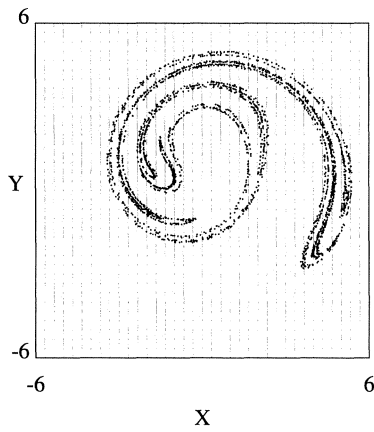
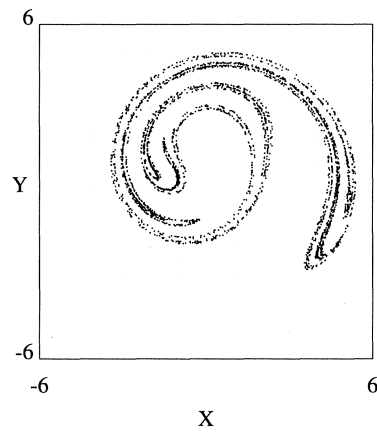
En, een opblaasfactor  $\frac{1}{2}$  geeft (nog steeds op het  $S(0.1)$  rooster):

$$\text{limcap}(S(0.5)) \approx -\frac{\log 25}{\log 0.1} = 1.397\dots,$$

Dat zijn nogal verschillen. Waar ligt dat aan? Laten we nog eens naar benadering (2) kijken en deze in de vorm van (3) schrijven:

$$\text{limcap}(\mathcal{A}) \approx -\frac{\log \tilde{\mathcal{N}}(\varepsilon; \mathcal{A}) - \log C}{\log \varepsilon}.$$

Dus, benadering (3) is veel ‘grover’ dan (2). Natuurlijk komen beide uitdrukkingen overeen als  $\varepsilon$  echt heel klein wordt, maar voor niet al te kleine  $\varepsilon$  heeft



FIGUUR 5. Twee overdekkingen van de attractor  $\mathcal{A}$  uit figuur 1. De eerste overdekkingen geeft de schatting  $\text{limcap}(\mathcal{A}) \approx 1.62$ , de tweede (meer nauwkeurige)  $\text{limcap}(\mathcal{A}) \approx 1.58$ .



de extra term ‘ $-\log C$ ’ een significante invloed op de uitkomst. In de bovenstaande voorbeelden is het natuurlijk duidelijk wat de waarden voor  $C$  moeten zijn:  $C_{S(2)} = 4$ ,  $C_{S(1)} = 1$  en  $C_{S(0.5)} = 0.25$ . Geen wonder dat  $S(1)$  de beste benadering geeft:  $\log C_{S(1)} = 0$ ! Met deze extra correctie-term volgt nu

$$\text{limcap}(S(2)) \approx -\frac{\log 400 - \log 4}{\log 0.1} = 2.0 = -\frac{\log 25 - \log 0.25}{\log 0.1} \approx \text{limcap}(S(0.5)).$$

Dit betekent in de praktijk dat we er voor moeten proberen te zorgen dat we onze attractor  $\mathcal{A}$  zo schalen dat de waarde  $C_{\mathcal{A}}$  – die we niet kennen en eigenlijk ook niet willen kennen – zo dicht mogelijk bij 1 ligt. Het bovenstaande voorbeeld geeft aan dat dit betekent dat we de absolute afmeting van  $\mathcal{A}$  niet ‘te groot’ (zoals bij  $S(2)$ ) en niet ‘te klein’ (zoals bij  $S(0.5)$ ) mogen maken, tenminste, als we de  $\varepsilon$  niet zo klein kunnen maken als we maar willen (hetgeen eigenlijk altijd het geval is). In de praktijk betekent dit dat we  $\mathcal{A}$  zo schalen dat  $\mathcal{A}$  ‘ongeveer passend’ overdekt wordt door één vierkant  $S(1)$ .

We kunnen nu de tot dusver ontwikkelde theoretische kennis toepassen op de attractor  $\mathcal{A}$  in figuur 1.

Eerst gaan we kort in op de achtergrond van deze attractor. Deze is verkregen met behulp van een computer-simulatie van een model, in de vorm van een dynamisch systeem, dat de modulatie van de getijde-golf in een zeker geïdealiseerd bekken (denk hierbij aan een baai, een fjord of een haven) beschrijft (zie [5]). Met modulatie bedoelen we hier dat de punten in figuur 1 niet de exacte waterstanden representeren, maar de veranderingen ten opzichte van de door ‘iedereen’ verwachte perfect periodieke – en dus zeker niet chaotische – eb en vloed cyclus. Grofweg gesproken kan je zeggen dat ieder puntje in figuur 1 (of figuur 5) een ‘meting’ van de waterstand tijdens springvloed representeert. Dit betekent dat de computer het dynamische systeem, in dit geval een (gewone) differentiaalvergelijking, simuleert en, in de fysische tijd, eens in de 28-dagen een waarde van  $X$  en  $Y$  geeft; zo’n waarde bepaalt een puntje in figuur 1. Het volgende puntje volgt 28 dagen later. Het getal 28 staat natuurlijk voor de periode van de springvloed-cyclus. In figuur 1 staan meer dan 5000 van deze puntjes, dus, deze figuur beschrijft de eb en vloed dynamica in het bekken over een periode van zo’n 400 jaar. De waarden  $X$  en  $Y$  representeren de amplitude en de fase van (een getransformeerde en geschaalde versie van) het getij; de amplitude, of wel de geschaalde hoogte van de waterstand, volgt uit  $\sqrt{X^2 + Y^2}$  (zie [6] voor meer details en voor de oorspronkelijke versie van figuur 1). De voor de hand liggende vraag ‘Waarom zou het getij chaotisch zijn?’ kan hier niet zo een, twee, drie worden beantwoord (en het antwoord valt ook volledig buiten de context van dit verhaal). In de artikelen [6] en [2] wordt uitgelegd hoe de nietlineariteit in het model die veroorzaakt wordt door de geometrie van het bekken (ofwel de vorm en structuur van de bodem) aanleiding kan geven tot chaotisch, en dus onregelmatig, gedrag.

De assen in figuur 1 en 5 geven aan dat  $X, Y \in (-6, 6)$ . Dit betekent dat

de attractor  $\mathcal{A}$  ‘behoorlijk groot’ is: grof gezegd wordt  $\mathcal{A}$  zo’n beetje passend overdekt door het vierkant  $S(12)$ . Om de zojuist geconstateerde vertroebelende invloed van de constante  $C_{\mathcal{A}}$  zoveel mogelijk te reduceren schalen we  $\mathcal{A}$  zodat we  $S(12)$  kunnen vervangen door  $S(1)$ . Dit komt in figuur 5 neer op het negeren van de  $\pm 6$ -en langs de assen: we definiëren de punten links onderaan de twee (grote) vierkanten in figuur 5 als  $(0, 0)$  en rechts bovenaan als  $(1, 1)$ .

In het eerste plaatje in figuur 5 is het vierkant overdekt met 144 kleinere vierkantjes. Dit betekent dat  $\varepsilon = \frac{1}{12}$ :  $S(1)$  wordt overdekt door een rooster van 144  $S(\frac{1}{12})$ 's. Enig telwerk levert dat 56 van deze  $S(\frac{1}{12})$ 's een of meer elementen van  $\mathcal{A}$  bevatten:  $\tilde{\mathcal{N}}(\frac{1}{12}; \mathcal{A}) = 56$ . Dit geeft met behulp van (3) de eerste benadering van  $\text{limcap}(\mathcal{A})$ :

$$\text{limcap}(\mathcal{A}) \approx \frac{\log 56}{\log 12} = 1.6199\dots$$

Fijnere roosters geven:  $\tilde{\mathcal{N}}(\frac{1}{24}; \mathcal{A}) = 158$  (het bijbehorende plaatje staat niet in deze tekst) en  $\tilde{\mathcal{N}}(\frac{1}{36}; \mathcal{A}) = 289$  (het tweede plaatje in figuur 5). Dit geeft de volgende, meer nauwkeurige, benaderingen:

$$\text{limcap}(\mathcal{A}) \approx \frac{\log 158}{\log 24} = 1.5929\dots \quad \text{en} \quad \text{limcap}(\mathcal{A}) \approx \frac{\log 289}{\log 36} = 1.5812\dots$$

Deze 3 opeenvolgende benaderingen komen dus uitstekend overeen. Zonder na te denken over de nauwkeurigheid van de benaderingen lijkt in ieder geval de conclusie dat  $\text{limcap}(\mathcal{A}) \approx 1.6$  gerechtvaardigd.

Heeft het nu nog zin om een nog fijnere overdekking te beschouwen, en zo een nog betere benadering van  $\text{limcap}(\mathcal{A})$  proberen te vinden? Het antwoord hierop is ‘nee’. Er zijn namelijk al een aantal vierkantjes  $S(\frac{1}{36})$  aan te wijzen waarin geen van de ruim 5000 punten van  $\mathcal{A}$  liggen, en die dus niet meetellen voor  $\tilde{\mathcal{N}}(\frac{1}{36}, \mathcal{A})$ , maar waarvan het duidelijk is dat er wel punten van  $\mathcal{A}$  in zouden liggen als we meer punten van  $\mathcal{A}$  hadden gehad (zie bijvoorbeeld in figuur 5 de  $S(\frac{1}{36})$ 's in de 6-de rij, 29-ste kolom en de 8-ste rij, 31-ste kolom). Dus, terwijl je zou denken dat er zoveel punten zijn dat een fijner rooster zeker nog zou kunnen, blijkt voor deze waarde van  $\varepsilon$  al dat we ‘last hebben’ van het feit dat we zo ‘weinig’ punten hebben. Dit probleem wordt natuurlijk alleen maar groter als we  $\varepsilon$  nog kleiner kiezen. Dit betekent dat  $\tilde{\mathcal{N}}(\varepsilon; \mathcal{A})$  voor kleinere  $\varepsilon$  (veel) te klein gaat worden en we dus een steeds grotere fout in de benadering van  $\text{limcap}(\mathcal{A})$  gaan maken. Zie ook de discussie in de volgende paragraaf.

OPMERKING 7. Voor een wiskundig preciesere definitie van de box-counting dimension dienen we  $\mathcal{N}(\varepsilon; \mathcal{A})$  in (1) te vervangen door  $\tilde{\mathcal{N}}(\varepsilon; \mathcal{A})$ . Natuurlijk is het a priori weer niet duidelijk of de limiet ‘altijd’ dezelfde uitkomst zal geven. We gaan hier wederom niet verder op in en verwijzen naar [1, 7, 9] voor meer details.

## 7. DISCUSSIE

We hebben in dit verhaal laten zien hoe het intuïtieve begrip dimensie gematematiseerd en gekwantiseerd, ofwel berekenbaar, kan worden. Met behulp van de formules (1), (2) en (3) kunnen we nu voor allerlei verzamelingen de dimensie bepalen. Soms kan dat analytisch, soms is hiervoor een benaderingsprocedure en een computer nodig. Het is ook duidelijk dat er eigenlijk geen goede reden meer is om aan te nemen dat de dimensie van een ‘object’/verzameling geheel-talig moet zijn.

Voor de laatste formule (3) geeft ons ook de mogelijkheid om de dimensie van een door de computer gegenereerde of zelfs van een experimenteel waargenomen ‘data-set’ te schatten. Met behulp van (3) en gedetailleerde kaarten kan nu zelfs een schatting worden gemaakt van een ‘natuurlijk fractaal object’ als de kustlijn van Engeland (zie [7]). Bij alle toepassingen van (3) dient men wel ten alle tijde bewust te zijn van de beperkte geldigheid van deze uitdrukking. We hebben gezien dat je zelfs bij een behoorlijk grote en nauwkeurige data-set als die van figuur 1 al vrij snel ‘in de problemen’ komt. Bij experimentele geografische of biologische ‘toepassingen’ van de box-counting dimension aanpak zijn deze problemen nog eens veel groter: er is meestal een relatief grote meetfout en het aantal betrouwbare ‘waarnemingen’ is vaak klein. In dit soort situaties is het dan ook maar de vraag wat de betrouwbaarheid of waarde van uitspraken over, bijvoorbeeld, ‘de dimensie van de kustlijn van Engeland’ is. Echter, dat neemt niet weg dat het met de hier gepresenteerde wiskundige ideeën mogelijk is om gestructureerd en nauwkeurig over dit soort kwesties na te denken.

Je zou verwachten dat de begrippen limiet-capaciteit en box-counting dimension vooral binnen de theorie en de praktijk van het interdisciplinaire vakgebied dynamische systemen van groot belang zijn. Echter, om attractors in dynamische systemen goed te beschrijven en te begrijpen zijn subtielere en meer geavanceerde begrippen noodzakelijk. In die zin is het begrip limiet-capaciteit slechts de eerste stap in de richting van de ontwikkeling van een ‘dynamisch’ dimensie begrip. Een van de belangrijkste redenen hiervoor is het feit dat de limiet-capaciteit geen rekening kan houden met inhomogeniteiten in de verdeling van punten over een attractor  $\mathcal{A}$  als in figuur 1. Het is zeer duidelijk dat sommige gebieden van  $\mathcal{A}$  veel vaker ‘bezocht’ worden dan anderen (zie figuur 5). De limiet-capaciteit kan dit niet ‘meten’ en geeft daarom vaak een onbetrouwbaar beeld ‘de dimensie’ van een attractor. We verwijzen naar inleidende, maar vaak toch nog redelijk geavanceerde, teksten als [1, 3, 9, 10] waarin meer geschikte en dynamische begrippen als correlatie-dimensie en Lyapunov exponent worden geïntroduceerd.

VERANTWOORDING/DANKWOORD. Een belangrijk gedeelte van het ‘materiaal’ in dit verhaal is gebaseerd op Barteld Braaksma’s hoofdstuk ‘*Gebroken dimensies en fractale verzamelingen*’ in [1], dat dieper op de materie ingaat. De figuren zijn gemaakt door Harmen van der Ploeg.

## LITERATUUR

1. H.W. BROER & F. VERHULST, *Dynamische Systemen en Chaos*, Epsilon Uitgaven **14** (1990).
2. A. DOELMAN, A.F. KOENDERINK & L.R. M. MAAS, Quasi-periodically forced nonlinear Helmholtz oscillators, preprint (2000).
3. J. GUCKENHEIMER & P. HOLMES, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Third edition, Applied Mathematical Sciences **42**, Springer (1990).
4. H. LAUWERIER, *Fractals. Meetkundige Figuren in Eindeloze Herhaling*, Aramith (1987).
5. L.R.M. MAAS, On the nonlinear Helmholtz response of almost-enclosed tidal basins with sloping bottoms, *J. Fluid Mech.* **349**, 361–380 (1997).
6. L.R.M. MAAS & A. DOELMAN, Chaotic tides, te verschijnen in *J. Phys. Oceanogr.* (2001).
7. B. MANDELBROT, *The Fractal Geometry of Nature*, Freeman (1982).
8. W. RUDIN, *Principles of Mathematical Analysis*, Third edition, McGraw-Hill (1976).
9. H.G. SCHUSTER, *Deterministic Chaos. An Introduction*, Second edition, VCH Verlag (1988).
10. S. WIGGINS, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Texts in Applied Mathematics **2**, Springer (1990).

## Uitputting en Evenwicht – Officiële en Experimentele Bepalingen van Oppervlakte en Inhoud Sinds de Oudheid

Jan van Maanen  
Rijksuniversiteit Groningen  
e-mail: maanen@math.rug.nl

In hoeverre was er in de Oudheid 'experimentele' wiskunde? Wanneer noemen we wiskunde eigenlijk 'experimenteel'? Als er stevig gerekend wordt? Als de aanpak niet meteen in een vakje te plaatsen is? Was chaostheorie twintig jaar geleden 'experimenteel' en is het nu officiële wiskunde geworden?

In de Griekse Oudheid leefden vergelijkbare vragen al. In elk geval was in de vlakke meetkunde duidelijk welke methoden acceptabel waren en welke niet. Neem de vraag naar de omvang van een gesloten vlakke figuur, een driehoek bijvoorbeeld. Wij beantwoorden deze vraag tegenwoordig door een reël getal te noemen: de oppervlakte, ofwel het aantal eenheidsvierkanten waarmee de driehoek precies bedekt kan worden. In navolging van Euclides deden Griekse meetkundigen dat anders. De omvang van een driehoek gaven ze aan in de vorm van een quadratuur, dat is de constructie van een vierkant dat even groot was als de driehoek. En die constructie moest plaatsvinden volgens de spelregels die Euclides in de *Elementen* had vastgelegd: met passer en liniaal en in een eindig aantal stappen.

Bij dit soort vragen, die in de Oudheid voor allerlei gesloten vlakke en ruimtelijke figuren gesteld werden, verlieten sommige wiskundigen het officiële pad. Archimedes bijvoorbeeld permitteerde zich het gebruik van illegale methoden, meer in het bijzonder het splitsen van een figuur in oneindig veel delen die hij aan een balans in evenwicht bracht met delen van een andere figuur. Maar hij deed dat alleen om een oplossing op het spoor te komen. De juistheid bewees hij vervolgens met 'nette' meetkunde en een redenering in een eindig aantal stappen.

### 1. RIJST KOKEN

Zomer 1996. Ik zit in een restaurant in Braga (Portugal). Er staat, naast andere heerlijkheden, rijst op tafel, en ik spreek met Eduardo over hoe wij zelf rijst koken. We blijken dezelfde regel te hanteren: op één kop rijst doen we twee koppen water in de pan. Meteen zegt Eduardo: "op deze manier vond Archimedes de inhoud van de kegel". In schaakpartijen heet dat een nieuwtje, iets wat iemand op het bord tovert om de ander te verrassen –heel toepasselijk in een restaurant trouwens. Het duurde even voor Eduardo aan mijn glimlach kon zien dat zijn opmerking aangekomen was. Waarop doelde hij?

In de eerste plaats is er een vuistregel, die in keukens ontstaan is en die van generatie op generatie doorgegeven wordt: *op één deel rijst doe je twee delen water.*

In de tweede plaats is er het historische gegeven hoe de klassieke meetkundigen, zeg Euclides en Archimedes, werkten met het begrip inhoud. Ze drukten de inhoud van een lichaam uit door het te vergelijken met een standaard-lichaam, net zoals wij eigenlijk. Als wij zeggen

$$\text{Inhoud kegel} = \frac{1}{3}\pi r^2 h$$

dan geven we aan hoe vaak een eenheids-kubus in de kegel past, namelijk het reële getal  $\frac{1}{3}\pi r^2 h$  keer. Dankzij de integraalrekening is het tegenwoordig geen probleem meer om een lichaam met gebogen grensvlakken te vergelijken met een kubus. Bij Archimedes lag dat heel anders, sterker nog: twee- en driedimensionale objecten met gebogen grenzen vergelijken met objecten met rechte of vlakke grenzen was de kern van zijn wiskundige onderzoek. Archimedes duidde de inhoud van de kegel in eerste instantie aan door hem te vergelijken met een even hoge cilinder op dezelfde grondcirkel:

$$\text{Inhoud kegel} = \frac{1}{3} \times \text{Inhoud cilinder met zelfde grondvlak en hoogte}$$

Hiermee hebben we de sleutel voor de interpretatie van Eduardo's opmerking. Als je rijst in een ronde pan gooit, vormt het een kegel op de bodem van de pan, even aangenomen natuurlijk dat er niet te veel rijst is. Bij het koken wil je dat de rijst helemaal onder water staat. Rijst en water samen vormen een cilinder, waarvan de rijst volgens de vuistregel één derde deel uitmaakt. De vuistregel komt dus precies overeen met Archimedes' bepaling van de inhoud van een kegel. De vraag of Archimedes inderdaad een dergelijke heuristiek gebruikte, en of hij er bewijskracht aan toedichtte, staat verder centraal in dit verhaal.

## 2. DE QUADRATUUR VAN DE CIRKEL

Een goed voorbeeld van de officiële aanpak van oppervlakteproblemen via uitputting is de eerste stelling uit de *Cirkelmeting* van Archimedes. Deze luidt:

**Stelling 2.1.** *Een willekeurige cirkel is even groot als een rechthoekige driehoek waarvan één rechthoekszijde gelijk is aan de straal van de cirkel en de andere rechthoekszijde gelijk is aan de omtrek van de cirkel.*

Deze stelling verplaatst het probleem van de quadratuur van de cirkel naar de rectificatie –rechttrekking– van de cirkel. Stel dat je met passer en liniaal een lijnstuk kunt construeren dat even lang is als de omtrek van  $C$ . Dan verander je eerst de cirkel  $C$  in de rechthoekige driehoek  $D$ , want volgens Archimedes zijn die twee even groot. En van elke driehoek is de quadratuur elementair (zie de stappen in Figuur 1), voor een rechthoekige driehoek al helemaal. Maar verder dan het aantonen dat de quadratuur en de rectificatie van de cirkel equivalent waren, kwam Archimedes niet met een exacte oplossing van het quadratuurprobleem. Zijn benadering  $3\frac{10}{71} < \frac{\text{omtrek}}{\text{middellijn}} < 3\frac{10}{70}$  was natuurlijk een fenomenaal

en heel praktisch alternatief voor de exacte oplossing, die trouwens zeer lang een open probleem bleef. Pas in 1882 bewees Lindemann dat de exacte quadratuur van de cirkel met passer en liniaal onmogelijk is.

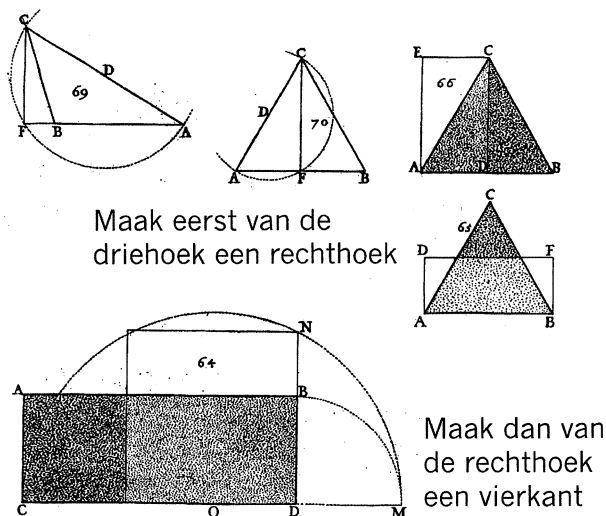
Laten we Archimedes' bewijs bekijken (Figuur 2).

*Bewijs.* Stel dat de cirkel  $C$  groter of kleiner is dan de rechthoekige driehoek  $D$ . Ik ga laten zien dat dat beide veronderstellingen tot een tegenspraak leiden, dus de enige mogelijkheid die overblijft is dat de cirkel even groot is als de driehoek.

Stel nu eerst dat  $C$  groter is dan  $D$ . Maak dan het ingeschreven vierkant  $ABCD$ , deel de bogen  $AB$ ,  $BC$ ,  $CD$  en  $DA$  middendoor, deel de helften opnieuw middendoor, en ga zo door totdat de cirkelsegmenten die buiten de ingeschreven veelhoek liggen, samen minder zijn dan het verschil tussen cirkel  $C$  en driehoek  $D$ .

Dus is de veelhoek groter dan  $D$ .

Maar  $ON$ , de afstand van het middelpunt tot de zijde van de veelhoek, is kleiner dan de straal van  $C$  en dus ook kleiner dan de kortste rechthoekszijde van  $D$ . En de omtrek van de veelhoek is kleiner dan de omtrek van de cirkel en dus ook kleiner dan de andere rechthoekszijde van  $D$ . Dus is de veelhoek



FIGUUR 1. De quadratuur van de driehoek, uit [7]. In de laatste stap wordt gebruikt dat hoek  $CNM$  recht is, zodat  $CD : DN = DN : DM = DN : DB$  en dus  $DN^2 = CD \times DB$ .

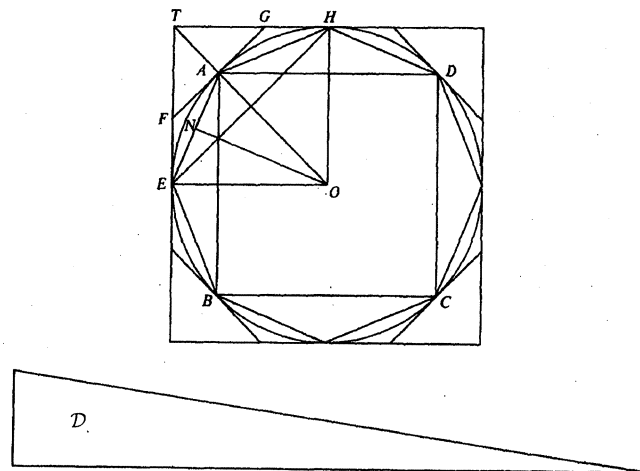
kleiner dan  $D$ , maar dit is in tegenspraak met de eerdere constatering dat de veelhoek groter is dan  $D$ .  $C$  kan dus niet groter zijn dan  $D$ .

De andere veronderstelling, dat  $C$  kleiner is dan de driehoek  $D$  leidt met een soortgelijke redenering over een eindige rij omgeschreven veelhoeken tot

een tegenspraak. Zie bijvoorbeeld [4, pp. 148–150], [6, pp. 91–93] of [8] voor het complete bewijs.  $\square$

Het proces dat Archimedes hier gebruikt wordt exhaustie genoemd, uitputting. De achtergrond van deze naam is duidelijk: je kunt de cirkel met ingeschreven veelhoeken zo ver uitputten (zo vol maken, of met omgeschreven veelhoeken zo nauw insluiten) als je wilt. Dijksterhuis vond exhaustie trouwens “wel de slechtste naam, dien men kon bedenken.” [2, p. 122]. Het principe is namelijk dat de figuur helemaal niet uitgeput wordt, want Archimedes doet het juist zo dat het proces na een eindig aantal stappen stopt. Vanaf Euclides (rond 300 v.Chr.) was de bewijsmethodiek op dit punt heel precies. Euclides geeft in de *Elementen* een stelling (boek X, stelling 1), die bij het onderzoek naar het afbreken van het proces zeer goed van pas komt:

**Stelling 2.2.** *Indien, wanneer twee ongelijke grootheden uitgezet zijn, van de grootste een stuk, groter dan de helft, wordt afgenomen en van de rest een stuk, groter dan de helft, en indien dit steeds zo doorgaat, dan zal er een grootte overblijven, die kleiner zal zijn dan de uitgezette kleinste grootte. [in de vertaling van Dijksterhuis]*



FIGUUR 2. Archimedes' figuur bij de *Cirkelmeting*

Bij het opvullen van  $C$  gaat het om de ruimte tussen  $C$  en de ingeschreven veelhoek, die kleiner moet worden dan het verschil tussen  $C$  en  $D$ . Bij de overstap van ingeschreven vierkant naar ingeschreven achthoek worden vier driehoeken afgenomen van de ruimte tussen  $C$  en het ingeschreven vierkant. Driehoek  $ABE$  is er één van; deze is gelijk aan de helft van een rechthoek met basis  $AB$  en dezelfde hoogte als het cirkelsegment. Aangezien de rechthoek groter is dan het cirkelsegment, is driehoek  $ABE$  groter dan de helft van het segment, en



de vier driehoeken samen nemen dus meer dan de helft van de ruimte buiten het ingeschreven vierkant weg. Bij de volgende stappen blijft dit argument van kracht, en dus kan stelling 2.2 hier toegepast worden (ook al doet Archimedes dat niet expliciet, wellicht heeft hij gevonden dat het hier wel erg voor de hand lag): in een eindig aantal stappen is de ruimte buiten de veelhoek kleiner geworden dan het verschil van  $\mathcal{C}$  en  $\mathcal{D}$ .

Archimedes' aanpak heeft het grote voordeel dat het begrip 'oneindig' (of 'oneindig klein') vermeden wordt. Het heeft tegelijk het grote nadeel dat een bewijs uit het ongerijmde nodig is, hetgeen alleen mogelijk is als het resultaat al bekend is. Nu is voor de cirkel de analyse die tot het resultaat leidt, niet zo heel ingewikkeld. Dit wordt bijvoorbeeld mooi toegelicht door een figuur uit 1698 van de Japanse wiskundige Satō Moshun (Figuur 3), die aannemelijk maakt dat  $\mathcal{C}$  is gelijk aan de rechthoek met als zijden de straal en halve omtrek van  $\mathcal{C}$ ; deze figuur is hier gekozen wegens haar visuele aantrekkelijkheid – historisch gezien is het waarschijnlijk dat Satō Moshun het werk van Archimedes gekend heeft – en deze figuur is dus geen analyse vooraf maar een toelichting achteraf bij het resultaat van Archimedes. Maar bij andere figuren, zoals bijvoorbeeld het parabolsegment, ligt het meteen veel ingewikkelder.

### 3. DE QUADRATUUR VAN HET PARABOOLSEGMENT

In de zestiende eeuw werden de eerste werken van Archimedes gedrukt. Na enkele publicaties van losse werken, verschenen verzamelde werken in Basel (1544) verzorgd door Venetorius en in Venetië (1558) door Commandino. In deze edities bevond zich een tractaat over *De quadratuur van het parabolsegment*. De vraag was: als van een parabool door een koorde een segment wordt afgesneden, wat is daarvan dan de quadratuur?

Archimedes gebruikt in zijn oplossing wat we zullen noemen de *standaard-ingeschreven* driehoek: bij het segment dat door koorde  $AB$  is afgesneden zijn  $A$  en  $B$  hoekpunten van deze driehoek, en het derde hoekpunt  $\mathcal{C}$  is het punt waar de raaklijn aan de parabool evenwijdig is aan  $AB$ . Archimedes bewees:

**Stelling 3.1.** *Het parabolsegment is  $\frac{4}{3}$  keer de standaard-ingeschreven driehoek.*

Het principe van het bewijs zal intussen duidelijk zijn: Archimedes werkt ook in dit geval met uitputting. Maar het raadsel zit hem in de vraag hoe hij aan het resultaat gekomen is, want dat ligt hier niet zo voor de hand als in het geval van de cirkel. Straks meer hierover, eerst de grote lijnen van het bewijs van de stelling.

*Bewijs.* In Figuur 4 is  $V$  het midden van koorde  $qQ$ , en  $PV$  is de toegevoegde middellijn. Deze deelt alle koorden die evenwijdig zijn aan  $qQ$  middendoor, en in het eindpunt  $P$  is de raaklijn evenwijdig aan  $qQ$ . Driehoek  $qQP$  is dus de standaard-ingeschreven driehoek bij koorde  $qQ$ . Tevens zijn bij de koorden  $qP$  en  $PQ$  de standaard-ingeschreven driehoeken  $qPr$  en  $PQR$  getekend.



FIGUUR 3. De cirkelmeting volgens Satō Moshun, overgenomen uit [10]

Uit de eigenschappen van de parabool leidt Archimedes eerst af dat

$$\text{opp. } \triangle PQq = 8 \times \text{opp. } \triangle PQR.$$

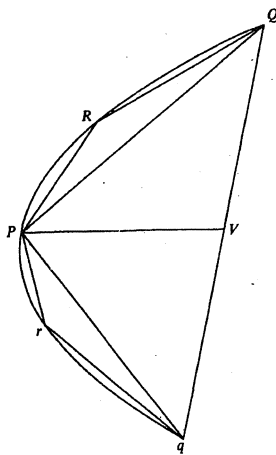
Dit geldt ook als je  $\triangle qPr$  neemt in plaats van  $\triangle PQR$ . De twee standaard-ingeschreven driehoeken in de tweede stap zijn dus samen een vierde deel van  $\triangle PQq$  uit de eerste stap. De oppervlakte van de in het segment ingeschreven veelhoek in de  $n$ -de stap is dus

$$\left(1 + \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \dots + \left(\frac{1}{4}\right)^n\right) \times \triangle PQq.$$

Voor deze meetkundige reeks leidt Archimedes af dat de som gelijk is aan

$$\frac{4}{3} \times \triangle PQq - \frac{1}{3} \times \left(\frac{1}{4}\right)^n \times \triangle PQq.$$

Stel nu, zegt hij, dat het segment groter is dan  $\frac{4}{3} \times \triangle PQq$ . Ga dan zo ver met de ingeschreven figuur dat die ook groter is dan  $\frac{4}{3} \times \triangle PQq$ . Dat kan, want het



FIGUUR 4. De quadratuur van het parabolsegment via uitputting

verschil tussen segment en veelhoek neemt per stap met minstens de helft af, dus als het segment groter is dan  $\frac{4}{3} \times \triangle PQq$ , dan is de ingeschreven veelhoek dat vanaf een bepaalde stap ook. Het geeft wel een tegenspraak, want de oppervlakte van elke ingeschreven figuur is de som van de meetkundige reeks, en die is minder dan  $\frac{4}{3} \times \triangle PQq$ .

