

Inhoud | Contents

Artikelen | Features

- Nieuws | News**
 229 'Open Up Math': boegbeeldproject wiskunde
Annoesjka Cabo
- Vakantiecursus | Summerschool**
 230 Grafentheorie en communicatie
Jop Briët
- Reis naar het bewijs | Trip to the proof**
 235 Mixing, wine, and serendipity
Thomas Ward
- Onderzoek | Research**
 237 Morley chains of osculant curves
Jan van de Craats
- Column | Column**
 247 Column van het jaar
Casper Albers
- Onderzoek | Research**
 249 Sneller zoeken met quantum random walks
Koen Groenland, Tom Bannink
- Evenement | Event**
 255 Royal decoration for Frank den Hollander
Remco van der Hofstad
- Evenement | Event**
 257 *KK*-theory, gauge theory and topological phases
Francesca Arici, Domenico Monaco
- In Memoriam | Obituary**
 265 Fred Steutel: Wiskundige, columnist, ingezonden brieven-schrijver en sportman
Frans Schurer, Klaas van Harn
- Interview | Interview**
 272 Simon Singh: It would be great if politicians understood what science is
Julia Kuhn
- Onderzoek | Research**
 274 Inplannen van afspraken voor een groot aantal patiënten
Madelon de Kemp
- Boekbespreking | Book Review**
 277 Over groei en vorm
Nicolaos Starreveld
- Onderzoek | Research**
 280 On the Fermat point of a triangle
Jakob Krarup, Kees Roos

Rubrieken | Departments

- 223 Redactioneel
Jan van Neerven
- 224 Agenda
Nicolaos Starreveld
- 226 Nieuws
Nicolaos Starreveld
- 263 Het keerpunt van Marc van Loo
Jan Beuving
- 287 In de verdediging
Geertje Hek
- 289 Boekbesprekingen
- 295 Problemen

Uitgelicht



Francesca Arici en Domenico Monaco doen verslag van een Lorentz Center workshop over *KK*-theorie.



Nicolaos Starreveld bespreekt het boek *Over groei en vorm* van D'Arcy Thompson.

Colofon

Het Nieuw Archief voor Wiskunde is een uitgave van het Koninklijk Wiskundig Genootschap en verschijnt vier maal per jaar. Het tijdschrift richt zich op eenieder die zich beroepsmatig met wiskunde bezighoudt, als academisch of industrieel onderzoeker, student, leraar, journalist of beleidsmaker. Het stelt zich ten doel te berichten over ontwikkelingen in de wiskunde in het algemeen en in de Nederlandse wiskunde in het bijzonder.

ISSN 0028-9825

Adres

Nieuw Archief voor Wiskunde
Centrum Wiskunde & Informatica
Postbus 94079, 1090 GB Amsterdam
tel. 020-5924288
redactie@nieuwarchief.nl
www.nieuwarchief.nl

Hoofredactie

Robbert Fokkink (r.j.fokkink@tudelft.nl)
Jan van Neerven (j.m.a.m.vanneerven@tudelft.nl)

Eindredactie/Bladmanager

Fred Drissen (fred@nieuwarchief.nl)

Redactie

Hans Cuypers (TU/e), Jan Draisma (Universität Bern en TU/e), Geertje Hek (UvA), Jan Hogendijk (UU), Teun Koetsier (VU), Sjoerd Rienstra (TU/e), Nicolaos Starreveld (UvA), Hans Sterk (TU/e), Mark Timmer (UT)

Problemenrubriek

problems@nieuwarchief.nl

Bureauredactie

Claartje Ottens

Illustraties

Ryu Tajiri

Vormgeving

Susanne Laws (omslag, begeleiding binnenwerk)
Kitty Molenaar (ontwerp)

Druk

Veldhuis Media

Advertenties

advertenties@nieuwarchief.nl

Koninklijk Wiskundig Genootschap

www.wiskgenoot.nl

Platform Wiskunde Nederland

www.platformwiskunde.nl

European Mathematical Society

www.euro-math-soc.eu

International Mathematical Union

www.mathunion.org

Koninklijk Wiskundig Genootschap

Het Koninklijk Wiskundig Genootschap (KWG) is een landelijke vereniging van beoefenaars van de wiskunde. In 1778 opgericht onder de zinspreuk: 'Een onvermoeide arbeid komt alles te boven' is het 's werelds oudste nationale wiskundegenootschap. Leden ontvangen het Nieuw Archief voor Wiskunde als onderdeel van hun lidmaatschap.

Bestuur

Erik van den Ban (UU, voorzitter), Joke Blom (CWI, penningmeester), Sonja Cox (UvA, secretaris), Theo van den Bogaart (HU), Marije Elkenbracht (ABN AMRO), André Ran (VU), Herman te Riele (CWI), Mark Veraar (TUD), Jan Wiegerinck (UvA)

Secretariaat

Joke Blom, CWI
Postbus 94079, 1090 GB Amsterdam
wiskgenoot@wiskgenoot.nl
www.wiskgenoot.nl

Ledenadministratie KWG / Abonnementen

Postbus 1064, 7940 KB Meppel
admin@wiskgenoot.nl, tel. 085-0160252

Contributie

De contributie voor gewone leden bedraagt € 90 per jaar. Voor gepensioneerden en werklozen € 45. Voor studenten, aio's en oio's van een Nederlandse universiteit of hbo-opleiding € 35. Studenten en pas afgestudeerden kunnen een jaar lang gratis lid worden. Voor leden van de wiskundige vakverenigingen VvS+OR, NVvW en NVORWO geldt het gereduceerd tarief van € 70. Leden die het Nieuw Archief voor Wiskunde niet willen ontvangen, betalen slechts € 50 contributie. Losse nummers zijn te bestellen voor € 20.

Instituten in Nederland en buitenland kunnen zich abonneren op het Nieuw Archief voor Wiskunde voor € 100 (Nederland), € 110 (Europa) of € 112,50 (buiten Europa).

Members of foreign societies

Members of SIAM, the *Société Mathématique de France*, the *Gesellschaft für Angewandte Mathematik und Mechanik*, the *Deutsche Mathematiker-Vereinigung*, and of the *American, Australian, Belgian, Indian, London, Spanish, and South-African Mathematical Societies* living outside of the Netherlands pay € 70 instead of the full membership fee of € 90 a year. Conversely, these foreign societies, as well as the VvS+OR and the NVORWO, have reduced membership fees for KWG members. A postage charge of € 10 is added for members living abroad but in Europe, and a charge of € 12,50 for members living outside of Europe.

Exchange subscriptions

Koninklijk Wiskundig Genootschap Library
Centrum Wiskunde & Informatica
P.O. Box 94079, NL-1090 GB Amsterdam
KWG-exchange@cw.nl

Informatie voor auteurs

Kopij

Het Nieuw Archief voor Wiskunde is een geredigeerd tijdschrift. Kopij dient per e-mail te worden aangeboden aan de hoofdredactie. Uitgebreide informatie en auteursinstructies kunt u vinden op www.nieuwarchief.nl.

Volgende nummers

nummer	maand	verschijnen	deadline kopij
19(1)	maart	2018	verstrekken
19(2)	juni	2018	01-02-2018
19(3)	september	2018	01-05-2018
19(4)	december	2018	01-08-2018

Instituutsleden

De publicaties van het Koninklijk Wiskundig Genootschap worden mede mogelijk gemaakt door de bijdragen van de volgende instituten.



Centrum Wiskunde & Informatica



Rijksuniversiteit Groningen



Universiteit van Amsterdam



Universiteit Leiden



Vrije Universiteit



Universiteit Utrecht



Technische Universiteit Eindhoven



Eurandom



Technische Universiteit Delft



Universiteit Twente



Radboud Universiteit Nijmegen



Freudenthal Instituut

Op het omslag

Communicatienetwerken kunnen wel eens last hebben van ruis. Door geschikt te coderen kan een foutloze communicatie worden gerealiseerd met maximale snelheid. Lees het artikel over grafentheorie en communicatie van Jop Briët op p. 230.

Uitdagingen

Als u van de diensten van Google gebruikmaakt, zal het u zijn opgevallen dat sinds enige tijd automatisch afspraken worden toegevoegd in uw agenda. Op basis van binnenkomende berichten in Gmail met bevestigingen van boekingen (vluchten, hotelovernachtingen, e.d.) verschijnt in uw Google Calendar netjes een afspraak op de betreffende datum en tijd. Ze zijn te herkennen aan de tekst: “Deze afspraak is automatisch gemaakt op basis van een e-mail.” Uw partner, die u permissie heeft gegeven uw agenda in te zien (reuze handig bij het afstemmen van afspraken), dient natuurlijk onwetend te blijven over eventuele escapades. Gelukkig is daar aan gedacht: “Deze afspraak is alleen zichtbaar voor jou, ongeacht de instellingen voor delen van je agenda.” Naar verluidt (zelf gebruik ik nog steeds een wegens) gebeurt hetzelfde als u navigeert met behulp van Google. Als in uw agenda staat dat u straks een vlucht vanaf Schiphol heeft, stelt het navigatieprogramma vanzelf voor om naar Schiphol te rijden.

Al jaren gebruik ik Gmail, gewoon omdat het zo overzichtelijk en gebruiksvriendelijk is, veel opslagruimte biedt en overal ter wereld gemakkelijk toegankelijk is. Dat neemt niet weg dat ik dat eigenlijk tegen beter weten in doe — zie de onthullingen over de NSA en ook de huidige discussie over de sleepwet. Voor mijn searchopdrachten gebruik ik een zoekmachine met meer privacy en in een bevlogen moment heb ik ooit zelfs een Protonmailaccount ingericht. Maar ja, dat kwam met slechts een halve gigabyte opslag en bovendien zouden mijn uitgaande mails dan opeens vanuit Zwitserland lijken te komen. Nog nooit gebruikt dus. Toch was het even slikken toen ik die automatische koppeling tussen de agenda en e-mail voor het eerst zag. Big Brother in levende lijve.

Dat wil geenszins zeggen dat big data niet ook veel nuttige toepassingen heeft. Dat is zonder twijfel het geval. Data science kan veel bijdragen aan het efficiënt doen verlopen van verkeers- en reizigersstromen, de veiligheid van zelfrijdende auto's, het afstemmen van vraag en aanbod bij de energieproductie in intelligente energienetwerken, het verbeteren van de medische diagnostiek: de lijst met bestaande en mogelijke toepassingen is bijna eindeloos. Hier ligt veel uitdagend werk voor wiskundigen. In het PWN-visiedocument *Formulas for Insight and Innovation* uit 2014 werd data science terecht als een van de ‘grand challenges’ voor de wiskunde aangemerkt en werd de beschikbaarheid van big data, naast de steeds verder toenemende rekenkracht, als een van de oorzaken van de toenemende reikwijdte en relevantie van de wiskunde genoemd.

Inmiddels begint data science ook uit te waaieren binnen de wiskunde. Een voorbeeld uit mijn eigen vakgebied. Onlangs was ik op een doodgewone conferentie over stochastische analyse in Engeland. Meestal gaat het dan om existentie en eenduidigheid, regulariteit, invariante maten, u kent het wel. Maar deze keer was er voor het eerst een discussieavond over data science:

hoe ons vakgebied eraan kan bijdragen en ervan profiteren. Een vooraanstaand wiskundige, bekend van zeer fundamentele bijdragen, bleek in samenwerking met mensen uit de machine learning, op basis van zijn ‘rough path theory’ een app te hebben ontwikkeld voor handschriftherkenning van Chinese karakters die beter blijkt te werken dan alles wat er op dit moment op de markt is. De vraag vanuit het publiek of wij als wiskundegemeenschap het vakgebied ‘data science’ niet eerst nader zouden moeten definiëren wimpelde hij overigens simpel weg: ‘data science’ is alles waarvoor je op dit moment het gemakkelijkst het meeste onderzoeksgeld kunt krijgen. Een jonge collega, bekend van zijn bijdragen aan de numerieke analyse van stochastische partiële differentiaalvergelijkingen, bleek in samenwerking met Google een deep learning-algoritme te hebben ontwikkeld dat de ‘curse of dimensionality’ (het fenomeen dat de rekentijd exponentieel schaalst met de dimensie) omzeilt: in tien minuten tijd lost het een aantal 100-dimensionale vergelijkingen op met een precisie van circa 0,5 procent, iets wat met gangbare numerieke algoritmes ondenkbaar zou zijn. Beide sprekers waren het erover eens dat hier grote wiskundige uitdagingen liggen: het werkt, maar een wiskundig kader dat uitlegt waarom ontbreekt.

Data science is niet de enige groeidiscipline voor wiskundigen. Onlangs is NWO Zwaartekracht-financiering toegekend aan het Quantum Software Consortium, dat informatici, wiskundigen en natuurkundigen bijeenbrengt om de eerste toepassingen te vinden voor toekomstige kwantumcomputers. Dit komt bovenop het eerder dit jaar gestarte Europese flagship-programma Quantum Technologies met een budget van een miljard euro. Dankzij de vele diepe verbanden tussen kwantum(informatie)theorie en wiskundige disciplines als operatorentheorie, representatietheorie en niet-commutatieve meetkunde liggen hier ook veel nieuwe wiskundige uitdagingen. In het nummer dat voor u ligt leggen Koen Groenland en Tom Bannink uit wat een ‘quantum random walk’ is en laten zien hoe een van de allereerste kwantumalgoritmes, Grovers sorteeralgoritme, hiermee beschreven kan worden. Dat ook de ‘klassieke’ informatietheorie nog steeds een bron van interessante wiskundige problemen is laat Jop Briët zien. Dit zijn slechts twee van de vele lezenswaardige bijdragen in dit nummer, die variëren van de geschiedenis van de meetkunde (Fermatpunten en de stelling van Morley) en de mathematische biologie (over groei en vorm) tot moderne toepassingen van de wiskunde in de natuurkunde (KK-theorie).

Met dit nummer neem ik na vier en een half jaar hoofdredacteurschap afscheid van u, lezer, en van dit prachtige blad en wens u wederom veel leesplezier toe!

✎

Jan van Neerven, hoofdredacteur
Delft Institute of Applied Mathematics, TU Delft

Agenda

| Upcoming Events

december 2017

4–8 december 2017

□ Random Dynamical Systems

Een workshop mede georganiseerd door Ale Jan Homburg (UvA).

plaats Lorentz Center@Oort, Leiden

info lorentzcenter.nl

11–15 december 2017

□ Future and Emerging Mathematical Technologies in Europe

Een workshop mede georganiseerd door Wil Schilders (TU/e), Evgeny Verbitskiy (UL) en Kees Vuik (TUD).

plaats Lorentz Center@Oort, Leiden

info lorentzcenter.nl

11–15 december 2017

□ YEQT XI 'Winter school on energy systems'

Elfde editie van de Young European Queueing Theorists (YEQT) workshop.

plaats Eurandom, Eindhoven

info eurandom.nl

15 december 2017 en 26 januari 2018

□ Van Dantzig Seminar

Landelijke serie van lezingen in de statistiek, waarvoor gerenommeerde sprekers uit zowel binnen- als buitenland worden uitgenodigd.

plaats VU Amsterdam resp. TU Delft

info few.vu.nl/~bkk320/vandantzig

januari 2018

8–12 januari 2018

□ The role of Science Education in a Changing World

Een workshop mede georganiseerd door Lucy Avraamidou (RUG) en Martin Goedhart (RUG).

plaats Lorentz Center@Snellius Leiden

info lorentzcenter.nl

13 januari 2018

□ Wintersymposium KWG 2018

Jaarlijks symposium van het KWG. Dit jaar is het thema 'WiskundeTaal'.

plaats Academiegebouw, Utrecht

info wiskgenoot.nl

17–19 januari 2018

□ LNMB-conferentie

43ste conferentie over de wiskunde van besluitkunde georganiseerd door het Landelijk Netwerk Mathematische Besluitkunde (LNMB).

plaats Congrescentrum De Werelt, Lunteren

info www.lnmb.nl

19 januari 2018

□ Conferentie rekenen-wiskunde aansluiting po naar vmbo

Conferentie bedoeld voor leerkrachten uit de bovenbouw van het po en wiskundeleraars uit de onderbouw van het vmbo.

plaats Bilderberg Parkhotel Rotterdam

info nvvw.nl

22–24 januari 2018

□ 17th Winter School Mathematical Finance

Winterschool over financiële wiskunde met onder andere minicursussen van Emmanuel Gobet (Palaiseau) en Sebastian Jaimungal (Toronto).

plaats Congrescentrum De Werelt, Lunteren

info staff.fnwi.uva.nl/p.j.c.spreij/winterschool/winterschool.html

22 januari–1 februari 2018

□ Nederlandse Wiskunde Olympiade

De eerste ronde van de jaarlijkse landelijke wiskundewedstrijd voor leerlingen uit de onderbouw van havo en vwo.

plaats eigen school

info wiskundeolympiade.nl

25 januari 2018

□ Conferentie vmbo en onderbouw havo/vwo

Jaarlijkse conferentie georganiseerd door de Nederlandse Vereniging van Wiskundeleraren.

plaats Domstad, Utrecht

info nvvw.nl

25–26 januari 2018

□ 36ste Panama-conferentie

De jaarlijkse Panama-conferentie richt zich op alle aspecten van leren en onderwijzen van rekenen-wiskunde.

plaats Congrescentrum Koningshof, Veldhoven

info panamaconferentie.sites.uu.nl

Suggesties voor deze agenda zijn welkom.