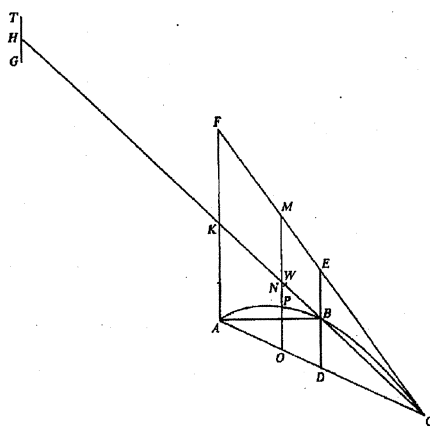
Dan zou het segment nog kleiner kunnen zijn dan  $\frac{4}{3} \times \triangle PQq$ . De ingeschreven veelhoek zou dan zeker ook kleiner moeten zijn dan  $\frac{4}{3} \times \triangle PQq$ . Maar door ver genoeg door te gaan met het opvullen komt er een ingeschreven veelhoek waarvan de oppervlakte zo dicht tegen  $\frac{4}{3} \times \triangle PQq$  aan ligt dat hij groter is dan het parabolsegment. Dat geeft ook een tegenspraak.

De enige mogelijkheid die overblijft is: opp. segment  $PQq = \frac{4}{3} \times \triangle PQq$ .  $\square$

Stelling en bewijs hebben wiskundigen eeuwenlang voor raadselen geplaatst. Je kunt deze uitputting namelijk alleen gebruiken als je al weet aan welke, door rechte lijnen begrensde figuur het segment gelijk is, want anders weet je niet welke twee alternatieven je moet uitsluiten. Hoe kwam Archimedes aan zijn resultaten?

Zoals het vaker gaat met historische vragen kwam het antwoord doordat er in 1906 in Constantinopel een nieuw document opdook. Het was een religieuze tekst uit de 13de eeuw, maar die bleek geschreven te zijn op perkament waar in de 10de eeuw al eerder op geschreven was, in Byzantijns Grieks, een zogenaamde *palimpsest*. De Deense classicus en wiskundige Johan Ludvig Heiberg had begrepen dat de oudste, weggekrabde tekst nog net leesbaar was, en dat het een wiskundig document was. Hij kreeg een beurs, trok naar Constantinopel en ontcijferde met veel moeite en grote deskundigheid een onbekend werk van Archimedes, de *Methode van de mechanische stellingen* (1907). Daarin bleek dat Archimedes oppervlakte (inhoud) onderzocht door de figuren (lichamen) in

strookjes (schijfjes) op te delen, en die aan een balans in evenwicht te brengen met figuren (lichamen) waarvan hij de oppervlakte (inhoud) al kende. Onder meer bevatte het nieuwe Archimedes-handschrift, dat trouwens op 29 oktober 1998 door Christie's in New York voor 2 000 000 US Dollar geveild werd, een brief van Archimedes aan Eratosthenes over de manier waarop hij de quadratuur van het parabolosegment had ontdekt.



FIGUUR 5. Quadratuur via evenwicht

De figuur bij Archimedes' analyse (Figuur 5) is als volgt opgebouwd. Eerst is er segment  $ABC$  van een parabool;  $B$  op de parabool is het eind van de toegevoegde middellijn  $BD$ ,  $D$  is dus het midden van koorde  $AC$ . Verder zijn in  $C$  de raaklijn aan de parabool en in  $A$  de lijn evenwijdig aan  $BD$  (en dus ook evenwijdig aan de as van de parabool) getrokken. Deze lijnen snijden elkaar in  $F$ , het verlengde van  $BD$  snijdt de raaklijn in  $E$ . Het verlengde van  $CB$  snijdt  $AF$  in  $K$ , en wordt nog een keer verlengd met een stuk  $KH$  dat gelijk is aan  $CK$ . Archimedes beschouwt  $CH$  als een balans, met het midden  $K$  als draaipunt. Vervolgens deelt hij het parabolosegment op in lijnstukken evenwijdig aan  $DB$ . Een ervan, willekeurig gekozen, is  $PO$ . Het verlengde van  $PO$  snijdt de raaklijn in  $M$  en  $CK$  in  $N$ . Nu past hij zijn kennis van de theorie van kegelsneden toe. Hieraan ontleent hij een **lemma** dat we hier verder niet zullen bewijzen, maar dat bijvoorbeeld analytisch direct te verifiëren is:

**Lemma 3.2.**  $MO : OP = CA : AO$ .

Dit geeft, toegepast op  $ED$  ook:  $ED : BD = AC : AD = 2 : 1$ , zodat  $EB = BD$ , waaruit weer volgt:  $MN = NO$  en  $FK = KA$ .

Uit het lemma en gelijkvormigheid volgt:  $MO : OP = CK : KN = HK : KN$ . Hang nu een lijnstuk, even lang als  $PO$  in  $H$  aan de balans (Archimedes noemt dit lijnstuk  $TG$ ). Omdat  $OP \times HK = MO \times KN$  maakt  $PO$  (hangend in  $H$ ) evenwicht met  $MO$  (hangend in  $N$ ). Je kunt dus lijnstuk voor lijnstuk

de parabool ophangen in  $H$ , en houdt evenwicht met de verzameling van alle lijnstukken  $MO$ , als die in hun midden vastgemaakt zijn aan  $KC$ .

Dan vervangt Archimedes de verzameling van alle lijnen  $MO$  door de hele driehoek  $ACF$ , die hij in het zwaartepunt  $W$  aan  $KC$  vastmaakt.  $W$  ligt dus zo op de zwaartelijn  $KC$  dat  $CW : WK = 2 : 1$ , en de arm  $KW$  waaraan driehoek  $ACF$  hangt is dus  $\frac{1}{3}KC$ . Dus:

$$\triangle ACF : \text{segment } \triangle ABC = 3 : 1$$

maar  $\triangle ACF = \triangle ACK$  (omdat  $K$  het midden is van  $AF$ ) en dus  $\triangle ACF = 4 \triangle ABC$  (omdat  $B$  het midden is van  $CK$ ). Met elkaar gecombineerd geeft dit

$$\text{segment } ABC = \frac{1}{3} \triangle ACF = \frac{4}{3} \triangle ABC$$

Maar Archimedes zag dit niet als bewijs, want “deze redenering geeft een soort bevestiging dat de conclusie juist is. Omdat we zien dat de stelling niet bewezen is, maar we wel de indruk hebben dat de conclusie juist is, moeten we ons baseren op een meetkundig bewijs dat ik zelf ontdekt en reeds bekend gemaakt heb.”

Balanceren van lijnstukken mag je wel op een idee brengen, maar je moet dat idee vervolgens met een meetkundig bewijs hard maken. Experiment gaat bij Archimedes dus samen met het streven naar formele correctheid.

#### 4. BOL, KEGEL EN CILINDER

Dat inhoud van de cilinder drie maal de inhoud van de kegel met zelfde grondvlak en hoogte is, had Euclides al bewezen, langs de officiële weg der uitputting (*Elementen* boek XII, stelling 10). Archimedes gaat er met zijn mechanische methode mee verder, en vindt de relatie tussen de inhoud van een bol en de inhoud van de kegel die een grote cirkel van de bol als grondcirkel heeft en de straal van de bol als hoogte.

Deze analyse volgt hier, als toegift, in een tekst voor 5/6 VWO, bewerkt door Anne van Streun (Figuur 6).

#### LITERATUUR

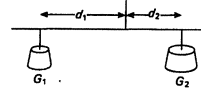
1. BARON, M. (1969). *The origins of the infinitesimal calculus*. Oxford etc.: Pergamon Press.
2. DIJKSTERHUIS, E.J. (1938). *Archimedes*. (eerste deel) Groningen-Batavia: P. Noordhof.
3. EDWARDS, C.H. JR. (1979). *The historical Development of the Calculus*. New York etc.: Springer.
4. FAUVEL, J.; GRAY, J. (1987). *The History of Mathematics – a reader*. Basingstoke etc.: MacMillan.
5. GRAY, S.I.B. (1999). ‘Heiberg and the Lost Palimpsest of Archimedes.’ *Newsletter of the British Society for the History of Mathematics* **39** (Summer 1999), 26–31.

## ONDERZOEK Toepassingen van de integraalrekening

redeneringen. Hij had niet de beschikking over de algebraïsche notaties die wij af en toe gebruiken om de redeneringen duidelijk te maken. Archimedes maakte veelvuldig gebruik van de 'hefboomwet' die in het kort op het volgende neerkomt:

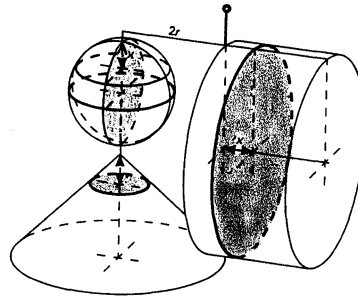
Als een balans in evenwicht is geldt:

$$d_1 \cdot G_1 = d_2 \cdot G_2$$



Deze wetmatigheid gebruikte hij om de inhoud van een bol te berekenen.

Er was reeds bekend dat de inhoud van een cilinder met grondvlak  $G$  en hoogte  $h$  gelijk was aan  $G \cdot h = \pi r^2 h$ , als  $r$  de straal is. Tevens wist men dat de inhoud van een kegel met grondvlak  $G$  en hoogte  $h$  gelijk was aan  $\frac{1}{3}Gh = \frac{1}{3}\pi r^2 h$ , waarbij  $r$  de straal van de grondcirkel is.



In de getekende balansopstelling van Archimedes hangen een bol met straal  $r$  en een kegel met hoogte = straal v.d. grondcirkel =  $2r$  aan een punt op afstand  $2r$  van het ophangpunt. Een cilinder met hoogte = straal =  $2r$  is vast bevestigd aan de andere balansarm.

- Toon aan dat de drie gekleurde schijfjes evenwicht maken voor elke waarde van  $x$ .
- Toon aan dat (bij gelijke dichtheid van de lichamen) uit het evenwicht volgt:  $2r \cdot (V_{\text{bol}} + V_{\text{kegel}}) = r \cdot V_{\text{cilinder}}$ . Leid hieruit de inhoudsformule af voor een bol met straal  $r$ .

36

FIGUUR 6. De methode van Archimedes bestemd voor 5/6 VWO, [11, p. 36]

- HEATH, T.L. (1897). *The works of Archimedes*. Cambridge: Cambridge University Press.
- MAROLOIS, SAMUEL. (1629). *Geometrie Ofte Meet-const.* Amsterdam.
- RUDIO, F. (1892). *Archimedes*, Huygens, Lambert, Legendre: 4 Abhandlungen über die Kreismessung Leipzig: Teubner.
- SATŌ MOSHUN. (1698). Tengen Shinan.
- SMITH, D.E.; MIKARNI, Y. (1914). *A history of Japanese mathematics*, Chicago: Open Court.
- BOS, D.J.P. (e.a.) (1992). *Wiskundelijn 516 vwo, Wiskunde B Analyse*. Groningen: Jacob Dijkstra.



## Snel Oplossen Is een Experiment Waard

Henk A. van der Vorst  
Universiteit Utrecht  
e-mail: vorst@math.uu.nl

Bij wetenschappelijk rekenwerk wordt de meeste rekentijd verstoekt aan het oplossen van stelsels lineaire vergelijkingen. Die stelsels kunnen behoorlijk groot van omvang zijn, zoals bijvoorbeeld bij stromingsproblemen waarbij elke vergelijking weergeeft hoe de waarde van een lokale (onbekende) grootte (zoals de stroomsnelheid) afhangt van de waarden in de directe omgeving. Men beperkt zich daarbij tot waarden over een van te voren vastgelegd rooster en het aantal roosterpunten bepaalt de omvang van het lineaire stelsel. In Figuur 1 zien we zo'n rooster schematisch aangegeven voor het doorrekenen van oceaanstromingen. Bij elk knooppunt hoort een onbekende en een vergelijking waarin deze onbekende gekoppeld wordt aan onbekenden uit de naaste omgeving in het rooster. Omdat we meestal veel roosterpunten nodig hebben om een redelijke nauwkeurigheid te halen, hebben we dus ook veel vergelijkingen. Een prettige bijkomstigheid is wel dat er in elke vergelijking maar een paar onbekenden voorkomen. De matrix van het stelsel staat dus grotendeels vol met nullen. We zullen daar later in dit verhaal nog dankbaar gebruik van maken.

We zien dat het rooster in de figuur uit een viertal verschillend aangegeven deelroosters bestaat. Dat komt omdat dit model in onze Utrechtse groep daadwerkelijk gebruikt is om oceaanstromingen door te rekenen, om te zien hoe onze oplosschema's zich gedroegen. Het probleem was voor een enkele computer echter te groot en daarom werd het rekenprobleem over vier samenwerkende computers verdeeld. De rekenschema's die we zullen introduceren lenen zich redelijk voor een dergelijke parallele verwerking.

Zoals aanstonds duidelijk zal worden leidt het bepalen van de onbekenden uit grote selsels problemen tot zeer omvangrijk rekenwerk. De aanpak via directe (Gauss-)eliminatie van onbekenden is vaak onaantrekkelijk, dat zag Gauss in 1823 ook al in, zij het om andere redenen dan pure rekensnelheid [2]. Hij stelde in dat jaar een iteratieve methode voor om stelsels van 4 vergelijkingevergelijkingenn met 4 onbekenden, die voortkwamen uit driehoeksmetingen, nauwkeurig op te lossen.

Laten we eerst maar eens kijken hoe dat elimineren ook al weer in zijn werk gaat. We bekijken als voorbeeld het stelsel

$$\begin{bmatrix} 10 & 0 & 1 \\ \frac{1}{2} & 7 & 1 \\ 1 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 21 \\ 9 \\ 8 \end{bmatrix}.$$



FIGUUR 1. Het rekenrooster voor een oceaanstroming

Het veeg-proces verloopt als volgt. Men trekt  $\frac{1}{20}$  keer de eerste rij van de tweede rij af en daarna nog eens  $\frac{1}{10}$  keer de eerste rij van de derde. Daardoor zijn er in de eerste kolom nullen gekomen en ziet het stelsel er als volgt uit

$$\begin{bmatrix} 10 & 0 & 1 \\ 0 & 7 & 1 - \frac{1}{20} \\ 0 & 0 & 6 - \frac{1}{10} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 21 \\ 9 - \frac{21}{20} \\ 8 - \frac{21}{10} \end{bmatrix}.$$

Toevallig staat er nu ook in de tweede kolom een 0 onder het diagonaal element en kan het stelsel zonder veel moeite opgelost worden. Dat leidt tot de oplossing  $x_3 = 1$ ,  $x_2 = 1$ , en  $x_1 = 2$ . Merk op dat we exact gerekend hebben. Dat is in dit geval niet zo lastig omdat er ‘mooie’ getallen uitkomen. Echter, wanneer er, zoals in meer realistische situaties, getallen met een paar decimalen achter de punt optreden, dan kan het exact rekenen tot zeer gecompliceerd rekenwerk aanleiding geven. De kans op rekenfoutjes is niet denkbeeldig en het is dan na controle lastig om op te sporen waar de rekenfout zit. Gauss had daar ook veelvuldig last van.

Gauss had een behoorlijk fysisch inzicht en wist dat de gewenste oplossingen voor zijn stelsels (voortvloeiend uit driehoeksmeting) componenten van dezelfde grootteorde hadden. Omdat in zijn stelsels ook de diagonaalelementen sterk overheersten, komt de grootste bijdrage van de oplossing in het rechterlid van de componenten die met een diagonaal element vermenigvuldigd zijn; in ons voorbeeld dus  $10x_1$ ,  $7x_2$ , en  $6x_3$ . Dat betekent dat als we in het stelsel de



buitendiagonaal coëfficiënten verwaarlozen, dus

$$\begin{bmatrix} 10 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 21 \\ 9 \\ 8 \end{bmatrix},$$

we van het verstoorde stelsel een redelijke benadering voor de oplossing van het ongestoorde stelsel mogen verwachten, in ons geval:  $x_1 = 2.1$ ,  $x_2 = \frac{9}{7}$ , en  $x_3 = \frac{8}{6}$ . Dat is inderdaad een, zij het wat grove, benadering voor de gewenste oplossing. Deze vorm van benadering staat tegenwoordig ook wel bekend als de Gauss-Jacobi benadering, omdat de wiskundige-astronoom Jacobi er ook gebruik van maakte bij het berekenen van baanafwijkingen voor de planeten in ons zonnestelsel.

Gauss pakte het nog wat slimmer aan. We kunnen het oorspronkelijke stelsel wat meer intact laten door alleen de coëfficiënten in de rechter bovendriehoek door nullen te vervangen. Dat leidt tot het stelsel

$$\begin{bmatrix} 10 & 0 & 0 \\ \frac{1}{2} & 7 & 0 \\ 1 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 21 \\ 9 \\ 8 \end{bmatrix},$$

en dit stelsel heeft als oplossing  $x_1 = 2.1$ ,  $x_2 = \frac{7.95}{7}$ , en  $x_3 = \frac{5.9}{6}$ . Dit is inderdaad nauwkeuriger (hoewel dat niet altijd hoeft te gelden; er zijn situaties waarbij deze aanpak geen verbetering geeft).

Al met al hebben we voor ons kleine probleempje tegen een slechts geringe besparing een beroerdere oplossing gekregen. Voordat we verder gaan, past daarom een beschouwing over de kosten van deze verschillende aanpakken. Voor een stelsel met  $n$  vergelijkingen en  $n$  onbekenden zijn we bij het vegen van de eerste kolom  $2(n-1)^2$  rekenoperaties kwijt (als we geen gebruik maken van de nullen die in het stelsel zelf al voorkomen). Daarna, voor de tweede kolom, bedragen de rekenkosten  $2(n-2)^2$  operaties. Hieruit volgt dat voor het vegen van de gehele onderdriehoek, de rekenkosten ongeveer  $\frac{2}{3}n^3$  bedragen. Het oplossen van het resterende bovendriehoeksstelsel kost dan nog ruwweg  $n^2$  operaties, en voor grotere  $n$  valt dat relatief in het niet. De kosten voor het verkrijgen van de exacte oplossing lopen dus op met de derde macht van  $n$ . Men gaat gemakkelijk na dat de benaderingen die met het benedendriehoeksgedeelte van het stelsel berekend worden, verkregen worden tegen kosten die evenredig zijn met de tweede macht van  $n$ . Voor grote problemen ligt hier dus een enorm voordeel voor het grijpen.

Echter: Hoe nu deze benaderde oplossing goedkoop te verbeteren? Ook daar had Gauss natuurlijk aan gedacht. Om zijn aanpak te beschrijven, gaan we over op matrixotatie. Deze was in ons' tijd nog niet bekend en dat maakte de beschrijving van zijn rekenaanpak wat omslachtig. We schrijven het stelsel als

$$Ax = b,$$

met

$$A = \begin{bmatrix} 10 & 0 & 1 \\ \frac{1}{2} & 7 & 1 \\ 1 & 0 & 6 \end{bmatrix},$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \text{en} \quad \begin{bmatrix} 21 \\ 9 \end{bmatrix}.$$

Het benedendriehoeksgedeelte van  $A$  noteren we met  $L$ :

$$L = \begin{bmatrix} 10 & 0 & 0 \\ \frac{1}{2} & 7 & 0 \\ 1 & 0 & 6 \end{bmatrix},$$

en dan is de benaderde oplossing dus verkregen uit het oplossen van

$$L\tilde{x} = b.$$

Voor verbetering van deze oplossing zoeken we naar het ontbrekende deel  $\Delta x$ :

$$A(\tilde{x} + \Delta x) = b$$

en hiervoor geldt dus

$$A\Delta x = b - A\tilde{x} \equiv r.$$

Het ligt nu voor de hand om  $\Delta x$  ook weer met een Gauss benadering te berekenen, dus we bepalen  $\tilde{\Delta x}$  uit

$$L\tilde{\Delta x} = r$$

en we corrigeren met deze benadering onze eerste schatting  $\tilde{x}$ :

$$\tilde{\tilde{x}} = \tilde{x} + \tilde{\Delta x}.$$

We kunnen deze truc natuurlijk blijven herhalen en dat levert ons de volgende iteratieformule op

$$x^{(i+1)} = x^{(i)} + L^{-1}(b - Ax^{(i)}),$$

waarbij de vector  $y = L^{-1}(b - Ax^{(i)})$  berekend wordt door

$$Ly = b - Ax^{(i)}$$

op te lossen.

We voeren dit proces nu uit voor ons kleine stelseltje. Bij gebrek aan betere informatie starten we maar met  $x^{(0)} = 0$ . In Tabel 1 staan de resultaten voor de eerste drie iteratiestappen weergegeven.

TABEL 1. Resultaten voor drie Gauss-iteraties

stap	1	2	3
$x_1$	2.1000	2.0017	2.000028
$x_2$	1.1357	1.0023	1.000038
$x_3$	0.9833	0.9997	0.999995

Zoals we zien, neemt het aantal correcte decimalen met ongeveer 2 per stap toe. Dat is natuurlijk niet altijd het geval: het hangt af van de mate waarin de diagonaal in  $A$  overheerst.

De rekenkosten per iteratie stap bedragen ruwweg  $2n^2$  operaties (optellingen, aftrekkingen, vermenigvuldigingen) voor het berekenen van  $Ax^{(i)}$ , plus nog eens  $n^2$  voor het oplossen van het benedendriehoeksstelsel met  $L$ , dus totaal  $\approx 3n^2$  operaties per stap. De oplossing via de directe veegmethode vergt  $\approx \frac{2}{3}n^3$  operaties, dus als je met het benaderde resultaten na minder dan

$$\left(\frac{2}{3}n^3\right)/(3n^2) = \frac{2}{9}n$$

slagen tevreden bent, dan heb je rekenwinst geboekt.

Voor Gauss had het rekenen met deze iteratie methode ook enorme voordelen: het was weliswaar, voor  $n = 4$  niet zo efficiënt in termen van rekenoperaties als het directe oplossen, maar het grote voordeel voor hem was dat je de benaderingen niet nauwkeurig hoeft uit te rekenen. Reken- en afrondfouten corrigeren zichzelf als het ware in het verdere verloop van de iteratie. Aan het residu  $b - Ax^{(i)}$ , dat je toch moet uitrekenen, kan je in iedere stap zien hoe goed de gevonden benadering bij het gegeven stelsel past. Hij schreef dan ook opgetogen in een brief aan zijn collega Gerling dat je dit half in je slaap kon uitvoeren, danwel dat je onderwijl aan andere zaken kon denken.

Gauss had het geluk dat zijn stelsels sterk diagonaal dominant waren, waardoor bij hem het iteratieproces redelijk snel convergeerde, ongeveer zoals in het voorbeeld dat wij bekeken hebben. Bij de zeer grote lineaire stelsels, zoals die bijvoorbeeld in oceaanstromingen optreden, is dat vrijwel nooit het geval en dan convergeert de Gaussmethode akelig langzaam of zelfs helemaal niet. Men heeft dus al heel lang uitgekeken naar snellere alternatieven.

We keren terug naar de basisiteratieformule

$$\begin{aligned} x^{(i+1)} &= x^{(i)} + L^{-1}(b - Ax^{(i)}) \\ &= x^{(i)} + L^{-1}r^{(i)}, \end{aligned} \tag{1}$$

met  $r^{(i)} = b - Ax^{(i)}$ . We zullen verder, voor het gemak, maar aannemen dat de startvector  $x_0$  gelijk is aan de vector met allemaal nullen.

Door vergelijking (1) te vermenigvuldigen met  $-A$  en er links en rechts  $b$  bij

op te tellen, krijgen we

$$\begin{aligned} b - Ax^{(i+1)} &= b - Ax^{(i)} - AL^{-1}r^{(i)} \\ r^{(i+1)} &= r^{(i)} - AL^{-1}r^{(i)} \\ &= (I - AL^{-1})r^{(i)}. \end{aligned} \tag{2}$$

Vergelijking (2) geeft een relatie tussen twee opeenvolgende resiductoren en we kunnen deze formule herhaald toepassen. We krijgen dan

$$\begin{aligned} r^{(i+1)} &= (I - AL^{-1})^{i+1}r^{(0)} \\ &= (I - AL^{-1})^{i+1}(b - Ax^{(0)}) \\ &= (I - AL^{-1})^{i+1}b. \end{aligned}$$

Kennelijk kan het residu in de  $i + 1$ -ste stap geschreven worden als allerlei machten van  $AL^{-1}$  maal de vector  $b$ . Merk op dat we nooit, maar dan ook echt nooit, de matrix  $L^{-1}$  zelf uitrekenen. Dat zou veel te inefficiënt zijn. Omdat deze matrix steeds voorkomt in een product dat wordtwordt losgelaten op een vector, komt het er dus op neer dat we in staat moeten zijn de vector  $y = L^{-1}z$  uit te rekenen voor zekere  $z$ . Dat gaat op de inmiddels bekende manier, door  $y$  op te lossen uit

$$Ly = z.$$

We kijken nu nog eens naar de oorspronkelijke iteratie en we zien dat we deze herhaald kunnen toepassen, zodat we de volgende vorm overhouden

$$\begin{aligned} x^{(i+1)} &= x^{(i)} + L^{-1}r^{(i)} \\ &= x^{(i-1)} + L^{-1}r^{(i-1)} + L^{-1}r^{(i)} \\ &\vdots \\ &= x^{(0)} + \sum_{j=0}^i L^{-1}r^{(j)} \\ &= \sum_{j=0}^i L^{-1}r^{(j)}. \end{aligned} \tag{3}$$

We kunnen dat nog wat handiger herschrijven:

$$\begin{aligned} L^{-1}r^{(j)} &= L^{-1}(I - AL^{-1})^j b \\ &= (I - L^{-1}A)^j L^{-1}b. \end{aligned}$$

Uit vergelijking (3) leren we dat  $x^{(i+1)}$  geschreven kan worden als een combinatie van de vectoren

$$L^{-1}b, L^{-1}AL^{-1}b, (L^{-1}A)^2L^{-1}b, \dots, (L^{-1}A)^iL^{-1}b.$$

Zo'n ruimte, die opgespannen wordt door opeenvolgende machten van een matrix (in ons geval  $L^{-1}A$ ), losgelaten op een vaste vector (in ons geval  $L^{-1}b$ ),

heet een Krylov-ruimte, naar de Russische wiskundige *A.N. Krylov*. Krylov maakte voor zijn berekeningen in de dertiger jaren van de vorige eeuw gebruik van deze ruimten, maar het duurde tot rond 1950 voordat dat voor lineaire stelsels een beetje succesvol begon te worden.

Lanczos kwam in 1952 op het idee om te onderzoeken of de Krylov-ruimte betere benaderingen bevatte dan de benadering die door het Gauss-proces wordt opgeleverd [4]. Hij merkte op dat het werk dat gedaan moest worden om de basis vectoren te berekenen ongeveer net zo veel was als het werk dat in het Gauss-iteratie proces verstookt moest worden: namelijk voor uitbreiding van de basis is een operatie met  $A$  nodig en een keer oplossen met  $L$ . De vraag rijst wel meteen of deze basis geschikt is om mee te rekenen en Lanczos had al snel in de gaten dat de 'Krylov'-basis vectoren niet erg orthogonaal waren; erger nog, ze hebben na een aantal stappen meestal een zeer kleine hoek met elkaar. Hij kwam daarom op het idee om van meet af aan een orthogonale basis voor de Krylov-ruimte te construeren. Dat gaat als volgt. De eerste vector is  $L^{-1}b$  en deze delen we door zijn lengte:

$$v_1 = L^{-1}b / \|L^{-1}b\|_2.$$

Vervolgens laten we hier de operatie  $L^{-1}A$  op los

$$w_1 = L^{-1}Av_1.$$

We hebben nu twee vectoren en we trekken van  $w$  de projectie op  $v_1$  af<sup>1</sup>:

$$z_1 = w_1 - (w_1, v_1)v_1.$$

We normeren dit resultaat en dat levert ons de tweede orthogonale basisvector

$$v_2 = z_1 / \|z_1\|_2.$$

Dit proces zetten we voort. Stel dat we  $j$  stappen hebben uitgevoerd, dan hebben we dus de onderling orthogonale vectoren  $v_1, v_2, \dots, v_j$ . De volgende vector krijgen we door de operatie  $L^{-1}A$  op  $v_j$  los te laten en er vervolgens de projecties op alle voorgaande vectoren vanaf te trekken:

$$\begin{aligned} w_j &= L^{-1}Av_j \\ z_j &= w_j - \sum_{k=1}^j (w_j, v_k)v_k \\ v_{j+1} &= z_j / \|z_j\|_2 \end{aligned}$$

De set vectoren  $v_1, v_2, \dots, v_j$  vormt een set van onderling loodrechte basisvectoren (met lengte 1) voor de Krylov-ruimte van dimensie  $j$ . We moeten nu nog in deze ruimte de beste benadering voor de oplossing bepalen. Dit deed Lanczos op een zeer handige manier. We vatten de basis vectoren op als de kolommen van een matrix:

$$V_j = [v_1 | v_2 | \dots | v_j],$$

<sup>1</sup> Met  $(x, y)$  noteren we het inproduct tussen de vectoren  $x$  en  $y$

dan kunnen we het resultaat van het orthogonalisatieproces schrijven als

$$L^{-1}AV_j = V_{j+1}H_{j+1,j}, \quad (4)$$

met  $H_{j+1,j}$  een  $j+1$  bij  $j$ -dimensionale matrix met de volgende elementen

$$H_{j+1,j} = \begin{bmatrix} (w_1, v_1) & (w_2, v_1) & (w_3, v_1) & \cdots & (w_j, v_1) \\ \|z_1\|_2 & (w_2, v_2) & (w_3, v_2) & \cdots & (w_j, v_2) \\ 0 & \|z_2\|_2 & (w_3, v_3) & \cdots & (w_j, v_3) \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \ddots & (w_j, v_j) \\ 0 & 0 & 0 & 0 & \|z_j\|_2 \end{bmatrix}.$$

Nu merken we op dat elke vector, en dus ook de gezochte benaderde oplossing, in de  $j$  dimensionale Krylov ruimte geschreven kan worden als  $x^{(j)} = V_j y$ , waarbij  $y$  een vector is met  $j$  coördinaten. Nu moeten we nog aangeven wat we met beste benadering zullen bedoelen. Lanczos vond het prima als het residu dat bij die benadering past loodrecht op de Krylov-ruimte stond. Om die Krylov-ruimte bij het stelsel te laten passen, keek hij naar het residu voor de vergelijking

$$L^{-1}Ax = L^{-1}b.$$

Dan eisen we dus dat

$$L^{-1}b - L^{-1}Ax^{(j)} \perp \{v_1, v_2, \dots, v_j\},$$

ofwel

$$L^{-1}b - L^{-1}AV_j y \perp \{v_1, v_2, \dots, v_j\}.$$

Door nu keurig de loodrechtheid voor alle vectoren uit te schrijven en gebruik te maken van de observatie dat

$$L^{-1}b = \|L^{-1}b\|_2 v_1$$

en van vergelijking (4), komen we tot het verrassende resultaat dat de vector  $y$  oplossing is van het  $j$  bij  $j$  stelsel

$$\begin{bmatrix} (w_1, v_1) & (w_2, v_1) & (w_3, v_1) & \cdots & (w_j, v_1) \\ \|z_1\|_2 & (w_2, v_2) & (w_3, v_2) & \cdots & (w_j, v_2) \\ 0 & \|z_2\|_2 & (w_3, v_3) & \cdots & (w_j, v_3) \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \|z_{j-1}\|_2 & (w_j, v_j) \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_j \end{bmatrix} = \begin{bmatrix} \|L^{-1}b\|_2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Met de gevonden oplossing  $y$  construeren we de benaderde oplossing  $x_j$  simpelweg als  $x_j = V_j y$ .

Een aantal opmerkingen is hier op zijn plaats. Lanczos bedacht deze methode oorspronkelijk voor het geval dat  $L^{-1}A$  symmetrisch, of liever symmetriseerbaar, was. In dat geval kunnen er enorme vereenvoudigingen worden aangebracht vanwege het feit dat dan  $(w_k, v_j) = 0$  voor alle  $j < k - 1$ . Dat betekent een grote winst in efficiency. De resulterende methode is in zijn uiteindelijke vorm gepresenteerd in 1954 door Hestenes en Stiefel, onder de naam *geconjugeerde gradiënten* methode [3]. Pas veel later werd de niet-symmetrische variant populair. De meest bekende en gebruikte variant voor niet-symmetrische matrices is *GMRES*, in 1986 voorgesteld door Saad en Schultz [5].

```

Bereken  $r_0 = b - Ax_0$  voor start vector  $x_0$ 
for  $i = 1, 2, \dots$ 
  Los op  $z_{i-1}$  uit  $Kz_{i-1} = r_{i-1}$ 
   $\rho_{i-1} = r_{i-1}^* z_{i-1}$ 
  if  $i = 1$ 
     $p_1 = z_0$ 
  else
     $\beta_{i-1} = \frac{\rho_{i-1}}{\rho_{i-2}}$ ;
     $p_i = z_{i-1} + \beta_{i-1} p_{i-1}$ 
  endif
   $q_i = Ap_i$ 
   $\alpha_i = \frac{\rho_{i-1}}{p_i^* q_i}$ 
   $x_i = x_{i-1} + \alpha_i p_i$ 
   $r_i = r_{i-1} - \alpha_i q_i$ 
  stop indien residu klein genoeg
end;

```

FIGUUR 2. De geconjugeerdegradiëntenmethode

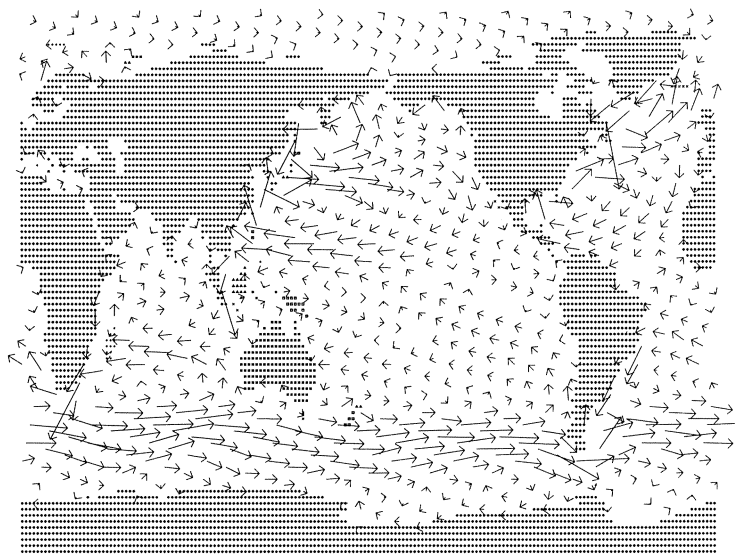
Laten we nog eens op een rijtje zetten wat we gedaan hebben:

- We zijn uitgegaan van de simpele Gauss-iteratie (ook wel Gauss-Seidel-iteratie genoemd).
- Die iteratie levert benaderingen in Krylov-ruimten op.
- Met net zoveel werk kan ook meteen een basis voor die Krylov-ruimten geconstrueerd worden.
- De constructie levert meteen de coëfficiënten van een klein stelsel op.
- Oplossing van dat kleine stelsel leidt tot een *optimale* benaderde oplossing in de Krylov-ruimte van de gegeven dimensie.