Redacteur: Nicolaos Starreveld

agenda@nieuwarchief.nl

29 januari–2 februari 2018

Study Group Mathematics with Industry
Jaarlijkse studiegroep waarin wiskunde en industrie samen problemen oplossen, dit jaar georganiseerd door TU/e.

plaats Technische Universiteit Eindhoven

info swi-wiskunde.nl

februari 2018

2–3 februari 2018

Nationale Wiskunde Dagen

24ste editie van de tweedaagse conferentie voor wiskundeleraren.

plaats Noordwijkerhout

info uu.nl/nationale-wiskunde-dagen

2 februari, 2 maart, 6 april, 4 mei en 1 juni 2018

Mark Kaç Seminar

Maandelijks seminar voor onderzoekers in kansrekening en statistische mechanica.

plaats Utrecht

info www.win.tue.nl/markkac

7 februari 2018

Onderbouw Wiskunde Dag

Activiteit van het Freudenthal Instituut. Leerlingen van 3 havo/vwo werken in teams gedurende een dag aan een ‘grote’ wiskundige (denk)opdracht waarin probleemoplossen centraal staat.

plaats eigen school

info uu.nl/onderwijs/onderbouwwiskundedag

11 februari 2018

Paradisolezing – De Nieuwe Mens

Eerste lezing in een serie over De Nieuwe Mens wordt gegeven door Ronald Hanson (QuTech), met als titel ‘Slecht nieuws voor luistervinken – over het onkraakbare quantuminternet.’

plaats Paradiso, Amsterdam

info paradiso.nl

maart 2018

7–9 maart 2018

YES IX, Scalable Statistics: On Accuracy and Computational Complexity

Negende editie van de Young European Statisticians (YES) workshop. Georganiseerd door Rui Castro, Paulo Serra (TU/e) en Botond Szabó (UL).

plaats Eurandom, Eindhoven

info eurandom.nl

15 maart 2018

W4 Kangoeroe

WereldWijdeWiskundeWedstrijd voor basis- en middelbare scholieren.

plaats eigen scholen

info w4kangoeroe.nl

16 maart 2018

Tweede wedstrijdronde Nederlandse wiskunde olympiade

De tweede ronde van de Nederlandse wiskunde olympiade.

plaats allerlei universiteiten

info wiskundeolympiade.nl

28 maart 2018

Grote Rekendag

Dag voor alle basisscholen in Nederland en Vlaanderen (groep 1 t/m 8) die helemaal in het teken staat van rekenen. Het thema van deze 16de editie is ‘De school als pakhuis’.

plaats basisscholen Nederland en Vlaanderen

info groterekendag.nl

april 2018

3–4 april 2018

Nederlands Mathematisch Congres 2018

Groots opgezet jaarlijks congres, dit jaar voor het eerst in Veldhoven. Er wordt een aantrekkelijk programma geboden voor heel wiskundig Nederland.

plaats Congrescentrum Koningshof, Veldhoven

info nmc.wiskgenoot.nl

9–15 april 2018

European Girls' Mathematical Olympiad

Europese wiskundewedstrijd speciaal voor meisjes.

plaats Florence, Italië

info egmo.org

16–20 april 2018

Uncertainty Quantification in Complex, Non-parametric Statistical Models

Een workshop mede georganiseerd door Botond Szabó (UL) en Aad van der Vaart (UL).

plaats Lorentz Center@Oort, Leiden

info lorentzcenter.nl

mei 2018

14–18 mei 2018

Fixed-Parameter Computational Geometry

Een workshop mede georganiseerd door Mark de Berg (TU/e) and Hans Bodlaender (UU).

plaats Lorentz Center@Oort, Leiden

info lorentzcenter.nl

22–25 mei 2018

The Computational Mathematics Aspects of Porous Media, and Fluid Flow

Een workshop mede georganiseerd door Cornelis Oosterlee (TUD) and Fred Vermolen (TUD).

plaats Lorentz Center@Oort, Leiden

info lorentzcenter.nl

28 mei – 1 juni 2018

Franco-Dutch YEP Workshop

Workshop met als onderwerp mathematisch statistische mechanica en random grafen, georganiseerd door Remco van der Hofstad (TU/e) en Arnaud Le Ny (Université Paris-Est, gast TU/e).

plaats Eurandom, Eindhoven

info eurandom.nl

juni 2018

18–22 juni 2018

Effective Methods for Diophantine Problems

Een workshop mede georganiseerd door Bas Edixhoven (UL) and Robin de Jong (UL).

plaats Lorentz Center@Oort, Leiden

info lorentzcenter.nl

Nieuws

| News

Deze rubriek is een kroniek van wiskundige activiteiten in Nederland. Toekomstige activiteiten worden aangekondigd en van voorbije activiteiten wordt verslag gedaan. Wilt u uw aankondiging of verslag in deze rubriek geplaatst zien? Stuur ons dan uw bijdrage, zo mogelijk met illustratie. De redactie behoudt zich het recht voor berichten te weigeren of in te korten.

Redacteur: Nicolaos Starreveld
 nieuws@nieuwarchief.nl

Zweeds eredoctoraat voor Gerard van der Geer

De universiteit van Stockholm heeft op 29 september een eredoctoraat verleend aan prof.dr. Gerard van der Geer, hoogleraar algebra aan het Korteweg-de Vries Instituut voor Wiskunde van de Universiteit van Amsterdam. Van der Geer ontvangt het eredoctoraat voor zijn bijdrage aan het onderzoek en het onderwijs van de Universiteit van Stockholm evenals voor zijn invloedrijke rol als redacteur van *Composito Mathematica*. In zijn wetenschappelijke carrière heeft Van der Geer een substantiële bijdrage geleverd aan de algebraïsche meetkunde en de differentiaalmeetkunde (moduli of curves, abelian varieties and K3-surfaces, Hilbert modular surfaces, and curves over finite fields).

kdvi.nl



Stieltjesprijs 2016 voor Richard Kraaij en Djordjo Milovic

Elk jaar wordt namens de Stichting Compositio Mathematica de Stieltjesprijs uitgereikt voor het beste Nederlandse proefschrift. Het betreft dit jaar de proefschriften van Richard Kraaij (TUD), *Semigroup Methods for Large Deviations of Markov Processes* (promotoren Frank Redig en Frank den Hollander, UL) en Djordjo Milovic (UL), *On the 16-Rank of Class Groups of Quadratic Number Fields* (promotor Peter Stevenhagen, UL). Kraaij werkt op het grensvlak van de waarschijnlijkheidsrekening en de functionaal-analyse; Milovic is gepromoveerd op een onderwerp uit de getaltheorie. Beiden zijn cum laude gepromoveerd, beiden hebben voornamelijk single author papers en beiden hebben zeer sterke referentiebrieven gekregen. Gezien de zeer hoge kwaliteit van deze twee proefschriften en tegelijkertijd het grote verschil in vakgebied en type onderzoek, stelt de jury voor om zowel aan Richard Kraaij als aan Djordjo Milovic de Stieltjesprijs toe te kennen. De prijs wordt uitgereikt tijdens het NMC 2018, dat van 3–4 april 2018 wordt gehouden in congrescentrum Koningshof te Veldhoven. Beide prijswinnaars zullen daar ook een voordracht houden over hun werk.

wiskgenoot.nl

Wiskunde helpt brandwondengenezing

Daniël Koppenol ontwikkelde samen met Fred Vermolen van de TU Delft diverse wiskundige modellen waarmee de genezing van (brand)wonden kan worden nagebootst. Daniël Koppenol promo-

veerde op donderdag 15 juni 2017 op dit onderwerp aan de TU Delft. Hij richtte zijn onderzoek onder meer op hypertrofische littekens: verdikte, vaak rode en zeer stugge littekens, die pijnlijke aanvoelen, jeuken en kunnen leiden tot bewegingsbeperkingen. “Helaas zijn de mechanismen achter het ontstaan van contracturen en hypertrofisch littekenweefsel op dit moment niet bekend”, zegt Koppenol. Dit wordt gedeeltelijk veroorzaakt door de complexiteit van het wondgenezingsproces, maar het komt ook doordat het nu eenmaal erg moeilijk is om de interacties tussen verschillende componenten van het wondgenezingsproces te onderzoeken met experimenten. Door gebruik te maken van wiskundige modellen, kan met dit probleem worden omgegaan. Koppenol ontwikkelde wiskundige modellen waarmee de genezing van verschillende type wonden en het ontstaan van littekens nagebootst kan worden.

tudelft.nl

Promovendus blaast nieuw leven in vakgebied in de wiskunde

Wiskundige en TU/e-promovendus Jorn van der Pol laat een onuitwisbare indruk achter in het vakgebied in de wiskunde dat gaat over zogeheten matroïden. In zijn promotieonderzoek kwam hij tot een aantal nieuwe, slimme technieken waarmee hij vraagstukken kon aanpakken die tot enkele jaren geleden ver buiten bereik leken, en blies hij zodoende nieuw leven in het vakgebied. Op 20 september 2017 promoveerde hij cum laude op zijn werk aan de Faculteit Wiskunde en Informatica van de TU/e. Een van de belangrijke vraagstukken rond grote matroïden draait om het bepalen van de theoretische onder- en bovengrens van het aantal mogelijke matroïdes bij een gegeven aantal elementen. Van der Pol vond tijdens zijn masteronderzoek al een bewijs voor een bovengrens dat werd gezien als een doorbraak. Voor dit werk, dat hij uitvoerde onder begeleiding van prof.dr. Remco van der Hofstad, dr. Rudi Pendavingh en prof.dr. Nikhil Bansal, kreeg hij in 2014 een TU/e Academic Award. In de rubriek ‘In de verdediging’ verderop in dit nummer kunt u meer lezen over zijn onderzoek.

tue.nl



Internationale wiskundetoernooien

Dit jaar waren vier Nederlandse universiteitsteams zeer succesvol bij de International Mathematics Competition for University Students, zoals de IMC voluit heet. In een veld met 71 teams uit de hele wereld waren de plaatsen 11, 12, 13 en 14 voor respectievelijk

Universiteit Utrecht, Universiteit Leiden, de Radboud Universiteit en de Universiteit van Amsterdam.

Voor het eerst deden dit jaar ook de Technische Universiteit Delft (42ste) en de Technische Universiteit Eindhoven mee, dat op de 54ste plaats eindigde.

De IMC is in 1994 begonnen in het Bulgaarse Plovdiv, met voornamelijk Bulgaarse deelnemers. Het evenement is een jaarlijkse wiskundewedstrijd voor wiskundestudenten die niet ouder dan 23 zijn. Na uitstapjes in het buitenland vindt de IMC sinds 2010 altijd plaats in Blagoevgrad (uitgesproken als: Bla-go-jev-grad), ook in Bulgarije. Inmiddels hebben studenten van meer dan tweehonderd universiteiten uit meer dan vijftig landen eraan deelgenomen. Hoewel het een individuele wedstrijd is, doen de meeste studenten mee in een team van hun eigen universiteit. Er zijn twee wedstrijddagen, elk met vijf vragen, waarvoor de studenten steeds vijf uur de tijd hebben. De opgaven komen uit de gebieden algebra, analyse, combinatoriek en meetkunde.

Het Nederlandse succes kwam niet uit de lucht vallen. In 2015 waren de Nederlandse universiteitsteams voor het eerst in bredere zin actief en succesvol bij de IMC. Behalve de UvA en de UU, die al sinds 2006 deelnamen, deed toen ook een team van de Radboud Universiteit mee, dat meteen achttiende van de 68 teams werd. Het team van de UU werd zelfs zevende, pal achter een team uit Barcelona en een top vijf met twee Russische teams, een Hongaars team, een Pools team en het nationale team van Israël.

Wat de Nederlandse prestaties van 2017 zo aansprekend maakt, is dat ze van zo veel universiteiten zijn en de opgaande lijn van de afgelopen jaren voortzetten. Dat belooft nog wat voor 2018!

Jaap de Jong en wiskundeolympiade.nl

Regeerakkoord: rekentoets telt niet meer mee en wordt vervangen

De rekentoets telde bij het afgelopen eindexamen voor de tweede keer mee in de zak-/slaagregeling van het vwo — leerlingen moesten er minimaal een 5 voor halen en mochten in geval van een 5 geen onvoldoende meer staan voor Nederlands, Engels en wiskunde. Voor het mbo, het vmbo en de havo telde de rekentoets nog niet mee.

Met het recentelijk gepresenteerde regeerakkoord is er gehoor gegeven aan kritiek binnen het werkveld en de politiek op vorm en inhoud en komt er een eind aan de rekentoets in de huidige vorm. De toets moet nog steeds gemaakt worden, maar vermoedelijk reeds met ingang van dit schooljaar zal het behaalde resultaat niet meer van belang zijn voor het behalen van een diploma; een ontwikkeling waar veel leerlingen blij mee zullen zijn, maar sommigen wellicht ook niet aangezien het cijfer daarmee waarschijnlijk ook niet meer zal meetellen in de cumlauderegeling.

Het regeerakkoord spreekt al over een alternatief, dat reeds in het schooljaar 2019–2020 in werking moet gaan als ‘geïntegreerd onderdeel van het examen’ voor alle niveaus van het voortgezet onderwijs. Naast het aanpassen van de toetsing is de regering ook voornemens om het rekenonderwijs zelf te versterken en verbeteren, maar hoe precies is nog onbekend.

Momenteel is de rekentoets overigens nog steeds onderdeel van de zak-/slaagregeling van het vwo, totdat hier een officieel besluit over is genomen.

Mark Timmer

Plimpton 322 blijkt een trigonometrisch tabel te zijn

Plimpton 322 is een gedeeltelijk gebroken kleitablet. Uitgever George A. Plimpton uit New York kocht het tablet in 1922 van een archeologische handelaar, Edgar J. Banks. Volgens Banks kwam het tablet uit Senkereh, een plaats in Zuid-Irak, waarschijnlijk was dit de oude stad Larsa. Er wordt verondersteld dat dit kleitablet geschreven is tussen 1800 en 1650 v. Chr. De exacte betekenis van de tekst op het tablet was tot voor kort onbekend. D.F. Mansfield en N.J. Wildberger, twee onderzoekers van de School of Mathematics and Statistics, UNSW, Sydney, Australia, hebben nu de mysterieuze tekst van Plimpton 322 kunnen ontcijferen. Hun bevindingen zijn gepubliceerd in het artikel 'Plimpton 322 is Babylonian exact sexagesimal trigonometry' in *Historia Mathematica*. Ze hebben ontdekt dat Plimpton 322 een trigonometrische tabel is waarvan de complexiteit uitstijgt boven wat tot nu toe gedacht werd dat het wiskundenniveau was van die tijd. Door de bevindingen van Mansfield en Wildberger is deze tabel nu de oudst bekende trigonometrische tabel. Tot voor kort was dat de 'table of chords' van Hipparchus.