Dit lijkt tot ingewikkelde code aanleiding te zullen geven, maar we zien in Figuur 2 dat dit, voor de geconjugeerdegradiëntenmethode reusachtig meevalt.

Is deze aanpak nu werkelijk zoveel goedkoper dan directe oplosmethoden of dan de oorspronkelijke Gauss methode? Dat hangt af van het aantal iteratieslagen dat nodig is. Dat aantal hangt weer af van de gekozen basis-iteratie. In plaats van  $L$  kan je natuurlijk ook andere delen  $K$  van de matrix  $A$  afsplitsen, zolang dat maar een goedkoop te berekenen benadering oplevert. Dat gebeurt tegenwoordig vaak met onvolledig uitgevoerde veegprocessen. Hoe onvolledig je dat nog mag doen om toch een efficiënte methode te krijgen is niet bekend, het hangt op een onbekende manier van het probleem af. Dit aspect vereist veel ervaring en veel geëxperimenteer, maar het levert dan ook vaak groot succes op. De oceaanstromingen kunnen meer dan een factor  $n$  sneller opgelost worden in vergelijking met slim uitgevoerde veegmethoden (die gebruik maken van de vele nullen in het stelsel). Bij NASA rekende men ooit uit dat de krachtverdeling, die optreedt bij de lancering van de ruimteshuttle, met het Gauss-veegproces meer dan 520,000 jaar rekentijd op een supercomputer zou vereisen. Met de standaard Gauss-iteratie treedt vrijwel geen convergentie op en die is dus onbruikbaar, maar de geconjugeerdegradiëntenmethode, in combinatie met een onvolledig veegproces, vroeg slechts een tiental minuten rekentijd.

Voor het oceaanprobleem uit Figuur 1 kon het stromingsprofiel uiteindelijk in luttele seconden op een normaal werkstation berekend worden: de grootte van de pijlen in Figuur 3 geeft de sterkte van de stroming aan.



FIGUUR 3. De berekende oceaanstroming

De zogenaamde Krylov-iteratie methoden worden op uiteenlopende gebieden met succes toegepast: Oceaanstromingen, Oliereservoir simulatie, elektrische



circuits, mechanische constructies, tunnelboringen, vliegtuigontwerp, medische toepassingen, etc.

Ondanks het feit dat de methode steeds toegesneden moet worden op een bepaalde klasse van problemen bieden ze zoveel snelheidswinst dat men niet meer zonder kan. Er vindt nog steeds zeer veel onderzoek plaats op dit gebied. Er worden hele congressen gewijd aan het ontwerp van bruikbare afsplitsingen  $K$  van de matrix van het stelsel, en de publikaties op dit gebied behoren tot de meest geciteerde artikelen uit de wiskunde. Een artikel over een efficiënte Krylov-methode voor niet-symmetrische stelsels (in feite een combinatie van het geconjugeerdegradiënten-idee met de GMRES methode) was het meest geciteerde wiskunde artikel uit de jaren negentig [6].

Voor een goed overzicht van iteratieve methode en verdere verwijzingen zie het zogenaamde *Templates* boek [1].

#### LITERATUUR

1. R. BARRETT, M. BERRY, T. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE, and H. VAN DER VORST. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, Philadelphia, PA, 1994.
2. C. F. GAUSS. *Werke, Band IX*. Teubner, Leipzig, 1903.
3. M. R. HESTENESI and E. STIEFEL. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, **49**:409–436, 1954.
4. C. LANZOS. Solution of systems of linear equations by minimized iterations. *J. Res. Natl. Bur. Stand.*, **49**:33–53, 1952.
5. Y. SAAD and M. H. SCHULTZ. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, **7**:856–869, 1986.
6. H. A. VAN DER VORST. Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of non-symmetric linear systems. *SIAM J. Sci. Statist. Comput.*, **13**:631–644, 1992.





## Cabri Geeft Oude Meetkunde Tweede Jeugd

Aad Goddijn

Freudenthal Instituut, Universiteit Utrecht

e-mail: A.Goddijn@fi.uu.nl

### 1. INLEIDING

CABRI is een computerprogramma om te tekenen, meer niet. Okee, het is een handig gereedschap om de figuren die gewoonlijk de meetkundeboeken bevolken op scherm of papier te krijgen. Maar dat verandert niets aan de meetkunde waar die boeken eigenlijk over gaan.

Waar, allemaal waar. CABRI is inderdaad alleen maar gereedschap. Maar wel een gereedschap waardoor het makkelijker wordt splinternieuwe vermoedens in de vlakke meetkunde op te sporen, bewijzen ervoor te vinden, of nieuwe verbanden tussen oude bekenden te leggen. Daarover gaat dit verhaal: CABRI als onderzoek- en experimenteergereedschap in de vlakke meetkunde.

De voorbeelden in dit artikel zijn geselecteerd om juist dit gebruik van CABRI te illustreren. Het programma is ook nuttig bij volgzzaam bestuderen van 'bestaande' meetkunde, bij het maken van mooie tekeningen bij de grote klassieken als de Feurbachcirkel en de deltoïde van Steiner, maar in dit artikel zoeken we expliciet naar de méérwaarde van CABRI bij eigen, bescheiden of geavanceerd, onderzoek.

#### 1.1. Verkrijgbaarheid CABRI



De officiële naam van het programma en de ontwerpers staan uiteraard vermeld op het infoscherm dat bij starten van CABRI vertoond wordt. In Nederland wordt CABRI gedistribueerd door Daedalus Onderwijsproducties. Zie <http://www.educadvb.nl>

### 1.2. Inhoudelijk overzicht

Er zijn vier bijna los van elkaar staande delen.

#### *Deel I: Draaiende lijnen.*

Eigenlijk is de basis voor dit deel een stelling die een belangrijke plaats heeft in het Voortgezette Meetkunde programma van wiskunde B2: de stelling van de constante omtrekshoek. De stelling wordt bekeken en toegepast vanuit een standpunt dat dicht aansluit bij de dynamische mogelijkheden van CABRI waardoor bewijzen zich laten stroomlijnen.

#### *Deel II: Tweemaal onderzoek doen met CABRI*

Een recent probleem en een inleiding op een spectaculair stukje 19<sup>e</sup>-eeuwse vlakke meetkunde. In beide gevallen gaat het om meetkundige plaatsen of banen van bewegende punten. Bij het eerste geval is de baan onverwacht van vorm en is deze zelf onderwerp van onderzoek; in het tweede geval gebruiken we ‘meetkundige plaats’ als gereedschap bij zoeken naar een bewijs.

#### *Deel III: Hoe betrouwbaar is het JA van CABRI?*

Bij werken met CABRI vallen de mooie resultaten je soms zomaar in de schoot. Dit artikel bevat diverse voorbeelden daarvan. CABRI levert echter geen bewijzen, laat alleen verbanden zien. Maar als CABRI zegt dat drie punten op een lijn liggen, is dat dan zeker zo? Die betrouwbaarheidsvraag is belangrijk voor wie iets ontdekt. Het is frustrerend te zoeken naar een bewijs voor iets dat alleen maar waar lijkt en het niet is.

CABRI zegt in dit gedeelte JA tegen een bewering, waarvan wij al bij voorbaat weten dat het NEE moet zijn. Maar CABRI gaat niet zomaar de fout in; een achttiende-eeuws mechaniek uit de wereld van de stoommachine moet uit de kast komen om CABRI tot dit valse getuigenis te brengen. CABRI is eigenlijk heel betrouwbaar onder ‘normale’ omstandigheden.

#### *Deel IV: Wat is construeren nu eigenlijk?*

Van de oude Grieken moest het met passer en liniaal. Waarom eigenlijk? En wat gebeurt er als we de constructiemiddelen uitbreiden? In dit gedeelte construeren we de regelmatige zevenhoek, voeren we de trisectie van de willekeurige hoek uit en ‘gebruiken’ we de kwadratuur van de cirkel als nieuw constructiemiddel.

### 1.3. Doe het zelf!

De lezer wordt uitgedaagd met de voorbeelden zelf te gaan experimenteren. Daartoe zijn de CABRI-bestanden die gebruikt worden, beschikbaar gemaakt. Ze zijn op te halen op de downloadpagina van het wisweb, dat te bereiken is via de www-site van het Freudenthal Instituut: <http://www.fi.uu.nl>. Klik jezelf via *nederlands naar wisweb* en *downloads* en haal daar *CWI-2001-CABRI* op. De

gevonden bestandsnamen corresponderen met vermeldingen tussen blokhaken als [bestand.fig] in dit artikel.

Er is geen inleiding-CABRI-voor-beginners opgenomen. Zulke inleidingen zijn beschikbaar bij enkele van de grotere wiskundemethoden, zoals Moderne Wiskunde en Getal en Ruimte. Zie [18], [7]. Een zeer korte ‘hands-on’ inleiding is opgenomen in de bundel ‘Werkstukken wiskunde B’, [14]. Veel extra informatie is ook te vinden via de homepage van Dick Klingens, [11].

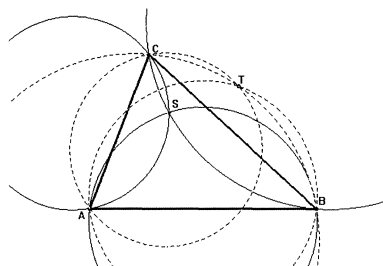
Bij enkele van de voorbeelden worden minder vanzelfsprekende mogelijkheden gebruikt; in die gevallen wordt verwezen naar constructietips. Deze staan aan het einde van dit artikel, vóór de literatuurverwijzingen. Bijvoorbeeld over HELP: zie constructietip 1.

## Deel I: Draaiende lijnen

### 2. KLEINE ONTDEKKING OVER GESPIEGELDE CIRKELS

Het volgende is bekend: als we de omgeschreven cirkel van een driehoek in de zijden van de driehoek spiegelen, dan gaan de drie spiegelbeelden door één punt en dat punt is het hoogtepunt van de driehoek.

Wat zou er gebeuren als we *uitgaan* van een willekeurig punt  $S$  in driehoek  $ABC$  – liever niet het hoogtepunt van  $ABC$  – en dan de cirkels door  $S$  waarvan de driehoekszijden koorden zijn in die koorden spiegelen? Was  $S$  wel het hoogtepunt van  $ABC$ , dan vonden we driemaal dezelfde omgeschreven cirkel, maar wat gebeurt in andere situaties? Het is typisch een vraag, die roept om CABRI-aanpak omdat het probleem dan snel getekend kan worden en er gekeken kan worden naar eventuele opmerkelijkheden.



FIGUUR 1.

Maar zie ook constructietip 2 en constructietip 3. De beloning kwam direct. De drie spiegelbeelden blijken verrassenderwijs door één punt te gaan,  $T$  in de figuur. Een fraai resultaat; het verscheen ‘toevallig’ bij het voorbereiden van deze cursus. Zonder CABRI was het al een hele toer geweest ook maar één ligging van deze configuratie nauwkeurig te tekenen. En dan nog zou het door één punt gaan van de drie cirkels bij  $T$  een bijna-toevalligheid kunnen zijn. Nú kan ik in de CABRI-tekening door verslepen van punten gemakkelijk vele

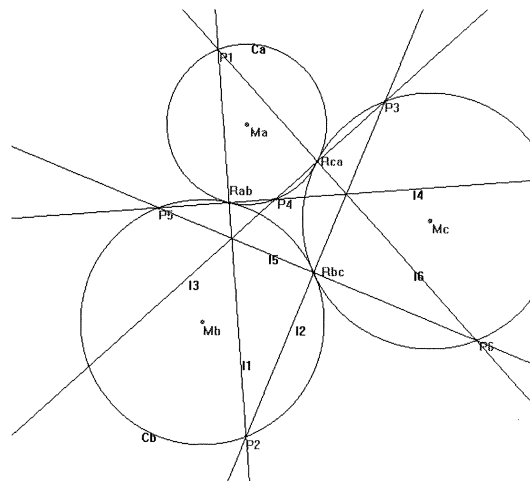
liggingen onder ogen krijgen.  $T$  dertelt het vlak over als we  $S$  verslepen, maar  $T$  is er steeds wél, als we tenminste afzien van voor de handliggende uitzonderingen, zoals  $S$  als hoogtepunt van de driehoek of  $S$  op de omschreven cirkel. Zo raak je overtuigd: er is iets nieuws ontdekt. Of iemand anders dat eerder heeft ontdekt? Ach, ik hoop het voor hem of haar, want zo'n ontdekkingsmoment is een bijzonder prettige ervaring, die je iedereen graag gunt.

Na de overtuigende experimenten – want dat waren het – komt de vraag: hoe bewijs je zo iets?

Ervaren meetkundepuzzelaars komen hier wel uit. Ze kennen hun gereedschap: hoeken en bogen zal het wel moeten worden en al snel staat de figuur vol tekenjes en extra lijnen waarbij hoeken vergeleken, opgeteld of afgetrokken worden, naar gelang de ligging van  $S$  ten opzichte van de driehoek.

Voor ons de vraag: kan CABRI hier helpen een bewijs te vinden en liefst een dat correct, overzichtelijk en volledig is? Dat kan, via de omweg van een CABRI-geïnspireerde dynamische versie van de hoeken-en-bogen-theorie. Ter inleiding een ogenschijnlijk totaal ander voorbeeld.

### 3. VENIJNIG STAARTJE BIJ DE 'DOORSTEEKCIRKELS'



FIGUUR 2.

In Figuur 2 zijn drie onderling rakende cirkel  $C_a$ ,  $C_b$  en  $C_c$  gegeven.  $P_1$  ligt – beweegbaar – op  $C_a$  en lijn  $l_1$  is getrokken door  $P_1$  en raakpunt  $R_{ab}$  van cirkels  $C_a$  en  $C_b$ .  $l_1$  snijdt  $C_b$  nogmaals in  $P_2$ . vanuit  $P_2$  wordt op dezelfde manier  $P_3$  op  $C_c$  bepaald.  $P_4$  ligt weer op  $C_a$  maar valt niet samen met  $P_1$ , maar doorgaan tot  $P_6$  en  $P_7$  loont:  $P_7$  blijkt weer wel op  $P_1$  uit te komen. Dat is bijzonder.

Dit was onderdeel van een praktische opdracht meetkunde met CABRI in een VWO 6 klas. Sommige leerlingen wisten de sluitingseigenschap te bewijzen door te kijken naar de pijlen  $M_a P_1$  en  $M_b P_2$ . Die zijn evenwijdig, maar tegengesteld gericht. Daarna is het bewijs dat  $P_7 = P_1$  niet moeilijk meer: er

treedt zes keer omkering op. Bij inwendig raken bij andere liggingen van de cirkels treedt geen omkering op van de pijlrichting, en zo weten we genoeg om ook bij andere en grotere kringen van rakende cirkels te voorspellen of sluiting na één of pas na twee rondgangen optreedt. Zie de volledig opdracht in [14] en een verslag in [8].

Sommige leerlingen merkten op dat de lijnen  $l_1, l_3$  en  $l_5$  door één punt gaan en dat  $l_2, l_4$  en  $l_6$  dat ook doen. Ze beten hun tanden stuk bij het zoeken naar een bewijs; dat kan gebeuren als er de kans wordt gegeven iets te ontdekken en er de aanwijzing ligt: probeer te bewijzen wat je ontdekt.

De onverwachte concurrentie van  $l_1, l_3$  en  $l_5$  nemen we nu onder de loep en gaan optimaal gebruik maken van de dynamische mogelijkheden van CABRI.

$P_1$  ligt vrij op cirkel  $C_a$ . Dat betekent dat we  $P_1$  over  $C_a$  kunnen rondslepen en dat we dat ook kunnen automatiseren. Als  $P_1$  rondgaat blijken de geheimzinnige snijpunten ook rond te lopen op een cirkel, de cirkel die door de drie raakpunten gaat. (Hoe dat zichtbaar wordt gemaakt? Gebruik constructietip 4.) Bovendien is  $Q_R$  steeds een middellijn van die cirkel. Een goed waarnemer ziet ook nog dat alle hoeken van de lijnen  $l_1$  tot en met  $l_6$  onderling constant zijn.

Door de dynamiek van CABRI te gebruiken is ons oorspronkelijk vermoeden dat  $l_1, l_3$  en  $l_5$  door één punt gaan, uitgegroeid tot een hele lijst vermoedens die bewezen moeten worden. Dat lijkt ongunstig, maar dat is het niet: al het ontdekte hangt uiteraard samen en van het een komt vast het ander.

Cirkels en bewegende lijnen die constante hoeken maken in punten op die cirkels; dat thema diepen we eerst wat uit.

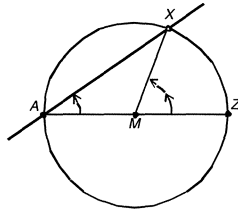
#### 4. HOEKEN EN BOGEN IN DYNAMISCHE HERINTERPRETATIE

Op de cirkel met middelpunt  $M$  van Figuur 3 ligt een vast punt  $A$  en een lopend punt  $X$ . Als  $X$  tegenkloks een hele ronde aflegt, draait de lijn  $AX$  om  $A$ , maar slechts over een halve ronde. Uiteindelijk ligt de lijn wél weer als geheel op zichzelf. Als  $X$  met vaste hoeksnelheid rondgaat, draait  $AX$  ook met vaste snelheid, zij het met halve snelheid. Dit alles is direct aan te tonen als we  $X$  tegenover  $A$  op de cirkel laten beginnen, in  $Z$  dus in Figuur 3, en hoek  $\angle XMZ$  met  $\angle XAZ$  vergelijken. Er geldt  $\angle XMZ = 2\angle XAZ$ , omdat  $\angle XMZ$  buitenhoek is van de gelijkbenige driehoek  $AMB$ .

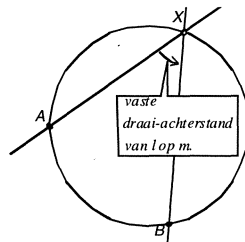
Vandaag houden we het op de dynamische beschrijving:

*Als  $A$  een vast punt op een cirkel is, en  $X$  een bewegend punt, dan is de hoeksnelheid van de om  $A$  draaiende lijn  $AX$  de helft van de hoeksnelheid van  $X$  over de cirkel.*

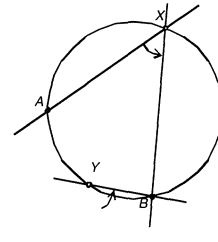
Ook als  $X$  het punt  $A$  passeert blijft dit gelden, al moeten we het voor een kort moment met de raaklijn in  $A$  tevreden zijn.



FIGUUR 3.



FIGUUR 4.



FIGUUR 5.

In Figuur 4 herhalen we het spel met twee vaste punten  $A$  en  $B$  en één lopend punt  $X$ . Nu draaien lijnen  $AX$  en  $BX$  met beide dezelfde snelheid om hun vaste punten  $A$  en  $B$  – dat volgt immers uit de dynamische beschrijving – en daarom moet de hoek bij  $X$  constant zijn. De draaiachterstand van de ene lijn op de ander zal immers constant zijn als beide lijnen met dezelfde snelheid ronddraaien.

Een bekend resultaat natuurlijk, maar wie het kent komt wellicht met een bezwaar, namelijk de ligging van Figuur 5. We gaan hier toch niet beweren dat de hoeken  $\angle AXB$  en  $\angle AYB$  gelijk zijn?

De hoeken, gemeten op de gewone manier, die zijn inderdaad niet gelijk. Maar dat is iets anders dan de hier beschreven ‘draaiachterstanden’! De beide pijlen laten de tegenklokse draaiachterstanden zien. Daarom toch maar deze fraaie ‘dynamische’ vorm:

*Als  $X$  over een vaste cirkel loopt die door  $A$  en  $B$  gaat, dan is de draaiachterstand van lijn  $XA$  op  $XB$  constant.*

De bedoelde draaiachterstand geven we voortaan aan met  $\Delta(AXB)$ . Lees dat als een beschrijving van wat bij  $X$  te zien is: hoeveel moet de lijn  $AX$  rond  $X$  draaien om op  $BX$  te vallen.

Met behulp van dat begrip kunnen we de **stelling van de constante omtrekshoek** mooi als volgt formuleren:

*Laten  $A, B$  en  $C$  drie punten zijn die niet op een lijn liggen. een punt  $X$  ligt dan en slechts dan op de cirkel door  $A, B$  en  $C$  als  $\Delta(AXB) = \Delta(ACB)$ .*

Drie opmerkingen:

- Het belangrijkste verschil met de traditionele stelling is dat we geen onderscheid hoeven te maken voor de twee bogen waarin  $AB$  de cirkel verdeelt. Anders gezegd: de koordenvierhoekstelling en de constante omtrekshoekstelling zijn dezelfde.
- Pas op, er geldt niet  $\Delta(AXB) = (BXA)$ . Integendeel,  $\Delta(AXB) = -\Delta(BXA)$ !
- De lezer die een betere onderbouwing wil van dit nogal intuïtieve verhaal heeft het exacte hart op de juiste plaats. In de volgende paragraaf volgen suggesties en literatuurverwijzingen in die richting, maar overslaan van de volgende paragraaf beïnvloedt de leesbaarheid van het daarna volgende niet.



## 5. THEORETISCH UITSTAPJE: GEORIËNTEERDE HOEKEN

Na de voorgaande verkenning is de volgende definitie voor de hand liggend. *De georiënteerde hoek van lijnen twee  $l$  en  $m$  is:*

- de draaiingshoek modulo  $\pi$  van een rotatie die  $l$  op  $m$  afbeeldt, als  $l$  en  $m$  elkaar snijden,
- 0, als  $l$  en  $m$  evenwijdig zijn.

We noteren de georiënteerde hoek van de lijnen  $l$  en  $m$  met:  $\Delta(l, m)$ .

Deze hoeken worden gerekend modulo  $\pi$  omdat na een draaiing over  $\pi$  de lijn  $l$  met de oorspronkelijke lijn samenvalt of ermee evenwijdig is.

Er moet nu wel het een en ander bewezen worden, of misschien had dat vóór de definitie al gemoeten. Bijvoorbeeld dat het er in de definitie niet toe doet welke rotatie gekozen wordt. Want er bestaan erg veel rotaties die  $l$  op  $m$  leggen!

Ook willen we zeker weten dat allerlei elementaire regels gelden. Bijvoorbeeld, alles modulo  $\pi$ :

- a.  $\Delta(l, m) = -\Delta(m, l)$ .
- b.  $\Delta(l, m) + \Delta(m, n) = \Delta(l, n)$ .
- c. als  $l$  en  $l'$  elkaar niet snijden geldt voor alle  $m$  :  $\Delta(l, m) = \Delta(l', m)$ .

b kunnen we met gebruikmaking van a ook herschrijven als

$$d. \Delta(l, m) + \Delta(m, n) + \Delta(n, l) = 0.$$

Een herformulering van de bekende stelling over de hoekensom van de driehoek, want de bewering geldt ook als de drie lijnen niet door een punt gaan.

Vaak zullen we  $\Delta(AB, CD)$  schrijven om de georiënteerde hoek van de lijnen  $AB$  en  $CD$  aan te geven. Het maakt niet uit welke punten van de lijnen worden genomen. Voor  $\Delta(AB, AC)$  zullen we kortweg schrijven  $\Delta(BAC)$ . Dat lijkt erg op de gewone hoeknotatie. Maar pas op,  $\Delta(BAC)$  en  $\Delta(CAB)$  zijn nu in het algemeen niet gelijk! Wie op de volgorde past, mag best ook  $\angle ABC$  noteren als echt  $\angle ABC$  bedoeld wordt en niet  $\angle CBA$ .

De stelling van de constante omtrekshoek ziet in de taal van de georiënteerde hoeken net zo uit als eerder geformuleerd.

Voor een systematische behandeling van georiënteerde hoeken zij verwezen we naar de uiterst grondige *Geometry* van M. BERGER [3]. Voor het krijgen van 'feeling' voor het werken met dit type hoeken: kijk naar uw computerscherm met geanimeerd draaiende beelden; zie weer constructietip 4.

De georiënteerde hoeken treden ook op natuurlijke wijze op als de vlakke cirkel-meetkunde met behulp van complexe getallen wordt beschreven. Een cirkel door de punten (complexe getallen)  $z_1, z_2$  en  $z_3$  bestaat uit juist de punten  $z$ , waarvoor de uitdrukking

$$\left(\frac{z - z_1}{z - z_2}\right) / \left(\frac{z_3 - z_1}{z_3 - z_2}\right)$$

een reële waarde heeft. De quotienten links en rechts van de schuine deelstreep staan voor argumentverschillen van de complexe getallen in teller en noemer; die geven elk dus een hoek aan. Alleen als de argumenten van linker en rechterbreuk gelijk zijn of  $\pi$  verschillen, is hun quotiënt reëel. Anders gezegd, maar slordig genoteerd: dan is  $\Delta z z_1 z_2 = \Delta z_3 z_1 z_2$ .

Een aantrekkelijke recente tekst over deze vorm van analytische meetkunde is [9].

#### 6. TOEPASSING OP DE GESPIEGELDE CIRKELS VAN FIGUUR 1

Als een cirkel  $c$  en twee punten  $A$  en  $B$  op die cirkel gegeven zijn, dan heeft  $\Delta(AXB)$  een vaste waarde als  $X$  op  $c$  ligt. Laten we die vaste waarde in dat geval ook maar aanduiden met  $\Delta(A, c, B)$ .

In het geval van Figuur 1 hadden we te maken met drie cirkels die gaan door respectievelijk  $A$  en  $B$ ,  $B$  en  $C$ ,  $C$  en  $A$ . Laten we ze  $c_{AB}$ ,  $c_{BC}$  en  $c_{CA}$  noemen.

Als  $X$  een willekeurig punt is, niet gelijk aan  $A$ ,  $B$  en  $C$ , dan geldt zeker:

$$\Delta(AXB) + \Delta(BXC) + \Delta(CXA) = 0.$$

Links staat immers de draaiing die  $AX$  via de tussenstanden  $BX$  en  $CX$  moet maken om op zich zelf uit te komen. Zonder veel moeite vinden we nu de **drie-cirkel-stelling**:

*Als  $c_{AB}$ ,  $c_{BC}$  en  $c_{CA}$  cirkels zijn respectievelijk gaande door  $A$  en  $B$ ,  $B$  en  $C$ ,  $C$  en  $A$ , dan geldt:*

*$c_{AB}$ ,  $c_{BC}$  en  $c_{CA}$  gaan dan en slechts dan door één punt, als*

$$\Delta(A, c_{AB}, B) + \Delta(B, c_{BC}, C) + \Delta(C, c_{CA}, A) = 0.$$

Het dan-en-slechts-dan van deze stelling hangt uiteraard nauw samen met dat van de constante hoekstelling; de scrupuleuze lezer gaat zelf wel na hoe dit zit.

De koppeling met het spiegelen van  $c_{AB}$  in  $AB$  moet nu nog gelegd worden. Geef het spiegelbeeld van  $c_{AB}$  in  $AB$  aan met  $\bar{c}_{AB}$ . Nu geldt  $\Delta(A, c_{AB}, B) = -\Delta(A, \bar{c}_{AB}, B)$ , want het spiegelen keert de draairichting om.

Als we in de situatie van Figuur 1 nu de drie cirkels door  $S$  maar weer  $c_{AB}$ ,  $c_{BC}$  en  $c_{CA}$  noemen, hebben we volgens de ene richting van de drie-cirkel-stelling direct

$$\Delta(A, c_{AB}, B) + \Delta(B, c_{BC}, C) + \Delta(C, c_{CA}, A) = 0$$

en voor de drie spiegelbeelden dus:

$$-\Delta(A, \bar{c}_{AB}, B) - \Delta(B, \bar{c}_{BC}, C) - \Delta(C, \bar{c}_{CA}, A) = 0.$$

Waarop de andere richting van drie-cirkel-stelling ons garandeert dat er een punt  $T$  is dat op de drie spiegelbeelden ligt. QED.

Het kost even moeite aan de nieuwe hoekbenadering en de drie-cirkel-stelling te wennen, maar het bewijs wordt er bijzonder overzichtelijk door. Laten we met enkele opgaven proberen of het beklijft!

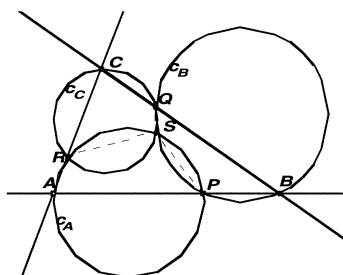
#### 7. OPGAVER EN TIPS BIJ 'DRAAIENDE LIJNEN'

1. Bewijs het 'venijn van de doorsteekcirkels'.

Tip: Laat eerst zien dat  $l_1$  en  $l_3$  een constante (draaings-, georiënteerde) hoek met elkaar maken. Gebruik  $(l_1, l_2) + (l_2, l_3) = (l_1, l_3)$ . Laat van de bijhorende cirkel zien dat die ook door  $R_{bc}$  gaat. Dat is het lastigst, maar nu kan feest gevierd worden door ook  $l_5$  in de rondedans uit te nodigen.

2. Bewijs met behulp van de drie-cirkel-stelling en de stelling van de constante omtrekshoek de bekende stelling van Miquel:

*Gegeven een driehoek  $ABC$  met een punt  $P$  ligt op de lijn  $AB$ ,  $Q$  op de lijn  $BC$  en  $R$  op de lijn  $CA$ . Zij  $c_A$  de cirkel door  $A, P$  en  $R$ ,  $c_B$  de cirkel door  $B, Q$  en  $P$ ,  $c_C$  de cirkel door  $C, R$  en  $Q$ . Dan gaan  $c_A, c_B$  en  $c_C$  door één punt.*



FIGUUR 6.

Merk op dat er geen enkele beperking is op de ligging van de punten  $P, Q$  en  $R$ , zolang ze niet samenvallen met  $A, B$  of  $C$ .

3. Teken een driehoek  $ABC$ . Geef de middens van de zijden aan met  $Ma, Mb$  en  $Mc$  ( $Ma$  op  $BC$ , etc.). Geef de voetpunten van de hoogtelijnen van  $ABC$  aan met  $Ha, Hb, Hc$ . (Idem:  $Ha$  op  $BC$ , etc.) Teken zes evenwijdige lijnen door deze zes punten en zorg dat ze 'gezamenlijk' draaibaar zijn [zesparall.fig].
  - a. Spiegel lijn  $AB$  in de lijn door  $Mc$ , lijn  $BC$  in de lijn door  $Ma$ , lijn  $CA$  in de lijn door  $Mb$ .  
Toon aan dat de drie spiegelbeelden door één punt – noem het  $P$  – gaan en toon aan dat  $P$  over de cirkel door  $Ma, Mb$  en  $Mc$  beweegt als de zes evenwijdige lijnen samen ronddraaien.
  - b. Spiegel de lijn door  $Ha$  in  $BC$ , de lijn door  $Hb$  in  $CA$ , de lijn door  $Hc$  in  $AB$ .  
Toon aan dat de drie spiegelbeelden door één punt – noem het  $Q$  – gaan en toon aan dat  $Q$  over de cirkel door  $Ha, Hb$  en  $Hc$  beweegt als de zes evenwijdige lijnen samen ronddraaien.

- c. (lastiger!) Laat zien dat de gevonden cirkels bij onderdeel **a** en **b** dezelfde zijn en dat de lijn  $PQ$  evenwijdig is aan de zes oorspronkelijke lijnen. (Opmerking: de cirkel is de zogenaamde *negenpunts*cirkel van  $ABC$ . Als  $H$  het hoogtepunt van de driehoek is, liggen de middens van  $HA$ ,  $HB$  en  $HC$  er ook op, vandaar *negenpunts*cirkel.)
- d. (véél lastiger) Laat zien dat  $PQ$  een van de lijnen van Wallace is van driehoek  $ABC$ .