Historia Mathematica



Christiaan Huygensprijs voor wiskunde 2017

De jury reikt de jaarlijkse Huygensprijs bij toerbeurt uit voor een promotieonderzoek in de astronomie, natuurkunde of wiskunde, de drie terreinen waarop Christiaan Huygens excelleerde. In dit wiskundejaar won econometrist Jochem de Bresser van de Universiteit Tilburg, voor zijn statistische analyse van de kloof tussen perceptie en realiteit wat betreft pensioenen. Opmerkelijk: mensen gaan minder sparen als ze hebben nagedacht over hun pensioen, en vrouwen denken vijf jaar korter te leven dan ze echt doen, terwijl mannen dit wel goed inschatten.

nemokennislink.nl

De 'Wiskundeclub' voor scholieren

In september 2017 heeft het Korteweg-de Vries Instituut voor Wiskunde aan de Universiteit van Amsterdam een wiskundeclub gestart voor leerlingen uit klas 4, 5 en 6 vwo. De Wiskundeclub is er om getalenteerde en gemotiveerde leerlingen van de intellectuele uitdaging en schoonheid van wiskunde te laten genieten. De club staat onder leiding van Professor Sergey Shadrin, geholpen door student-assistenten.

kdvi.nl

AlphaGo Zero verplettert AlphaGo

In 2016 versloeg AlphaGo voor het eerst Go-spelers van wereldklasse. Dit computerprogramma beruiste op meerdere neurale netwerken en supervised learning; het werd 'opgevoed' met door mensen gespeelde partijen. Nu is er AlphaGo Zero, gebaseerd op reinforcement learning. Dit neurale netwerk leerde zichzelf, zonder enige menselijke input, in drie dagen Go spelen op bovenmenselijk niveau. Waarna het voorganger AlphaGo met 100–0 versloeg.

nemokennislink.nl

Neerlands Hoop voor Jan Beuving

Op 2 oktober 2017 ontving cabaretier en wiskundige Jan Beuving de Neerlands Hoop, de prijs voor een veelbelovende theatermaker met het grootste toekomstperspectief, met zijn voorstelling 'Raaklijn'. De prijs is in 2003 in het leven geroepen door de Vereniging van Schouwburg- en Concertgebouwdirecties (VSCD).

Uit het juryrapport: "Door zijn achtergrond als wiskundige in te zetten biedt Jan Beuving in 'Raaklijn' een uniek perspectief. Beuving integreert de meest complexe wiskundige theorieën met veel inzicht maar ook humor in zijn voorstelling. Daarnaast zet Beuving ook een lijn uit over zijn amateurvoetbalelftal en de vele PVV-stemmers die hij onder zijn teamgenoten heeft. Hij zet zichzelf neer als de raaklijn die langs twee uiteenlopende werelden scheert, wat leidt tot een frisse kijk op onze verdeelde samenleving. Muzikaal is het hoogstaand, dankzij de pianobegeleiding van Tom Dicke en Beuving's teksten, die met mathematische precisie in elkaar zijn gezet."

Lezers van dit blad kennen Jan Beuving ook van de rubriek 'Het keerpunt'.

vscd.nl

Koninklijk Wiskundig Genootschap

Wintersymposium 2018

Op zaterdag 13 januari vindt het jaarlijkse Wintersymposium van het KWG plaats. Thema is 'WiskundeTaal', aanmelden kan via www.wiskgenoot.nl.

NMC 2018

Op 3 en 4 april 2018 vindt het 54ste Nederlands Mathematisch Congres plaats in congrescentrum Koningshof in Veldhoven. De *early bird*-registratie sluit op 27 december 2017. Zie voor meer informatie en aanmelden nmc.wiskgenoot.nl.

Call for submissions for the KWG PhD prize

The yearly KWG PhD prize is awarded to the PhD student in mathematics at a Dutch university giving the best presentation on his/her research at NMC 2018. To compete for the KWG PhD prize please send an extended abstract (1–2 pages) and a brief CV (max. 1 page) to kwg.phd.prize@gmail.com by 28 January 2018. Please also ask your supervisor to send a concise letter of recommendation to the same address.

Recent verschenen:

Epsilon Uitgaven (www.epsilon-uitgaven.nl)

Zebra 51. Symmetrie in telproblemen en puzzels, Rogier Bos en Susanne Tak, € 10, 2017.

Zebra 50. De wiskunde van Rubik's kubus, Leroy Soesman, Marko van Eekelen en Bernard van Houtum, € 10, 2017.

Annoesjka Cabo

directeur Interfacultair Onderwijs
TU Delft
a.j.cabo@tudelft.nl

Nieuws

'Open Up Math': boegbeeldproject wiskunde

"In 2025 moet al het leermateriaal aan hbo en wo open beschikbaar zijn voor studenten en docenten." Dit is een van de ambities van het ministerie van Onderwijs geformuleerd in het programma Open en Verbonden Hoger Onderwijs. Om te onderzoeken wat ervoor nodig is om die ambitie te realiseren heeft de minister in 2017 twee boegbeeldprojecten toegekend: een aan de hogescholen en een aan de universiteiten. Bij de hogescholen betreft het een project van verschillende verpleegkundeopleidingen, bij de universiteiten een project van verschillende wiskundeafdelingen. Het wiskundeproject heeft de naam 'Open Up Math' gekregen.

De projectleiders van Open Up Math zijn Hans Cuypers (TU/e) en Annoesjka Cabo (TUD). Het projectteam bestaat verder uit Martijn Ouwehand (Open education specialist TUD), Julie Beardsell (Innovation expert for education TUD), Caroline de Wit (projectsecretaris) en Bart van den Dries (docent aan de TUD). Doel van het project is om te inventariseren wat docenten vinden dat nodig is om eigen leermateriaal open te stellen voor hergebruik en vooral ook waaraan materiaal van anderen moet voldoen om te worden hergebruikt. Ook wordt onderzocht hoe een gebruikersgemeenschap kan worden opgezet om hierover met collega's van gedachten te wisselen en om materiaal van elkaar te bekijken en mogelijk te gebruiken. In eerste instantie wordt dit bekeken binnen 4TU-verband, dat wil zeggen, onder wiskundigen van de vier technische universiteiten in Twente, Wageningen, Eindhoven en Delft. De bedoeling is om later ook de rest van de Nederlandse wiskundegemeenschap hierbij te betrekken.

Enquête

Het projectteam van Open Up Math ziet dat er al veel gebeurt op het gebied van delen en hergebruiken van onderwijsmaterialen. Dit gebeurt vaak op informele wijze met collega's onder elkaar. Om hierin inzicht te krijgen, is dit najaar een enquête uitgezet. Hierbij een oproep aan allen die de vragenlijst hebben ontvangen: laat het projectteam weten wat je belangrijk vindt omtrent open onderwijsmaterialen. Met deze informatie kan het boegbeeldproject de — niet te stuiten — ontwikkelingen op het gebied van open onderwijs in goede banen leiden zodat de wiskundegemeenschap er echt iets aan heeft!

Bijeenkomst

Op 11 december a.s. vindt een workshop plaats in Utrecht van 15.00–17.30 uur in het gebouw van Surf (Kantoren Hoog Overborch, Hoog Catharijne, Moreelsepark 48, 3511 EP Utrecht.) Daar worden de resultaten uit de enquête besproken met geïnteresseerden. Naar aanleiding van de uit-

komsten daarvan wordt een vervolgplan ontwikkeld met de aanwezigen. Wil je deelnemen aan deze (gratis) workshop? Vind je het leuk om collega's op informele wijze te ontmoeten en om mee te praten over dit onderwerp? Geef je dan op door een mailtje te sturen naar c.m.dewit@tudelft.nl.

Informatie

Ter kennismaking met de boegbeeldprojecten is een korte introductievideo gemaakt. Deze kun je bekijken op:

<https://youtu.be/OTRDvXbF3tM>

Ideeën, opmerkingen, vragen en voorbeelden graag sturen naar de projectsecretaris: Caroline de Wit (c.m.dewit@tudelft.nl).

Biografie Annoesjka Cabo

Annoesjka Cabo studeerde wiskunde aan de UvA, werkte aan het CWI en promoveerde aan de TU Delft in de stochastische meetkunde toegepast op beeldanalyse. Tegelijkertijd rondde zij haar opleiding tot violiste af. Daarna speelde zij bij onder andere het Rotterdams Philharmonisch Orkest en is zij actief in de kamermuziek en als vioolpedagoge. Sinds 2012 is Annoesjka werkzaam als docent aan de TU Delft. Met ingang van 1 augustus 2016 is zij directeur Interfacultair Onderwijs.

Jop Briët

Centrum Wiskunde & Informatica
Amsterdam
jop.briet@cwi.nl

Vakantiecursus

Grafentheorie en communicatie

Op de vakantiecursus 2016 van het Platform Wiskunde Nederland hield Jop Briët een voordracht over het optimaliseren van communicatie over netwerken met ruis. Door geschikt te coderen kan een foutloze communicatie worden gerealiseerd met maximale snelheid, een en ander afhankelijk van de vorm van het netwerk.

Claude Shannon, vader van de moderne informatietheorie, zette rondom 1950 wiskundige modellen uiteen waarmee communicatie over kanalen met ruis formeel bestudeerd kan worden [8]. Hij liet onder andere zien dat het extreme model waarin berichten altijd perfect moeten aankomen volledig gekarakteriseerd wordt door eigenschappen van simpele grafen [9]. Zo hoort bij ieder kanaal een graaf waarvan de ‘Shannon-capaciteit’ gelijk is aan de maximale communicatiesnelheid die behaald kan worden door middel van codes die berichten moeten beschermen tegen ruis. Het aantal gevolgen dat deze observatie heeft gehad voor de ontwikkeling van de grafentheorie is moeilijk te overschatten. Een belangrijke tegenhanger van communicatie onder ruis is *bruncodering*, waarin gescheiden partijen gebruik maken van bestaande patronen in de informatie waarover ze beschikken om efficiënt informatie uit te wisselen. Analoog aan Shannons observatie merkte Witsenhausen op dat ook dit probleem volledig bestudeerd kan wor-

den aan de hand van eigenschappen van grafen [10]. Het doel van deze tekst is om deze verbanden bloot te leggen en de relevante graafparameters en -concepten te bestuderen. We dekken slechts een minuscule deel van de reusachtige hoeveelheid literatuur over dit onderwerp (‘zero-error information theory’) en verwijzen naar Körner en Orlitsky [3] voor een uitgebreid overzicht en naar Lubetzky [5] voor meer recente resultaten.

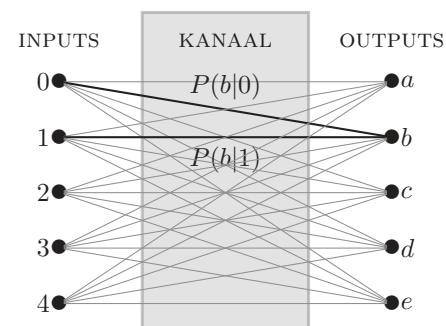
Kanaalcodering

Een discreet communicatiekanaal met ruis wordt gemodelleerd als een drietal $\mathcal{N} = (S, R, P)$ bestaande uit een eindige verzameling *inputs* S , een eindige verzameling *outputs* R en een kansverdeling $P(\cdot | s)$ over de outputverzameling R voor elke input $s \in S$; zie Figuur 1. Het ruismodel dat hiermee vertegenwoordigd wordt dicteert dat als een verzender input $s \in S$ door het kanaal $\mathcal{N}$ stuurt, de ontvanger het signaal $r \in R$ ontvangt met kans $P(r | s)$. We zullen aannemen dat kanalen *geheugenloos* zijn,

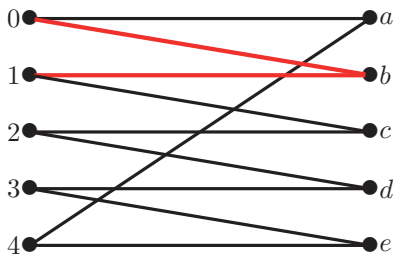
wat wil zeggen dat als de verzender opeenvolgend inputs s en t (in S) verstuurt, de kans dat de ontvanger respectievelijk signalen q en r (in R) ontvangt gegeven is door het product $P(q | s)P(r | t)$.

Ruisloos communiceren

Shannon vroeg hoe efficiënt er gecommuniceerd kan worden als de kans op error, misinterpretatie van het verzonden bericht, *nul* moet zijn. Dit is onmogelijk als alle transitiekansen $P(r | s)$ strikt groter zijn dan nul, aangezien de ontvanger dan nooit zeker kan weten wat het oorspronkelijke bericht was. Om te zien wanneer er wel iets te halen valt bekijken we nu de versie van



Figuur 1 Een discreet kanaal met 5 vijf inputs en 5 outputs. De kans op output b na input 0 of 1 is respectievelijk $P(b | 0)$ of $P(b | 1)$.



Figuur 2 De bipartiete graaf van het kanaal uit Figuur 1.

Figuur 1 waarin input s en output r alleen met elkaar verbonden zijn als $P(r|s) > 0$.

Dit resulteert in een bipartiete graaf zoals bijvoorbeeld die in Figuur 2. Als de ontvanger het signaal b ontvangt dan kan die afkomstig zijn geweest van beide inputs 0 en 1. In eerste instantie lijkt het er dus misschien op alsof zelfs in dit geval niet perfect gecommuniceerd kan worden. Dit kan echter wel als de partijen van tevoren afspreken dat slechts een subset van de inputs gebruikt wordt. Bijvoorbeeld, als de verzender alleen inputs 0 en 2 gebruikt, dan zal de ontvanger nooit in verwarring zijn omdat er geen output is die van beide deze inputs afkomstig kan zijn. Dit laat zien dat elke keer dat dit kanaal gebruikt wordt er minstens één bit perfect verstuurd kan worden.

Shannon realiseerde zich dat het aantal berichten dat perfect verstuurd kan worden precies overeenkomt met een bekende graafparameter van de volgende zogeheten *verwarringsgraaf*. De knopen van de verwarringsgraaf $G = (V, E)$ zijn de inputs van het kanaal, dus $V = S$. Twee knopen $u, v \in V$ vormen een kant als er een $r \in R$ is zodat beide $P(r|v) > 0$ en $P(r|w) > 0$ geldt (zie Figuur 3). Met andere woorden, de kanten zijn precies de paren inputs die niet met zekerheid van elkaar onderscheiden kunnen worden.

De relevante graafparameter is het *onafhankelijkheidsgetal* $\alpha(G)$: de maximale grootte van een verzameling knopen die onderling geen kanten vormen, ofwel een onafhankelijke verzameling vormen. Zo komen we op de eerste stelling.

Propositie 1. Voor een kanaal met verwarringsgraaf G is $\alpha(G)$ het maximale aantal berichten dat zonder ruis gecommuniceerd kan worden.

De Shannon-capaciteit

In het *kanaalcoderingsprobleem* heeft de verzender een verzameling van m mogelijke berichten die ze zonder ruis naar een

ontvanger moet kunnen sturen met behulp van een geheugenloos kanaal N met verwarringsgraaf $G = (V, E)$. Hierboven zagen we dat als $m \leq \alpha(G)$ geldt, dit probleem opgelost kan worden door elk bericht te coderen als een unieke input van een onafhankelijke verzameling van grootte m in G . Voor elk bericht hoeft het kanaal dan maar één keer gebruikt te worden.

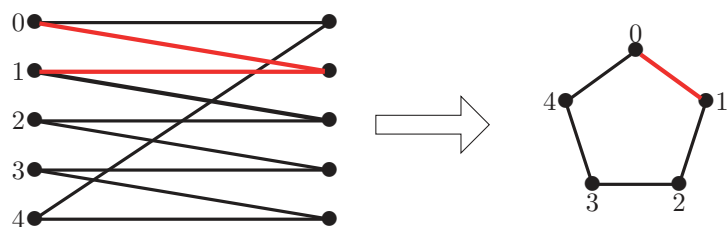
Wat als $m > \alpha(G)$? Een makkelijke oplossing is om de berichten te coderen als reeksen van n inputs uit een onafhankelijke verzameling van grootte $\alpha(G)$, waar de *bloklengte* n gekozen wordt zodat $\alpha(G)^n \geq m$. Elke reeks zal ontvangen worden als een unieke reeks outputs, waardoor de ontvanger de verzender luid en duidelijk kan verstaan. Van het voorbeeld uit Figuur 3 zien we bijvoorbeeld dat als $m = 5$, we met bloklengte $n = 3$ toekunnen; er kunnen dan acht mogelijke berichten verstuurd worden. Maar dit kan efficiënter! De vraag die we moeten stellen is:

“Wanneer kunnen twee inputreeksen $u_1, \dots, u_n \in S$ en $v_1, \dots, v_n \in S$ met zekerheid onderscheiden worden?”