### 7.1. Noot over de lijn van Wallace:

Neem een punt op de omgeschreven cirkel van driehoek  $ABC$  en laat loodlijnen vanuit dit punt neer op de zijden van  $ABC$ . De drie voetpunten liggen op een lijn, de zogenaamde lijn van Wallace. Zie [4], [1] en [10]. In de oorspronkelijke druk van [4] (Antiquarisch, N.V. Servire, Den Haag 1944) stond nog verklaard dat de snijpunten van de lijn van Wallace met de negenpunts cirkel zich precies gedragen als  $P$  en  $Q$  van opgave 3 en ook dat de omhullende van  $PQ$  een deltoïde is. Zie [wallace.fig]; gebruik de locusoptie, zie constructietip 4.

### 7.2. Over opgave 3

Opgave 3 is geïnspireerd door een artikel van Floor van Lamoen in de American Mathematical Monthly, [12]. daar worden voor de lijnen door de  $H$ -punten een andere richting genomen dan die door de  $M$ -punten, waardoor weer andere mogelijkheden ontstaan.

- 4 In de inleiding op het probleem van de drie gespiegelde cirkels spelen  $H$  (het hoogtepunt van  $ABC$ ) en de omgeschreven cirkel van  $ABC$  een rol. Zie de eerste regels onder het kopje ‘kleine ontdekking over gespiegelde cirkels’. Enkele verdergaande vragen over het verband tussen  $T$  en  $S$ .
- [cirkspieg2.fig] Vermenigvuldig de omgeschreven cirkel van  $ABC$  eens met een factor vanuit  $H$ . Bekijk de baan van  $T$  (de locus of meetkundige plaats) als  $S$  over zo’n cirkel loopt. Het vervormen van die baan als we de vermenigvuldigfactor wijzigen is instructief. Er is veel te onderzoeken hier!
  - [cirkspieg2.fig] Onderzoek waar  $T$  ligt als  $S$  op een van de hoogtelijnen van  $ABC$  ligt.
  - [cirkspieg2.fig] Laat  $S$  lopen over een lijn door het hoogtepunt van de driehoek. De meetkundige plaats ziet er eenvoudiger uit. Onderzoek de invloed van de richting van de lijn.
  - Teken het midden  $Z$  van  $TS$  en beweeg  $S$ .  $T$  en uiteraard  $Z$  bewegen dan mee. De beweging van  $T$  is ronduit grillig, maar let eens op  $Z$ !  $Z$  blijkt steeds op een vaste cirkel te liggen; de cirkel horend bij de factor  $1/2$  van onderdeel a van zoëven, de negenpunts cirkel. Geen tips hier, want op het moment van schrijven is er nog geen bewijs bekend bij de auteur dezes.

## 8. SLOTOPMERKING DRAAIENDE LIJNEN

We zijn begonnen met twee voorbeelden waarin iets te bewijzen viel met behulp van de hoeken en bogentheorie. Die hebben we – geïnspireerd door de mogelijkheden van CABRI – in dynamische vorm geherformuleerd. Daardoor waren beslist niet triviale stellingen overzichtelijk bewijsbaar geworden. Cabri stimuleerde in zekere zin deze theorievorming, of, bescheidener gezegd, het herontdekken van het zeer functionele begrip georiënteerde hoek.

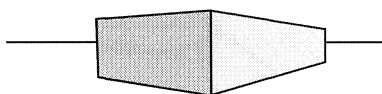
Het eerste voorbeeld, dat van de in de zijden gespiegelde cirkels  $SAB$ ,  $SBC$  en  $SCA$ , werd (althans door de auteur deze) muisgewijs gevonden, doordat CABRI zo makkelijk dergelijke experimenten toelaat. Dat deze drie cirkels door één punt  $T$  gaan is nu bewezen.

Een kwalijke (?) zaak is dat we gemakkelijk dingen ontdekken die op het scherm evident waar zijn, maar waar een bewijs niet zo gauw gevonden is. Zoals bij 4d!

## Deel II: Tweemaal onderzoek doen met Cabri

## 9. SYMMETRISCHE FOTO'S VAN EEN BERUCHTE LOODS

Vorig jaar stond een bijzondere foto in de krant van de beruchte ontplofte vuurwerkloods in Enschede. De loods was rechthoekig en niet vierkant, maar op de foto lagen linker- en rechterkant even ver uit het midden; ongeveer zoals in Figuur 7. Omdat er veel aandacht was voor de analyse van alle beelden van de ramp, vooral voor het exact bepalen van de feitelijke explosieplek, kwam er meer meetkunde dan gebruikelijk in de kranten terecht.

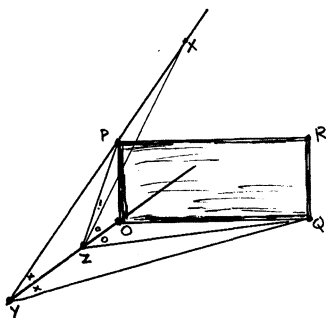


FIGUUR 7. Schets naar foto

In Figuur 8 is de rechthoekige (en niet vierkante) loods  $OQRP$  in bovenaanzicht geschetst. Hierin zijn de hoeken  $\angle PYO$  en  $\angle OYQ$  aan elkaar gelijk.  $Y$  is dus zo'n punt waarvandaan de foto gemaakt zou kunnen zijn. Karel Knip vroeg zich in zijn rubriek *Dagelijkse Wetenschap* van 3 juni 2000 af, of de punten  $Y$  met deze eigenschap op een rechte lijn liggen.

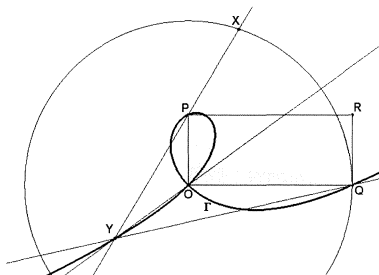
Daar kunnen we kort over zijn. Als het waar is, moet het de lijn door  $Y$  en  $O$  zijn, want vlak bij  $O$  (hoe vlak bij dan ook) liggen geschikte standpunten. Zou  $Z$  op  $OY$  nu óók zo'n punt kunnen zijn? Natuurlijk niet, want de hoeken  $\angle QZO$  en  $\angle OZP$  zijn ongelijk,  $\angle QZO$  is immers gelijk aan  $\angle OZX$ , waarbij  $X$  het spiegelbeeld van  $Q$  in  $OY$  is.

Niet op een rechte lijn dus! Maar waar liggen die punten  $Y$  dan wel? Het bijzondere punt  $X$  van het simpele schetsje dat de doorslag gaf maken we tot drievende kracht van een CABRI-schets. Zie Figuur 9.



FIGUUR 8. Bovenaanzicht

Na tekenen van de rechthoekige loods is *begonnen* met de cirkel rond  $O$  door  $Q$  en daar is punt  $X$  op gelegd.  $OX$  is in de schets van Figuur 8 gelijk aan  $OQ$ , vandaar dat  $X$  op die cirkel thuis hoort. De lijn  $XP$  en de middelloodlijn van  $XQ$  snijden elkaar nu in een voor de fotograaf geschikt punt  $Y$ .  $YQ$  is ter voltooiing van de schets toegevoegd, maar speelt voor het bepalen van  $Y$  natuurlijk geen rol.  $YQ$  en  $YP$  liggen symmetrisch [loads.fig] ten opzichte van  $YO$  en daarom is  $Y$  een geschikt punt. We tekenen de meetkundige plaats van  $Y$ , dwz. de baan die  $Y$  doorloopt als we  $X$  laten lopen. Zie constructietip 4.



FIGUUR 9. CABRI-constructie

De fotograaf kan alleen het stuk linksonder gebruiken, maar wij onderzoeken uiteraard de kromme als geheel. Het onderzoekspel met CABRI kan beginnen.

We onderzoeken eerst  $\Gamma$  als geheel, gewoon door  $X$  over de cirkel te schuiven, waardoor  $Y$  over  $\Gamma$  loopt. Bijzondere liggingen van  $Y$  worden dan zichtbaar, met bijhorende bijzondere liggingen van  $X$ .

- 5a. Waar moet  $X$  liggen om  $Y$  op de posities  $P, O, Q$  en 'oneindig ver' te plaatsen?
- b. Het is ook mogelijk de vorm van de rechthoek door slepen te veranderen, door bijvoorbeeld  $OP$  geleidelijk te vergroten.  $\Gamma$  past zich dan aan. Breng het vervormen van  $\Gamma$  globaal in kaart.
- c. In het vierkante geval splitst de kromme zich in een lijn en een cirkel. Welke lijn en cirkel? Waarom? Wat gebeurt er rond het overgangsmoment?

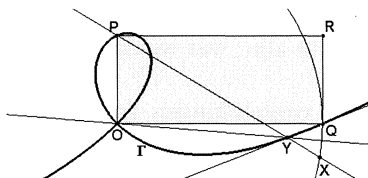
In Figuur 10 ligt  $X$  dicht bij  $Q$  en nu laten we  $X$  naar  $Q$  naderen. De lijn  $YQ$ , die het spiegelbeeld van lijn  $PY$  in  $OY$  is, nadert dan tot een limietstand. De volgende opgave gaat dan ook over raaklijnen aan  $\Gamma$  in enkele bijzondere punten.

- 6a. De raaklijn in  $Q$  aan  $\Gamma$  is de lijn door  $Q$  die evenwijdig is aan diagonaal  $OR$ . Vind daar een exact argument voor terwijl je  $X$  langzaam naar  $Q$  sleept.

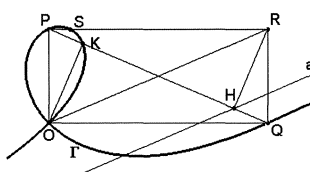
Dit is een aardig resultaat, en het is vlot gevonden door met het uitgangschetsje te spelen.

Verder met  $X$  schuiven en met ‘gelijksoortige’ middelen redeneren als hierboven:

- b. de raaklijn in  $P$  aan  $\Gamma$  is evenwijdig is aan  $OR$ ,  
 c. de twee raaklijnen in  $O$  aan  $\Gamma$  zijn de binnen- en buitendeellijnen van hoek  $\angle POQ$ .



FIGUUR 10.



FIGUUR 11.

We beginnen  $\Gamma$  nu al een beter te leren kennen, maar we gaan een stapje verder en de problemen worden wat uitdagender. In Figuur 11 zijn enkele nieuwe punten aangegeven:

- $S$ , het tweede snijpunt van  $\Gamma$  met rechthoekszijde  $PR$ ,
- $K$ , het voetpunt van de loodlijn uit  $O$  op  $PQ$  dat ook  $\Gamma$  ligt,
- $H$ , voetpunt van de loodlijn uit  $R$  op  $PQ$ .

Met ‘soortgelijke middelen’ kan nu aangetoond worden:

- d. de raaklijn in  $S$  aan  $\Gamma$  gaat door  $Q$ ,  
 e.  $\Gamma$  heeft een asymptoot, namelijk de lijn door  $H$  die evenwijdig is aan  $OR$ .

De laatste twee zijn minder eenvoudig dan de voorgaande onderdelen.

Voor de asymptoot volgt hier een forse tip. Stel eerst vast dat het lengteverschil van  $YP$  en  $YQ$  als  $Y$  ver weg loopt, nadert tot de lengte van  $KH$ . De bewering over de asymptoot klinkt nu al een stuk minder vreemd en met het oude deuntje ‘de deellijn uit een hoek van een driehoek verdeelt de overstaande zijde in de verhouding der aanliggende zijden’ in het hoofd kan het gaan lukken.

Wat die raaklijn in  $S$  betreft: misschien bent u al tevreden als u bewijst dat de hele lus  $PSKQ$  aan één kant van lijn  $SQ$  ligt, maar u mag ook twee in  $S$  aan  $SQ$  rakende cirkels vinden waar het boogje  $PSK$  tussengeklemd zit. Wie zover komt, zoekt natuurlijk ook de derde raaklijn aan  $\Gamma$  die door  $Q$  gaat op!

Spelen met de tekening, reflecteren, onderdelen toevoegen aan de tekening, weer weghalen, een bijzondere ligging proberen. Experimenteren en redeneren: afwisselend, maar ook vaak op het zelfde moment. Een tussentijds hypothese bedenken en die snel even toetsen, weerleggen of versterkt weten. Dat is de juiste werkwijze hier. En blijven kijken:

- f. De kromme lijkt op die van opgave 4c. Is er een verklaring? een mogelijke test is het onderzoeken van de asymptoot bij 4c en die vergelijken met wat in 6e gevonden/bewezen is.

Lukt het niet, dan is er nog het vangnet van de analytische meetkunde. Jan van de Craats onderzocht de meetkundige plaats en de kromme  $\Gamma$  al eerder met analytische middelen, zie [5]. Hij vertaalde het Enschedese fotoprobleem een beetje anders naar de wiskunde toe; namelijk door de absolute grootte van de hoeken  $\angle QYO$  en  $\angle OYP$  aan elkaar gelijk te stellen. De lijn  $PQ$  speelt dan ook een rol en voor de meetkundige plaats moeten stukken worden geselecteerd van lijn en kromme; maar de vloeiende kromme waar  $Y$  op mag liggen, die we  $\Gamma$  noemden, speelt natuurlijk de hoofdrol in het wiskundig vervolgonderzoek.  $\Gamma$  blijkt dan een derdegraadskromme te zijn, wat we op grond van 5c al vermoedden.

CABRI kan wel coördinaten afdrukken en vergelijkingen geven van getekende cirkels, lijnen en kegelsneden, maar kan deze objecten niet algebraïsch manipuleren. Dit doen we dus zelf of we laten een computerprogramma dit doen.

Ook de nu volgende analytische aanpak laten we aansluiten bij ons eerste schetsje. Net als in [5] nemen we  $O$  als centrum,  $OQ$  als  $x$ -as en  $OP$  als  $y$ -as, waarbij  $P$  en  $Q$  respectievelijk de punten  $(0, p)$  en  $(q, 0)$  zijn. Punt  $X$  is nu het lopende punt  $(q \cos \varphi, q \sin \varphi)$ , waarin  $\varphi$  steeds  $\angle XOQ$  is. De vergelijkingen van de lijn  $XP$  en van de middelloodlijn van  $XQ$  kunnen nu zonder veel moeite opgesteld worden:

$$(y - p) \cdot (q \cos \varphi) = x \cdot (q \sin \varphi - p) \text{ en } y \sin \varphi = x \cdot (1 - \cos \varphi).$$

Halen we  $x$  en  $y$  uit hun ineenstrengelingen los, dan vinden we een parametervoorstelling van  $\Gamma$ . MAPLE – maar voor de computer warm is doet u het zelf – doet dat via het commando ‘solve’:

$$x = \frac{pq \sin \varphi \cos \varphi}{q(\cos(\varphi) - 1) + p \sin \varphi}, \quad y = \frac{pq \cos \varphi(1 - \cos \varphi)}{q(\cos(\varphi) - 1) + p \sin \varphi}.$$

Geduldig inpraten op mijn algebra-slaaf MAPLE (ik ben nu niet sneller dan MAPLE maar ik moet MAPLE wel begeleiden onnodige zijpaden te mijden) levert tenslotte

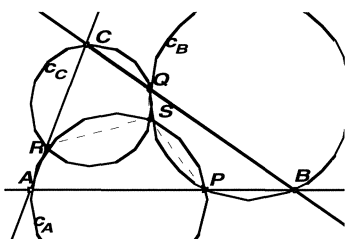
$$px^3 - qx^2y - pqx^2 + pxy^2 - qy^3 + pqy^2 = 0.$$

Zie in [5] hoe daar bijvoorbeeld de asymptoot (en meer) mee gevonden kan worden.

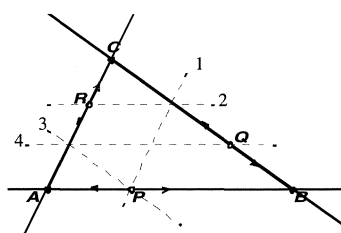


## 10. IN HET SPOOR VAN EEN MAGISCHE CIRKEL

Voor ons tweede onderzoek nemen we de figuur bij de stelling van Miquel als uitgangspunt. Zie Figuur 12 hiernaast. De 'deelpunten' liggen vrij op de lijnen  $AB$ ,  $BC$  en  $CA$ . De drie cirkels (door  $R - A - P$ ,  $P - B - Q$ ,  $Q - C - R$ ) gaan door het 'Miquel'-punt  $S$ ; dat was een eenvoudige toepassing van de drie cirkelstelling uit deel I, zie ook opgave 2. We merken nog op dat de drie hoeken die  $SP$ ,  $SQ$ , en  $SR$  maken met de drie zijden  $AB$ ,  $BC$  en  $CA$  gelijk zijn. Dit alles geldt in het algemene geval. Het plan is nu  $P$ ,  $Q$  en  $R$  te laten bewegen, maar in samenhang, en dan te onderzoeken wat de baan van  $S$  wordt.



FIGUUR 12.



FIGUUR 13.

Limietgevallen worden hierbij niet geschuwd. Laat  $P$  bijvoorbeeld eens tot  $A$  naderen; de cirkel door  $R$ ,  $A$  en  $P$  nadert dan tot de cirkel die  $AB$  in  $A$  raakt en door  $R$  gaat. CABRI laat dat allemaal moeiteloos zien, tenzij we bij toeval  $P$  precies op  $A$  leggen.

Een voor de hand liggende koppeling van de bewegingen van  $P$ ,  $Q$  en  $R$  is nu die, waarbij de drie punten met eigen constante snelheden elk hun eigen zijde in dezelfde tijd doorlopen.  $P$ ,  $Q$  en  $R$  delen dan op elk moment hun eigen zijde, dus  $AB$ ,  $BC$  en  $CA$ , in dezelfde verhouding in, ze vormen een *evenredig delend drietal*. Door één punt als uitgangspunt te nemen,  $P$ , maken we middels de evenwijdige lijnen optie moeiteloos zo'n drietal. Zie de constructie in Figuur 13; de cijfers geven de volgorde der lijnen aan. Deze constructie is in deze paragraaf voortdurend op de achtergrond werkzaam. [evenredig3.fig] en zie constructietip 6 over het verbergen van storende lijnen.

Nu voegen we de Miquelcirkels toe en we kunnen gaan experimenteren.

In Figuur 14 zien we bij I de figuur in een willekeurig stand [evremique1.fig]. Bij II nadert  $P$  to  $A$ . De punten  $Q$  en  $R$  naderen tot  $B$  en  $C$ . De drie cirkels naderen tot aan de zijden rakende cirkels. De limietsituatie zelf is bijzonder, die hangt namelijk niet meer van de keuze van  $P$  af en is geheel en alleen bepaald door de driehoek  $ABC$ .  $S$  van het limietgeval is dan ook een bijzonder punt van de driehoek. We vinden – maar hebben nog niet klassiek meetkundig bewezen – de volgende stelling:

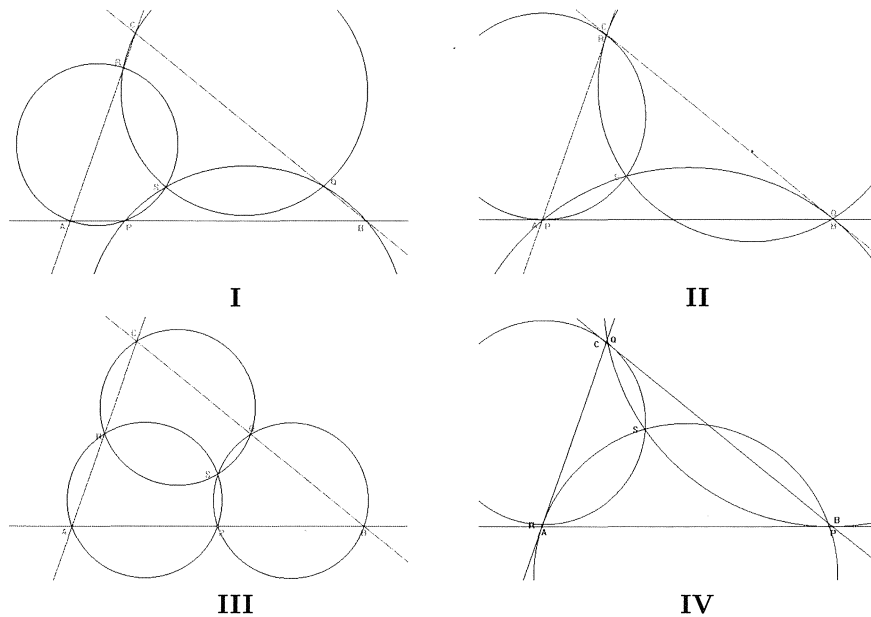
*Gegeven driehoek ABC.*

*De volgende drie cirkels gaan door één punt:*

*die in A raakt aan AB en door C gaat,*

*die in B raakt aan BC en door A gaat,*

*die in C raakt aan CA en door B gaat.*



FIGUUR 14.

Dit bijzondere punt heet het eerste punt van Brocard en wordt meestal met  $\Omega$  aangeduid.

De eerdere genoemde hoekgelijkheden van  $SP$ ,  $SQ$ , en  $SR$  met de drie zijden  $AB$ ,  $BC$  en  $CA$  suggereren nu:

$$\triangle B A \Omega = \triangle A C \Omega = \triangle C B \Omega.$$

De gemeenschappelijke hoek heet de *Brocardhoek* van de driehoek en wordt met  $\omega$  aangeduid.

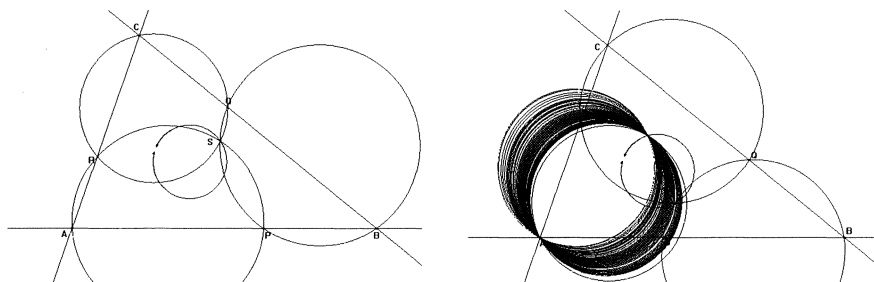
Nauwkeurige bewijzen van dat alles kunnen we opstellen, als we de stelling van de constante omtrekshoek uitbreiden met het raakgeval, waarin het lopende punt  $X$  eigenlijk met  $A$  samenvalt; in plaats van lijn  $XA$  nemen we dan de raaklijn. Dat voeren we verder niet uit; de lezer kan het makkelijk zelf invullen.

In geval **III** is  $P$  het midden van  $AB$ . Ook  $Q$  en  $R$  zijn dan middens van hun zijden en  $S$  is het midden bij uitnemendheid van de driehoek zelf: het middelpunt  $M$  van de omgeschreven cirkel.

In geval **IV** nadert  $P$  tot  $B$  en we vinden, geheel analoog aan **II**, het *tweede Brocardpunt*  $\Omega'$  en de bijhorende *tweede Brocardhoek*  $\omega'$ . Als extra bijzonderheid zij vast vermeldt dat  $\omega = \omega'$ . Daar zijn vele bewijzen voor; [10] geeft een zeer elegant bewijs.

Maar misschien geeft CABRI ons straks nog een tip...

Na het slepen van  $P$  over  $AB$  tekenen we natuurlijk de meetkundige plaats van  $S$ , als  $P$  de lijn  $AB$  doorloopt. Het antwoord van CABRI verschijnt als een mysterieuze in de driehoek zwevende cirkel. Zie Figuur 15.



FIGUUR 15. De meetkundige plaats van  $S$       FIGUUR 16. Het spoor van de  $A$ -cirkel.

Dit is echt bijzonder, dat is duidelijk. Maar eerst testen en kijken!

Verslepen van de hoekpunten van de driehoek leert ons dat het steeds om een cirkel gaat, dat die cirkel niet altijd geheel binnen de driehoek ligt (maar dat wisten we al:  $M$  hoort ligt op de meetkundige plaats en ligt niet altijd binnen de driehoek), dat de punten van Brocard, die er uiteraard op liggen, wel altijd binnen de driehoek liggen en samen met  $M$  een gelijkbenig driehoek vormen (met tophoek  $2\omega$ ) overigens. Enzovoort.

De cirkel door  $\Omega$ ,  $\Omega'$  en  $M$  staat bekend als de Brocardcirkel van de driehoek. Onze meetkundige plaats lijkt dus de Brocardcirkel te zijn, maar dat vraagt om een bewijs.

Over de Brocardpunten, de Brocardcirkel en alles wat daar mee samenhangt is erg veel bekend geraakt in de 19e eeuw. Brocard zelf was een Franse luitenant, anders gezegd: een amateurwiskundige als U en ik. In het boek van EMMERICH [6] uit 1891 wordt een serie andere amateurs vermeldt, die schouder aan schouder met de groten van de 19e eeuw dit meetkundegebied exploreerden. Dit boek is een goudmijn; het is geschreven door een Duitse Gymnasiumprofessor, die overduidelijk ook zijn leerlingen inwijdde in deze exotische geheimen van de driehoek. In het boek staat weliswaar een stelling die wel tot een bewijs van ons vermoeden leidt, namelijk dat de meetkundige plaats van  $S$  juist de Brocardcirkel is, maar laten we eerst eens kijken hoever we zelf al doorexplorerend kunnen komen.

Tussen de gegevens en het te bewijzen ligt hier een gapende kloof. De gegevens gaan over punten op de driehoek, de gevonden cirkel zweeft schijnbaar vrij rond. Toen ik – alleen met kennis van de stelling van Miquel en de Brocardpunten zelf, maar in totale onwetendheid van het bestaan van de Brocardcirkel – Figuur 15 op het scherm kreeg, was de eerste reactie dan ook: dit kan niet, het lukt me nooit dit te bewijzen, dit wordt opzoeken in weet ik wat voor boek. Het viel mee doordat ik proberenderwijs op een ijzersterke algemeen bruikbare strategie stuitte, die ik me pas achteraf bewust werd. Ik laat nu direct het moment zien waarop ik dacht: hiermee moet het lukken, en leg dan uit wat die ijzersterke strategie is.

In Figuur 16 heeft de cirkel door  $A, P$  en  $R$  een spoor getrokken terwijl  $P$  een stukje versleept is. De opeenvolgende standen van de cirkel zijn zo nog te zien en het signaal dat hier vanuit gaat is niet mis te verstaan. Die cirkels gaan door één vast punt, dat bovendien op de meetkundige plaats van  $S$  ligt.

Verbazing, maar ook meteen het gevoel: een spoor is gevonden. We geven dit (vermoedelijke) punt vast de naam  $T_a$ , omdat het via  $A$  tot stand komt. Laten we de  $B$ -cirkels een spoor trekken, dan vinden we evenzo  $T_b$ , en analoog ook  $T_c$ .

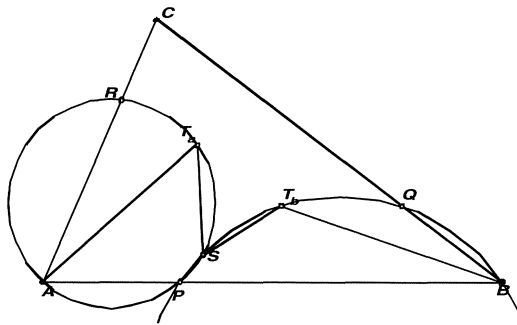
De belangrijke strategische overweging is nu dat het nieuw ontdekte fenomeen slechts met één bewegende cirkel en twee bewegende punten ( $P$  en  $R$ ) te maken heeft. Ofwel: we zochten naar een verklaring voor de beweging van  $S$  en we keken naar de beweging van een simpeler deel van de figuur. Waarschijnlijk is het ontdekte fenomeen dan ook eenvoudiger aan te tonen maar bruikbaar voor het vervolg. Het zou daarom deel één van ons bewijs kunnen zijn, waarbij deel twee zich dan baseert op de bewezen aanwezigheid van  $T_a$ , en uiteraard op die van de twee andere vaste punten, waar respectievelijk de  $B$ -cirkels en  $C$ -cirkels dóór zullen gaan. Een mogelijk bewijsplan is eigenlijk gevonden, al is er nog geen millimeter uitgeschreven bewijs in zicht.

We kunnen de bewijshelften apart aanpakken; ik geef de voorkeur aan helft twee eerst. Als dat faalt, begin ik misschien niet aan helft een. Bovendien ben ik erg blij dat de gevonden punten  $T_a, T_b$  en  $T_c$  vaste punten op de cirkel lijken te zijn, waardoor ik al in bekend vaarwater kom.

De volgende figuur geeft aan dat uit het bestaan van vaste punten  $T_a$  en  $T_b$  de cirkelvormigheid van de baan van  $S$  afgeleid kan worden.

$T_a$  en  $T_b$  zijn vast, dat is dus de aanname. Maar dan zijn  $\triangle PAT_a$  en  $\triangle T_bBP$  ook vast en wegens de constante omtrekshoekstelling zijn  $\triangle PST_a$  en  $\triangle T_bSP$  dan ook vast. Gevolg:  $\triangle T_aST_b$  is vast en  $S$  loopt dus over een cirkel die door  $T_a$  en  $T_b$  gaat.

Deel twee van het bewijs is nu al in kannen en kruiken.



FIGUUR 17.

Deel één lijkt wat moeilijker te zijn, maar de motivatie stijgt na het voorgaande succes. Bovendien geeft Figuur 17 nog een aanwijzing. Als  $T_a$  inderdaad vast is, is ook de andere hoek bij  $A$  vast. We concluderen dat  $\triangle RPT_a$  ook vast

is, vervolgens dat  $\triangle PRT_a$  vast is en uiteindelijk dat alle driehoeken  $\triangle PT_aR$  onderling gelijkvormig moeten zijn.

Dit is zoeken vanuit wat we willen bewijzen, namelijk vanuit het bestaan van  $T_a$ . Dat is een perfecte weg, als die maar niet als bewijs wordt gepresenteerd, maar als zoektocht. Uiteindelijk vinden we vast wel een omgekeerde logisch verantwoorde weg, misschien kunnen we bijvoorbeeld vanuit het evenredig delen van  $P$  en  $R$  (dat we nog steeds ongebruikt hebben gelaten!) komen tot die gelijkvormigheid, en dan: enzovoort!

Daarmee zijn nog niet alle details van het bewijs gevonden, geenszins, er zal nog fors naar allerlei hoekgelijkheden en vaste verhoudingen gezocht moeten worden, maar een plan voor een bewijs is er nu wel.

Voor mijn betoog, dat CABRI een onderzoekshulpmiddel is, is het niet nodig het bewijs af te maken. Het kan, het is aan de lezer.

Goed mogelijk is dat die lezer tijdens het zoeken nog andere schoonheden ontdekt, bijvoorbeeld door ook de lijn  $SP$  een spoor te laten trekken, waardoor weer drie andere stabiele punten op de Brocardcirkel worden gevonden. Bovendien wordt zichtbaar waarom  $\omega = \omega'$ . En ook wordt duidelijk waarom de meetkundige plaats bestaat uit de héle Brocardcirkel minus één punt. Dat ene punt is het symmediaanpunt, het snijpunt van de spiegelbeelden van de zwaartelijnen in de deellijnen. Zie ook weer [10], waar ook al die tot nu toe genoemde bijzondere punten bij name worden genoemd. Dat zijn er nú al tien. We stoppen.

### Deel III: Hoe betrouwbaar is het JA van Cabri?

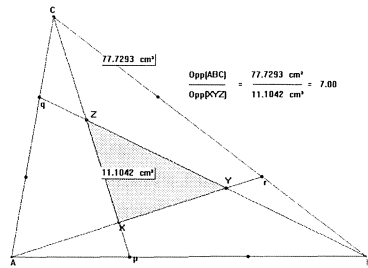
CABRI heeft opties waarmee gemeten kan worden: lengtes van lijnstukken, cirkels en bogen, grootte van hoeken, oppervlakte van gesloten figuren als veelhoeken en cirkels. CABRI kan ook testen of lijnen evenwijdig zijn, of een punt op een lijn ligt, enzovoort.

Hier volgen drie voorbeelden waarin we CABRI uitspraken laten doen over enkele meetkundige vragen.

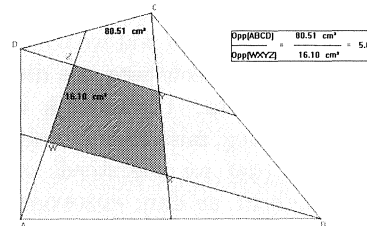
#### 11. VOORBEELD 1: OPPERVLAKTE DEELDRIEHOEK

CABRI kan de oppervlakte van een veelhoek geven en rekenen met zulke ‘gemeten’ getallen.

Verdeel de drie zijden van een driehoek  $ABC$  eens cyclisch in de verhouding 1: 2 en verbind de deelpunten met de hoekpunten tegenover de betreffende zijden. Zie Figuur 18. Een binnendriehoek ontstaat,  $XYZ$ . De verhouding  $\text{Opp}(XYZ) : \text{Opp}(ABC)$  is vast, namelijk 1 : 7. Aldus CABRI in [driedeel17.fig] zie ook constructietip 5.



FIGUUR 18.



FIGUUR 19.

### 12. VOORBEELD 2: OPPERVLAKTE DEELVIERHOEK

Verbind in een vierhoek  $ABCD$  eens de middens van elke zijde met een van de punten van de overliggende zijden; kies steeds op dezelfde manier met de klok mee. Zie Figuur 19. Een binnenvierhoek  $WXYZ$  ontstaat. De verhouding  $\text{Opp}(WXYZ) : \text{Opp}(ABCD)$  is vast, namelijk  $1 : 5$ . Aldus CABRI.

Dit komt uit het waardevolle *Rethinking Proof*, van MICHAEL D. DE VILLIERS, [17].

Door flink te slepen met de hoekpunten raak je meer en meer overtuigd, tot er in één van de twee gevallen iets misgaat: er blijken hier toch andere verhoudingen tussen de oppervlakten op te treden. Dan komen de echte problemen, want de twijfel slaat toe: geldt de constantheid in het andere geval wel? En hoe moet dat dan bewezen worden?

Maar het voorbeeld waarbij de gelijkheid verworpen moet worden levert het grootste probleem. Een ongelijkheid, bijvoorbeeld dat de verhouding altijd kleiner is dan  $1 : 3$  is of altijd groter dan  $1 : 5$ , lijkt echt wél stand te houden. En dat verbeterde vermoeden is tot op heden niet weerlegd, maar ook niet bewezen.

Bij de voorgaande voorbeelden gaf CABRI uiteindelijk toch nog wel eerlijk uitsluitsel. Nu een voorbeeld waar we eerst aantonen dat iets niet waar is, waarna we vervolgens CABRI via verfijnde foltermethodes een vals getuigenis laten afleggen. De conclusie zal zijn: CABRI doet zijn beweringen op numerieke gronden en beredeneert niets maar gaat normaliter niet zo makkelijk de fout in.

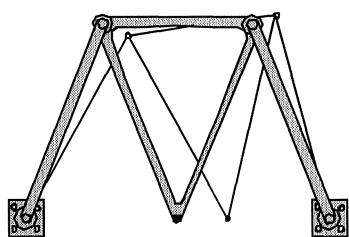
### 13. VOORBEELD 3: HET MECHANISME VAN ROBERT MANCHESTER.

Bij het bouwen van een stoommachine was het van belang dat via een mechaniek van scharnierende stangen een bewegend punt gemaakt kon worden, dat zonder zelf langs een rechte staaf of gleuf te glijden, toch alleen rechtlignig kon bewegen.

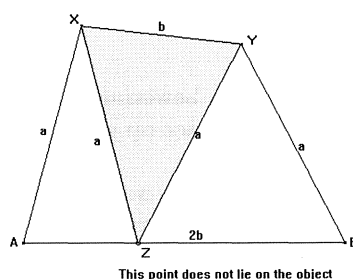
Dat is moeilijker dan op het eerste gezicht te verwachten is. Een cirkelbeweging heb je al te pakken als je een ruwe tak met één spijker ergens vastzet. De tak kan dan nog draaien om de spijker en elk punt van de tak beschrijft een

cirkel, zonder dat er vooraf een cirkel aanwezig was. Zo'n stangenmechanisme, daar stapelen de cirkels zich op elkaar; maar rechte lijnen? Echt niet vanzelf.

Een rechte lijn teken je in het algemeen met een liniaal. Maar hoe maak je een rechte liniaal? Met een andere liniaal? Maar hoe werd die dan ooit recht gemaakt? Na deze inleiding kan de titel van een beroemd essay van A.B. Kempe uit 1877 beter op waarder geschat worden: *How to Draw a Straight Line?*



FIGUUR 20.



FIGUUR 21.

Een van de pogingen tot rechtgeleiding in de 18e eeuw is het stangenmechanisme van Robert Manchester. Zie Figuur 20. Er zijn twee vaste punten onderaan. Deze liggen op afstand  $2b$  van elkaar. Aan die vaste punten zitten draaibare staven van lengte  $a$ ; aan hun einden zit, ook weer draaibaar, een driehoek met twee zijden  $a$  en een zijde  $b$ . ( $a$  is groter dan  $b$ ). Het is duidelijk dat het geheel nog flexibel is.

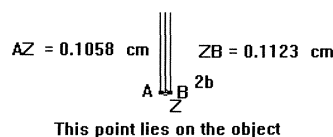
Kijk nu naar het zwarte punt van de driehoek. Dat kan zich midden tussen de twee vaste punten bevinden, maar ook naar de vaste punten toe bewegen. Een tweede mogelijke stand is met dunne lijntjes aangegeven.

Nabouwen van het mechanisme met CABRI, Figuur 21 of [manchester.fig] ondersteunt sterk de indruk dat het bewuste punt  $Z$  over de rechte lijn tussen de twee vaste punten  $A$  en  $B$  beweegt. CABRI helpt ons uit de droom: ook al zien we het niet,  $Z$  ligt niet op  $AB$ . (Constructietip 7.) De testoptie 'is element van' zegt NEE.

Omdat het toch wel een foto-finish is en na eerdere negatieve ervaringen willen we graag ook een bewijs. Het volgende elegante bewijs uit het ongerijmde is bedacht door Sieb Kemme.

Veronderstel dat  $Z$  zich inderdaad op het inwendige van lijnstuk  $AB$  bevindt. Teken nu de hoogtelijnen uit  $X$  en  $Y$  in de driehoeken  $AXZ$  en  $ZYB$ . Deze zijn dan evenwijdig. De voetpunten van de hoogtelijnen – noem ze  $C$  en  $D$  – delen  $AZ$  en  $ZB$  middendoor en daarom geldt  $|CD| = b$ . Ofwel: die parallelle hoogtelijnen liggen op een afstand  $b$  van elkaar. Maar dat wringt met de schuine stand van  $XY$ !  $XY$  kan dus alleen evenwijdig aan  $AB$  zijn, en makkelijk is na te gaan dat dit alleen in de symmetrische stand van het mechanisme kan plaatsvinden. Bewezen is eigenlijk: als  $Z$  op (het inwendige van)  $AB$  ligt, is  $Z$  het midden van  $AB$ .

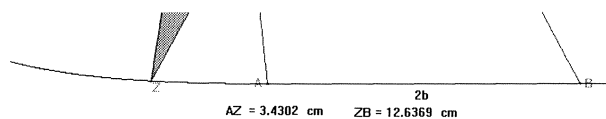
Maken we  $b$  heel klein ten opzichte van  $a$ , dan kunnen we in situaties waarin  $AZ$  niet gelijk is aan  $ZB$ , CABRI toch laten beweren dat  $Z$  op  $AB$  ligt. Zie de screendump in Figuur 22.



FIGUUR 22.

CABRI doet zijn bewering ‘ $Z$  ligt op  $AB$ ’ blijkbaar op grond van afgeronde berekeningen en niet op grond van geautomatiseerd bewijzen in een of andere vorm.

Maar we hadden wel zoiets subtiels als het mechanisme van Manchester nodig, om dat boven water te krijgen. De ervaring leert dat, als CABRI beweert dat iets *niet* geldt, dat dit door handig verslepen in de figuur meestal wel snel werkelijk zichtbaar in beeld gebracht kan worden, en dat numerieke beweringen als die in voorbeeld 1 en 2, met succes op hun waarde getoetst kunnen worden door CABRI meer decimalen te laten vertonen. (constructietip 8).



FIGUUR 23.

Er is echter in gevallen als die van het mechanisme van Manchester ook fraaiere manieren van redeneren.

Laat  $Z$  eens een spoor trekken, en laat daarbij het punt  $Z$  ook eens ‘buiten’  $AB$  lopen. De baan van  $Z$  blijkt buiten  $AB$  helemaal geen rechte lijn te zijn. De baan moet wel een stuk van een algebraïsche kromme zijn, want we construeren  $Z$  door snijden van cirkels. Maar een algebraïsche kromme gaat heus niet plotseling van rechtlijnig in niet-rechtlijnig over, dus het stuk tussen  $A$  en  $B$  zal niet werkelijk rechtlijnig zijn en dus ligt  $Z$  in het algemeen niet op  $AB$ .

Een exacte onderbouwing van deze schetsmatige argumentatie is wiskundig goed mogelijk en kan bijvoorbeeld rusten op de stelling uit de theorie van de complex-analytische functies die zegt dat een complex differentieerbare functie, die constant is op een interval, constant is op zijn hele definitiegebied. Maar dit gaat de perken van dit artikel wat te buiten en is ook niet nodig: we gebruiken de schetsmatige argumentatie alleen als een heuristiek, in dit geval als een aanwijzing of het onderzoek richting bewijs of richting weerlegging van een vermoeden zal worden voortgezet.



## Deel IV: Wat is construeren nu eigenlijk?

### 14. ALLEEN MET PASSER EN LINIAAL?

In het recente examen Wiskunde B2 nieuwe stijl zat een opgave waarin gevraagd werd iets te tekenen en de werkwijze toe te lichten. Een leerling schreef: ‘Ik teken de conflictlijn van de cirkel en punt  $F$ , dat is een ellips en die snijdt ik met de lijn  $SM$ .’ Dat gaf discussie bij de examenbespreking: ‘Ho, ho, dat is geen construeren!’ Beantwoord met: ‘Waarom eigenlijk niet, die ellips leren ze toch juist op deze manier kennen? Dat is toch precies tekenen en toelichten?’

Traditioneel zou construeren moeten gebeuren ‘met passer en liniaal’, anders was het geen goede meetkunde. Maar in het veruit belangrijkste boek van de Griekse meetkunde, namelijk *De Elementen* van Euclides, komen de woorden passer en liniaal absoluut niet voor!

We kijken daarom eens naar het begin van *De Elementen*. Euclides’ eerste drie postulaten, volgend na de lijst van 23 definities, zijn:

*Het volgende wordt geëist:*

1. Een rechte lijn te trekken van elk punt naar elk punt.
2. Een eindige rechte lijn onbepaald door te zetten in een rechte lijn.
3. Een cirkel te beschrijven met elk middelpunt en elke afstand.

De eerste propositie, iets verder, stelt de opgave:

*Propositie 1:*

*Op een gegeven eindige rechte lijn een gelijkzijdige driehoek construeren.*

Wat in de eerste propositie staat moet gedaan worden, maar wat er vóór staat, dus de postulaten, mag gebruikt worden. Het derde postulaat wordt nu twee keer gebruikt. Als de eindige lijn met  $AB$  is aangeduid, geeft dat postulaat ons de cirkel met middelpunt  $A$  en ‘straal’  $AB$  en die met middelpunt  $B$  en straal  $AB$ . Uit Euclides’ tekst maken we op dat hij vindt dat het (de) snijpunt(en) van de cirkels er nu vanzelf is (zijn). Wat Euclides zo eigenlijk doet, is de meetkundige wereld vanuit enkele eerste beginselen opbouwen. In het begin is er als het ware slechts een lijnstuk, en propositie 1 gerandeert dat dan de gelijkzijdige driehoek erop ook bestaat. Op de zijden zouden we weer een gelijkzijdige driehoek kunnen bouwen, zo vinden we dan nieuwe punten, die weer goed zijn om –postulaat 1– nieuwe rechte lijnen te vinden. De meetkundige wereld wordt gemaakt, geconstrueerd. Redeneringen en bewijzen op grond van logische grondregels garanderen daarbij de juistheid van de constructies.

Constructies, ja; maar puur mentaal, puur uitgevoerd in de geest. De passer en de liniaal, gereedschappen van timmerman en smid, dienden de wiskunde slechts als ondersteuning van de gedachte, die in zekere zin de eigenlijke wiskundige werkelijkheid is. Hetzelfde geldt voor CABRI. CABRI, uitgedoekt tot op de opties lijn en cirkel (dat kan, zie [elementen.fig]), is slechts een passer

aan het stopcontact of een liniaal op batterij.

Construeerbaar in de zin van *De Elementen* is wat via de postulaten vanuit gegeven punten ‘mentaal’ maakbaar is. Daar hoort de diagonaal van een vierkant met gegeven zijde bij, daar hoort de hoek van 15 graden bij, daar hoort het verdelen van een gegeven lijnstuk in 790888540138 gelijke delen bij, want die constructie kan goed beschreven worden, ook al zou een echte passer halverwege het materieel uitvoeren van het proces wel versleten, volledig verroest, of in de zon gesmolten zijn.

De Griekse meetkundige slaagden er niet in bepaalde constructies te vinden: die van de trisectie van een willekeurige gegeven hoek bijvoorbeeld. Intuïtief geloofde iedereen heus dat dat derde deel van de hoek er was, maar het lukte niet dat derde deel te ‘construeren’ vanuit de postulaten. Wij weten tegenwoordig dat zo’n constructie vanuit de postulaten van Euclides onmogelijk is. Het in theorie beschrijven van wat Euclidisch in de vlakke meetkunde construeerbaar is, is ons gemakkelijker gemaakt omdat we het Griekse meetkundig construeren kunnen vertalen in de twee millennia jongere taal van de algebra. Vatten we het platte vlak op als beschreven door complexe getallen, en beginnen we met de punten 0 en 1 als gegeven, dan laten de Euclidische constructies alleen opbouwen via optellen, aftrekken, vermenigvuldigen en delen en worteltrekken toe om nieuwe objecten te verkrijgen.

Het getal  $z = 3 + \sqrt{\frac{5+\sqrt{3}}{6-\sqrt{17}}}$  is dus construeerbaar. Onder andere door wortels weg te kwadrateren vinden we dat  $z$  voldoet aan een (irreducibele) veeltermvergelijking met gehele coëfficiënten van graad 8:

$$130321z^8 - 3127704z^7 + 32017812z^6 - 182229912z^5 + 600239754z^4 \\ - 998170776z^3 + 262503468z^2 + 1351418472z + 370356037 = 0.$$

De graad van de vergelijking is een macht van 2. Dat komt door het herhaald wegkwadrateren.

Noodzakelijk is die 2-macht als graad wel, maar:

- we hebben niet aangetoond dat  $z$  niet aan een simpeler vergelijking voldoet.
- we hebben ook niet aangetoond, dat als  $z$  aan zo’n vergelijking voldoet,  $z$  ook inderdaad met passer en liniaal construeerbaar is. Dat is in het algemeen ook niet zo!

Voor een gedetailleerde theorie is hier geen plaats; goede moderne algebraboeken bespreken het construeren met passer en liniaal met liefde en zorg in het licht van de Galoistheorie. Een recente uiteenzetting is *Galois Theory* van IAN STEWART [15]. Dit boek heeft het immense voordeel dat het de subtiliteiten van de algemene theorie illustreert met goed gekozen rekenvoorbeelden en opgaven die vitale punten van het betoog belichten. Een wiskundig-didactisch hoogstandje.

Gauss zag in dat we een regelmatige  $n$ -hoek kunnen construeren als we de  $n$ -de eenheidswortel  $z = e^{\frac{2\pi i}{n}}$  kunnen construeren. Dit getal is wortel van een

zogenaamde cirkedelingsvergelijking; die irreducibele vergelijking heeft juist  $\varphi(n)$  als graad, waarin de Euler-totiënt functie is. Wil de regelmatige  $n$ -hoek construeerbaar zijn, dan moet  $\varphi(n)$  een macht van 2 zijn. In het geval van de cirkedelingsvergelijkingen kan aangetoond worden dat de construeerbaarheid dan ook inderdaad mogelijk is. In feite berust dat op het feit dat de andere wortels van de cirkedelingsvergelijking machten zijn van de eerst aangegeven wortel  $z$ , hetgeen garandeert dat de lichaamsuitbreiding  $\mathbb{Q}(z)$  van  $\mathbb{Q}$  een zogenaamde *normale* uitbreiding is, een noodzakelijke voorwaarde voor het inzetten van de hoofdstelling van de Galoistheorie. Zie hierover weer [15].

Als nu  $n = p_1^{a_1} \cdot p_2^{a_2} \cdot \dots \cdot p_k^{a_k}$  de ontbinding van  $n$  in priemfactoren is, dan weten uit de elementaire getaltheorie:

$$\varphi(n) = p_1^{a_1-1} \cdot p_2^{a_2-1} \cdot \dots \cdot p_k^{a_k-1} \cdot (p_1 - 1) \cdot (p_2 - 1) \cdot \dots \cdot (p_k - 1).$$

Daarom mag dus alleen de multipliciteit van priemfactor 2 groter dan 1 zijn en moeten de andere priemfactoren van de vorm  $2^m + 1$  zijn.

Er is nog een klein extraatje dat de stelling fraaier maakt. Namelijk het volgende: als  $m$  een oneven factor  $q$  heeft, is  $2^m + 1$  zeker niet priem, maar deelbaar door  $2^{m/q} + 1$ . Als bijvoorbeeld  $m = 10$ , bestaat de ontbinding  $2^{10} + 1 = (2^2 + 1)(2^8 - 2^6 + 2^4 - 2^2 + 1)$ .

Daarom geldt de stelling van Gauss in de volgende vorm:

*Precies die regelmatige  $n$ -hoeken zijn construeerbaar waarvoor en waarin de priemfactoren  $p_2$  tot en met  $p_k$  van de vorm  $2^{2^s}$  zijn.*

De enige bekende priemgetallen van die vorm zijn: 3, 5, 17, 257 en 65537 en er is weinig hoop dat er nog een is; onder de  $10^{40000}$  is er zeker geen te vinden.

De stelling van Gauss beschrijft precies welke rationale veelvouden van  $\pi$  als hoek construeerbaar zijn. Maar vergis je niet: veel méér hoeken zijn construeerbaar, bijvoorbeeld de hoek van Pythagoras zelf, de scherpste hoek van de 3-4-5 driehoek; die is overigens géén rationaal veelvoud van  $\pi$ .

#### 7 Opgave: Hoezo irrationaal?

Tip: Begin met een hoek  $\alpha$  waarvan de tangens 2 is. Construeer die met CABRI door gebruikmaking van het rooster. Zie constructietip 9. Het rooster helpt om het veelvouden van die hoek te maken door een serie gelijkvormige driehoeken te maken, de oude hypotenusa is steeds de nieuwe kortste zijde. Laat na één stap zien dat de Pythagorashoek gelijk is aan  $2\alpha - \pi/4$ . Stapel door met  $a$  en laat zien dat je nooit op de  $x$ -as terugkeert. Weet en gebruik dat  $\sqrt{5}$  irrationaal is.

8 Toon aan dat de trisectie van de hoek in het algemeen niet mogelijk is met passer en liniaal. Doe dit hier door een regelmatige  $n$ -hoek te noemen die wel construeerbaar is, terwijl de  $3n$ -hoek dat niet is.

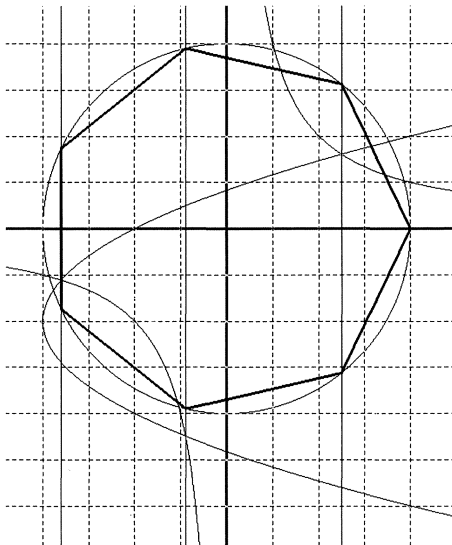
(Zie [15] voor een instructiever bewijs dat los staat van de stelling van Gauss.)

9 Wat is de laagste  $n$  waarvoor de regelmatige  $n$ -hoek niet met passer en liniaal construeerbaar is?

#### 15. CONSTRUCTIE VAN DE REGELMATIGE ZEVENHOEK MET CABRI

Zie [zevenhoek.fig] of Figuur 24. Met de optie TEKEN3>>KEGELSNEDE zijn een orthogonale hyperbool en een parabool getekend, beide door vijf eerder

geconstrueerde punten met eenvoudige rationale coordinaten. Dat kan met CABRI.



FIGUUR 24.

Door de snijpunten van hyperbool en parabool zijn verticale lijnen getekend. De bewering is dat de snijpunten van deze lijnen met de cirkel zes van de zeven punten van een regelmatige zevenhoek zijn.

Laten we eerst maar eens nagaan of dat klopt!

De coordinaten zijn zo gekozen dat de cirkel de vergelijking  $x^2 + y^2 = 1$  heeft.

De hyperbool heeft dan als vergelijking  $4xy = 1$ .

En de parabool:  $(2y + 1)^2 = 2(x + 1)$ .

Elimineer nu  $y$  om de  $x$ -coordinaten van de snijpunten te vinden. Dat levert de vergelijking:

$$8x^3 + 4x^2 - 4x - 1 = 0.$$

De wortels van deze vergelijking zouden juist

$$\cos\left(\frac{2\pi}{7}\right), \cos\left(\frac{4\pi}{7}\right) \text{ en } \cos\left(\frac{6\pi}{7}\right) \text{ moeten zijn.}$$

We zouden dus

$$(x - \cos\left(\frac{2\pi}{7}\right))(x - \cos\left(\frac{4\pi}{7}\right))(x - \cos\left(\frac{6\pi}{7}\right)) = 0$$

tot

$$8x^3 + 4x^2 - 4x - 1 = 0$$

moeten kunnen herleiden. Dat kan vast wel en wie durft gaat zijn gang maar.

Er is echter een mooiere weg. De punten van de zevenhoek zijn de oplossingen van de vergelijking  $z^7 - 1 = 0$ . We willen de niet triviale oplossingen hebben en we willen met de relatie

$$x = \frac{z + z^{-1}}{2}$$

gaan werken, vanwege het verband

$$\cos(\vartheta) = \frac{e^{i\vartheta} + e^{-i\vartheta}}{2};$$

daarom schrijven we de vergelijking voor de niet triviale oplossingen alvast als in de vorm:

$$\frac{z^7 - 1}{z - 1} + z^3 + z^2 + z + 1 + z^{-1} + z^{-2} + z^{-3} = 0.$$

Nu hoeven we slechts de formule  $2x = z + z^{-1}$  te kwadrateren en in de derde macht te verheffen

$$4x^2 = z^2 + 2z^{-2}$$

$$8x^3 = z^3 + 3z + 3z + 3z^{-1} + z^{-3}$$

en de vergelijking voor  $z$  vertaalt zich vlekkeloos naar:

$$8x^3 + 4x^2 - 4x - 1 = 0$$

Ja, dat is hem. Inderdaad is de zevenhoek geconstrueerd.

Er zijn twee mogelijke bezwaren.

- a. Aan de constructie is op geen enkel moment ‘meetkundig’ te zien dat we op de zevenhoek afstevenen. Alleen na wat gecijfer blijkt het te kloppen, dat is geen inzichtelijke verklaring.

Begrijpelijk. Met een maar erbij. Bij de constructie van de regelmatige driehoek (it *Elementen*, propositie 1, zie boven) construeerden we inderdaad direct de gelijkzijdigheid met de passer; dat was duidelijk. Maar dit bezwaar van de niet-meetkundigheid kleeft al aan de bekende constructies van de regelmatige vijfhoek! Die berusten er in het algemeen op dat de verhouding van zijde en diagonaal van de regelmatig vijfhoek de gulden snedeverhouding is. Die verhouding wordt eigenlijk steevast eerst apart geconstrueerd, los van enige notie van vijfhoekigheid, en dan gebruikt om de vijfhoek te maken. Correct, maar even zo via de omweg van de algebra.

- b. De constructie gebruikt tot tweemaal toe het tekenen van een kegelsnede door vijf eerder gevonden punten.

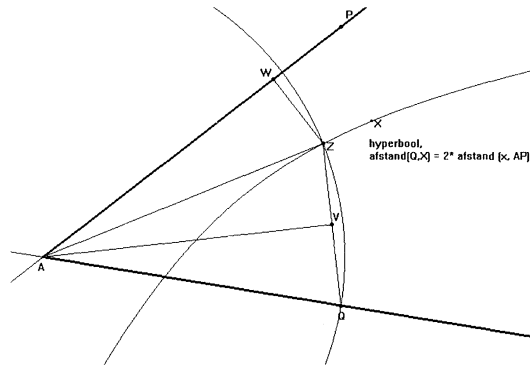
Dit is een bezwaar als de spelregel zou zijn: alleen passer en liniaal. Maar het is een kloeke stap naar een nieuw onderzoeksterrein van de meetkunde als we *De Elementen* uitbreiden met postulaat 6:

6. Een kegelsnede te tekenen door 5 gegeven punten waarvan er geen 4 op een lijn liggen.

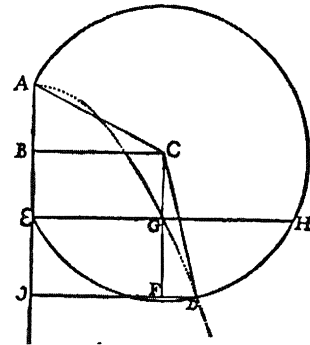
En dat gaan we nu doen.

## 16. CONSTRUCTIES MET KEGELSNEDEN

Sporadisch werd dit al door de Grieken gedaan, veelvuldig in de zeventiende eeuw door bijvoorbeeld Descartes en Roberval, systematisch in de laatste jaren van de twintigste eeuw onder invloed van gebruik van CABRI.



FIGUUR 25.



FIGUUR 26.

Een oude constructie voor de trisectie van een willekeurige hoek zien we in Figuur 25.  $\angle PAQ$  is de te dielden hoek. Teken de meetkundige plaats van de punten  $X$  waarvoor de afstand  $XQ$  het dubbele is van de afstand van  $X$  tot  $PA$ . We weten dat dit een hyperbool is en we kunnen (traditioneel) gemakkelijk vijf punten van die meetkundige plaats construeren. Volgens postulaat 6 kan de hyperbool worden *geconstrueerd*. CABRI kan het ook, zie de optie CONSTRUEER1 >> KEGELSNEDE. Teken de cirkel met middelpunt  $A$  die door  $Q$  gaat. Die snijdt de hyperbool nog in  $Z$ . Teken de loodlijn vanuit  $Z$  op  $AP$ . Dit geeft  $W$  als voetpunt. Teken de middelloodlijn van  $ZQ$ ; het midden van  $ZQ$  noemen we  $V$ . Nu zijn de driehoeken  $AWZ$ ,  $AZV$  en  $AVQ$  congruent. Daarmee is de trisectie van de hoek  $\angle QAP$  een feit.

De tekening in Figuur 26 is van Roberval en is opgenomen in de *Harmonie Universelle* van Marin Mersenne uit 1637. Mersenne was een voorvechter van algehele invoering van de verdeling van het muzikale octaaf in twaalf gelijke delen, een verdeling die voor het stemmen van instrumenten met fretten (luit, gitaren en gamba's) praktisch onvermijdelijk was. Om de posities van de fretten te bepalen was een exacte waarde nodig voor  $\sqrt[3]{2}$ ; die moest dus geconstrueerd worden.

Neem  $AB = BE$  en  $AB = EG$ . De parabool en de cirkel snijden elkaar in  $D$ . Ga na dat  $JD = \sqrt[3]{2}EG$ . Daarna kan wel met bekende traditionele constructies (die in dit muziektractaat worden beschreven) ook  $\sqrt[3]{2}$  worden bepaald/geconstrueerd. Het construeren van  $\sqrt[3]{2}$  stond bekend als het Delische probleem, de opgave om de (gegeven) altuurkubus in Delos in inhoud te verdubbelen. Mersenne zegt dat hij het Delische probleem muzikaal, met de stemsleutel, heeft opgelost; ook een mooie uitbreiding van de constructiemiddelen.

In 1997 verscheen een artikel van CARLOS VIDELA [16] waarin een variant op de stelling van Gauss wordt bewezen voor de situatie die ontstaat als postulaat

6 wordt aangenomen. Eerst wordt vastgesteld dat alles wat dan geconstrueerd wordt, algebraïsch bereikt kan worden door alleen het trekken van derdemachtswortels aan de constructie-apparatuur toe te voegen. Voor de uitbreiding van de stelling van Gauss blijkt nog nodig dat een normale lichaamsuitbreiding van  $\mathbb{Q}$  met een graad van de soort  $2^a 3^b$  inderdaad een toren is van tweede- en derdegraads uitbreidingen. Dat berust op een stelling van Burnside uit 1904, die de oplosbaarheid van eindige groepen garandeert waarvan de orde slechts door twee priemgetallen deelbaar is. Over Burnside, zie [13], voor de bewuste stelling moet u een behoorlijk gevorderd boek over groepentheorie inzien; dit gaat het bereik van [15] ver te buiten.

Hier is Videla's resultaat, enigszins informeel geformuleerd:

*Bij toelating van het tekenen van kegelsneden door vijf gegeven punten bij de constructiemiddelen, kunnen juist de regelmatige  $n$ -hoeken geconstrueerd worden waarvoor de priem ontbinding van  $n$  van de vorm  $2^s 3^t p_3 \cdot \dots \cdot p_k$  is, waarbij de priemfactoren  $p_j$  voor  $3 \leq j \leq k$  van de vorm  $2^a 3^b + 1$  zijn.*

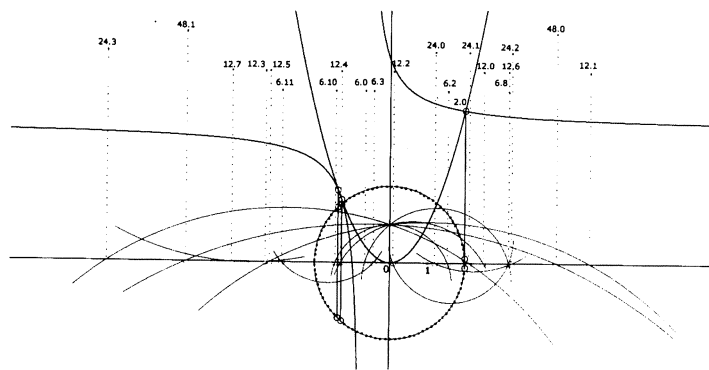
BAINVILLE en GENEVÉS [2] gaan van hieruit actief aan de slag om (met CABRI) systematisch veelhoeken te construeren. De voorbeeldconstructie van hierboven met de zevenhoek is van hen.

Ze geven ook een lijst van nieuwe geschikte priemgetallen onder de 1000000, waarbij dus  $b > 0$ .

De vooruitgang op Gauss is aanzienlijk:

7, 13, 19, 37, 73, 97, 109, 163, 193, 433, 487, 577, 769, 1153, 1297, 1459, 2593, 2917, 3457, 3889, 10369, 12289, 17497, 18433, 39367, 52489, 139969, 147457, 209953, 331777, 472393, 629857, 746497, 786433, 839809, 995329. Onder de  $10^{30}$  zijn er in totaal 231 te vinden.

In Figuur 27 zien we de constructie van 97-hoek, zoals afgebeeld in [2].



FIGUUR 27.

$97 = 2^5 \cdot 3 + 1$ . Vijf maal zijn er kwadraatwortels bepaald met snijdende cirkels en in de laatste stappen van de constructie is eenmaal het snijden van kegelsneden gebruikt, op de manier van de constructie van de zevenhoek.

## 17. KWADRATUUR VAN DE CIRKEL EN VERDER

Omdat het getal  $\pi$  transcendent is, dat wil zeggen niet aan een veeltermvergelijking met gehele coëfficiënten voldoet, kan de oppervlakte of omtrek van een cirkel met gegeven straal ook niet met behulp van kegelsneden geconstrueerd worden.

Maar CABRI laat ons wel lengte van bogen meten, laat gegeven lengten op een cirkel passen en meer zaken die echt binnen de wereld der gewone kegelsnedenconstructies niet mogelijk zijn.

Daardoor kunnen we bijvoorbeeld nu ook als meetkundige plaats construeren:

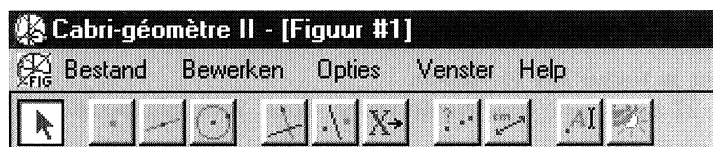
- allerlei cycloïden; zowel de gewone, de epi-cycloïden en de hypocycloïden. Zie [cyclo.gif, epihypcyclo.fig].
- de spiraal van Archimedes [archi.fig]
- de klassieke quadratrix waarmee onder andere de trisectie van de hoek én de kwadratuur van de cirkel gerealiseerd kan worden. Zie [quadra.fig].

Hier ligt nog een heel onderzoeksgebied open!

### A. CONSTRUCTIETIPS CABRI

#### A.1. Constructietip 1: HELP en menuindeling

Als Help is aangezet, wordt na het aanklikken van een menukeuze onder op het scherm informatie over de mogelijkheden gegeven. Het is geen gek idee een overzicht van de menukeuzen te maken, of dit te kopiëren uit [14]. In deze tekst hebben de menubuttons op de hoofdbalk de namen van [14]. Dat zijn:



FIGUUR 28.

#### A.2. Constructietip 2: Omgeschreven cirkel

De cirkels  $ABS$  etc. worden het makkelijkst getekend met behulp van een al aanwezige Macro. Die wordt via FILE>OPEN opgehaald uit de map Macros. Zorg dat de .mac files zichtbaar zijn en kies circscr.mac. Na openen kan de macro aangeroepen worden als optie onder CONSTRUEER 3.