Als deze reeksen verstuurd worden ontvangt de ontvanger respectievelijk signaalreeksen $x_1, \dots, x_n \in R$ en $y_1, \dots, y_n \in R$. Merk op dat als er een $i \in \{1, \dots, n\}$ is waarvoor u_i en v_i verschillend zijn en géén kant vormen in de verwarringsgraaf, de signalen x_i en y_i dan nooit hetzelfde zijn. Het antwoord op de vraag luidt daarom:

“Als er een coördinaat $i \in \{1, \dots, n\}$ is zodat $u_i \neq v_i$ en $\{u_i, v_i\} \notin E$.”

Twee reeksen kunnen dus alleen *verward* worden als voor *elke* coördinaat geldt dat $u_i = v_i$ of $\{u_i, v_i\} \in E$. Dit motiveert de verwarringsgraaf $G^{\boxtimes n} = (V^n, E_n)$ voor inputreeksen van lengte n , waarvan de knopenverzameling bestaat uit het n -voudig cartesisch product van V met zichzelf en de kantenverzameling E_n bestaat uit de paren $(u_1, \dots, u_n), (v_1, \dots, v_n)$ waarvoor voor elke i geldt dat $u_i = v_i$ of $\{u_i, v_i\} \in E$.



Figuur 3 Van een communicatiekanaal naar zijn verwarringsgraaf.

De notatie $G^{\boxtimes n}$ komt voort uit het feit dat het sterke graafproduct $G \boxtimes H$ van twee grafen G en H associatief is en dat $G^{\boxtimes n} = G \boxtimes G \boxtimes \dots \boxtimes G$ (n maal). De graaf $G \boxtimes H$ heeft $V \times W$ als knopenverzameling en twee knopen $(v, w), (v', w') \in V \times W$ vormen een kant als $v = v'$ of $\{v, v'\} \in E$ en $w = w'$ of $\{w, w'\} \in F$. Zo komen we op onze tweede stelling:

Propositie 2. Voor een kanaal met verwarringsgraaf G is het maximale aantal berichten dat met inputreeksen van lengte n zonder ruis verzonden kan worden gelijk aan $\alpha(G^{\boxtimes n})$.

De voorgaande discussie suggereerde al de volgende eenvoudige stelling over het gedrag van het onafhankelijkheidsgetal onder het sterke graafproduct.

Propositie 3. Zij $G = (V, E)$ en $H = (W, F)$ twee grafen. Dan geldt

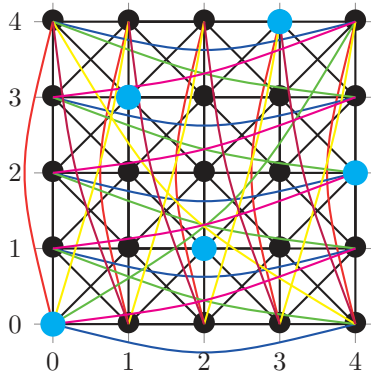
$$\alpha(G \boxtimes H) \geq \alpha(G) \alpha(H).$$

Een direct bewijs volgt uit het feit dat het cartesisch product $I \times J$ van een onafhankelijke verzameling $I \subseteq V$ in G en een onafhankelijke verzameling $J \subseteq W$ in H een onafhankelijke verzameling in $G \boxtimes H$ vormt. Door Propositie 3 herhaaldelijk toe te passen krijgen we ook de volgende voor de hand liggende gevolgtrekking.

Propositie 4. Zij G een graaf. Dan geldt

$$\alpha(G^{\boxtimes k}) \geq \alpha(G)^k.$$

De 5-cykel laat al zien dat *blokcoderen*, het coderen van berichten in goed gekozen inputreeksen, kan leiden tot efficiënter gebruik van een kanaal. Shannon merkte op dat de graaf $C_5 \boxtimes C_5$ een onafhankelijke verzameling van grootte 5 heeft: de knopen $(0,0), (2,1), (4,2), (1,3), (3,4)$ vormen onderling geen kanten (zie Figuur 4). Over een kanaal met C_5 als verwarringsgraaf

Figuur 4 $C_5 \boxtimes C_5$.

kunnen er gemiddeld per input dus geen één, maar minstens $\log_2(5)/2 = 1,16\dots$ bits gecommuniceerd worden.

De Shannon-capaciteit van een graaf geeft de maximale communicatiesnelheid wanneer de bloklengthe naar oneindig gaat.

Definitie 1 (Shannon-capaciteit). De *Shannon-capaciteit* van een graaf $G = (V, E)$ is gedefinieerd als

$$\Theta(G) = \lim_{n \rightarrow \infty} \alpha(G^{\boxtimes n})^{1/n}.$$

Wanneer de bloklengthe naar oneindig gaat kunnen er gemiddeld per input $\log \Theta(G)$ bits ruisloos verstuurd worden. Helaas is de Shannon-capaciteit in het algemeen extreem lastig te bepalen. Zelfs voor de 5-cykel was dit niet eenvoudig. Gegeven dat

$$\sqrt{5} = \alpha(C_5^{\boxtimes 2})^{1/2} > \alpha(C_5) = 2$$

is een natuurlijke vervolgvraag of er wellicht een natuurlijk getal $n \geq 3$ bestaat zodat $\alpha(C_5^{\boxtimes n})^{1/n} > \sqrt{5}$. De vraag naar de exacte waarde van $\Theta(C_5)$ stelde Shannon zelf al in zijn beroemde artikel [9]. Het duurde echter meer dan twintig jaar voordat Lovász [4] met een ingenieus bewijs liet zien dat Shannons ondergrens van $\sqrt{5}$ niet verbeterd kan worden.

Stelling 1 (Lovász). *Voor elk natuurlijk getal n geldt*

$$\alpha(C_5^{\boxtimes n})^{1/n} \leq \sqrt{5}.$$

In het bijzonder hebben we $\Theta(C_5) = \sqrt{5}$.

Voor het bewijs verwijzen we naar *Proofs from THE BOOK* [1, Hoofdstuk 37], vernoemd naar het geheime boek waarin, volgens Paul Erdős, God de perfecte bewijzen voor wiskundige stellingen bewaart.

De capaciteiten van k -cyclen met oneven $k \geq 9$ zijn evenmin bekend. *Even* cyclen zijn een ander verhaal. De reden daarvoor is dat deze grafen bipartiet zijn en daarom behoren tot de klasse *perfecte grafen*. De definitie van deze klasse is gebaseerd op de volgende basisconcepten.

Definitie 2 (Clique). Een *clique* is een deelverzameling knopen waarvan elke twee een kant vormen. Het *cliquegetal* van een graaf G , genoteerd als $\omega(G)$, is het aantal knopen in een clique van maximale grootte.

Definitie 3 (Geldige knopenkleuring). Een *geldige knopenkleuring* van een graaf is een toewijzing van een kleur aan elke knoop zodanig dat elke kant tweekleurig is. Het *chromatisch getal* van een graaf G , genoteerd als $\chi(G)$, is het minimum aantal kleuren voor een geldige knopenkleuring.

Merk op dat het chromatisch getal altijd minstens het cliquegetal is.

Definitie 4 (Geïnduceerde deelgraaf). Zij $G = (V, E)$ een graaf en $S \subseteq V$ een deelverzameling. De *deelgraaf geïnduceerd door S* is de graaf met knopenverzameling S wiens kantenverzameling bestaat uit de kanten in $\{u, v\} \in E$ zodanig dat $u, v \in S$.

De Petersen-graaf (Figuur 5) heeft bijvoorbeeld het pentagon als geïnduceerde deelgraaf. De definitie van een perfecte graaf is nu als volgt.

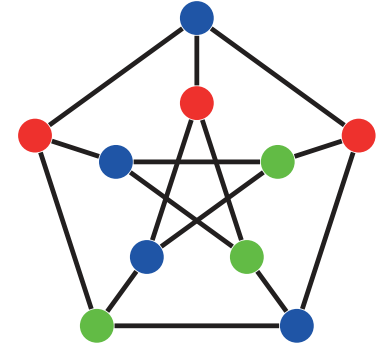
Definitie 5 (Perfekte graaf). Een graaf is *perfect* als voor elke geïnduceerde deelgraaf geldt dat het chromatisch getal gelijk is aan het cliquegetal.

De Petersen-graaf (Figuur 5) is een voorbeeld van een graaf die *niet* perfect is, omdat het cliquegetal van de graaf zelf verschilt van zijn kleurgetal. Bipartite grafen zijn perfect omdat ze chromatisch getal 2 hebben. We hebben de volgende stelling over de Shannon-capaciteit van perfecte grafen.

Stelling 2. *Zij G een perfecte graaf, dan geldt*

$$\Theta(G) = \alpha(G).$$

Voor het bewijs verwijzen we naar [2, Hoofdstuk 5].



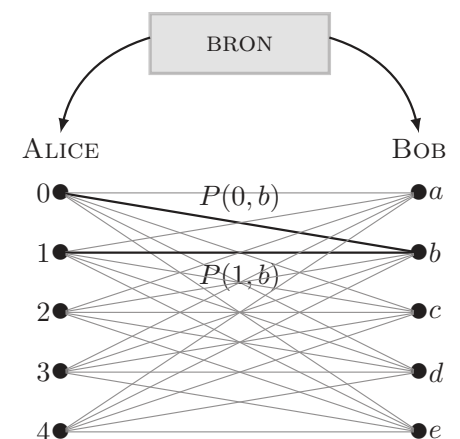
Figuur 5 De Petersen-graaf heeft cliquegetal 2 en chromatisch getal 3.

Broncodering

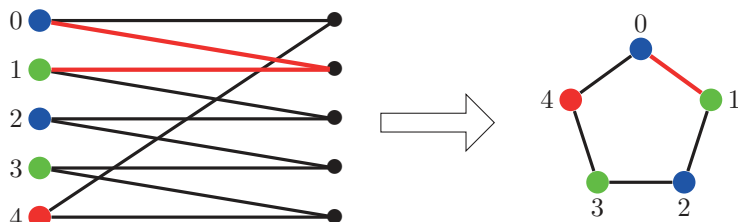
In het broncoderingsprobleem krijgen twee partijen, Alice en Bob, allebei een input van een externe bron en is hun doel om met zo min mogelijk communicatie Alices input aan Bob duidelijk te maken. Alice kan hiervoor uiteraard haar hele input naar Bob sturen, maar soms kan er communicatie bespaard worden als Bobs input al wat informatie over die van Alice bevat.

Vergelijkbaar met een discreet kanaal is een *discrete bron* een drietal $\mathcal{M} = (U, W, P)$ bestaande uit een eindige inputverzameling U voor Alice, een eindige inputverzameling W voor Bob en een kansverdeling P over alle mogelijke paren in $U \times W$. Het model dicteert dat de kans dat Alice input $u \in U$ en Bob input $w \in W$ krijgt gegeven is door $P(u, w)$. Zoals voor kanalen met ruis wordt aangenomen dat de bron geheugenloos is, wat wil zeggen dat de kansverdeling P onveranderd blijft na elke keer dat de bron inputs uitdeelt.

Zoals Shannon voor het kanaalcoderingsprobleem, beschouwde Witsenhausen [10] de minimale communicatie die nodig is in het geval waarin er geen fouten getole-



Figuur 6 Een discrete bron met 5 mogelijke inputs voor Alice en Bob.



Figuur 7 Van een bron naar zijn karakteristieke graaf.

reerd kunnen worden, oftewel de maximale snelheid waarmee het probleem opgelost kan worden in het foutloze model. Als voorheen doen de waarden van de kansen $P(u,v)$ er dan niet meer toe en is het alleen van belang of ze nul zijn of niet. Voor de bron van Figuur 6 krijgen we zo bijvoorbeeld een graaf die er zo uitziet als die van Figuur 2.

Het probleem van foutloze broncodering kan wederom bestudeerd worden in graaftheoretische termen. Hiervoor associëren we met een bron $\mathcal{M} = (U, W, P)$ een zogeheten *karakteristieke graaf* $G = (V, E)$ met knopenverzameling $V = U$ en waarin twee knopen $u, v \in V$ een kant vormen als er een $w \in W$ is zodat beide $P(u, w) > 0$ en $P(v, w) > 0$ gelden (zie Figuur 7). De gedachte achter deze graaf is dat als Bob $w \in W$ ontvangt, hij niet weet of Alices input u of v is en Alice hem dus van extra informatie moet voorzien. Met andere woorden, de kanten in de karakteristieke graaf G identificeren precies Alices inputparen die Bob niet altijd kan onderscheiden. Zoals voor het kanaalcoderingsprobleem is er voor elke graaf een bron met die graaf als karakteristieke graaf. Om het broncoderingsprobleem op te lossen kijken we echter naar een andere graafparameter dan het onafhankelijkheidsgetal. Het oplossen van een broncoderingsprobleem betekent dat Alice voor elk van haar mogelijke input u een zo kort mogelijk bericht voor Bob heeft waaruit hij samen met zijn eigen input v haar input kan herleiden. Als we deze berichten zien als kleuren dan moet de kleuring van Alices inputverzameling er aan voldoen dat paren van inputs die door Bob mogelijk verward kunnen worden verschillende kleuren hebben. Met andere woorden, de kleuring moet een geldige kleuring van de karakteristieke graaf zijn! Andersom zien we ook dat een geldige kleuring van de karakteristieke graaf gebruikt kan worden om het broncoderingsprobleem op te lossen. De eerste stelling voor het broncoderingsprobleem luidt dus:

Stelling 3. *Het oplossen van één instantie van het broncoderingsprobleem voor een bron met karakteristieke graaf G is equivalent met het vinden van een geldige kleuring van G .*

Als gevolgtrekking zien we ook direct het volgende feit.

Propositie 5. *Het minimale aantal bits dat Alice Bob kan sturen bij het coderen voor een bron met karakteristieke graaf G is $\lceil \log_2 \chi(G) \rceil$.*

De Witsenhausen-snelheid

Broncodering kan soms efficiënter gedaan worden door blokken inputreeksen samen te coderen. Hierbij wachten Alice en Bob totdat ze n paren inputs hebben ontvangen, waarna Alice haar reeks codeert in een zo kort mogelijk bericht waaruit Bob, samen met zijn inputreeks, Alices reeks kan achterhalen. Natuurlijk kan Alices bericht bestaan uit de reeks van n kleuren die ze haar inputs in de individuele instanties zou geven, zodat Bob elk van haar inputs onafhankelijk van elkaar kan achterhalen. In graaftheoretische termen zegt dit dat blokken van n inputs gecodeerd kunnen worden met $\chi(G)^n$ verschillende berichten. Maar, zoals al aangekondigd, kan dit soms efficiënter. De vraag die we moeten stellen is wanneer Bob twee inputreeksen $u_1, \dots, u_n \in V$ en $v_1, \dots, v_n \in V$ van Alice niet met zekerheid kan onderscheiden. Merk op dat dit gebeurt wanneer voor elke $i \in \{1, \dots, n\}$ geldt dat $u_i = v_i$ of $\{u_i, v_i\} \in E$, omdat als er één coördinaat i is waarvoor geen van beide dingen gelden, Bob u_i en v_i , en daarom de hele reeksen, kan onderscheiden aan de hand van zijn i de input. De mogelijk te verwarren paren zijn dus weer precies de kanten van de graaf $G^{\boxtimes n}$! Zo komen we bij de tweede stelling.

Stelling 4. *Het minimum aantal verschillende kleuren (codewoorden) dat Alice kan gebruiken is $\chi(G^{\boxtimes n})$.*

De volgende definitie volgt daarom als een natuurlijke tegenhanger van de Shannon-capaciteit.