#### A.3. Constructietip 3: Tweede snijpunt van cirkels

CABRI gaat soms onverwacht met snijpunten van cirkels om. Teken we het snijpunt in dit geval gewoon, dus via Teken1 of door aanwijzen, dan blijkt  $T$  na verschuiven soms aan het andere snijpunt (dus  $A$ ,  $B$  of  $C$ ) te blijven hangen. Als  $A$  al snijpunt is van twee cirkels en je wilt echt  $T$  als het andere snijpunt vinden, spiegel dan  $A$  in de verbindingslijn van de middens der cirkels. Dat werkt zeker.



*A.4. Constructietip 4: meetkundige plaatsen van bewegende punten of lijnen*

CABRI heeft diverse mogelijkheden om bewegingen te laten zien.

Met ANIMATE kunnen we de figuur automatisch laten bewegen. Kies de optie en probeer het aandrijvende punt te verslepen. het werkt als de veer van een flipperkast. Laat maar los en het gaat. Met Undo wordt de oorspronkelijke stand hersteld.

Met SPOOR AAN kunnen we daarbij het bewuste punt een spoor laten trekken. Klik het te sporen object aan. En later weer UIT, want deze optie veroorzaakt anders veel chaos op het scherm. Gebruik Refresh onder EDIT.

Tenslotte kunnen we met MEETKUNDIGE PLAATS dat spoor in één keer krijgen. Vooral die laatste mogelijkheid gebruikt de ervaren onderzoeker erg veel. Gebruik de optie CONSTRUEER3>LOCUS (of meetkundige plaats) en klik eerst het punt (of de lijn) aan waarvan we de baan willen zien, daarna het aandrijvende punt. In het geval van de doorsteekcirkels du eerst het snijpunt van  $l_1$  en  $l_3$ , daarna  $P_1$ . In het geval van loods.fig dus eerst  $Y$ , dan  $X$ . Bij de meetkundige plaats van lijnen zijn er twee mogelijkheden: een serie lijnen tekenen of de omhullende van de lijnen tekenen. Instelbaar bij OPTIONS>>PREFERENCES.

*A.5. Constructietip 5: de rekenmachine en het doorgeven van getalvariabelen*

Getallen die door meting ontstaan, kunnen in de rekenmachine worden geschoven. Het resultaat van een berekening kan uit de rekenmachine worden geschoven en desnoods in een tekst worden opgenomen. Zulke getallen, indien buiten de rekenmachine gebracht, kunnen ook worden doorgegeven via ‘Maat overbrengen’ onder CONSTRUCTIE1. Men experimentere zelf; wie weet dat het kan, vindt na deze informatie de weg wel.

*A.6. Constructietip 6: verbindingen verbergen*

Bij ingewikkelder constructies is het heel onhandig als steeds alle hulplijnen en dergelijke zichtbaar blijven. In het voorbeeldbestand voor de evenredig delende drietallen zijn de lijnen dan ook verborgen.

De optie EXTRA2>>HIDE/SHOW geeft daar gelegenheid voor. Verberg/Verschijn: door aanklikken van een object wordt de status omgeschakeld van het een naar het ander.

*A.7. Constructietip 7: CABRI een test laten uitvoeren.*

Het gaat om de opties onder MEET1. Klik de juiste vraag aan, daarna de objecten waar de vraag over gaat en schuif de antwoordrechthoek naar de juiste plek en klik tot slot.

*A.8. Constructietip 8: De nauwkeurigheid van afgedrukte berekeningen wijzigen.*

Zoek bij OPTIES >> VOORKEUREN.

A.9. Constructietip 9: Werken met het rooster.

Kies eerst ‘Toon assenstelsel’ onder menu Extra2 en kies dan ‘Definieer rooster’. Klik het assenstelsel aan. Je kunt nu makkelijk zorgen dat gekozen punten tot het rooster horen.

LITERATUUR

1. J.M. AARTS. *Meetkunde, facetten van de planimetrie en stereometrie*. Epsilon uitgaven, Utrecht 2000.
2. ERIC BAINVILLE, BERNARD GENEVÉS. Constructions using conics. *Mathematical Intelligencer*, Volume 22, number 3, 2000.
3. M. BERGER. *Geometry*. Springer Verlag Berlin (1987), twee delen.
4. O. BOTTEMA. *Hoofdstukken uit de elementaire meetkunde*. Epsilon uitgaven, Utrecht 1997.
5. JAN VAN DE CRAATS. Alledaagse wiskunde. *Nieuw Archief voor Wiskunde*, september 2000.
6. EMMERICH, DR. A. *Die Brocardschen Gebilde und ihre Beziehungen zu den verwandten merkwürdigen Punkten und Kreisen des Dreiecks*. Berlin, Georg Reimer Verlag, 1891.
7. REDACTIE GETAL EN RUIMTE: *Practicumboekje Computerpracticum VWO NT 6. Onderdeel van deel NT6 van Getal en Ruimte*. Bevat informatie over CABRI en een inleidend practicum.
8. GOSSELINK, I. Praktische opdrachten Voortgezette Meetkunde. *Nieuwe Wiskrant. Tijdschrift voor Nederlands Wiskundeonderwijs*, 8, 4–9, (1999).
9. HAHN, LIANG-SHIN. *Complex numbers & Geometry*. The Mathematical Association of America (1994).
10. HONSBERGER, R. *Episodes in in Nineteenth and Twentieth Century Euclidean Geometry*. The Mathematical Association of America (1995).
11. DICK KLINGENS. *Homepage, meetkundepagina*.  
<http://www.pandd.demon.nl/meetkunde.htm>
12. FLOOR VAN LAMOEN. Morley Related Triangles on the Nine-point Circle. *The American Mathematical Monthly*; Volume 107, number 10, december 2000.
13. THE MAC TUTOR HISTORY OF MATHEMATICS ARCHIVE.  
<http://www-groups.dcs.st-and.ac.uk:80/history/>.
14. PROFI TEAM. *Werkstukken wiskunde B*. Freudenthal Instituut, 2001.
15. IAN STEWART. *Galois Theory*, second edition. Chapman and Hall, 1987.
16. CARLOS R. VIDELA. On points constructible from conics. *Mathematical Intelligencer*, Volume 19, number 2, 1997.
17. MICHAEL D. DE VILLIERS. *Rethinking Proof*. Key Curriculum Press, 1999.
18. ZIJLSTRA, J. *CABRI voor Windows*. Wolters Noordhoff, 1999.



## De Wondere Wereld van de Poisson Kansverdeling

Henk Tijms  
Vrije Universiteit, Amsterdam  
e-mail: [tijms@econ.vu.nl](mailto:tijms@econ.vu.nl)

## 1. INLEIDING

In 1837 verscheen het werk *Recherche sur la Probabilité des Jugement en Matière Criminelle et en Matière Civile* van de beroemde Franse wiskundige Siméon-Denis Poisson (1781-1840). Terloops wordt in dit werk een kansverdeling geïntroduceerd die later naar Poisson vernoemd zou worden en die zou uitgroeien tot één van de belangrijkste verdelingen uit de kansrekening. Poisson zelf onderkende niet het grote praktische belang van deze kansverdeling. Het was L. von Bortkiewicz die als eerste het belang van de Poisson verdeling onderkende in zijn boek *Das Gesetz der Kleinen Zahlen* uit 1898 met de onvergetelijke analyse van het aantal doden dat in het Pruisische leger viel door een klap met een paardenhoef. Wiskundig ontstaat de Poisson verdeling als limietverdeling van de binomiale kansverdeling met de parameters  $n$  en  $p$ . Als in de binomiale verdeling  $n \rightarrow \infty$  en  $p \rightarrow 0$  zodat  $np = \lambda$  constant blijft, dan geldt

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0}} \binom{n}{k} p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{voor } k = 0, 1, \dots$$

Het bewijs is in elk elementair boek over kansrekening te vinden, zie bijvoorbeeld TIJMS [3]. Voor toepassingen gaat het echter om de fysische interpretatie van het limietresultaat:

*Als een zeer groot aantal onafhankelijke uitvoeringen van een kans-experiment met een zeer kleine succeskans wordt gedaan, dan wordt de kansverdeling van het totale aantal successen bij goede benadering gegeven door de Poisson verdeling  $\{e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots\}$ , waarbij de getalwaarde van  $\lambda$  gegeven wordt door het product van het aantal proefnemingen en de succeskans.*

Deze interpretatie is een direct gevolg van het feit dat de binomiale kans  $\binom{n}{k} p^k (1-p)^{n-k}$  te zien is als de kans op  $k$  successen in  $n$  onafhankelijke uitvoeringen van een Bernoulli experiment met succeskans  $p$ , waarbij een Bernoulli experiment een kansexperiment is met ‘succes’ en ‘mislukking’ als de enige twee mogelijke uitkomsten.

De bovenstaande interpretatie verklaart waarom in vele situaties uit het dagelijkse leven de Poisson verdeling naar voren komt. Stochastische grootheden die bij goede benadering een Poisson verdeling volgen zijn het aantal ernstige

verkeersongelukken dat jaarlijks in een bepaald gebied plaatsvindt, het aantal schadeclaims dat een verzekeringsmaatschappij jaarlijks binnenkrijgt, het aantal oproepen dat dagelijks bij een alarmcentrale binnenkomt, het aantal postbodes dat jaarlijks door een hond gebeten wordt, etc. In dit soort toepassingen is het een uiterst nuttig gegeven dat de Poisson verdeling volledig vastgelegd is door zijn verwachtingswaarde  $\lambda$  en de eigenschap heeft dat de spreiding van de verdeling gelijk aan  $\sqrt{\lambda}$  is. Stel dat in de krant staat dat afgelopen jaar het aantal verkeersdoden met 20% gestegen is bij een gemiddeld aantal verkeersdoden van 1000 per jaar in de voorgaande jaren. Hoe moet je dit beoordelen? Het is redelijk een Poisson model te veronderstellen voor het aantal verkeersdoden per jaar. Een stijging van 20% op een gemiddelde van 1000 is dan een toename van 200, oftewel een toename van  $200/\sqrt{1000} = 6.3$  standaarddeviaties ten opzichte van het gemiddelde. Zo'n grote afwijking van de verwachtingswaarde kan geen toevalsfluctuatie zijn (de Poisson verdeling is te benaderen door de normale verdeling als  $\lambda \geq 25$ : de kans dat een normaal verdeelde stochast een waarde aanneemt die 6.3 of meer standaarddeviaties boven de verwachtingswaarde ligt is  $1.49E-10$ ). Hier is geen sprake van een toevalsfluctuatie maar moet er iets bijzonders aan de hand zijn. Anders ligt het wanneer de krant meldt dat het afgelopen jaar het aantal moorden in de stad met 20% gestegen is waar het gemiddeld aantal moorden per jaar in die stad gemiddeld 25 was in de voorgaande jaren. Hier betekent een stijging van 20% een toename van 5, oftewel een toename van  $5/\sqrt{25} = 1$  standaarddeviatie boven het gemiddelde. Dit valt heel wel te verklaren als een toevalstreffer. De Poisson verdeling speelt ook een hoofdrol in de probabilistische analyse van praktijkproblemen op het gebied van onder meer voorraadbeheer (grootwinkelbedrijven, distributiecentra) en wachtrijsystemen (luchtvaartmaatschappijen, telecommunicatie).

In deze bijdrage zullen we eerst laten zien dat de Poisson verdeling ook uiterst bruikbaar is in situaties die op het eerste gezicht niets te maken lijken te hebben met de fysische achtergrond van de Poisson verdeling. Het beroemde verjaardagsprobleem zal als uitgangspunt dienen. Vervolgens laten we zien hoe bruikbaar de Poisson verdeling is om loterijen te analyseren. Tenslotte wordt aandacht besteed aan geschikte praktische opdrachten in het studiehuis.

## 2. HET VERJAARDAGSPROBLEEM

Een simpel probleem met een uiterst verrassende oplossing: wat is de kans dat in een willekeurig gevormde groep van  $m$  personen twee of meer personen op eenzelfde dag jarig zijn? Hierbij veronderstellen we dat alle dagen van het jaar even waarschijnlijk zijn en nemen we aan dat het jaar 365 dagen telt (29 februari wordt buiten beschouwing gelaten). In het bijzonder is het interessant te weten hoe groot een willekeurig gevormde groep van personen moet zijn opdat de kans dat twee of meer personen op eenzelfde dag jarig zijn tenminste 50% is. Het verrassende antwoord op deze laatste vraag is 23. Vrijwel iedereen die voor het eerst van het verjaardagsprobleem hoort, geeft veel hogere waarden als antwoord. Een veelgehoord antwoord is 183.

Eerst zullen we een exacte formule voor het verjaardagsprobleem geven,

maar deze formule geeft weinig inzicht waarom een groep van 23 personen volstaat om een fifty-fifty kans te hebben dat in zo'n groep twee of meer personen op eenzelfde dag jarig zijn. Dit inzicht is wel te verkrijgen door een Poisson benadering voor de verjaardagskans af te leiden door op een 'handige' wijze het probleem op te vatten als een serie deexperimenten. Om de exacte oplossing van het verjaardagsprobleem te vinden, passen we een veelgebruikte aanpak uit de kansrekening toe. We berekenen de complementaire kans dat in een willekeurig gevormde groep van  $m$  personen elk van de personen op een verschillende dag jarig is. In totaal zijn er  $365^m$  mogelijke uitkomsten voor de verjaardagen van de  $m$  personen. Elk van deze uitkomsten is even waarschijnlijk. Het aantal uitkomsten waarin alle  $m$  verjaardagen op een verschillende dag vallen is  $365 \times 364 \times \dots \times (365 - m + 1)$ . De kans dat alle  $m$  personen op een verschillende dag jarig zijn is dus gelijk aan  $365 \times 364 \times \dots \times (365 - m + 1)$  gedeeld door  $365^m$ . Noteren we met

$$p_m = \text{de kans dat in een willekeurig gevormde groep van } m \text{ personen twee of meer personen op eenzelfde dag jarig zijn,}$$

dan vinden we dus de exacte formule

$$p_m = 1 - \frac{365 \times 364 \times \dots \times (365 - m + 1)}{365^m}.$$

Hoe vind je de Poisson benadering? Daarvoor is enige creativiteit vereist. Je moet het kansexperiment zien op te vatten als een reeks van deexperimenten. Het idee is om alle mogelijke combinaties van twee personen te beschouwen en na te gaan of er een combinatie is waarin beide personen op eenzelfde dag jarig zijn. Alleen dan als er zo'n combinatie is, zijn er twee of meer personen uit de hele groep op eenzelfde dag jarig. Op  $\binom{m}{2}$  manieren kun je twee personen uit een groep van  $m$  personen kiezen. Je voert dus  $\binom{m}{2} = \frac{1}{2}m(m-1)$  deexperimenten uit. Een deexperiment noem je succesvol als de twee personen op eenzelfde dag jarig zijn. Elk deexperiment heeft dezelfde kans  $p = \frac{1}{365}$  op succes. De uitkomsten van de deexperimenten zijn niet onafhankelijk van elkaar. De afhankelijkheid is echter als 'licht' op te vatten omdat het aantal mogelijke geboortedagen groot is. Laat de stochastische variabele  $X$  het aantal deexperimenten aangeven dat tot een succes leidt. In de groep van  $m$  personen zijn twee of meer personen op eenzelfde dag jarig alleen dan als tenminste één van de  $n = \frac{1}{2}m(m-1)$  deexperimenten tot een succes leidt. Dit betekent dat de gezochte  $p_m$  gegeven wordt door  $P(X \geq 1)$ . Op grond van de ontstaanswijze van de Poisson verdeling is het redelijk om de kansverdeling van  $X$  te benaderen door een Poisson verdeling met verwachtingswaarde  $\lambda = np = \frac{1}{2}m(m-1)/365$ . In het bijzonder vind je dan de benadering

$$P(X = 0) \approx e^{-\frac{1}{2}m(m-1)/365},$$

waarbij het  $\approx$ -teken staat voor 'is ongeveer gelijk aan'. De kans  $p_m$  is gelijk aan  $1 - P(X = 0)$ . Hiermee vind je voor de gezochte kans  $p_m$  de benadering

$$p_m \approx 1 - e^{-\frac{1}{2}m(m-1)/365}.$$

TABEL 1. De kans  $p_m$  in het verjaardagsprobleem

$m$	exact	Poisson	$m$	exact	Poisson
5	0.0271	0.0270	23	0.5073	0.5000
10	0.1169	0.1160	30	0.7063	0.6963
15	0.2529	0.2500	40	0.8920	0.8820
22	0.4757	0.4689	50	0.9704	0.9651

Dit blijkt na numerieke vergelijking een uitstekende benadering te zijn. In Tabel 1 geven we de exacte waarde en de Poisson benaderingswaarde van de kans  $p_m$  voor verschillende waarden van  $m$ .

De Poisson benadering verklaart ook waarom het minder verrassend is dan het lijkt dat een groep van 23 personen volstaat om een kans van tenminste 50% te hebben dat twee of meer personen op eenzelfde dag jarig zijn. In deze situatie voer je tenslotte  $\binom{23}{2} = 253$  deelexperimenten uit die elk een succeskans van  $\frac{1}{365}$  hebben!

De Poisson benaderingsaanpak geeft niet alleen inzicht, maar is ook een robuuste aanpak waarmee allerlei varianten van het verjaardagsprobleem kunnen worden geanalyseerd. Neem het bijna-verjaardagsprobleem: wat is de kans dat in een willekeurig gevormde groep van  $m$  personen twee of meer personen binnen één dag van elkaar jarig zijn? Was het vrij eenvoudig om het verjaardagsprobleem exact op te lossen, voor het bijna-verjaardagsprobleem ligt dit anders. De afleiding van de exacte oplossing van het bijna-verjaardagsprobleem is verre van eenvoudig en zal hier achterwege gelaten worden. De Poisson benadering voor de oplossing van het bijna-verjaardagsprobleem is echter even eenvoudig te geven als in het verjaardagsprobleem. Opnieuw beschouw je  $n = \binom{m}{2}$  deelexperimenten, waarbij je in elk deelexperiment de verjaardagen van twee personen vergelijkt. Een deelexperiment wordt nu succesvol genoemd als de twee personen binnen één dag van elkaar jarig zijn. Het enige wat verandert is de waarde van de succeskans. De kans dat twee gegeven personen binnen één dag van elkaar jarig zijn, is gelijk aan  $p = \frac{3}{365}$ . Noteren we

$p_m^*$  = de kans dat in een willekeurig gevormde groep van  $m$  personen twee of meer personen binnen één dag van elkaar jarig zijn,

dan is  $p_m^*$  gelijk aan de kans dat tenminste één deelexperiment succesvol is. De kansverdeling van het aantal succesvolle deelexperimenten benaderen we door een Poisson verdeling met verwachtingswaarde

$$\lambda = np = \frac{1}{2}m(m-1) \times \frac{3}{365}.$$

Dit leidt tot de benaderingsformule

$$p_m^* \approx 1 - e^{-\frac{3}{2}m(m-1)/365}.$$

TABEL 2. De kans  $p_m^*$  in het bijna-verjaardagsprobleem

$m$	exact	Poisson	$m$	exact	Poisson
5	0.0797	0.0789	20	0.8045	0.7902
10	0.3147	0.3092	25	0.9263	0.9151
13	0.4829	0.4733	30	0.9782	0.9720
14	0.5375	0.5267	35	0.9950	0.9925

Opnieuw blijkt uit numerieke vergelijking dat dit een uitstekende benadering is. In Tabel 2 geven we de exacte waarde en de Poisson benaderingswaarde van  $p_m^*$  voor verschillende waarden van  $m$ . In de boeken Tijms (2001) en Tijms, Heierman en Nobel (2000) komen verschillende andere varianten van het verjaardagsprobleem aan de orde. Het materiaal in deze boeken leent zich voor praktische opdrachten.

### 3. LOTERIJEN EN HET VERJAARDAGSPROBLEEM

In de Boston Globe van 6 februari 1978 stond een interview met een official van de Massachusetts Daily Lottery. Deze loterij bestond toen twee jaar. In de loterij wordt elke avond (behalve zondag) een viercijferig getal getrokken uit de 10 000 getallen 0000, 0001, ..., 9999. In het interview werd aan de official gevraagd of het in de voorbije 625 trekkingen was voorgekomen dat een getal meer dan één keer getrokken was. De loterijbobo antwoordde hierop ontkenkend en voegde daaraan toe dat het opnieuw vallen van een eerder getrokken getal pas te verwachten was na ongeveer 5000 trekkingen. Naar aanleiding van deze uitspraak ontving de krant verschillende ingezonden brieven waarin lezers te kennen gaven dat de loterijofficial hartstikke fout zat. Inderdaad waren meerdere getallen meer dan één keer gevallen in de voorbije 625 trekkingen van de Massachusetts Daily Lottery. De kans dat geen enkel getal twee of meer keer valt in 625 trekkingen van de Massachusetts Daily Lottery is onvoorstelbaar klein en is gelijk aan  $3.4 \times 10^{-9}$ . Hoe kun je deze kans berekenen? Simpelweg door te bedenken dat de vraagstelling terug te brengen is tot het verjaardagsprobleem met 10 000 mogelijke geboortedagen waarbij een willekeurig gevormde groep van 625 personen wordt beschouwd. Het enige wat je hoeft te doen is in bovenstaande formule voor het verjaardagsprobleem 365 te vervangen door 10 000. In het verjaardagsprobleem met 10 000 mogelijke geboortedagen en een willekeurig gevormde groep van  $m = 625$  personen is de kans op twee of meer personen met eenzelfde geboortedag gelijk aan de Poisson kans

$$1 - e^{-\frac{1}{2}m(m-1)/10\,000}$$

Oftewel de kans dat in 625 trekkingen van de loterij er niet een getal is dat twee of meer keer valt, is dus bij goede benadering gelijk aan

$$e^{-\frac{1}{2}(625 \times 624)/10\,000} = 3.4 \times 10^{-9}.$$

Achteraf gezien is het ook niet zo verwonderlijk dat deze kans zo enorm klein is: In feite is sprake van  $\binom{625}{2} = 195\,000$  deexperimenten elk met de succeskans  $\frac{1}{10\,000}$  zodat het verwachte aantal successen gelijk aan 19.5 is.

#### VERJAARDAGSVERRASSING IN DE QUEBEC SUPER LOTTO

Een soortgelijke anekdote is te geven over de Quebec Super Lotto in Canada. Deze lotto verloot elk jaar uit niet-geclaimd prijzengeld 500 auto's als bonusprijzen onder de 2.4 miljoen abonnees van de lotto. De computer genereert daartoe 500 maal random een nummer uit de 2.4 miljoen registratienummers van de abonnees zonder erop te letten of een nummer al getrokken is. De ongesorteerde lijst van de 500 getrokken nummers wordt in de krant gepubliceerd. Op 28 juli 1982 meldde de Montréal Gazette dat tot verbijstering van de loterijofficials een zekere Antonio Gallardo zich gemeld had voor twee gewonnen Oldsmobile auto's, zie Hanley (1992). Het registratienummer van de heer Gallardo kwam twee keer voor onder de gepubliceerde lijst van de 500 winnende nummers en de heer Gallardo had blijkbaar geen moeite gehad dit te ontdekken in de lijst van ongesorteerde nummers. De loterijofficials verklaarden dat de kans ongeveer 1 op de 46 miljoen was dat de heer Gallardo's registratienummer twee of meer keer getrokken zou worden. Op zichzelf is deze uitspraak juist (met het Poisson model volgt inderdaad dat de kans dat een *gegeven* registratienummer twee of meer keer valt in 500 trekkingen gelijk is aan

$$1 - e^{-\lambda} - \lambda e^{-\lambda} = 2.17 \times 10^{-8}$$

waarbij  $\lambda = np$  met  $n = 500$  en  $p = 1/2\,400\,000$ ). Maar de uitspraak is echter weinig relevant: het gaat om de kans dat er *een of ander* registratienummer is dat twee of meer keer voorkomt onder de 500 random getrokken registratienummers. Bedenken we dat voor deze kans in feite sprake is van een verjaardagsprobleem met 2.4 miljoen mogelijke verjaardagen en een willekeurig gevormde groep van 500 personen, dan volgt dat de gezochte kans benaderd wordt door de Poisson kans

$$1 - e^{-\frac{1}{2}500 \times 499/2\,400\,000} = 0.05.$$

Hoewel een kans van 5% niet groot is, is zo'n kans toch niet te verwaarlozen. Zeker gelet op het feit dat de Quebec Super Lotto elk jaar 500 bonusprijzen verloot uit niet-geclaimd prijzengeld zodat je er op wachten kon dat een keer iemand's registratienummer twee keer getrokken werd!

#### 4. HET UITKERINGSPERCENTAGE VAN DE NEDERLANDSE LOTTO

Sinds 10 juni 2000 draait een totaal vernieuwde Nederlandse Lotto. Voor die datum werden zes getallen uit de getallen 1, ..., 45 getrokken. In de nieuwe opzet worden zes (reguliere) getallen plus één bonusgetal uit de getallen 1, ..., 45 getrokken en daarnaast één kleur uit zes mogelijke kleuren. Het prijzenschema is ook totaal vernieuwd, zie Tabel 3. De jackpot bijvoorbeeld wordt



TABEL 3. Prijzenschema vernieuwde Lotto

Nr.	Combinatie	Prijs (in gulden)	Kans ( $p_i$ )
0	Jackpot:6 + kleur goed	4 miljoen of meer *	2.046E-08
1	6 goed	1 miljoen *	1.023E-07
2	5 + Bonusgetal + kleur goed	150.000 *	1.228E-07
3	5 + Bonusgetal goed	100.000 *	6.139E-07
4	5 + kleur goed	1.500	4.665E-06
5	5 goed	1.000	2.333E-05
6	4 + Bonusgetal + kleur goed	375	1.166E-05
7	4 + Bonusgetal goed	250	5.832E-05
8	4 + kleur goed	37,50	0.000216
9	4 goed	25	0.001079
10	3 + Bonusgetal + kleur goed	15	0.000288
11	3 + Bonusgetal goed	10	0.001438
12	3 + kleur goed	7,50	0.003452
13	3 goed	5	0.017262
14	2 + Bonusgetal + kleur goed	2 gratis loten met jackpot	0.002589
15	2 + Bonusgetal goed	2 gratis loten	0.012946
16	2 + kleur goed	1 gratis lot met jackpot	0.022656
17	2 goed	1 gratis lot	0.113282

\* prijs wordt bij meerdere winnaars gedeeld

nu gewonnen op een rijtje met de zes getrokken reguliere getallen goed plus de kleur goed. De jackpot begint met 4 miljoen gulden en wordt elke week dat de jackpot blijft staan, verhoogd met een half miljoen gulden. Voor een ingevuld rijtje dat meedingt naar de jackpot is de prijs verhoogd van f 1 naar f 1,50. Het is nog mogelijk rijtjes zonder kleur in te vullen voor f 1, maar daar wordt in de praktijk vrijwel geen gebruik van gemaakt. In het vervolg nemen we aan dat elk meespelend rijtje een f 1,50 rijtje is (op de gratis f 1 loten na). Een interessante vraag is: wat is het percentage van de inleg dat gemiddeld als prijzengeld wordt uitgekeerd? De Nederlandse Lotto beweert zelf dat dit ongeveer 50% is (de Volkskrant, 30 december 2000).

Hoe kun je dit als buitenstaander controleren? De Lotto geeft geen informatie over de inzet van de deelnemers. Dit in tegenstelling tot de Duitse Lotto die op de fraaie website [www.toto-lotto-bw.de](http://www.toto-lotto-bw.de) een schat aan informatie geeft over de inzet en de aantallen winnaars in elke prijs categorie over alle gedane trekkingen. De website van de Nederlandse Lotto steekt hierbij armetierig af. De enige publiek toegankelijke informatie die de Nederlandse Lotto geeft betreft de wekelijkse publicatie in de krant van de aantallen winnaars van de jackpot, van de hoofdprijs en van de 2de tot en met de 5de prijs (voor het gemak nummeren we de 18 prijs categorieën in Tabel 3 als prijs 0, prijs 1, ..., prijs 17, waarbij prijs 0 de jackpot is en prijs 17 twee goed is). We zullen laten zien dat het uit-

TABEL 4. Gepubliceerde aantallen prijswinnaars

week	jackpot	hoofdprijs	2e prijs	3e prijs	4e prijs	5e prijs
32	0	0	0	0	15	77
33	1	0	0	1	7	105
34	0	0	0	0	11	38
35	0	0	0	0	13	113
36	0	0	1	0	16	54
37	0	0	0	2	17	77
38	0	0	0	1	8	29
39	0	0	0	2	9	84
40	0	0	0	4	26	116
41	0	1	0	1	13	60
42	0	0	0	0	13	69
43	0	1	0	1	13	81
44	0	0	0	2	19	72
45	0	0	1	3	18	61
46	0	1	1	0	17	92
47	0	0	0	0	6	46
48	0	0	0	1	8	60
49	0	0	0	0	7	72
50	0	0	0	0	5	45
52	0	0	1	4	21	156
1	0	0	1	0	22	102
2	0	0	0	0	11	74
3	0	0	0	3	20	106
4	0	0	0	1	9	52
5	0	1	0	1	14	91
6	0	1	0	1	15	71
7	0	1	0	1	16	82
8	0	0	3	2	21	91
9	1	0	0	2	9	96
10	0	0	1	3	12	102

keringspercentage niettemin geschat kan worden uit de gedeeltelijke informatie over de aantallen prijswinnaars. Elementaire maar interessante kansrekening volstaat voor dit leuke probleem. Het materiaal in deze paragraaf is daarom mede geschikt voor onderwijsdoeleinden, zie ook paragraaf 5.

#### SCHATTING VOOR HET AANTAL INGEVULDE RIJTJES

In Tabel 4 hebben we de gegevens verzameld over de aantallen prijswinnaars van de eerste zes van de 18 prijzen over de laatste 20 weken van het jaar 2000 en de eerste 10 weken van het jaar 2001. Het totale aantal winnaars van de eerste

zes prijzen over deze 30 weken is gelijk aan  $w = 2838$ . Laat  $n$  het onbekende aantal rijtjes zijn dat over deze 30 weken is ingevuld. Laat

$p_i =$  de kans om de  $i^{\text{de}}$  prijs te winnen met één ingevuld rijtje

voor  $i = 0, 1, \dots, 17$ . In eerste instantie nemen we aan dat elk rijtje een f 1,50 rijtje is. Dan wordt de onbekende  $n$  geschat uit de vergelijking

$$n(p_0 + p_1 + p_2 + p_3 + p_4 + p_5) = w \quad (1)$$

met  $w = 2838$ . De kansen  $p_i$  ( $i = 0, \dots, 17$ ) worden in Tabel 3 gegeven. Deze kansen zijn eenvoudig uit te rekenen met behulp van het hypergeometrische kansmodel met 6 rode ballen (de zes getrokken reguliere lottogetallen), 1 blauw getal (het getrokken bonusgetal) en 38 zwarte ballen (de niet-getrokken getallen). Bedenk dat het aankruisen van de zes getallen in een rijtje in feite correspondeert met het trekken van zes ballen in dit hypergeometrische kansmodel. Daarbij komt nog dat de aangekruiste kleur met kans  $\frac{1}{6}$  goed gekozen is en met kans  $\frac{5}{6}$  verkeerd gekozen is. Bijvoorbeeld,

$$p_{10} = \frac{1}{6} \times \binom{6}{3} \binom{1}{1} \binom{38}{2} / \binom{45}{6}.$$

De relatie (1) geeft de schatting

$$n = 98\,364\,597$$

voor het totale aantal ingevulde rijtjes over 30 weken. Het is redelijk om

$$R = \frac{98\,364\,597}{30} = 3\,278\,820$$

als schatting te nemen voor het aantal rijtjes dat in een willekeurig gekozen week wordt ingevuld. Onder de voorlopige aanname dat elk rijtje een f 1,50 rijtje is, vinden we voor de wekelijkse inleg als eerste schatting

$$\text{gemiddelde inleg per week} = f\,1,50 \times R = 4\,918\,320 \text{ gulden.} \quad (2)$$

De formules (1) en (2) zullen we later iets aanpassen om rekening te houden met het feit dat een aantal van de ingevulde rijtjes van gratis loten afkomstig zijn en het feit dat sommige van deze gratis loten niet voor alle prijzen meedingen.

#### VERWACHTTE UITBETALING

Elk van de 18 prijzen uit Tabel 3 geeft een bijdrage tot de verwachtingswaarde van de wekelijkse uitbetaling aan prijzengeld. Definieer de stochastische variabele  $X_i$  als

$X_i =$  het aantal winnaars van prijs  $i$  in een willekeurige week.

Voor de prijzen  $i = 4, \dots, 17$  is de bijdrage tot de verwachte wekelijkse uitbetaling

$$E(X_i) \times (\text{geldbedrag van prijs } i), \quad (3)$$

waarbij we voorlopig de afspraak maken om de als loten uitgekeerde prijzen 14, 15, 16 en 17 als de geldbedragen f 3, f 2, f 1,50 en f 1 te tellen. Later zullen we hiervoor corrigeren. Verder stellen we voorlopig

$$E(X_i) = Rp_i \quad \text{voor } i = 4, \dots, 17 \quad (4)$$

(ook hier zullen we straks corrigeren voor die prijzen  $i$  waarvoor de kleur meetelt omdat onder de  $R$  loten een aantal gratis loten van f 1 zit waarop alleen een prijs zonder kleur kan vallen). Voor de prijzen 0 (= jackpot), 1 (= hoofdprijs), 2 en 3 geldt dat deze bij meerdere winnaars gedeeld worden. Voor  $i = 1, 2, 3$  nemen we de volgende bijdrage tot de verwachte wekelijkse uitbetaling:

$$P(X_i > 0) \times (\text{geldbedrag van prijs } i). \quad (5)$$

Onder de aanname van random ingevulde rijtjes is  $P(X_i = 0) = (1 - p_i)^R \approx e^{-Rp_i}$ . Dit geeft de Poisson benadering

$$P(X_i > 0) \approx 1 - e^{-Rp_i} \quad \text{voor } i = 1, 2, 3, \quad (6)$$

in overeenstemming met het feit dat sprake is van een zeer groot aantal ( $R$ ) experimenten elk met een zeer kleine succeskans ( $p_i$ ). Voor de jackpot geldt ook dat deze bij meerdere winnaars gedeeld wordt, maar hier is tevens de complicatie dat de jackpot niet een vast bedrag is. De jackpot begint met 4 miljoen gulden en loopt elke week met een half miljoen gulden op zolang de jackpot niet valt. Noteren we met  $p_J$  de kans dat de jackpot valt in een willekeurig gekozen week, dan is een schatting voor de verwachte hoogte van de jackpot als deze valt gelijk aan

$$\sum_{k=1}^{\infty} \left[ 4 + \frac{1}{2}(k-1) \right] p_J (1 - p_J)^{k-1} = 3.5 + \frac{1}{2p_J} \text{ miljoen gulden.}$$

Voor het verwachte aantal weken dat het duurt voordat de jackpot valt nemen we als schatting de verwachtingswaarde  $1/p_J$  van de geometrische verdeling  $\{p_J(1 - p_J)^{k-1}, k \geq 1\}$ . Dit geeft de schatting

$$\frac{3.5 + 1/(2p_J)}{1/p_J} \text{ miljoen gulden} \quad (7)$$

voor de gemiddelde bijdrage van de jackpot aan de wekelijkse uitbetaling. Resteert nog de kans  $p_J$  op het vallen van de jackpot te schatten. Hiervoor nemen we analoog aan (6) de Poisson benadering

$$p_J \approx 1 - e^{-Rp_0}. \quad (8)$$

TABEL 5. Ongecorrigeerde bijdragen tot de uitbetaling

Prijs	Bijdrage (in guldens)	Prijs	Bijdrage (in guldens)
0	727 119	9	88 436
1	284 992	10	14 150
2	49 708	11	47 166
3	86 638	12	84 898
4	22 946	13	282 995
5	76 485	14	25 470
6	14 341	15	84 898
7	47 803	16	111 429
8	26 531	17	371 431

In de Nederlandse Lotto met ongeveer 3 miljoen rijtjes per week is de kans ongeveer 94% dat de jackpot blijft staan. Het is dan ook niet verwonderlijk als de jackpot een (te) groot aantal opeenvolgende weken blijft staan. Zo is de kans dat de jackpot niet valt in 27 of meer opeenvolgende weken ongeveer gelijk aan  $(0.94)^{27}$  oftewel ongeveer 19%. Een hoge jackpot bevordert deelname aan de Lotto, maar niet als deze gemiddeld slechts één keer in de 17 weken valt (in de Amerikaanse Powerball Lottery met twee trekkingen per week valt de jackpot gemiddeld na ongeveer 9 trekkingen).

In de afleiding van de formules (6) en (8) hebben we impliciet aangenomen dat de deelnemers hun rijtjes *random* invullen. In werkelijkheid is dit niet het geval. Mensen gebruiken geboortedata, geluksgetallen, rekenkundige rijen, etc. om hun getallen te kiezen, vgl. Tijms (2001). Het feit dat een groot deel van de rijtjes niet random wordt ingevuld, heeft gevolgen voor de kansverdeling van  $X_i$  (= het aantal winnaars van prijs  $i$ ). De kansverdeling als geheel kan niet door een Poisson verdeling worden benaderd. Het gaat echter alleen om de kans  $P(X_i = 0)$ . Uit een empirische studie in Kadell and Ylvisaker (1991) voor de California Lotto 6/53 blijkt echter dat de Poisson benadering voor deze kans redelijk accuraat is (de empirisch bepaalde kans is iets groter dan de Poisson benadering). De uitdrukkingen (1) t/m (8) leiden tot Tabel 5 waarin voor de verschillende prijzen de (ongecorrigeerde) bijdragen tot de verwachte wekelijkse uitbetaling aan prijzengeld worden gegeven. Optellen van de bedragen in Tabel 5 geeft de schatting  $f$  2 447 435 voor de verwachte wekelijkse uitbetaling aan prijzengeld. Dit leidt tot de eerste schatting:

$$\text{ongecorrigeerde uitkeringspercentage} = \frac{2\,447\,435}{4\,918\,230} \times 100\% = 49.8\%.$$

##### 5. GECORRIGEERDE UITKERINGSPERCENTAGE

De berekening van bovenstaande schatting voor het uitkeringspercentage dient iets aangepast te worden. De wekelijkse inleg moet gecorrigeerd worden voor

het feit dat een aantal van de ingevulde rijtjes afkomstig zijn van gratis loten die uitgekeerd zijn als de prijzen 14 t/m 17, zie Tabel 3. Verder is het zo dat de gratis  $f$  1 loten niet meedingen voor prijzen waarvoor ook de getrokken kleur goed moet zijn. Tenslotte hebben we de als gratis loten uitgekeerde prijzen 14 t/m 17 gerekend als uitbetalingen ter waarde van  $f$  3,  $f$  2,  $f$  1,50 en  $f$  1. Deze geldbedragen dienen uiteraard gecorrigeerd te worden met het uitkeringspercentage van de Lotto. Een en ander leidt tot de invoering van drie parameters die we iteratief gaan schatten. Deze parameters zijn

$$\begin{aligned} u &= \text{uitkeringsfractie} \\ c_1 &= \text{fractie rijtjes behorende bij gratis loten} \\ c_2 &= \text{fractie rijtjes behorende bij gratis } f \text{ 1 loten.} \end{aligned}$$

De prijzen 15 en 17 zijn gratis  $f$  1 loten en deze loten dingen alleen mee voor prijzen waar de getrokken kleur er niet toe doet, oftewel de gratis  $f$  1 loten dingen alleen mee voor de prijzen 1, 3, 5, 7, 9, 11, 13, 15 en 17. De kans dat op een  $f$  1 rijtje prijs  $i$  met  $i$  oneven valt is uiteraard gelijk aan  $p_{i-1} + p_i$ . De formules (1) t/m (8) vereisen nu een kleine aanpassing. In formule (1) moeten we  $np_i$  vervangen door

$$n(1 - c_2)p_i \text{ voor } i = 0, 2, 4$$

en door

$$n(1 - c_2)p_i + nc_2(p_{i-1} + p_i) \text{ voor } i = 1, 3, 5.$$

Opgeteld geeft dit echter weer dezelfde vergelijking (1), hetgeen achteraf ook logisch is. In formule (2) moeten we  $R$  vervangen door  $R(1 - c_1)$ , terwijl we  $R$  in formule (4) moeten vervangen door  $R(1 - c_2)$  voor die prijzen  $i$  met  $i$  even. In formule (6) met  $i = 2$  en in formule (8) wordt  $R$  ook vervangen door  $R(1 - c_2)$ . In formule (4) met  $i$  oneven en in formule (6) met  $i$  oneven moet  $Rp_i$  door

$$R(1 - c_2)p_i + Rc_2(p_{i-1} + p_i)$$

vervangen worden. Tenslotte in formule (3) wordt voor  $i = 14, 15, 16$  en  $17$  in plaats van de geldbedragen  $f$  3,  $f$  2,  $f$  1,50 en  $f$  1 de geldbedragen  $u \times f$  3,  $u \times f$  2,  $u \times f$  1,50 en  $u \times f$  1 genomen. Bij het berekenen van het ongecorrigeerde uitkeringspercentage nemen we in eerste instantie  $u = 1$  en  $c_1 = c_2 = 0$ . De eerste ronde van de berekeningen leiden tot de nieuwe schattingen  $u = 0.4976$ ,  $c_1 = 0.1670$  en  $c_2 = 0.1392$ . Vervolgens herhalen we de berekeningen op basis van de aangepaste formules (1) t/m (8) met deze schattingen. Dit leidt tot nieuwe schattingen voor  $u$ ,  $c_1$  en  $c_2$ . Vervolgens herhalen we de berekeningen met deze nieuwe schattingen net zo lang tot de schattingen voor  $u$ ,  $c_1$  en  $c_2$  voldoende geconvergeerd zijn. Dit gebeurde al na vier iteratiestappen en de uiteindelijke schattingen van  $u$ ,  $c_1$  en  $c_2$  zijn

$$u = 0.5175, \quad c_1 = 0.1670, \quad c_2 = 0.1432$$

bij  $n = 98\,364\,597$ . Ter controle: de verwachte waarden

$$\begin{aligned} n(1 - c_2)p_4 &= 393 \\ n(1 - c_2)p_5 + nc_2(p_4 + p_5) &= 2360 \end{aligned}$$

voor de totale aantallen winnaars van de prijzen 4 en 5 stemmen overeen met de werkelijke waarden 411 en 2374 uit Tabel 4. Overigens een logische vraag is of gegevens over 30 weken voldoende zijn om tot betrouwbare uitspraken te komen. Het gaat echter niet om het aantal weken, maar om het totale aantal ingevulde rijtjes. Het aantal van ongeveer honderd miljoen ingevulde rijtjes over de 30 weken is groot genoeg.

In Tabel 6 geven we de uiteindelijke bijdragen van de verschillende prijzen aan de verwachtingswaarde van de wekelijkse uitkering van de lotto. De verwachte waarde van de totale uitkering in een willekeurige week is 2 120 093 gulden. Voor de gemiddelde wekelijkse inleg vonden we uiteindelijk op grond van de aangepaste versie van formule (2) het inlegbedrag  $f$  4 096 837. Dit geeft de schatting

$$\text{gecorrigeerde uitkeringspercentage} = \frac{2\,120\,093}{4\,096\,837} \times 100\% = 51.7\%$$

Dit is een iets hoger percentage dan het ongecorrigeerde uitkeringspercentage van 49.8%. Het werkelijke uitkeringspercentage wordt echter naar beneden beïnvloed door de volgende twee andere factoren. Ten eerste het feit dat het merendeel van de rijtjes met de hand en dus niet random wordt ingevuld. Dit betekent dat de werkelijke kans dat de jackpot en andere grote prijzen vallen kleiner is dan in de modelsituatie van random ingevulde rijtjes (om dit in te zien denk het extreme geval in dat alle deelnemers hetzelfde rijtje kiezen, in welk geval er praktisch gesproken nooit winnaars van de hoogste prijzen zullen zijn). Een ander aspect dat het uitkeringspercentage iets zal drukken is het feit dat niet iedere winnaar van de gratis loten deze loten ook werkelijk gebruikt. Het is een bekend verschijnsel dat niet alle prijzen geclaimd worden (in de Amerikaanse Powerball Lottery wordt ongeveer 12% van het prijzengeld niet geclaimd). Zou slechts 75% van de gratis loten ingeleverd worden, dan zou het uitkeringspercentage van 51.7% dalen naar 49.2%.

#### CONCLUSIE

Onze berekeningen bevestigen de uitspraak van de Lotto dat zij ongeveer 50% van de inleg als prijzengeld uitkeert aan de deelnemers (een lager percentage aan prijzengeld dan het uitkeringspercentage van meer dan 65% bij de Staatsloterij, maar een stuk hoger dan het uitkeringspercentage van nog geen 25% bij de Nationale Postcode Loterij). Wat we echter niet kunnen onderschrijven is een eerdere reclameslogan van de Lotto: Lotto de meeste kans om miljonair te worden. De kans om met één rijtje de jackpot te winnen is  $2.05 \times 10^{-8}$ . Deze kans is nog kleiner dan de kans om bij 25 worpen met een zuivere munt alleen maar kop te gooien. Zou je elke week trouw 12 rijtjes invullen, dan zou je meer dan 54 duizend jaar van leven moeten hebben om met een kans van tenminste

TABEL 6. Gecorrigeerde bijdragen tot de uitbetaling

Prijs	Bijdrage (in guldens)	Prijs	Bijdrage (in guldens)
0	695 531	9	90 968
1	291 826	10	12 124
2	43 759	11	48 516
3	87 386	12	72 744
4	19 661	13	291 098
5	78 675	14	11 293
6	12 288	15	45 192
7	49 172	16	49 409
8	22 733	17	197 717

50% ooit in je leven een keer de jackpot te winnen! Dit laatste is eenvoudig na te rekenen met het Poisson kansmodel. Vul je  $m$  jaar lang elke week 12 rijtjes in, dan is de kans dat je ooit de jackpot wint gelijk aan  $1 - e^{-52m \times 12p_0}$  met  $p_0 = 2.046E - 8$ . Stel je dit gelijk aan 0.5, dan vind je  $m = 54\,286$  jaar.

#### 6. PRAKTISCHE OPDRACHTEN

Een vereenvoudiging van het Lotto-probleem uit de vorige paragraaf leent zich goed voor een praktische opdracht aan VWO-leerlingen. De voorgestelde vereenvoudigingen zijn:

- laat de prijzen 14 t/m 17 met de gratis loten weg
- neem in eerste instantie een vaste jackpot ter waarde van bijvoorbeeld 12 miljoen gulden.

Het gedeelte uit de vorige paragraaf waarin correcties op het uitkeringspercentage worden aangebracht wordt overbodig als de prijzen 14 t/m 17 niet in beschouwing genomen worden. De volgende vragen kunnen dan aan de leerlingen worden gesteld:

- laat zien hoe de kansen in Tabel 3 worden berekend,
- bereken het totale verwachte aantal winnaars van de prijzen 0 t/m 5 als in totaal  $n$  rijtjes worden ingevuld door de deelnemers,
- schat het onbekende aantal rijtjes ( $R$ ) dat gemiddeld per week wordt ingevuld,
- bereken voor  $i = 0, 1, 2, 3$  de kans dat prijs  $i$  valt in een week waarin  $R$  rijtjes worden ingevuld,
- schat de gemiddelde uitbetaling aan prijzen per week en het gemiddelde uitkeringspercentage (neem aan dat elk rijtje  $f$  1,50 kost),



TABEL 7. Het prijzenschema voor de Powerball Lottery

prijs	goed	bedrag	kans ( $p_i$ )
1(=jackpot)	5 witte en 1 rode	\$ miljoenen	1.249E-8
2	5 witte	\$ 100 000	5.119E-7
3	4 witte en 1 rode	\$ 5 000	2.747E-6
4	4 witte	\$ 100	1.126E-4
5	3 witte en 1 rode	\$ 100	1.181E-4
6	3 witte	\$ 7	0.004843
7	2 witte en 1 rode	\$ 7	0.001654
8	1 witte en 1 rode	\$ 4	0.008475
9	1 rode	\$ 3	0.013560

- f. pas de schattingen in vraag e aan voor het geval de uitkering van de jackpot niet vast is maar begint met 4 miljoen gulden en elke week dat de jackpot niet valt met een half miljoen gulden wordt verhoogd. Bereken daartoe eerst de kansverdeling van het aantal weken dat de jackpot blijft staan en de verwachte hoogte van de jackpot op het moment dat deze valt.

Oplossingen van praktische opdrachten ‘circuleren’ na verloop van tijd onder leerlingen. De voorgaande opdracht is echter simpel aan te passen, bijvoorbeeld door de Lotto 6/47 te beschouwen in plaats van de Lotto 6/45 (in dat geval is het aan te bevelen de aantallen winnaars in Tabel 4 iets te wijzigen). Een andere variant van de voorgaande praktische opdracht is het volgende lottoprobleem.

#### AMERIKAANSE POWERBALL LOTTERY

Deze loterij wordt twee keer per week in een groot aantal staten in Amerika gespeeld. Bij elke trekking worden vijf witte ballen uit een drum met 49 witte ballen en één rode bal (de Powerball) uit een drum van 42 rode ballen getrokken. Deelnemers kunnen op 9 manieren prijzen winnen, zie Tabel 7. In de Tabel wordt ook de getalwaarde van de kans  $p_i$  gegeven om met 1 rijtje prijs  $i$  te winnen ( $i = 1 \dots, 9$ ). De jackpot wordt gewonnen als de nummers van de vijf getrokken witte ballen en het nummer van de getrokken rode bal goed voorspeld zijn. De jackpot wordt bij meerdere winnaars gedeeld. Naast de variërende jackpot is er voor elk van de acht prijzen 2 t/m 9 een vast bedrag dat elke winnaar ontvangt, ongeacht het aantal winnaars van de betreffende prijs. De hoogte van de jackpot hangt zowel af van de inleg van de deelnemers als van het aantal weken dat de jackpot al staat. In principe wordt een vast deel van de inleg elke keer gevoegd bij het bedrag dat al in de jackpot zit (door deze opzet loopt de inleg sterk op als de jackpot een aantal weken niet gevallen is). De filosofie van de Powerball Lottery is te gaan voor een zeer hoge jackpot (bij de hoogste jackpot uit de historie bestond de keuze tussen 295 miljoen dollar in 25 jaarlijkse betalingen of 161 miljoen dollar direct in cash). Uit gegevens van de loterij blijkt dat het prijzengeld dat aan de jackpot wordt

besteed gemiddeld 1.4 keer het gemiddelde bedrag is dat aan de andere acht prijzen tezamen wordt besteed. Dit laatste gegeven komt straks van pas bij het berekenen van het uitkeringspercentage van de Powerball Lottery, waarbij we uitgaan van de gegevens van de aantallen prijswinnaars van de hoogste vier prijzen van de in totaal negen prijzen. Deze aantallen winnaars van de prijzen 1 t/m 4 over de 105 trekkingen in het jaar 2000 worden gegeven in Tabel 8 (in feite is ook bekend dat de totale aantallen winnaars van de prijzen 5 t/m 9 respectievelijk gelijk zijn aan 129 940, 5 377 236, 1 837 423, 9 458 371 en 15 175 051, maar voor de opdracht doen we net alsof deze aantallen niet bekend zijn). De gegevens in Tabel 8 zijn ontleend aan de website [www.musl.com](http://www.musl.com) van de Powerball Lottery. De praktische opdracht bestaat uit de volgende vragen:

- a. waarom is de kans op de prijs 9 niet gelijk aan  $\frac{1}{42}$  maar gelijk aan  $\frac{1}{74}$ ?
- b. reken de kansen  $p_1, \dots, p_9$  in Tabel 7 na,
- c. schat op grond van de gegevens in Tabel 8 het totale verwachte aantal rijtjes dat over de beschouwde 105 trekkingen is ingevuld en schat de gemiddelde inleg per trekking (elk ingevuld rijtje kost \$ 1),
- d. schat de gemiddelde uitbetaling per trekking over alleen de vaste geldprijzen 2 t/m 9,
- e. schat de gemiddelde uitbetaling per trekking over alle prijzen 1 t/m 9. Gebruik hierbij dat de gemiddelde uitbetaling aan de jackpot ongeveer 1.4 keer zo groot is als de gemiddelde uitbetaling over de andere acht prijzen tezamen,
- f. schat het uitkeringspercentage van de Powerball Lottery,
- g. uit gegevens van de Powerball Lottery blijkt dat ongeveer gemiddeld 9.3 trekkingen nodig zijn alvorens de jackpot valt en dat de gemiddelde hoogte van de jackpot ongeveer gelijk is aan 31.5 miljoen dollar wanneer deze valt. Is dit in overeenstemming met hetgeen hiervoor berekend is?

De berekeningen in deze opdracht moeten leiden tot de conclusie dat het uitkeringspercentage van de Amerikaanse Powerball Lottery ongeveer 50% is. Het gebruikelijke uitkeringspercentage van Lotto's!

#### LITERATUUR

1. J.A. HANLEY. Jumping to coincidences, defying odds in the realm of the preposterous, *American Statistical Association* **46**, 197–202, 1992.
2. D. KADELL, D. YLVIKAKER. awful, *Chance*, **4**, 22–25, 1991.
3. H.C. TIJMS. *Spelen met Kansen*, 2<sup>de</sup> druk, Epsilon Uitgaven, Utrecht, 2001.

TABEL 8. De aantallen winnaars van de prijzen 1 t/m 4

Prijs	1	2	3	4	Prijs	1	2	3	4
12/30	0	1	11	665	6/28	0	3	20	854
12/27	2	4	14	816	6/24	0	2	26	884
12/23	0	12	37	1,469	6/21	0	6	17	686
12/20	0	2	22	794	6/17	0	2	16	731
12/16	0	4	23	770	6/14	0	5	16	673
12/13	0	3	23	772	6/10	0	6	26	726
12/9	0	3	19	804	6/7	0	2	11	572
12/6	0	3	20	805	6/3	1	5	19	840
12/2	0	7	23	897	5/31	0	1	16	623
11/29	1	24	163	5,662	5/27	1	6	32	1,487
11/25	0	17	69	2,790	5/24	0	7	30	1,142
11/22	0	14	72	2,466	5/20	0	5	23	1,057
11/18	0	17	49	1,891	5/17	0	6	17	1,009
11/15	0	6	49	1,654	5/13	0	3	24	1,005
11/11	0	7	34	1,612	5/10	0	5	24	900
11/8	0	5	30	1,127	5/6	0	3	14	933
11/4	0	7	34	1,206	5/3	0	1	28	685
11/1	0	4	30	1,204	4/29	0	6	17	795
10/28	0	3	30	996	4/26	0	4	24	935
10/25	0	3	13	780	4/22	0	5	20	903
10/21	0	2	10	866	4/19	1	6	59	2,217
10/18	0	1	19	717	4/15	0	9	37	1,702
10/14	0	6	22	825	4/12	0	7	49	1,742
10/11	0	2	14	596	4/8	0	6	36	1,475
10/7	0	1	17	806	4/5	0	3	47	1,483
10/4	0	4	14	603	4/1	0	5	33	1,340
9/30	0	2	20	695	3/29	0	5	20	891
9/27	0	1	6	533	3/25	0	6	21	912
9/23	2	6	27	837	3/22	0	2	16	830
9/20	0	3	18	834	3/18	0	3	20	852
9/16	0	6	12	751	3/15	0	0	13	735
9/13	0	0	17	637	3/11	0	5	33	1,169
9/9	0	3	28	780	3/8	0	0	29	1,065
9/6	0	1	23	793	3/4	3	35	204	8,163
9/2	0	4	17	1,034	3/1	0	35	117	5,044
8/30	0	1	10	556	2/26	0	9	80	2,632
8/26	1	6	23	797	2/23	0	10	47	2,242
8/23	0	3	13	524	2/19	0	4	45	1,466
8/19	0	3	23	894	2/16	0	3	20	1,385
8/16	1	2	18	691	2/12	0	5	24	1,107
8/12	0	2	20	698	2/9	0	5	23	862
8/9	0	6	20	978	2/5	0	3	16	887
8/5	0	3	15	789	2/2	0	3	35	1,367
8/2	0	3	22	625	1/29	0	0	23	832
7/29	0	4	20	815	1/26	0	4	19	625
7/26	1	2	13	670	1/22	0	0	23	796
7/22	1	5	47	1,518	1/19	0	2	9	623
7/19	0	8	35	1,266	1/15	0	3	15	776
7/15	0	4	33	1,142	1/12	0	2	16	552
7/12	0	3	25	1,140	1/8	1	3	19	833
7/8	0	5	27	1,108	1/5	0	5	22	868
7/5	0	1	26	796	1/1	0	4	18	944
7/1	0	3	20	1,013					





## Wiskunde in Telecommunicatienetwerken

Rob van der Mei

KPN Research, Leidschendam  
Vrije Universiteit, Amsterdam  
e-mail: R.D.vanderMei@kpn.com

De afgelopen jaren is de informatie- en communicatietechnologie sterk in opkomst, en anno 2001 is het gebruik van diensten zoals het World-Wide Web (WWW), electronic mail en mobiel telefoneren niet meer uit het dagelijkse leven weg te denken. Telecommunicatienetwerken moeten daardoor steeds grotere hoeveelheden digitale informatie transporteren. Dit kan leiden tot enorme verkeersdrukke in het netwerk, waardoor informatie vaak te laat bij de eindgebruiker aankomt, als de eindgebruiker überhaupt al toegang tot het netwerk krijgt. Een bekend voorbeeld is het WWW, dat vaak spottend de bijnaam World-Wide Wait verdient. Om verkeersproblemen in telecommunicatienetwerken te kunnen voorspellen, te voorkomen en op te lossen speelt kennis van de wiskunde een belangrijke rol. In dit verhaal wordt ingegaan op de wiskundige modellering van het verkeer in telecommunicatienetwerken.

### 1. EEN STUKJE GESCHIEDENIS

Telecommunicatie is een vak met een rijke geschiedenis. De eerste vormen van communicatie over afstand hebben een sterke militaire component. De toneelschrijver Aeschylus (vijfde eeuw voor Christus) beschrijft al dat de val van Troje op snelle wijze door de overwinnaars aan Zeus wordt doorgegeven. Via grote vuren op enkel bergtoppen wordt de overwinning aan de oppergod gemeld. Op deze wijze kan snel een afstand van 500 kilometer worden overbrugd. Er is hier sprake van een zeer vroege vorm van digitale optische communicatie. Een verbeterd optisch communicatiesysteem vinden we in Europa aan het eind van de 18e eeuw. Het systeem is gebaseerd op een vinding van de Fransman Claude Chappe. Op torens worden beweegbare armen geplaatst. Iedere stand van de armen heeft een bepaalde betekenis. Op deze wijze kan teken voor teken een bericht worden doorgegeven. Dit systeem staat bekend als de luchttelegraaf. Een bericht van vijftig woorden doet er zes uur over om van Straatsburg naar Parijs te worden getransporteerd. Met de uitvinding van het Morse-toestel rond 1836 ontstaat er een nieuwe en veel snellere communicatiestructuur, waarmee

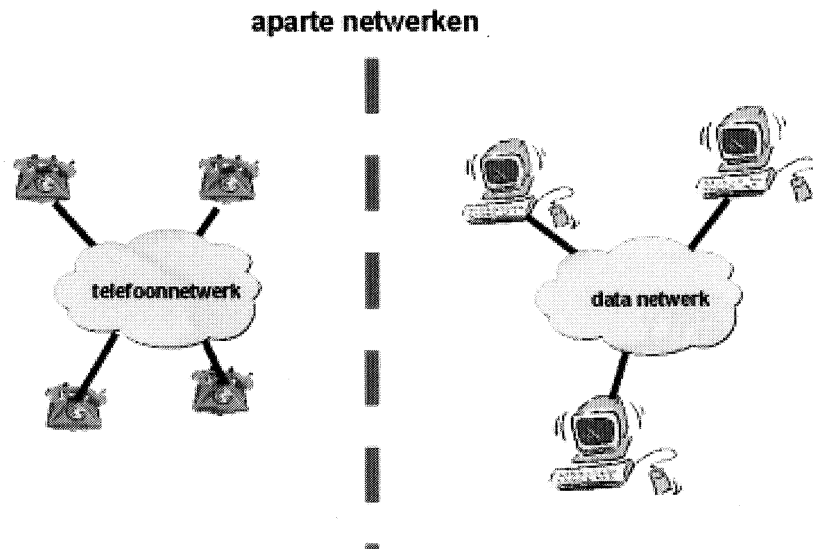
ook transatlantische communicatie mogelijk wordt. Hoewel deze communicatiemogelijkheid voornamelijk door overheden wordt gebruikt, ontstaat er ook een publieke dienst die voor commerciële doeleinden kan worden ingezet. Snelle communicatie blijkt een lucratieve faciliteit. Rond 1877 vindt Graham Bell de telefoon uit, die onder de naam klankentelegraaf in Nederland wordt geïntroduceerd. In de jaren twintig doet de radio zijn intrede, rond 1930 komt de televisie op de markt en in de jaren zestig begint mobiele telefonie op te komen. In de jaren zeventig zien we een nieuwe ontwikkeling: de computer doet zijn intrede in het communicatieveld. Via op afstand gekoppelde terminals kan de computer zonder tussenkomst van een computercentrum rechtstreeks worden gebruikt. Bovendien komen er minicomputers, en later zelfs microcomputers. Uiteindelijk resulteren de technische ontwikkelingen in de PC zoals we die nu kennen. Er ontstaat vervolgens behoefte aan telecommunicatie tussen deze systemen: de datacommunicatie doet zijn intrede. Door het groeiend gebruik van computers neemt het belang van datacommunicatie verder toe. Het gebruik van modems maakt het mogelijk telefonienetten te gebruiken voor datacommunicatie. Door de opkomst van het Internet aan het eind van de jaren zestig zijn de ontwikkelingen op het gebied van datacommunicatie nog eens versneld. Uiteindelijk is het de universitaire wereld die de mogelijkheden van de computer maximaal heeft benut om nieuwe communicatievormen en netstructuren te ontwikkelen. Het resultaat is de opkomst van multimedia-diensten, een communicatievorm waarin spraak, tekst en beeld zijn geïntegreerd. Het Internet wordt echter pas echt *booming* in de jaren negentig met de doorbraak van het World-Wide Web (WWW) en electronic mail. Anno 2001 is het Internet niet meer uit het dagelijks leven weg te denken.

## 2. TECHNOLOGISCHE ONTWIKKELINGEN

De opkomst van de informatie- en communicatietechnologie heeft grote gevolgen voor de maatschappij: voor het bedrijfsleven, de overheid en voor particulieren. Aan deze ontwikkeling ligt een aantal technologische ontwikkelingen ten grondslag, zoals de sterke toename van de rekenkracht van computers, de beschikbare hoeveelheid opslagruimte (geheugen, disk-ruimte) en de transmissiecapaciteit (bandbreedte), en de digitalisering van audio en video. Daarnaast zien we nog een aantal trends op het gebied van netwerken, zoals hun onderlinge integratie en de overgang van circuit- naar pakketgeschakelde netwerken.

### 2.1. Integratie van netwerken

Een belangrijke ontwikkeling die aan de basis heeft gestaan van de opkomst van multimedia-diensten is de integratie van netwerken. Tot die tijd werden voor het leveren van de meeste diensten *aparte* netwerken aangelegd. Zo bestond de infrastructuur uit een verzameling van fysiek en logisch gescheiden netwerken, voor verschillende toepassingen als telefoon-, data-, telex-, huurlijnen- en autotelefoonverkeer. Integratie van deze aparte netwerken heeft geleid tot de ontwikkeling van geïntegreerde netwerken. Geïntegreerde netwerken maken het mogelijk allerlei nieuwe diensten te leveren, zowel voor de consumenten-



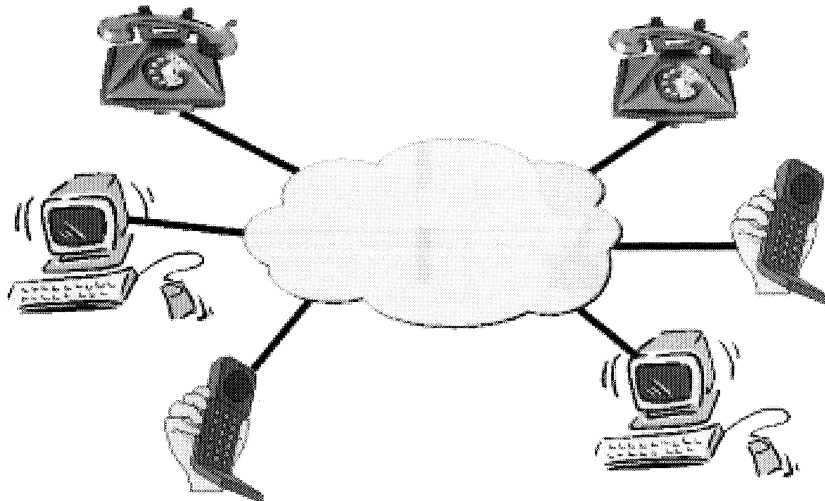
FIGUUR 1. Netwerkindegratie: overgang van gescheiden netwerken ...

markt (betaal-TV, telewerken, beeldtelefoon, video-vergaderen) als de zakelijke markt (digitale huurlijnen, koppeling van Local Area Networks, bandwidth-on-demand). De overgang van aparte naar geïntegreerde netwerken wordt geïllustreerd in Figuren 1 en 2.