Definitie 6. (Witsenhausen-snelheid). De *Witsenhausen-snelheid* van een graaf G is gedefinieerd als

$$R(G) = \lim_{n \rightarrow \infty} \chi(G^{\boxtimes n})^{1/n}.$$

Als de blokengte naar oneindig gaat, hoeven er gemiddeld $\lceil \log_2 R(G) \rceil$ bits per broninstantie gecommuniceerd te worden.

Hierboven gaven we al een intuïtief bewijs voor de volgende stelling over het gedrag van het chromatisch getal onder het sterke graafproduct.

Propositie 6. *Zij G een graaf. Dan geldt*

$$\chi(G^{\boxtimes n}) \leq \chi(G)^n.$$

Dat het coderen van blokken een verschil kan maken, met andere woorden, dat $\chi(G^{\boxtimes n}) < \chi(G)^n$ soms kan gelden, is weer te zien aan de hand van ons lopende voorbeeld, de 5-cykel. We weten al dat $\chi(C_5) = 3$. Witsenhausen liet zien dat strikt minder dan negen kleuren voldoende zijn voor $C_5^{\boxtimes 2}$.

Propositie 7 (Witsenhausen). $\chi(C_5^{\boxtimes 2}) \leq 5$.

De meest expliciete manier om dit te zien is door te verifiëren dat Witsenhausens kleuring (Figuur 8) geldig is. Merk allereerst op dat de tabel alle knopen van $C_5 \boxtimes C_5$ bevat. Merk ten tweede op dat elke rij een onafhankelijke verzameling vormt. Omdat elke onafhankelijke verzameling een eigen kleur heeft is de kleuring geldig.

Deze kleuring vond Witsenhausen niet door met veel geduld alle mogelijkheden te doorzoeken. Een vertaling van zijn elegante argument is gebaseerd op tellen modulo 5. Merk op dat twee knopen $x, y \in \{0, 1, 2, 3, 4\}$ van de graaf C_5 een kant vormen als $x - y = 1 \pmod{5}$ of $x - y = 4 \pmod{5}$. Twee knopen $(x, y), (x', y') \in \{0, 1, 2, 3, 4\} \times \{0, 1, 2, 3, 4\}$ van de graaf $C_5 \boxtimes C_5$ vormen daarom een kant als

(0, 0)	(2, 1)	(4, 2)	(1, 3)	(3, 4)
(0, 1)	(2, 2)	(4, 3)	(1, 4)	(3, 0)
(0, 2)	(2, 3)	(4, 4)	(1, 0)	(3, 1)
(0, 3)	(2, 4)	(4, 0)	(1, 1)	(3, 2)
(0, 4)	(2, 0)	(4, 1)	(1, 2)	(3, 3)

Figuur 8 Witsenhausens kleuring van $C_5 \boxtimes C_5$.

$(x, y) - (x', y') \pmod{5}$ behoort tot de verzameling

$(0,1) (1,0) (1,1) (0,4) (4,0) (1,4) (4,1) (4,4)$.

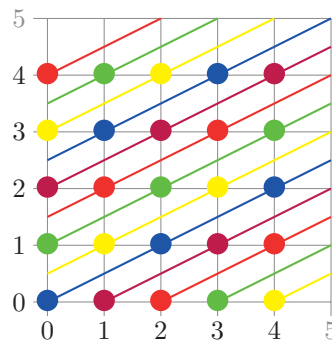
Beeld de knopen van $C_5^{\boxtimes 2}$ af op een tweedimensionaal grid (Figuur 9).

Witsenhausen observeerde dat alle knopen op een lijn wiens richting niet gegeven is door één van de punten uit de rij hierboven een onafhankelijke verzameling vormen. De reden hiervoor is dat het verschil van elke twee punten op een lijn in richting (a,b) gelijk is aan een veelvoud van (a,b) . Beschouw bijvoorbeeld de lijn door de oorsprong $(0,0)$ in richting $(2,1)$ (de blauwe lijn in Figuur 9). Modulo 5 bevat deze lijn de punten $(0,0)$, $(2,1)$, $2(2,1) = (4,2)$, $3(2,1) = (6,3) = (1,3) \pmod{5}$ en $4(2,1) = (8,4) = (3,4) \pmod{5}$, precies de eerste rij uit Figuur 8. Als twee knopen geen kant vormen dan mogen we ze dezelfde kleur

geven en omdat we alle knopen kunnen dekken met vijf parallelle lijnen zijn vijf kleuren dus voldoende.

Analoog aan de kanaalsetting gelden de volgende twee stellingen voor de Witsenhausen-snelheid van C_5 .

Propositie 8 (Lovász). $R(C_5) = \sqrt{5}$.



Figuur 9 Lijnkleuring van $C_5 \boxtimes C_5$.

Dit volgt direct uit Stelling 1, die zei dat $\alpha(C_5^{\boxtimes n}) \leq 5^{n/2}$, en door de volgende eenvoudige observatie toe te passen op de graaf $G = C_5^{\boxtimes n}$:

Propositie 9. Zij $G = (V, E)$ een graaf. Dan geldt

$$\alpha(G) \chi(G) \geq |V|.$$

Waarom is dit zo? Stel we hebben een geldige kleuring van een graaf G met $c = \chi(G)$ knopen. Noem de kleuren $1, 2, \dots, c$. Voor kleur i , laat n_i het aantal knopen met kleur i zijn. Elke knoop heeft een kleur, dus het totale aantal knopen is $n_1 + n_2 + \dots + n_c = |V|$. Knopen met eenzelfde kleur moeten een onafhankelijke verzameling vormen. Daarom kan elke kleurklasse uit hooguit $\alpha(G)$ knopen bestaan. We hebben dus dat $n_i \leq \alpha(G)$ voor elke kleur i waaruit de claim volgt. $\square$

Referenties

- 1 M. Aigner en G.M. Ziegler, *Proofs from THE BOOK*, Springer, 2014.
- 2 R. Diestel, *Graph Theory*, Graduate Texts in Mathematics, Vol. 173, Springer, 2005.
- 3 J. Körner en A. Orlitsky, Zero-error information theory, *IEEE Trans. Inform. Theory* 44(6), *Information Theory: 1948–1998* (1998) 2207–2229.
- 4 L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* 25(1) (1979) 1–7.
- 5 E. Lubetzky, *Graph Powers and Related Extremal Problems*, Ph.D. thesis, Tel Aviv University, 2007.
- 6 K.A. Mathew en P.R.J. Östergård, New lower bounds for the Shannon capacity of odd cycles, *Designs, Codes and Cryptography* 84 (2015), 13–22.
- 7 J. Nayak, E. Tuncel en K. Rose, Zero-error source-channel coding with side information, *Information Theory, IEEE Transactions on* 52(10) (2006), 4626–4629.
- 8 C.E. Shannon, A mathematical theory of communication, *Bell System Technical Journal* 27 (1948), 379–423, 623–656.
- 9 C.E. Shannon, The zero error capacity of a noisy channel, *Institute of Radio Engineers, Transactions on Information Theory* 2(3) (1956), 8–19.
- 10 H.S. Witsenhausen, The zero-error side information problem and chromatic numbers, *IEEE Trans. Information Theory* 22(5) (1976), 592–593.

Thomas Ward

School of Mathematics
University of Leeds
t.b.ward@leeds.ac.uk

Trip to the proof

Mixing, wine, and serendipity

Mixing for measure-preserving group actions is a fundamental notion in ergodic theory, with different phenomena arising for different acting groups. In 1993, Klaus Schmidt and Tom Ward proved that 2-mixing implies mixing of all orders for actions by commuting automorphisms of connected groups. Tom Ward explains how he became interested in this problem, and describes how a chance encounter with a paper on a seemingly unrelated problem in number theory played a key role.

During the 1980's, while I was a graduate student at Warwick University under the supervision of Klaus Schmidt, a specific kind of algebraic dynamical system was emerging as a surprisingly rich and relatively unexplored field. In hindsight, a small shift in how a key example constructed by Ledrappier is thought of might have predicted some of this, but prediction with the benefit of hindsight is a little too easy.

Mixing is a mathematical version of the idea of, well, mixing. If two ingredients of a cocktail are poured carefully into a glass — so carefully that perhaps they form individual layers — then the action of a stirrer is ‘mixing’ if after some time every mouthful tastes the same. That is, every part of the glass has the ingredients in the same proportion up to a negligible error. This becomes a mathematical concept by noticing that the volume may be viewed as a measure on the space consisting of the contents of the glass, and the action of the stirrer might be thought of as iteration of a map that preserves that measure (unless it is being stirred using a straw, and the person stirring is taking a crafty sip every now and then). Avoiding all the interesting and subtle physical and chemical issues involved — particularly egregious in the circumstances — we might as well assume the action of stirring is invertible, and for mathematicians the resulting structure of a measure-preserving action of the

integers might as well be an action of any group. Having no wish to trip up on any measure theory, let's say that the group is countable. So here is mixing: if a countable group G acts by transformations preserving a measure μ on a probability space, then it is called mixing if $\mu(A \cap gB)$ converges to $\mu(A)\mu(B)$ as g ‘goes to infinity’ in G . And why not be ambitious? Mixing on $k+1$ sets (or mixing of order k) means that for any measurable sets $A_0, \dots, A_k$ the measure

$$\mu(A_0 \cap g_1 A_1 \cap \dots \cap g_k A_k)$$

of the intersection converges to $\prod_{j=0}^k \mu(A_j)$ as the group elements g_j go to infinity and move apart from each other. So here is a mathematical question: given a measure-preserving action of a countable group, determine if it is mixing on k sets for a given k . When $G = \mathbb{Z}$ it is a long-standing question of Rokhlin as to whether mixing on 2 sets forces mixing on 3 sets.

Which brings us to Ledrappier's example [1] (simplified for convenience from his harmonic condition example): let X be the subset of $\{0,1\}^{\mathbb{Z}^2}$ consisting of the points x satisfying $x_{s,t} = x_{s+1,t} + x_{s,t+1}$ modulo 2 for every $(s,t) \in \mathbb{Z}^2$. This is a compact group, and the shift in $\mathbb{Z}^2$ defines an action of $\mathbb{Z}^2$ that preserves the natural Haar measure. The system is easily shown to be mixing on 2 sets, but the fact that the relation $x_{s,t} = x_{s+2^n,t} + x_{s,t+2^n}$ modulo 2 holds for all $n \geq 1$ (a direct consequence of the

properties of the Frobenius $a \mapsto a^2$ modulo 2 under iteration) forces a correlation between triples of sets separated by arbitrarily large distances — failure of mixing on 3 sets. Ledrappier also pointed out that any system like this built from automorphisms of compact groups has a property called ‘Lebesgue spectrum’. A productive shift in perspective is to think of this system as the dual group of the module $\mathbb{Z}[u_1^{\pm 1}, u_2^{\pm 1}] / \langle 1 + u_1 + u_2 \rangle$, with the action of $(a,b) \in \mathbb{Z}^2$ dual to multiplication by $u_1^a u_2^b$. Thus a version of the mixing question becomes this: describe the mixing properties of such a system built from a module M over the ring $R = \mathbb{Z}[u_1^{\pm 1}, \dots, u_d^{\pm 1}]$ in terms of properties of the module M — in the certain knowledge that the answer is non-trivial because it is for Ledrappier's example.

Work of Kitchens and Schmidt [4] probed the mixing properties of systems whose compact group is zero-dimensional, uncovering a complex collection of properties leading to many interesting questions. Schmidt [3] also showed that the way in which a ‘shape’ produced by the Frobenius automorphism witnesses failure of higher-order mixing as seen by Ledrappier could not take place if the compact group X is connected. That is, for a mixing $\mathbb{Z}^d$ action by automorphisms of a compact connected group, choosing the times $g_1, \dots, g_k$ to be dilates of a fixed shape in $\mathbb{Z}^d$ would never show failure of mixing, raising the question: for these connected systems, does mixing imply mixing of all orders?

By 1991 I was working at Ohio State University, and we were notified that some duplicate journals were being discarded. As life was then full of time for mathematics, I went into the basement and leafed



Doug Lind, Tom Ward and Klaus Schmidt at a Lorentz Center workshop in 2014

through piles of journals in recycling bins, tearing out any articles that looked vaguely interesting. I piled these up, and left to attend a workshop at CIRM in Luminy.

In that beautiful place, Klaus Schmidt reminded Doug Lind and me of this open problem over a splendid meal. Perhaps with the assistance of the generous provision of wine, I became sure that I had an argument, essentially using the Lebesgue spectrum property, that proved mixing of all orders for these connected systems. Not for the first, and not for the last, time, Klaus let me whither on for some time as we walked under the pine trees before politely pointing out that my suggested argument applied unchanged to Ledrappier's example.

Flying back to Columbus, the problem was firmly in my mind. The ' $\times 2, \times 3$ ' system, itself studied for other reasons, was the natural start. Ledrappier's salutary example showed that the result sought really couldn't come from the familiar toolbox of

spectral or entropy methods. Via Fourier analysis of indicator functions of sets, it seemed to come down to this: what can you say about solutions of $\sum_{j=1}^k a_j x_j = 1$ in a number field, where the variables x_j come from a finitely-generated multiplicative subgroup? For ' $\times 2, \times 3$ ' the field would be $\mathbb{Q}$, and the multiplicative subgroup $\{2^a 3^b \mid a, b \in \mathbb{Z}\}$. Failure of mixing of all orders seemed roughly equivalent to equations of this shape having too many — infinitely many — non-trivially different solutions. Trivially different solutions abound if a sub-sum vanishes, because that vanishing sub-sum can be scaled by powers of 2 and 3 arbitrarily.

After a few days back in Columbus, I sorted through the pile of torn-out papers on my desk. One was a (then) recent paper of Schlickewei [2] with a form of ' S -unit theorem'. For the finite-dimensional case at hand (it turned out later that the topological dimension of the compact group X plays a role) a simple reduction argu-

ment was suddenly completely clear. If a $\mathbb{Z}^d$ -action by automorphisms of a compact connected abelian group fails to be mixing on $k \geq 3$ sets, then by the Fourier analysis argument there is a linear equation with k terms over a field of characteristic zero that has infinitely many distinct solutions lying in a multiplicative group with d generators (corresponding to the automorphisms defining the action). By the S -unit theorem, this is impossible unless infinitely many of them come from a vanishing sub-sum: a linear equation with $j < k$ terms. But finding infinitely many solutions for that shorter linear equation is a witness to failure of mixing on $j < k$ sets. Thus mixing on 2 sets implies mixing of all orders. Some cleaning up was needed — algebra to reduce to cyclic modules, and a more subtle process needed to deal with infinite-dimensional compact groups which do not readily permit the translation into statements in number fields — but this quickly led to the proof of the full case with Klaus Schmidt [5].

There are some lessons to take from this strange coincidence and happy resolution. Certainly ideas produced under the influence of wine may eventually face the sobering reality of counter-examples — but can be motivating nonetheless. More importantly, Tramezzino's tale *Peregrinaggio di tre giovani figliuoli del re di Serendippo* in which 'accidents and sagacity' play such a role still has something for us. The rush of modern academic life, the growing use of online journals and their sophisticated and well-intentioned nudging towards related articles, the demarcation of subject areas, the overwhelming growth in the volume of the mathematical literature — all make the benefits of serendipity less easy to access. If you have the good fortune of time to spend on mathematics, spend some of it on the not 'suggested article', on the articles that readers are not 'also reading', and on items with the wrong subject classification codes. ☞

References

- 1 F. Ledrappier, Un champ markovien peut être d'entropie nulle et mélangeant, *C.R. Acad. Sci. Paris Sér. A-B* 287(7) (1978), A561–A563.
- 2 H.P. Schlickewei, An explicit upper bound for the number of solutions of the S -unit equation, *J. Reine Angew. Math.* 406 (1990), 109–120.
- 3 K. Schmidt, Mixing automorphisms of compact groups and a theorem by Kurt Mahler, *Pacific J. Math.* 137(2) (1989), 371–385.
- 4 K. Schmidt, *Dynamical systems of algebraic origin*, Birkhäuser, 1995.
- 5 K. Schmidt and T. Ward, Mixing automorphisms of compact groups and a theorem of Schlickewei, *Invent. Math.* 111(1) (1993), 69–76.

Jan van de Craats

Korteweg-de Vries Institute for Mathematics
University of Amsterdam
j.vandecraats@uva.nl

Research

Morley chains of osculant curves

At the end of the nineteenth century the algebraic geometer Frank Morley discovered a nice little theorem on the trisectors of a triangle, known as ‘Morley’s trisector theorem’. In the March issue Jan van de Craats and Jan Brinkhuis elucidated the role of cardioids in Morley’s discovery of his theorem. In this article Jan van de Craats aims to answer the question *why* Morley was interested in studying cardioids at all.

In 1899 the algebraic geometer Frank Morley (1860–1937) discovered a surprising result on the trisectors of a triangle. He mentioned it to friends, who spread it over the world in the form of mathematical gossip. Morley’s trisector theorem, as it later became known, reads as follows (see Figure 1): *in any triangle, the three points of intersection of the adjacent angle trisectors form an equilateral triangle*.

In their recent paper ‘Cardioids and Morley’s trisector theorem’ [1], Jan van de Craats

and Jan Brinkhuis elucidated the role of cardioids in Morley’s discovery of his theorem. Morley and his son Frank Vigor Morley also explained this connection on pages 239–244 of their book *Inversive Geometry* [4] from 1933, but this is no easy reading. The Morleys usually present their results in an informal way and rigorous proofs are seldom given. However, the book has been reprinted in 1954 and, again, in 2013, indicating that also the modern reader might find it valuable to study its contents.

In our paper [1], we took pains to explain Morley’s reasoning on cardioids and trisectors in an accessible way and to provide detailed proofs. However, we didn’t explain *why* Morley was interested in studying cardioids at all. The present paper, which in part is based on Morley’s 1929 article [3] and on chapter XXI of *Inversive Geometry* [4], aims to answer this question. Its main results are summarized in Theorems 2, 3 and 4 and their proofs. In a strict sense, these theorems do not contain new results, but perhaps our pre-

sentation may inspire modern readers to excavate more jewels from Morley’s work.

Morley considered a cardioid as a member of an infinite sequence of rational curves in the Argand plane (the Euclidean plane coordinated by complex numbers) $B_1, B_2, B_3, B_4, \dots$ in which B_1 is a point, B_2 a circle, B_3 a cardioid, and $B_4, B_5, \dots$ are ‘higher’ curves. The curves B_n have many interesting properties, leading to intriguing theorems of a general nature with pleasant special cases. Morley’s trisector theorem is just one instance. Another example is the five circles theorem illustrated in Figure 2. In Morley’s words [4, p.265]: *We place a ring of five circles with centers on a given circle and each intersecting the next on the circle. The five other intersections of the adjacent circles, being joined in order, form the five-line, and the salient thing is that the intersections of non-adjacent sides are also on the respective five circles*.

Later in this paper we will present a proof of this theorem as a special case of a more general result on curves of type B_4 , but first we need to introduce some of Morley’s idiosyncratic notations and terminology. The next two sections repeat in a condensed form a similar introduction in Van de Craats and Brinkhuis [1, pp. 26–28].

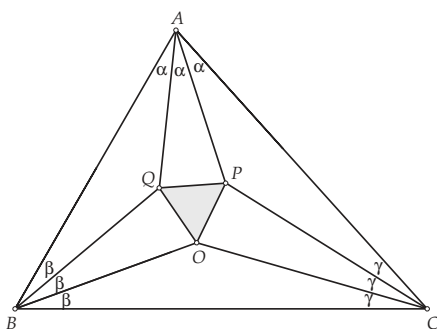


Figure 1 Morley’s trisector theorem.

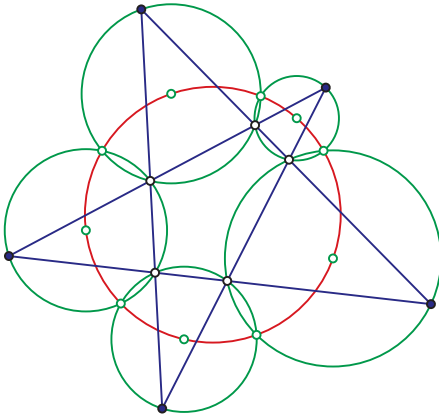


Figure 2 Morley's five circles theorem.

Line-equations and map-equations

Points in the plane, viewed as complex numbers, will be represented by lower case letters. The letter t will always be used for points on the unit circle, so $t\bar{t} = 1$ in other words, $\bar{t} = 1/t$. Such points will be called *turns*, since multiplication by t amounts to an anticlockwise rotation around the origin over an angle $\arg(t)$. Occasionally, also the Greek letter τ will be used for turns.

For any two distinct points x and c on a line L the vectors $x - c$ and $c - x$ both indicate the direction of L , but in opposite sense. However, the quotient $t = (x - c)/(\bar{x} - \bar{c})$, which is a point on the unit circle, is independent of the order of x and c . In fact, it only depends on L and not on the choice of x and c on L . It is a turn, called the *clinant* of L . Its argument equals *twice* the directed angle from the real axis to L . Two lines are parallel if and only if their clinants are equal. Furthermore, two lines are perpendicular if and only if their clinants differ by a factor -1 .

If we fix the point c on L and consider x as a variable, then the equation

$$x - c = t(\bar{x} - \bar{c}) \quad (1)$$

represents all points of L . Note that it is a *self-conjugate* equation: conjugation yields the same equation since $\bar{t} = 1/t$.

If L passes through the origin we may take $c = 0$, leading to the simple equation $x = t\bar{x}$. If the origin is not on L , the image b upon reflecting the origin in L completely determines L . Since the line through the origin and b is perpendicular to L , the clinant t of L equals $-b/\bar{b}$. Furthermore, the point $b/2$ is on L , so in equation (1) we may take $c = b/2$. The resulting equation

$$x - b/2 = -(b/\bar{b})(\bar{x} - \bar{b}/2)$$

can be written as

$$\bar{b}x + b\bar{x} = b\bar{b}. \quad (2)$$

An equivalent way to express the points x of the line L is

$$x - b - t\bar{x} = 0, \quad (3)$$

where $b = -t\bar{b}$. Note that equation (3) also holds if L passes through the origin: then $b = 0$ should be taken. Equations like (1), (2) and (3) will be called *line-equations*.

For any function $f(t)$ the equation $x = f(t)$ may be viewed as a parametric representation of a curve Γ in the plane with parameter t running through the unit circle. It will be called a *map-equation* of the curve Γ . In the sequel, $f(t)$ will always be a polynomial in t . In particular, the map-equation $x = c + at$ represents a circle with center c and radius $|a|$.

The line-equation of a cardioid

As explained in Van de Craats and Brinkhuis [1], for given $a \neq 0$ and c the map-equation

$$x = c + 2at - \bar{a}t^2 \quad (4)$$

describes a *cardioid* when the parameter t runs through the unit circle (see Figure 3, where we have taken $c = 0$ and $a = 1$). Its name is derived from its heart-like shape. The point c , which is not on the curve, is called the *center* of the cardioid. Centers of cardioids played an important role in Morley's discovery of his trisector theorem. The cardioid has a *cusp* when $dx/dt = 0$, which in Figure 3 occurs for parameter value $t = 1$ at the point $x = 1$, and in general for $t = a/\bar{a}$ at $x = c + a^2/\bar{a}$.

In [1] it is shown that the tangent line to the cardioid (4) at the point with parameter value t is given by the line-equation

$$(x - c) - 3at + 3\bar{a}t^2 - (\bar{x} - \bar{c})t^3 = 0. \quad (5)$$

Also this equation is self-conjugate, i.e.,

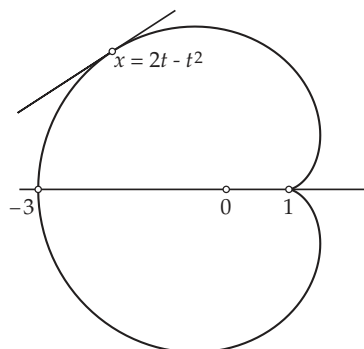


Figure 3 The cardioid $x = 2t - t^2$ with a tangent line.

conjugation (and multiplying both sides by $-t^3$) yields the same equation.

For a given turn t , equation (5) represents the tangent line to the cardioid at the point with parameter value t . For varying t we get the set of all tangent lines. The line-equation (5) thus yields an alternative representation of the cardioid (4), namely as the envelope of the set of its tangent lines.

The map-equation of the cardioid may be recovered from the line-equation by taking two 'neighboring' lines from this set, say, for parameter values t and τ , and subtracting their equations to get an equation for their point of intersection

$$\begin{aligned} -3a(t - \tau) + 3\bar{a}(t^2 - \tau^2) \\ - (\bar{x} - \bar{c})(t^3 - \tau^3) = 0. \end{aligned}$$

Dividing by $(t - \tau)$, taking the limit $\tau \rightarrow t$ and conjugating yields an equation for the point of tangency x ,

$$-3\bar{a} + 6a/t - 3(x - c)/t^2 = 0,$$

from which it follows that

$$x = c + 2at - \bar{a}t^2.$$

This, indeed, is the map-equation (4) of the cardioid. The method for obtaining the map-equation from the line-equation thus formally may be described as differentiation with respect to t , followed by conjugation and solving the resulting equation for x . In the sequel, we will always define curves by line-equations.

The curves B_n

In general, for $n \geq 1$ we define a curve B_n by the line-equation

$$\begin{aligned} B_n: (x - c) - \binom{n}{1}a_1t + \binom{n}{2}a_2t^2 - \dots \\ + (-1)^{n-1}\binom{n}{n-1}a_{n-1}t^{n-1} \\ + (-1)^n(\bar{x} - \bar{c})t^n = 0 \end{aligned} \quad (6)$$

where $a_{n-k} = \bar{a}_k$ for all k to ensure that the equation is self-conjugate. Note that for even n , say $n = 2m$, this implies that the middle coefficient a_m is real. The point c is called the *center* of the curve. The reason for including the binomial coefficients will become clear when, below, we will introduce the so-called 'polarized' equation of B_n .

For a fixed turn t , equation (6) represents a line. Taking two distinct points x_1 and x_2 on this line, subtraction yields $(x_1 - x_2) + (-1)^n t^n (x_1 - x_2) = 0$, so the clinant of this line is $-(-1)^n t^n$. For vary-

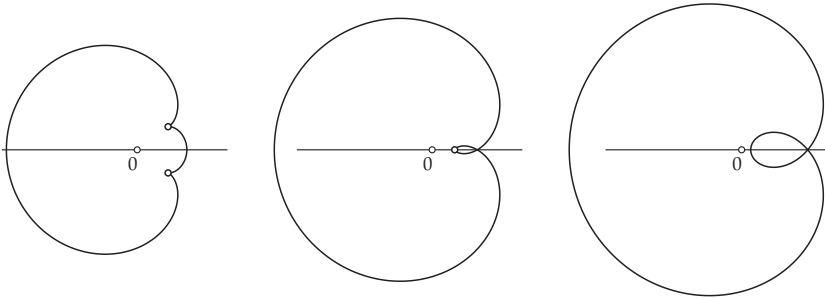


Figure 4 Curves B_4 with two cusps (left), two coinciding cusps (middle) and no cusps (right).

ing t we get a set of lines, the envelope of which is the curve B_n . The curve B_n is completely determined by its center c and the coefficients a_k .

As examples, we treat the cases $n = 2, 1$ and 4 in more detail.

For a curve B_2 we have the line-equation

$$B_2: (x-c) - 2\rho t + (\bar{x}-\bar{c})t^2 = 0 \quad (7)$$

where the coefficient ρ should be real to keep the equation self-conjugate.

Differentiation with respect to t yields $-2\rho + 2(\bar{x}-\bar{c})t = 0$, so by conjugation and solving for x we get as the map-equation of the curve $x = c + \rho t$, which indeed is a circle. Its center is c and its radius is $|\rho|$. For any fixed turn t , equation (7) represents the tangent line to the circle at the point with parameter value t . Its clinant is $-t^2$.

The line-equation of the 'curve' B_1 is

$$B_1: (x-c) - (\bar{x}-\bar{c})t = 0.$$

For varying turns t this is just the set of all lines through c , and the map-equation of B_1 is simply $x = c$. The clinant of a line from the line-equation is t .

For $n = 4$ we get the line-equation

$$B_4: (x-c) - 4at + 6\mu t^2 - 4\bar{a}t^3 + (\bar{x}-\bar{c})t^4 = 0.$$

Again, the middle coefficient μ must be real to keep the equation self-conjugate. Its clinant is $-t^4$. The map-equation now becomes

$$x = c + 3at - 3\mu t^2 + \bar{a}t^3.$$

The cusp-equation $dx/dt = 0$ is

$$a - 2\mu t + \bar{a}t^2 = 0.$$

The reader might like to verify that the roots of this equation are turns if and only if $\mu^2 - a\bar{a} \leq 0$. In Figure 4 curves B_4 have been drawn with $c = 0$, $a = 1$ and $\mu = 0.6$ (two cusps), $\mu = 1$ (two coinciding cusps at $x = 1$ for parameter value $t = 1$) and $\mu = 1.2$ (no cusps).

Examples of osculant curves

To explain the concept of *osculant curves* of a curve B_n , we first take as an example a cardioid B_3 . In its most simple form, its line-equation is given by

$$x - 3t + 3t^2 - \bar{x}t^3 = 0.$$

Its map-equation is

$$x = 2t - t^2,$$

so its center is $c = 0$ and its cusp is $x = 1$ (with parameter value $t = 1$). By a change of coordinates, any non-degenerate cardioid can be written in this form.

For any three turns t_1, t_2, t_3 , we associate to the line-equation of the cardioid the *polarized equation*

$$x - (t_1 + t_2 + t_3) + (t_1 t_2 + t_2 t_3 + t_3 t_1) - \bar{x} t_1 t_2 t_3 = 0.$$

This again is a self-conjugate equation, so it represents a line associated with the three parameter values t_1, t_2 and t_3 . If $t_1 = t_2 = t_3 = t$, it is a tangent line to the cardioid, but if only $t_2 = t_3 = t$ we get the equation

$$x - (t_1 + 2t) + (t^2 + 2t_1 t) - \bar{x} t_1 t^2 = 0.$$

This, for fixed t_1 and varying t may be viewed as the line-equation of a curve of type B_2 . Indeed, differentiating with respect to t , conjugating and solving for x yields

$$x = t_1 + t - t_1 t$$

which is the map-equation of the circle with center t_1 and radius $|1 - t_1|$.

For $t = t_1$, the map-equations of both the circle and the cardioid yield the point $x_1 = 2t_1 - t_1^2$, while also, in view of the line-equations, their tangent lines at this point coincide. Therefore, x_1 is a point where the two curves touch. The circle is called an *osculant circle* to the cardioid. Furthermore, the circle also passes through the cusp $x = 1$ (take $t = 1$). Thus we have

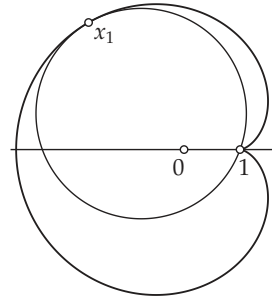


Figure 5 Any osculant circle of a cardioid passes through its cusp.

the result that *any osculant circle of a cardioid passes through its cusp* (see Figure 5).

We now take two turns t_1 and t_2 and a variable turn t in the polarized equation

$$x - (t_1 + t_2 + t) + (t_1 t_2 + t_2 t + t_1 t) - \bar{x} t_1 t_2 = 0. \quad (8)$$

This is the line-equation of a 'curve' of type B_1 , which, as a map-equation, is just the point

$$x_{12} = t_1 + t_2 - t_1 t_2.$$

It is the second intersection point of the osculant circles for t_1 and t_2 (their other intersection point is the cusp $x = 1$). The point x_{12} is an *osculant* of both the osculant circles; it is called a *second osculant* of the cardioid (the first osculants being the osculant circles).

For $t = t_3$ the line-equation (8) of the second osculant yields the fully polarized form

$$x - (t_1 + t_2 + t_3) + (t_1 t_2 + t_2 t_3 + t_3 t_1) - \bar{x} t_1 t_2 t_3 = 0.$$

This is a line through x_{12} , which, by symmetry, also passes through x_{23} and x_{31} . It is called a *third osculant* of the cardioid.

We thus have proved: *for any three osculant circles of a cardioid, their second intersection points are collinear* (as we have seen, all osculant circles also pass through the cusp). See Figure 6.