### 2.2. Overgang van circuit- naar pakketgeschakelde netwerken

Parallel aan de integratie van netwerken loopt de overgang van zogenaamde circuitgeschakelde naar pakketgeschakelde netwerken. Het traditionele telefoonnetwerk is *circuitgeschakeld*: om een telefoonverbinding tussen twee eindpunten mogelijk te maken wordt een bi-directionele verbinding (een circuit genoemd) opgezet die actief blijft totdat de telefoonverbinding wordt afgebroken. Daarbij wordt voor iedere verbinding een vaste hoeveelheid transmissiecapaciteit (meestal bandbreedte genoemd) in het netwerk gereserveerd. Voor standaard vaste telefonie is dat bijvoorbeeld 64 kilobits per seconde, en voor het mobiele GSM netwerk ongeveer 10 kilobits per seconde (afhankelijk van de gebruikte codering). In een *pakketgeschakeld* netwerk daarentegen wordt niet van tevoren een vaste hoeveelheid bandbreedte gereserveerd, maar wordt bandbreedte alleen gebruikt wanneer er daadwerkelijk informatie verstuurd wordt. Daartoe wordt een hoeveelheid digitale informatie die over het netwerk moet worden verstuurd (zoals bijvoorbeeld een Web-pagina, videobeelden of gedigitaliseerd stemgeluid) opgesplitst in een aantal pakketten (de maximale lengte van een pakket verschilt per netwerk), die vervolgens over het netwerk naar hun bestem-

### geïntegreerd netwerk

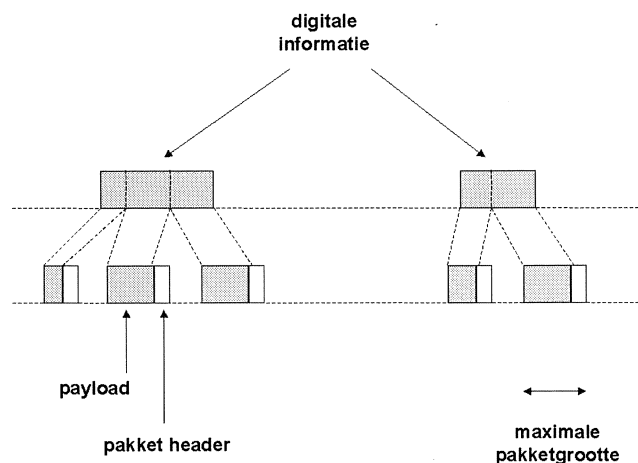


FIGUUR 2. ... naar geïntegreerde netwerken.

ming worden gestuurd. Een pakket bestaat uit twee delen: de *pakket header* (met informatie over het de lengte van het pakket, de bestemming van het pakket, enzovoorts) en de *payload* (de daadwerkelijke informatie), zie Figuur 3. Aan de ontvangende kant worden de pakketjes vervolgens weer netjes uitgepakt en verwerkt.

Het verschil tussen circuitgeschakelde en pakketgeschakelde netwerken vindt een aardige analogie met het wegennetwerk. In circuitgeschakelde netwerken heeft iedere weggebruiker een gereserveerde rijstrook ter beschikking, terwijl in pakketgeschakelde netwerken meerdere weggebruikers op dezelfde rijstrook kunnen rijden. Deze analogie met het wegennetwerk maakt ook meteen de wederzijdse voor- en nadelen duidelijk. Het grote voordeel van circuitgeschakelde netwerken is dat er feitelijk geen verkeersproblemen (congestie) in het netwerk voorkomen, terwijl er in pakketgeschakelde netwerken de kans op verkeersopstoppingen aanwezig is, zeker wanneer het netwerk zwaar belast is. De prijs die circuitgeschakelde netwerken hiervoor moeten betalen is verlies aan efficiëntie van het gebruik van transportcapaciteit. Zo kunnen in een pakketgeschakeld netwerk in principe meer verbindingen worden toegelaten. Figuur 4 illustreert een situatie waarin drie verbindingen zijn aangevraagd. In een circuitgeschakelde implementatie kunnen slechts twee verbindingen tegelijk actief kunnen zijn,





FIGUUR 3. Digitale informatie wordt in pakketjes over het netwerk gestuurd.

terwijl in een pakketgeschakelde implementatie alle drie de "verbindingen" (al kun je hier niet echt van een verbinding spreken) tegelijk actief kunnen zijn. Het verschil in efficiëntie tussen circuit- en pakketgeschakelde netwerken manifesteert zich vooral wanneer het verkeer dat door een applicatie gegenereerd wordt *grillig* is, dat wil zeggen dat de applicatie perioden van hoge en lage activiteit afwisselt. De meeste moderne applicaties, zoals World-Wide Web, electronic mail en Video-on-Demand, genereren grillig verkeer. Pakketgeschakelde netwerken zijn dan ook geschikter voor het ondersteunen van de meeste moderne applicaties dan circuitgeschakelde netwerken. Dat is dan ook een van de redenen dat pakketgeschakelde netwerken de laatste jaren steeds populairder zijn geworden.

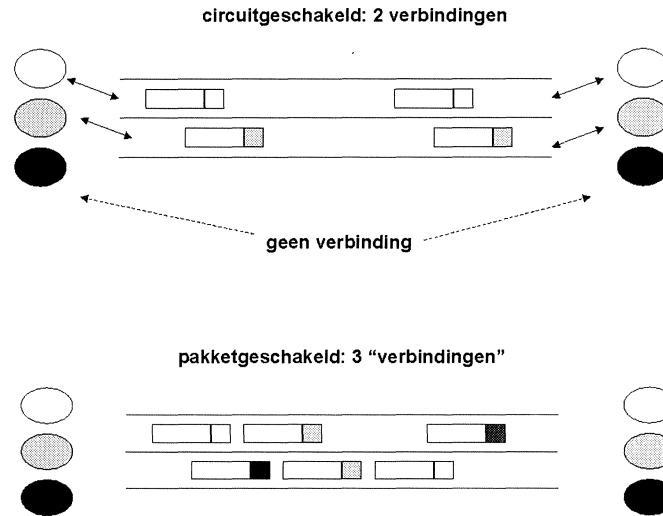
### 3. VERKEERSMODELLERING VAN TELEFOONNETWERKEN

Het verkeer in circuitgeschakelde telefoonnetwerken heeft twee belangrijke entiteiten: het aankomstproces van inkomende gesprekken (ook wel calls genoemd), en de verdeling van het lengte van een call.

#### 3.1. Aankomstproces van telefoongesprekken

Verkeersmetingen in telefoonnetwerken hebben experimenteel aangetoond dat calls over een tijdsperiode  $t$  ( $t = 1, 2, \dots$ ) bij benadering worden gegenereerd volgens een homogeen Poisson-proces<sup>2</sup> met aankomstintensiteit  $\lambda(t)$ . Dat wil

<sup>2</sup> Het Poisson proces is een natuurlijke manier om het onzekere karakter van de aankomsten van telefoontjes te beschrijven, en wordt is meer detail besproken tijdens de lezing van



FIGUUR 4. Circuit- versus pakketgeschakelde netwerken: verschil in efficiency.

zeggen dat de tijden tussen opeenvolgende calls onderling onafhankelijk en exponentieel verdeeld zijn met gemiddelde  $1/\lambda(t)$ : als  $X$  de tijd tussen twee opeenvolgende calls is, dan geldt voor  $t = 1, 2, \dots$ ,

$$\text{Prob}\{X > s\} = e^{-s\lambda(t)} \quad (s > 0). \quad (1)$$

Een interessante eigenschap van de exponentiële verdeling is dat deze *geheugenloos* is, dat wil zeggen

$$\text{Prob}\{X > s + t \mid X > s\} = \text{Prob}\{X > t\} \quad (s, t > 0). \quad (2)$$

De geheugenloosheidseigenschap van de exponentiële verdeling, die aan het Poisson-proces ten grondslag ligt, is in overeenstemming met de intuïtie dat telefoongesprekken "volledig random" gegenereerd worden. Stel bijvoorbeeld dat het gemiddeld aantal telefoongesprekken in een uur 60 is, maar dat gedurende twee minuten geen calls zijn aangekomen. Dan zegt deze informatie niets over hoe lang we nog moeten wachten totdat de volgende call aankomt.

Merk op dat het gemiddeld aantal calls per tijdseenheid over het algemeen varieert van uur tot uur, van dag tot dag, van week tot week, van maand tot maand, etcetera. Het tijdsinterval  $t$  waarover  $\lambda(t)$  constant verondersteld kan worden is in de orde van grootte van een kwartier tot een uur. In de praktijk is  $\lambda(t)$ , het gemiddeld aantal calls ter tijdseenheid voor tijdsinterval  $t$  ( $t = 1, 2, \dots$ ) vrij nauwkeurig te voorspellen op grond van historische gegevens.

### 3.2. Lengte van een telefoongesprek

Traditioneel wordt aangenomen dat de gespreksduren in telefoonnetwerken ook exponentieel verdeeld zijn. Recentelijke verkeersmetingen in telefoonnetwerken hebben echter experimenteel aangetoond dat de gespreksduur een kansverdeling met een *dikke staart* heeft.

#### DEFINITIE

Een niet-negatieve stochastische variabele  $X$  heeft een *dikke staart* als

$$\text{Prob}\{X > t\} = c(t)t^{-\alpha}, \quad t \rightarrow \infty, \quad 1 < \alpha < 2, \quad (3)$$

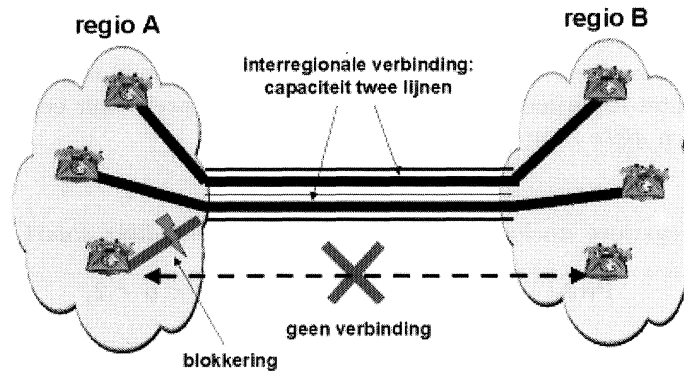
waarbij  $c(\cdot)$  de eigenschap heeft dat  $\lim_{t \rightarrow \infty} c(tx)/c(t) = 1$  voor alle  $x > 0$ . De functie  $c(\cdot)$  wordt ook wel *langzaam variërend* bij oneindig genoemd. Voorbeelden van langzaam variërende functies zijn de constante functie  $c(t) = c_0$  en de logaritmische functie  $c(t) = \log(t)$ . De bovengenoemde stochastische variabele  $X$  wordt ook wel *regulier variërend met index  $\alpha$*  genoemd.

De intuïtie achter stochastische variabelen met dikke staarten is dat de kans op uitschieters relatief groot is ten opzichte van kansvariabelen met dunne staarten (dat wil zeggen met een exponentieel dalend staartgedrag voor  $t \rightarrow \infty$ ). Merk op dat onder de bovengenoemde definitie een stochastische variabele met een dikke staart een eindige verwachting, maar een oneindige variantie, heeft.

De vraag rijst *waarom* de duur van een telefoongesprek een dikke staart heeft. Een interessante verklaring wordt gegeven door Bolotin [2], die stelt dat aan de lengte van telefoongesprekken een *logaritmisch principe* ten grondslag ligt. Het idee daarachter is dat tijdens conversaties de menselijke perceptie van tijd de Wet van Weber volgt, die zegt dat de "gelijkwaardige" perceptie van verschillen wordt gemeten op een logaritmische schaal: "*just noticeable difference*" in a stimulus bears a constant ratio to the stimulus. In de context van telefoonconversaties: een uitbreiding van een gesprek van 30 naar 40 seconden "voelt hetzelfde" als een uitbreiding van een gesprek van 3 naar 4 minuten.

### 3.3. Wiskundig intermezzo: de Erlang-blokkeringsformule

Een eenvoudige toepassing van elementaire wiskunde is het zogenaamde *dimensioneren* van telefoonnetwerken. Stel twee regio's,  $A$  en  $B$ , zijn met elkaar verbonden via een interregionale netwerkverbinding, die het mogelijk maakt dat klanten uit regio  $A$  kunnen bellen met klanten uit regio  $B$ , en vice versa. De verbinding heeft echter slechts twee lijnen, waardoor maximaal twee interregionale telefoongesprekken tegelijkertijd gevoerd kunnen worden. Als een klant uit regio  $A$  een verbinding aanvraagt met een klant uit regio  $B$  (of omgekeerd) terwijl beide lijnen bezet zijn, dan wordt de aanvraag van deze verbinding geweigerd. Dit fenomeen wordt ook wel *blokkering* genoemd. Figuur 5 illustreert de bovenstaande situatie. De kans dat een willekeurig telefoontje geblokkeerd wordt, noemen we de *blokkeringskans*.



FIGUUR 5. Alle lijnen bezet: nieuwe verbinding geblokkeerd.

Het zal duidelijk zijn dat de klant van een telefoonmaatschappij ontevreden wordt (en misschien wel wegloopt naar een concurrent!) wanneer te vaak blokkering optreedt, ofwel wanneer de blokkeringskans te hoog is. Natuurlijk kan een telefoonmaatschappij de blokkeringskans laag houden door maar genoeg telefoonlijnen neer te leggen. In het uiterste geval kan een aparte telefoonlijn tussen ieder tweetal eindgebruikers worden aangelegd, maar de kosten van aanleg van al die telefoonlijnen zijn natuurlijk erg hoog. Stel bijvoorbeeld dat een telefoonmaatschappij  $M$  klanten heeft, dan zouden  $M(M-1)/2$  telefoonlijnen moeten worden aangelegd. Dat loopt snel op: voor een maatschappij met 1000000 klanten zou dit neerkomen op maar liefst  $1/2 \times 1000000 \times 1000000 = 499999500000$  telefoonlijnen! En reken maar dat dat duur is! Daarom moet een telefoonmaatschappij een zorgvuldige afweging maken tussen een voldoende lage blokkeringskans enerzijds en de kosten van het aanleggen en onderhouden van telefoonlijnen anderzijds. Een interessante vraag is nu:

*"Hoeveel telefoonlijnen moeten minimaal worden neergelegd tussen regio A en regio B zodanig dat de blokkeringskans kleiner is dan een of andere kans  $\alpha$ ?"*

Dit heet het *dimensioneren* van het telefoonnetwerk. Merk op dat de maximale blokkeringskans  $\alpha$  (een getal tussen 0 en 1) iets zegt over de service-graad die de telefoonmaatschappij aan de klant wil of kan leveren - typische waarden van  $\alpha$  zijn 0.01%, 0.001% of zelfs lager.

Om de bovenstaande vraag te beantwoorden maken we aantal aannamen. Stel dat er interregionale telefoongesprekken tussen regio A en regio B worden aangevraagd volgens een zogenaamd Poisson-proces met een gemiddelde van  $\lambda$  telefoontjes per minuut, dat de gemiddelde gespreksduur van een telefoontje  $\mu$  minuten bedraagt, en dat er  $N$  telefoonlijnen tussen regio A en regio B zijn. De blokkeringskans, die afhangt van  $\lambda$ ,  $\mu$  en  $N$ , wordt genoteerd als  $p_{\text{blokkering}}(\lambda, \mu, N)$ .

Dan kan de vraag als volgt worden geformuleerd:

"Wat is de kleinste  $N$  zodanig dat, voor gegeven  $\lambda$  en  $\mu$ ,

$$p_{\text{blokkering}}(\lambda, \mu, N) < \alpha? \quad (4)$$

Deze vraag kan worden opgelost met behulp van de wiskundige theorie van zogenaamde Markov-ketens. Eenvoudig kan worden aangetoond dat de blokkeringskans wordt gegeven door de volgende formule:

$$p_{\text{blokkering}}(\lambda, \mu, N) = \frac{(\lambda\mu)^N / N!}{\sum_{i=0}^N (\lambda\mu)^i / i!}. \quad (5)$$

Deze formule wordt ook wel de *Erlang-blokkeringsformule* genoemd. Met behulp van deze formule kan voor iedere gewenste  $\alpha$ , en voor gegeven  $\lambda$  en  $\mu$ , direct worden bepaald hoeveel lijnen  $N$  er minimaal nodig zijn zodanig dat de blokkeringskans kleiner is dan  $\alpha$ .

#### 4. VERKEERSMODELLERING VAN INTERNET-VERKEER: INTRIGERENDE EIGENSCHAPPEN

Het verkeer op het Internet, en meer algemeen in pakketgeschakelde geïntegreerde netwerken, hebben een aantal fundamenteel andere eigenschappen dan het verkeer in circuitgeschakelde telefoonnetwerken. Op grond van uitgebreide verkeersmetingen is empirisch aangetoond dat het Internet een tweetal intrigerende eigenschappen bezitten die in de traditionele verkeersmodellen voor telefonie niet voorkomen: long-range dependence en schaalinvariantie. De definities, interpretaties en implicaties van deze eigenschappen worden in deze sectie besproken.

##### 4.1. Long range dependence

DEFINITIE

Zij  $X = \{X_t, t > 0\}$  een stationair stochastisch proces met autocovariantiefunctie

$$r_X(s) := \text{Cov}\{X_t, X_{t+s}\} \quad (s > 0), \quad (6)$$

onafhankelijk van  $t$ . Dan heet  $X$  *long-range dependent* (LRD) als

$$\int_{s=0}^{\infty} r_X(s) ds = \infty. \quad (7)$$

Anders heet  $X$  *short-range dependent* (SRD).

De niet-integreerbaarheid van de autocovariantiefunctie weerspiegelt de intuïtie achter LRD, dat ondanks het feit dat de autocovarianties  $r_X(s)$  klein zijn voor grote waarden van  $s$ , het cumulatieve effect van deze autocovarianties significant is. Het fenomeen LRD wordt ook wel het *Jozef-effect* genoemd, refererend

naar de persoon in het Oude Testament die de droom van de farao over "de zeven magere en zeven vette koeien" uitlegde als "zeven magere en zeven vette jaren" dat het oude Egypte in het vooruitzicht had. Intuïtief zegt long-range dependence iets over tijdsafhankelijkheden die in veel natuurlijke processen voorkomen, waarbij clusters van grote en kleine waarden voorkomen.

Een alternatieve karakterisering van LRD die in de literatuur op het gebied van telecommunicatiesystemen vaak wordt gebruikt is dat de autocovariantiefunctie  $r_X(s)$  hyperbolisch naar 0 convergeert voor grote waarden van  $s$ :

$$r_X(s) = c(s)s^{-\alpha}, \quad s \rightarrow \infty, \quad (8)$$

met  $0 < \alpha < 1$ , en waarbij  $c(\cdot)$  een langzaam variërende functie bij oneindig is. SRD processen daarentegen worden gekarakteriseerd door

$$r_X(s) = c(s)e^{-\eta s}, \quad s \rightarrow \infty, \quad (9)$$

met  $\eta > 0$ , ofwel door een autocovariantiefunctie die exponentieel naar 0 convergeert voor grote waarden van  $s$ . Merk op dat de alternatieve definitie van LRD in (8) zwakker is dan (7), en dat bijvoorbeeld periodieke processen wel aan (7) voldoet maar over het algemeen niet aan (8). In het vervolg zullen de alternatieve definities door elkaar worden gebruikt.

#### 4.2. Schaalinvariantie

Recente verkeersmetingen aan het Internet hebben nog een andere eigenschap van het verkeer aangetoond: *schaalinvariantie*. Simpel gezegd is het dynamische gedrag van een schaalinvariant proces onafhankelijk van de tijdsschaal waarop het bekeken wordt. Met andere woorden, in een schaalinvariant proces middelt de gemeten hoeveelheid verkeer niet uit over grote tijdsintervallen. Figuur 6 illustreert het verschil tussen een schaalinvariant proces en een niet-schaalinvariant proces door een realisatie te plotten van de gemiddelde hoeveelheid verkeer per tijdseenheid over verschillende tijdseenheden.

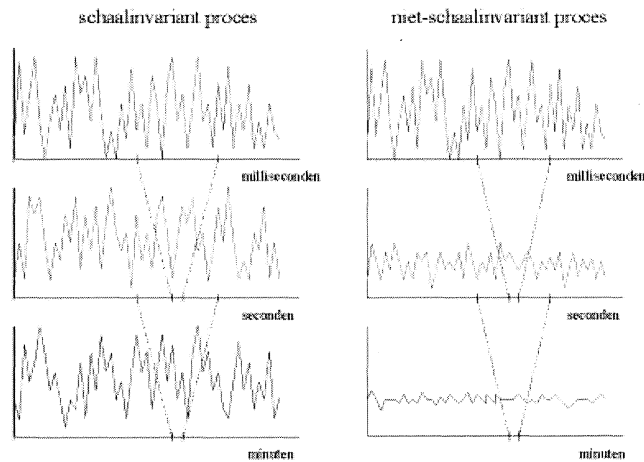
Op grond van verkeersmetingen op Internet-verbindingen heeft men empirisch aangetoond dat Internet-verkeer schaalinvariant is over verschillende tijdsschalen, variërend van milliseconden tot minuten (zie bijvoorbeeld [7]). Het is belangrijk op te merken dat schaalinvariante processen fundamenteel anders zijn dan niet-schaalinvariante processen. Zo is bijvoorbeeld het Poisson-proces, dat vaak wordt gebruikt om het aantal binnenkomende telefoongesprekken te modelleren (zie ook sectie 3) niet schaalinvariant: het aantal telefoontjes per tijdseenheid "middelt uit" over grote tijdsintervallen.

#### DEFINITIE

Zij  $X = \{X_t, t > 0\}$  een continuë tijds-proces met drift 0 en stationaire aangroeiingen. Dan heet  $X$  (strict) *schaalinvariant* wanneer voor alle  $a > 0$  geldt

$$X_{at} \stackrel{d}{=} a^H X_t \quad (t > 0), \quad (10)$$

waarbij de notatie " $\stackrel{d}{=}$ " staat voor gelijkheid voor alle eindig-dimensionale verdelingen.  $X$  heet *asymptotisch schaalinvariant* als  $X_{at}$  (genormaliseerd)



FIGUUR 6. Schaalinvariante en niet-schaalinvariantie processen.

zwak convergeert naar een strict schaalinvariant proces voor  $a \rightarrow \infty$ . In de context van telecommunicatienetwerken wordt vaak gekeken naar (de iets zwakkere eigenschap van) tweede-orde schaalinvariantie.  $X$  heet *tweede-orde schaalinvariant* als voor elke  $a > 0$   $X_{at}$  en  $a^H X_t$  dezelfde tweede orde karakteristieken hebben. Tenslotte heet  $X$  *asymptotisch tweede orde schaalinvariant* als de tweede orde karakteristieken  $X_{at}$  (genormaliseerd) convergeren naar de tweede orde karakteristieken van een tweede orde schaalinvariant proces als  $a \rightarrow \infty$ . De parameter  $H$  heet de *Hurst*-parameter. Het is eenvoudig na te gaan dat voor een niet-triviaal tweede orde schaalinvariant proces met stationaire aangroeiingen geldt dat  $H \in (0, 1)$ . Strict schaalinvariante processen met stationaire aangroeiingen en  $H \geq 1$  bestaan weliswaar, maar zijn niet praktisch relevant in de context van telecommunicatienetwerken. De Hurst parameter  $H$  is een indicatie voor de graad van schaalinvariantie.  $H \approx 1$  betekent dat bursts meer prominent aanwezig zijn op langere tijdsschalen. Het is eenvoudig in te zien dat  $H \in (1/2, 1)$  impliceert dat het proces  $X$  LRD is, terwijl  $X$  SRD is als  $H \in (0, 1/2]$ .

Schaalinvariantie (in de internationale vakliteratuur ook wel *self-similarity* genoemd) is een fenomeen dat voorkomt in verschillende fysische en wiskundige systemen. In de theorie van de *fractals* wordt getracht het voorkomen van schaalinvariante vormen in de natuur (zoals bijvoorbeeld de vorm van een slakkenhuis, de vorm van een blad een boom, de vorm van een bloemkool, of een kustlijn) te verklaren als gevolg van een basaal onderliggende mechanisme dat iteratief herhaald wordt. Een stuk prachtige wiskundige theorie over schaalinvariantie wordt beschreven in Peitgen et al. [6]. Een absolute must voor de

liefhebber!

De vraag rijst "Hoe ontstaat schaalinvariantie?" Het antwoord op deze vraag is niet eenduidig en is al jaren onderwerp van discussie tussen experts. Een interessante verklaring voor het bestaan van het fenomeen schaalinvariantie is gelegen in het feit dat individuele verkeersstromen feitelijk bestaan uit bursts van verkeer, waarbij de lengte van een burst een verdeling met een dikke staart heeft. Om precies te zijn, neem de superpositie van een  $N$  onderling onafhankelijke ON/OFF-bronnen, met waarden 1 (voor ON-perioden) en 0 (voor OFF-perioden), waarbij de lengte  $L_{ON}$  een dikke staart heeft (dat wil zeggen  $\text{Prob}\{L_{ON} > t\} = c(t)t^{-\alpha}$ ,  $1 < \alpha < 2$ ). Willinger et al. [7] bewijzen dat deze superpositie voor  $N \rightarrow \infty$  convergeert (in de zin van de eindig-dimensionale verdelingen) naar een Fractional Brownian Motion (FBM), een strict schaalinvariant proces, met Hurst parameter  $H = (3 - \alpha)/2$ . Met andere woorden, de totale hoeveelheid verkeer gegenereerd door een groot aantal ON-OFF bronnen met dikstaartige ON-perioden (die het burst-level verkeer gegenereerd door applicaties zoals WWW en electronic mail modelleren) is bij benadering schaalinvariant.

Een andere interessante vraag is "Wat zijn de implicaties zijn van schaalinvariantie?" Ook deze vraag is al jaren een punt van discussie. Een belangrijke issue is netwerk-dimensionering: hoeveel capaciteit moet aan de netwerkverbindingen worden toegekend. Als een verkeersstroom niet direct op een gegeven link kan worden toegelaten treedt buffering op. De vraag is: "Gegeven de karakteristieken van de verkeersstroom, hoeveel capaciteit moet de uitgaande link hebben, zodanig dat

$$\text{Prob}\{Q > x\} < \alpha?" \quad (11)$$

voor een of andere  $\alpha$ , en waarbij  $Q$  de buffervulling is. Of omgekeerd, "Gegeven de capaciteit van de uitgaande link, hoeveel verkeer met gegeven karakteristieken kan de link aan zodanig dat (11) geldt?" Omdat netwerkbuffers altijd eindig zijn kan de grootheid  $\text{Prob}\{Q > x\}$  bijvoorbeeld de kans representeren dat een binnenkomend pakket verloren gaat (omdat de buffer vol is), of de kans dat de wachttijd boven een bepaalde kritische waarde zit, hetgeen essentieel is voor de performance van real-time diensten, zoals beeldtelefonie. Met behulp van vrij elementaire wachttijdtheorie (een prachtig toepassingsgebied van de wiskunde!) kan worden aangetoond dat voor de traditionele SRD verkeersmodellen de buffervulling,  $Q$ , geldt (zie [1]):

$$\text{Prob}\{Q > x\} \approx e^{-\eta x} \quad (x \rightarrow \infty), \quad (12)$$

waarbij  $\eta > 0$ . Dat wil zeggen dat de staartkansen exponentieel snel naar 0 gaan als  $x \rightarrow \infty$ . Voor LRD modellen (in het bijzonder voor het FBM-model met parameter  $H$ ) geldt onder bepaalde aannamen dat (zie [5])

$$\text{Prob}\{Q > x\} \approx e^{-\gamma x^\beta} \quad (x \rightarrow \infty), \quad (13)$$



waarbij  $\gamma > 0$  en  $\beta = 2 - 2H \in (0, 1)$ . Met andere woorden, de staartkansen van de buffervulling voor LRD verkeersstromen gaan significant langzamer naar 0 als  $x \rightarrow \infty$  dan in SRD verkeersstromen! Deze resultaten impliceren dat het verwaarlozen van LRD kan resulteren in een significante onder-dimensionering van netwerk-buffers, hetgeen tot gevolg kan hebben dat de performance van het netwerk veel slechter is dan men zou verwachten op grond van SRD modellen. Dit illustreert een fundamenteel verschil tussen SRD en LRD modellen.

#### ACKNOWLEDGEMENT

De historische gegevens in de introductie zijn deels gebaseerd op [3]. Een meer gedetailleerde versie van dit verhaal is gepubliceerd in [4].

#### LITERATUUR

1. J. ABATE, G.L. CHOUDHURY and W. WHITT (1994). Asymptotics for steady-state tail probabilities in structures Markov chains. *Stochastic Models* **10**, 99-143.
2. V.A. BOLOTIN (1994). Telephone circuit holding time distributions. In: *The Fundamental Role of Teletraffic in the Evolution of Telecommunication Networks*, (eds. J. Labetoulle and J.W. Roberts), Elsevier Science, 125-134.
3. L.A.A.M. COOLEN (1998). *Er was telefoon*. Inaugurele rede Universiteit Maastricht.
4. R.D. VAN DER MEI (2000). Verkeersmodellering van netwerken. *Nieuw Archief voor Wiskunde* **5**, 390-396.
5. I. NORROS (1994). A storage model with self-similar input. *Queueing Systems* **16**, 387-396.
6. H.O. PEITGEN, H. JURGENS en D. SAUPE (1992). *Chaos and Fractals - New Frontiers of Science* (Springer-Verlag, New York).
7. W. WILLINGER, M.S. TAQQU, W.E. LELAND en D.V. WILSON (1995). Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements. *Statistical Science* **10**, 67-85.



## Medewerkers

Prof.dr. F. Beukers

Vakgroep Wiskunde, Universiteit Utrecht, Budapestlaan 6, 3584 CD Utrecht,  
030-2531419, beukers@math.uu.nl

Prof.dr. J. van de Craats

KMA, Postbus 90154, 4800 RG Breda, tel. 076-5273816, jcr@euronet.nl

Prof.dr. A. Doelman

Korteweg de Vries Instituut, Universiteit van Amsterdam, Plantage Muider-  
gracht 24, 1018 TV Amsterdam, 020-5255296, doelman@science.uva.nl

Drs. A.J. Goddijn

Freudenthal Instituut, Tiberdreef 4, 3561 GG Utrecht, 030-2611611,  
A.Goddijn@fi.uu.nl

M. Kindt

Freudenthal Instituut, Tiberdreef 4, 3561 GG Utrecht, 030-2611611,  
M.Kindt@fi.uu.nl

Dr. J.A. van Maanen

BE Initieel Onderwijs en  $\beta$ -didaktiek, Faculteit der Wiskunde en Natuurwe-  
tenschappen, Rijksuniversiteit Groningen, Postbus 800, 9700 AV Groningen,  
050-3637132, maanen@math.rug.nl

Dr. R.D. van der Mei

Vrije Universiteit, Wiskunde en Informatica, 1081HV Amsterdam  
KPN Research, afd. Netwerk planning, k. LC140, Postbus 421, 2260 AK Leid-  
schendam, 070-3326452, R.D.vanderMei@kpn.com

Prof.dr. H.C. Tijms

Faculteit der Economische wetenschappen en Econometrie, Vrije Universiteit,  
De Boelelaan 1105, 1081 HV Amsterdam, 020-4446013, tijms@econ.vu.nl

Prof.dr. H.A. van der Vorst

Vakgroep Wiskunde, Universiteit Utrecht, Budapestlaan 6, 3584 CD Utrecht,  
030-2533732, vorst@math.uu.nl

## Contacten Centrum voor Wiskunde en Informatica

Dr. M. Bakker

Centrum voor Wiskunde en Informatica, Kruislaan 413, Postbus 94079,  
1090 GB Amsterdam, 020 592 4172, Miente.Bakker@cwi.nl

Wilmy van Ojik

Centrum voor Wiskunde en Informatica, Kruislaan 413, Postbus 94079,  
1090 GB Amsterdam, 020 592 4200, Wilmy.van.Ojik@cwi.nl

## CWI SYLLABI

- 1 Vakantiecursus 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 Vacantiecursus 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 Vakantiecursus 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 Vakantiecursus 1987: *De personal computer en de wiskunde op school*. 1987.
- 15 Vakantiecursus 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 Vakantiecursus 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 Vakantiecursus 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 Vakantiecursus 1990: *Getallentheorie en haar toepassingen*. 1990.
- 28 Vakantiecursus 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 Vakantiecursus 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 Vakantiecursus 1993: *Het reële getal*. 1993.
- 36 Vakantiecursus 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 Vakantiecursus 1995: *Kegelsneden en kwadratische vormen*. 1995.
- 41 Vakantiecursus 1996: *Chaos*. 1996.
- 42 H.C. Doets. *Wijzer in Wiskunde*. 1996.
- 43 Vakantiecursus 1997: *Rekenen op het Toeval*. 1997.
- 44 Vakantiecursus 1998: *Meetkunde, Oud en Nieuw*. 1998.
- 45 Vakantiecursus 1999: *Onbewezen Vermoedens*. 1999.
- 46 P.W. Hemker, B.W. van de Fliert (eds.). *Proceedings of the 33<sup>rd</sup> European Study Group with Industry*. 1999.
- 47 K.O. Dzhaparidze. *Introduction to Option Pricing in a Securities Market*. 2000.
- 48 Vakantiecursus 2000: *Is wiskunde nog wel mensenwerk?* 2000.
- 49 Vakantiecursus 2001: *Experimentele wiskunde*. 2001.

## 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*. 1971.
- 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 der 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. Koksmä. *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. Pijs. *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. Pijs, 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 approximatiethorie*. 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 Jongh, 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; toepassing 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: eenstapmethoden*. 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. Koorwinder (ed.). *Representations of locally compact groups with applications, part I*. 1979.
- 38.2 T.H. Koorwinder (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. Koorwinder (ed.). *The structure of real semisimple Lie groups*. 1982.
- 50 H. Nijmeijer. *Inleiding systeemtheorie*. 1982.
- 51 P.J. Hoogendoorn (red.). *Cursus cryptografie*. 1983.