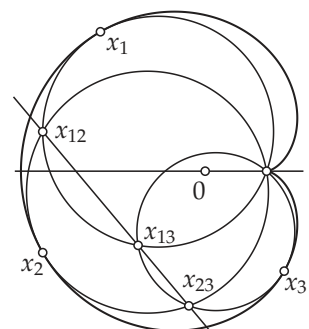


Figure 6 Three osculant circles of a cardioid. They all pass through the cusp, and their second intersections are collinear.

As a second example, consider the osculants of a circle given by the line-equation

$$(x-c) - 2\rho t + (\overline{x-c})t^2 = 0 \quad (\text{with real } \rho)$$

(cf. (7)). Recall that its map-equation is $x = c + \rho t$. For any two turns t_1 and t_2 the polarized equation yields the osculant line

$$L_{12}: (x-c) - \rho(t_1 + t_2) + (\overline{x-c})t_1 t_2 = 0.$$

For fixed t_1 and variable t we get the line-equation of a first osculant, which is a 'curve' of type B_1 :

$$(x-c) - \rho(t_1 + t) + (\overline{x-c})t_1 t = 0.$$

As a map-equation, this is just the point $x_1 = c + \rho t_1$ on the circle. Similarly, for variable t and fixed t_2 , we get the line-equation

$$(x-c) - \rho(t + t_2) + (\overline{x-c})t t_2 = 0$$

which, as a map-equation, is just the point $x_2 = c + \rho t_2$ on the circle. The osculant line L_{12} , therefore, is the line through x_1 and x_2 .

Osculant curves in general

Let $n \geq 2$. Take a general curve B_n with line-equation (6), or, written in a more compact form,

$$x - c + \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} a_k t^k + (-1)^n (\overline{x-c}) t^n = 0 \quad (9)$$

where $a_{n-k} = \overline{a_k}$ for all k .

For n turns $t_1, \dots, t_n$, the polarized form of the line-equation is

$$x - c + \sum_{k=1}^{n-1} (-1)^k a_k \sigma_k + (-1)^n (\overline{x-c}) \sigma_n = 0 \quad (10)$$

where the σ_k are the symmetric functions of $t_1, \dots, t_n$ defined by

$$\begin{aligned} \sigma_1 &= t_1 + t_2 + \dots + t_n \\ \sigma_2 &= t_1 t_2 + \dots + t_{n-1} t_n \\ \sigma_3 &= t_1 t_2 t_3 + \dots + t_{n-2} t_{n-1} t_n \\ &\vdots \\ \sigma_n &= t_1 t_2 \dots t_n. \end{aligned}$$

It will be clear how in general the osculant curves of B_n will be defined: fix m of the parameter values t_i in the polarized equation and take the others equal to t . Then a curve of type B_{n-m} is obtained, an m -th osculant of B_n . The n -th osculant is the line given by the fully polarized line-equation (10).

To obtain the map-equation of B_n , we differentiate (9) with respect to t ,

$$\sum_{k=1}^{n-1} (-1)^k k \binom{n}{k} a_k t^{k-1} + (-1)^n n (\overline{x-c}) t^{n-1} = 0.$$

By using $k \binom{n}{k} = n \binom{n-1}{k-1}$, division by n , conjugation and multiplication by t^{n-1} we get

$$\sum_{k=1}^{n-1} (-1)^k \binom{n-1}{k-1} \overline{a_k} t^{n-k} + (-1)^n (x-c) = 0.$$

Since $\overline{a_k} = a_{n-k}$, the map-equation of B_n can be written as

$$x = c - \sum_{k=1}^{n-1} (-1)^{n-k} \binom{n-1}{k-1} a_{n-k} t^{n-k}.$$

Using $\binom{n-1}{k-1} = \binom{n-1}{n-k}$ and writing k instead of $n-k$ we get

$$x = c - \sum_{k=1}^{n-1} (-1)^k \binom{n-1}{k} a_k t^k. \quad (11)$$

Its derivative with respect to t is

$$\begin{aligned} \frac{dx}{dt} &= - \sum_{k=1}^{n-1} (-1)^k k \binom{n-1}{k} a_k t^{k-1} \\ &= - (n-1) \sum_{k=1}^{n-1} (-1)^k \binom{n-2}{k-1} a_k t^{k-1} \end{aligned} \quad (12)$$

so the cusps of B_n (if any) satisfy the cusp-equation

$$\sum_{k=1}^{n-1} (-1)^k \binom{n-2}{k-1} a_k t^{k-1} = 0. \quad (13)$$

For a parameter value t_1 the first osculant curve is given by the line-equation that results from taking $t_2 = \dots = t_n = t$ in the polarized equation (10):

$$x - c + \sum_{k=1}^{n-1} (-1)^k \left\{ \binom{n-1}{k} t^k + \binom{n-1}{k-1} t_1 t^{k-1} \right\} a_k + (-1)^n (\overline{x-c}) t_1 t^{n-1} = 0. \quad (14)$$

Its map-equation, obtained in the usual way by differentiation, conjugation, solving for x and writing k instead of $n-k$, is

$$x = c - \sum_{k=1}^{n-1} (-1)^k \left\{ \binom{n-2}{k-1} t_1 t^{k-1} + \binom{n-2}{k} t^k \right\} a_k. \quad (15)$$

This is a curve of type B_{n-1} with center $x = c + a_1 t_1$, so all centers of the first osculants are on the *centric circle*, the circle with map-equation

$$x = c + a_1 t.$$

For $t = t_1$, both the line-equations (9) of the curve B_n and (14) of the first osculant curve yield the same line, while also the map-equations (11) and (15) yield the same point, so at this point the two curves

touch. But, as we have seen already in the case of the cardioid, *the first osculant also passes through the cusps (if any) of B_n* . To prove this, we write the map-equation (11) of B_n in the form

$$x = c - \sum_{k=1}^{n-1} (-1)^k \left\{ \binom{n-2}{k-1} + \binom{n-2}{k} \right\} a_k t^k.$$

For any t_c satisfying the cusp-equation (13), both the map-equations (11) of B_n and (15) of the osculant curve yield

$$x = c - \sum_{k=1}^{n-1} (-1)^k \binom{n-2}{k} a_k t_c^k$$

as desired.

The higher osculants yield more complicated formulas, but at this point we can collect a general result, starting from the bottom end, the osculant lines given by fully polarized equations like (10). Let a curve B_n and $n+1$ parameter values $t_1, \dots, t_{n+1}$ be given. Any n of these determine an osculant line, $n+1$ lines in total. Any $n-1$ parameter values determine an osculant point, the intersection of two osculant lines. Any $n-2$ parameter values determine an osculant circle, the circumscribed circle of three osculant points, which is also the circumscribed circle of the triangle formed by the corresponding three osculant lines. Any $n-3$ parameter values determine an osculant cardioid, with four osculant circles through its cusp. And so on.

Thus, a curve B_n and $n+1$ parameter values determine $n+1$ lines with lots of interesting properties. In a later section we will show that the situation with respect to the lines is completely general, since we will prove that *any $n+1$ lines, no two parallel, determine a unique curve B_n for which they are osculant lines*. But first we will study the osculants of curves B_4 in more detail, since these contain interesting special cases.

Osculants of a curve B_4

Let a curve B_4 be given. Without loss of generality, we may assume that its center is $c = 0$. Then its line-equation is

$$B_4: x - 4at + 6\mu t^2 - 4\overline{a}t^3 + \overline{x}t^4 = 0 \quad (16)$$

with real μ . Its map-equation is

$$x = 3at - 3\mu t^2 + \overline{a}t^3. \quad (17)$$

The cusp-equation $dx/dt = 0$ yields

$$a - 2\mu t + \overline{a}t^2 = 0. \quad (18)$$

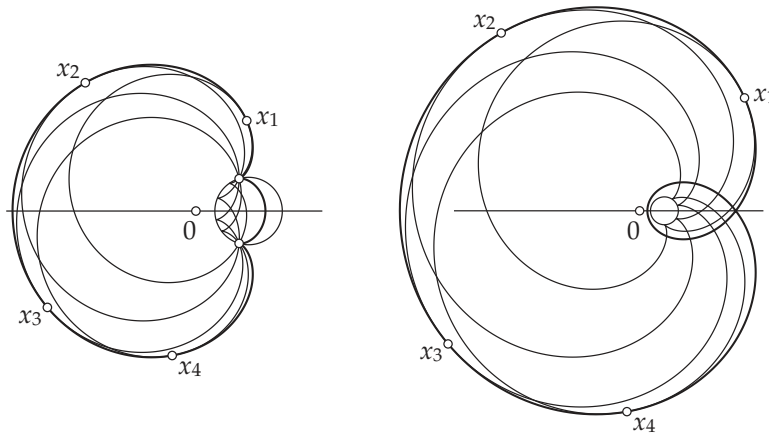


Figure 7 Curves B_4 with and without cusps, together with four osculant cardioids and the cuspidal circle.

If the discriminant $\mu^2 - a\bar{a}$ is negative or zero, the solutions of the cusp-equation are turns and B_4 has two (possibly coinciding) cusps.

For any four parameter values t_1, t_2, t_3, t_4 , the polarized form

$$x - a(t_1 + t_2 + t_3 + t_4) + \mu(t_1 t_2 + \dots) - \bar{a}(t_1 t_2 t_3 + \dots) + \bar{x} t_1 t_2 t_3 t_4 = 0$$

represents an osculant line L_{1234} . It contains the four osculant points $x_{123}, x_{124}, x_{134}, x_{234}$, where, e.g.,

$$x_{123} = a(t_1 + t_2 + t_3) - \mu(t_1 t_2 + t_2 t_3 + t_3 t_1) + \bar{a} t_1 t_2 t_3.$$

Any two parameter values determine an osculant circle like K_{12} , given by the map-equation

$$K_{12}: x = a(t_1 + t_2 + t) - \mu t_1 t_2 - \mu(t_1 + t_2)t + \bar{a} t_1 t_2 t.$$

Its center is $m_{12} = a(t_1 + t_2) - \mu t_1 t_2$ and its radius is $|a - \mu(t_1 + t_2) + \bar{a} t_1 t_2|$. The circle K_{12} contains the osculant points x_{123} and x_{124} and passes through the cusps of the osculant cardioids C_1 and C_2 , where, e.g., C_1 is given by the map-equation

$$C_1: x = at_1 + 2(a - \mu t_1)t + (\bar{a} t_1 - \mu)t^2. \quad (19)$$

As we have proved above, the cardioid C_1 passes through each of the cusps of B_4 (if any).

The center of C_1 is $x = at_1$ and the centric circle (the circle containing all centers of osculant cardioids) thus has map-equation $x = at$.

The parameter value t_c of the cusp of C_1 is the root of its cusp-equation

$$a - \mu(t_1 + t) + \bar{a} t_1 t = 0. \quad (20)$$

We will show now that if $\mu \neq 0$, the cusps of the osculant cardioids, together with

the cusps of B_4 (if any), are on a circle, the so-called cuspidal circle, given by the map-equation

$$x = (1/\mu)(a^2 + (a\bar{a} - \mu^2)t). \quad (21)$$

To prove this, note that for $\mu \neq 0$, the map-equation (19) of C_1 can be written as

$$\mu x = (\mu t - a)(a - \mu(t_1 + t) + \bar{a} t_1 t) + a^2 + (a\bar{a} - \mu^2)t_1 t$$

so if t_c is the root of the cusp-equation (20) of C_1 , then

$$\mu x = a^2 + (a\bar{a} - \mu^2)t_1 t_c$$

which, indeed, defines a point on the cuspidal circle (21).

It might happen that t_1 is a root of the cusp equation (18) of B_4 . Then, manifestly, t_1 is also the root of the cusp-equation (20) of the osculant cardioid C_1 . The cusps of B_4 (if any) therefore also occur as cusps of osculant cardioids, so the cuspidal circle also passes through the cusps of B_4 . See Figure 7, where curves B_4 are shown with $a = 1$ and, respectively, $\mu = 0.6$ (two cusps) and $\mu = 1.25$ (no cusps). In each case, four osculant cardioids have been drawn together with the cuspidal circle.

Morley's five circles theorem

It might happen that in the line-equation (16) of the curve B_4 we have $a = 0$, so that the map-equation of B_4 reduces to

$$x = -3\mu t^2$$

where μ is a real number. This is a circle described twice. If $\mu = 0$, then B_4 degenerates into a point. Leaving this case aside, we may suppose without loss of generality that $\mu = 1$ in which case B_4 is a circle with radius 3 described twice.

The first osculant C_1 then is the cardioid with map-equation

$$x = -2t_1 t - t^2$$

(cf. equation (19)). Its center is the origin, its cusp is $c_1 = t_1^2$, so the cuspidal circle is the unit circle. In Figure 8 we see B_4 with five osculant cardioids at points $x_1, \dots, x_5$, respectively, and the cuspidal circle.

The second osculant K_{12} is the circle with map-equation

$$x = -(t_1 + t_2)t - t_1 t_2.$$

It has center $-t_1 t_2$ and radius $|t_1 + t_2|$. The center thus is on the unit circle, which is also the cuspidal circle. Furthermore, for $t = -t_1$ we get the cusp t_1^2 of C_1 while $t = -t_2$ yields the cusp t_2^2 of C_2 , so, indeed, the circle K_{12} passes through the cusps of C_1 and C_2 .

The third osculant x_{123} is the point

$$x_{123} = -(t_1 t_2 + t_2 t_3 + t_3 t_1).$$

It is the common point of the osculant circles K_{12}, K_{23} and K_{31} .

For five parameter values $t_1, \dots, t_5$ we have five osculant cardioids $C_1, \dots, C_5$, ten osculant circles $K_{12}, \dots$, each tangent to two osculant cardioids, ten osculant points $x_{123}, \dots$, each on three osculant circles and, finally, five osculant lines $L_{1234}, L_{1235}, L_{1245}, L_{1345}, L_{2345}$, each containing four osculant points. For example, L_{1234} contains the points $x_{123}, x_{124}, x_{134}$ and x_{234} . Figure 9 shows the curve B_4 with $a = 0, \mu = 1$, the cuspidal circle in red, the five osculant cardioids with their cusps in yellow, the ten osculant circles with their centers on the cuspidal circle in green, the ten osculant points and the five osculant lines in blue. The reader is invited to identify the osculant cardioids with their cusps, the osculant circles with their cen-

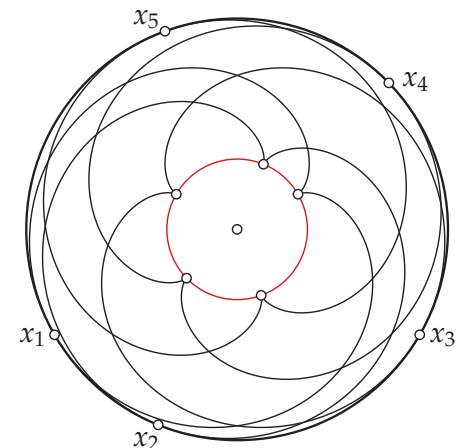


Figure 8 The curve B_4 with $a = 0$ and $\mu = 1$ with five osculant cardioids and the cuspidal circle.

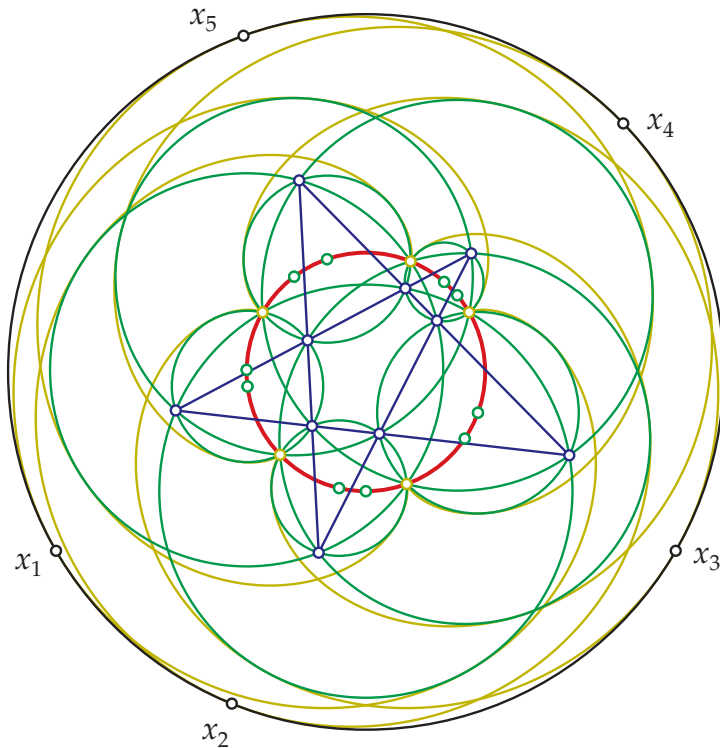


Figure 9 The curve B_4 with $a = 0$, $\mu = 1$ (black) with the cuspidal circle (red). Furthermore, for five parameter values we have drawn the five osculant cardioids with their cusps (yellow), the ten osculant circles with their centers (green), the ten osculant points and the five osculant lines (blue).

ters, the osculant points and the osculant lines.

In Figure 9, the five osculant lines form a pentagon with four osculant points on each (extended) side. For determining the pentagon, it is sufficient to draw the cuspidal circle and only five of the ten osculant circles, as shown in Figure 10. This yields Morley's five circles theorem, as announced in the introduction and illustrated in Figure 2.

Theorem 1 (Morley's five circles theorem).

If five circles are chosen with their centers on a given circle such that each intersects the next on the circle, then the five lines through the other intersection points of adjacent circles intersect again on the respective circles.

Proof. Let the given circle be the unit circle and let $K_{12}, K_{23}, K_{34}, K_{45}, K_{51}$ be the five circles with centers $m_{i,i+1}$ on the unit circle, each circle $K_{i-1,i}$ intersecting the next one $K_{i,i+1}$ on the unit circle in point c_i (indices modulo 5, see Figure 10). Note that $m_{i,i+1}/c_i = c_{i+1}/m_{i,i+1}$, so $m_{i,i+1}^2 = c_i c_{i+1}$.

We will show now that it is possible to choose turns t_1, t_2, t_3, t_4, t_5 such that the five circles $K_{i,i+1}$ are the second osculant curves of the curve

$$B_4: x + 6t^2 + \overline{x}t^4 = 0$$

for consecutive parameter pairs (t_i, t_{i+1}) . Then the centers of these circles should be $m_{12} = -t_1 t_2$, $m_{23} = -t_2 t_3$, $m_{34} = -t_3 t_4$, $m_{45} = -t_4 t_5$, $m_{51} = -t_5 t_1$, respectively, while

$c_1 = t_1^2$, $c_2 = t_2^2$, $c_3 = t_3^2$, $c_4 = t_4^2$, $c_5 = t_5^2$, the point c_i being the cusp of the osculant cardioid C_i of B_4 for parameter value t_i .

To determine the turns t_i from the centers $m_{i,i+1}$ and the intersection points c_i , we choose t_1 as one of the two square roots of c_1 and define $t_2 = -m_{12}/t_1$, $t_3 = -m_{23}/t_2$, $t_4 = -m_{34}/t_3$, $t_5 = -m_{45}/t_4$. It then follows from $c_1 = t_1^2$ and $m_{i,i+1}^2 = c_i c_{i+1}$ that also $c_i = t_i^2$ for $i = 2, 3, 4, 5$, as desired.

With these parameter values t_i it is possible to construct all osculant cardioids, circles, points and lines of B_4 as in Figure 9. Among these osculants, we find the five osculant circles, the ten osculant points and the five osculant lines of Figure 10, with three osculant points on each circle and four osculant points on each osculant line, as indicated above. For instance, circle K_{12} contains the osculant points x_{123} , x_{124} and x_{125} , while line L_{1234} contains x_{123} , x_{124} , x_{134} and x_{234} . This establishes Morley's five circles theorem. $\square$

A curve B_4 with a cuspidal segment

Take a curve B_4 given by (16) with $a = 1$ and $\mu = 0$. Then its map-equation is

$$x = 3t + t^3.$$

The curve B_4 has two cusps, at $x = \pm 2i$, taken for $t = \pm i$. We will prove that in this case the cuspidal circle degenerates into the segment connecting the cusps $\pm 2i$ of B_4 .

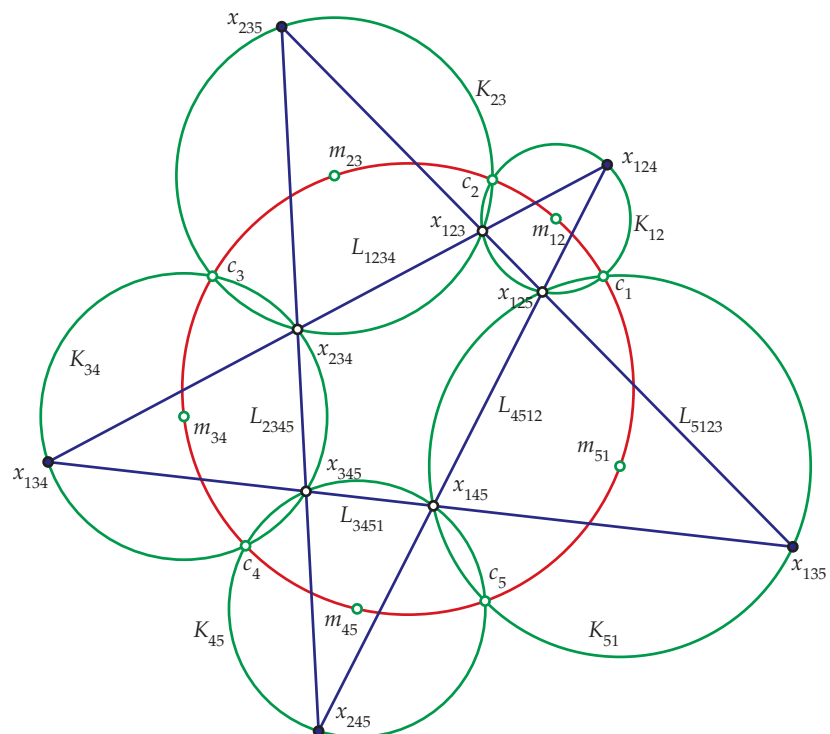


Figure 10 Morley's five circles theorem in relation to Figure 9.

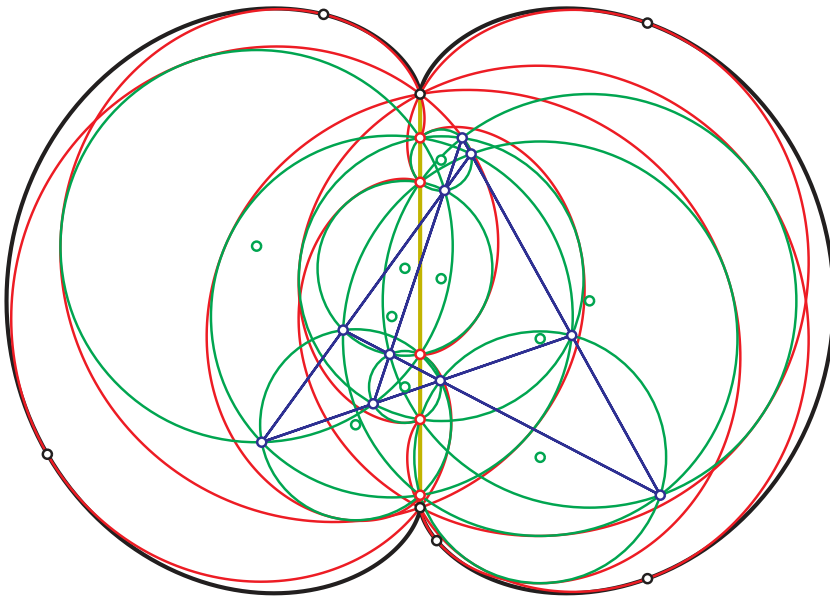


Figure 11 The curve B_4 with $a = 1$, $\mu = 0$, the cuspidal segment (yellow), five osculant cardioids with their cusps (red), the corresponding ten osculant circles with their centers (green), the ten osculant points and the five osculant lines (blue).

Any osculant cardioid C_1 with map-equation

$$x = t_1 + 2t + t_1 t^2$$

passes for $t = \pm i$ through the cusps $\pm 2i$ of B_4 . Furthermore, the cusp of C_1 is obtained from $dx/dt = 0$, which yields $t = -1/t_1$, so the cusp of C_1 is $t_1 - 1/t_1$. This, indeed, is a point on the segment $[-2i, 2i]$. Therefore, the cuspidal segment connects the cusps of B_4 (see Figure 11).

The osculant circle K_{12} with map-equation

$$x = t_1 + t_2 + t + t_1 t_2 t$$

has center $m_{12} = t_1 + t_2$ and radius $|1 + t_1 t_2|$. It contains the cusps of the cardioids C_1 and C_2 .

The osculant point x_{123} , given by

$$x_{123} = t_1 + t_2 + t_3 + t_1 t_2 t_3,$$

is on the osculant circles K_{12} , K_{23} and K_{31} .

The four osculant points x_{123} , x_{124} , x_{134} , x_{234} are on the osculant line L_{1234} , given by the line-equation

$$\begin{aligned} x - (t_1 + t_2 + t_3 + t_4) \\ - (t_1 t_2 t_3 + t_1 t_2 t_4 + t_1 t_3 t_4 + t_2 t_3 t_4) \\ + \overline{x} t_1 t_2 t_3 t_4 = 0 \end{aligned}$$

which is a fully polarized form of the line-equation of B_4 .

In Figure 11 we have drawn the curve B_4 (black) and the cuspidal segment (yellow) and, for five parameter values t_1, t_2, t_3, t_4, t_5 , the osculant cardioids $C_1, \dots$ with their cusps (red), the ten osculant circles $K_{12}, \dots$ with their centers (green), the ten osculant

points $x_{123}, \dots$ and the five osculant lines $L_{1234}, \dots$ (blue).

Curves B_n determined by $n + 1$ lines

Theorem 2. Suppose that $n + 1$ lines, no two parallel, are given. Then there exists a unique curve B_n for which the given lines are osculant lines.

Proof. For clearness we will give the proof for $n = 3$, but in such a way that it is obvious that the general proof for $n \geq 2$ proceeds in a similar way. Note that the case $n = 1$ is trivial: the ‘curve’ B_1 then is the intersection point of the two lines.

Therefore, let four arbitrary lines L_1, L_2, L_3, L_4 be given, no two parallel, and let their clinants be the (distinct) turns $\tau_1, \tau_2, \tau_3, \tau_4$. Without loss of generality, we may assume that $\tau_1 \tau_2 \tau_3 \tau_4 = 1$. Let b_i be the image of the origin upon reflection in line L_i . Then line L_i is given by the self-conjugate equation $x - b_i - \tau_i \overline{x} = 0$, where $b_i = -\tau_i \overline{b_i}$.

For ease of computation, we will work with the turns $t_i = \overline{\tau_i} = \tau_i^{-1}$. Then, in view of $\tau_1 \tau_2 \tau_3 \tau_4 = 1$, the clinants can also be written as $t_2 t_3 t_4, t_1 t_3 t_4, t_1 t_2 t_4, t_1 t_2 t_3$ and the four lines become

$$\begin{aligned} x - b_1 - \overline{x} t_2 t_3 t_4 &= 0, \\ x - b_2 - \overline{x} t_1 t_3 t_4 &= 0, \\ x - b_3 - \overline{x} t_1 t_2 t_4 &= 0, \\ x - b_4 - \overline{x} t_1 t_2 t_3 &= 0, \end{aligned}$$

where

$$\overline{b_i} = -b_i t_i.$$

The general line-equation of a cardioid is

$$x - c - 3at + 3\overline{a}t^2 - (\overline{x} - \overline{c})t^3 = 0$$

and if the four given lines are the osculant lines for parameter values t_1, t_2, t_3, t_4 of this cardioid, then their line-equations can also be written as

$$\begin{aligned} x - c - a(t_2 + t_3 + t_4) + \overline{a}(t_3 t_4 + t_4 t_2 + t_2 t_3) \\ - (\overline{x} - \overline{c})t_2 t_3 t_4 &= 0, \\ x - c - a(t_1 + t_3 + t_4) + \overline{a}(t_3 t_4 + t_4 t_1 + t_1 t_3) \\ - (\overline{x} - \overline{c})t_1 t_3 t_4 &= 0, \\ x - c - a(t_1 + t_2 + t_4) + \overline{a}(t_2 t_4 + t_4 t_1 + t_1 t_2) \\ - (\overline{x} - \overline{c})t_1 t_2 t_4 &= 0, \\ x - c - a(t_1 + t_2 + t_3) + \overline{a}(t_2 t_3 + t_3 t_1 + t_1 t_2) \\ - (\overline{x} - \overline{c})t_1 t_2 t_3 &= 0. \end{aligned}$$

Therefore, b_1, b_2, b_3 and b_4 should satisfy

$$\begin{aligned} b_1 &= c + a(t_2 + t_3 + t_4) - \overline{a}(t_3 t_4 + t_4 t_2 + t_2 t_3) \\ &\quad - \overline{c} t_2 t_3 t_4, \text{ et cetera.} \end{aligned}$$

In other words, the 4-tuple $(c, a, -\overline{a}, -\overline{c})$ should be a solution of the following system of four equations in the unknown z_1, z_2, z_3, z_4 ,

$$\begin{aligned} b_1 &= z_1 + z_2(t_2 + t_3 + t_4) \\ &\quad + z_3(t_3 t_4 + t_4 t_2 + t_2 t_3) + z_4 t_2 t_3 t_4, \\ b_2 &= z_1 + z_2(t_1 + t_3 + t_4) \\ &\quad + z_3(t_3 t_4 + t_4 t_1 + t_1 t_3) + z_4 t_1 t_3 t_4, \\ b_3 &= z_1 + z_2(t_1 + t_2 + t_4) \\ &\quad + z_3(t_2 t_4 + t_4 t_1 + t_1 t_2) + z_4 t_1 t_2 t_4, \\ b_4 &= z_1 + z_2(t_1 + t_2 + t_3) \\ &\quad + z_3(t_2 t_3 + t_3 t_1 + t_1 t_2) + z_4 t_1 t_2 t_3. \end{aligned}$$

Conjugating this system, using $t_1 t_2 t_3 t_4 = 1$ and $\overline{b_i} = -b_i t_i$, we get the equivalent system

$$\begin{aligned} -b_1 &= \overline{z_1} t_2 t_3 t_4 + \overline{z_2}(t_3 t_4 + t_4 t_2 + t_2 t_3) \\ &\quad + \overline{z_3}(t_2 + t_3 + t_4) + \overline{z_4}, \\ -b_2 &= \overline{z_1} t_1 t_3 t_4 + \overline{z_2}(t_3 t_4 + t_4 t_1 + t_1 t_3) \\ &\quad + \overline{z_3}(t_1 + t_3 + t_4) + \overline{z_4}, \\ -b_3 &= \overline{z_1} t_1 t_2 t_4 + \overline{z_2}(t_2 t_4 + t_4 t_1 + t_1 t_2) \\ &\quad + \overline{z_3}(t_1 + t_2 + t_4) + \overline{z_4}, \\ -b_4 &= \overline{z_1} t_1 t_2 t_3 + \overline{z_2}(t_2 t_3 + t_3 t_1 + t_1 t_2) \\ &\quad + \overline{z_3}(t_1 + t_2 + t_3) + \overline{z_4}, \end{aligned}$$

showing that with any solution (z_1, z_2, z_3, z_4) also $(-\overline{z_4}, -\overline{z_3}, -\overline{z_2}, -\overline{z_1})$ is a solution. In particular, when the system has *only one* solution, then these two solutions must be identical, in other words, then the solution must be of the form $(c, a, -\overline{a}, -\overline{c})$ for certain a and c , as desired. Therefore, we only have to show that the determinant of this system is non-zero. Using the symmetric functions

$$\begin{aligned}\sigma_1 &= t_1 + t_2 + t_3 + t_4, \\ \sigma_2 &= t_1 t_2 + t_1 t_3 + t_1 t_4 + t_2 t_3 + t_2 t_4 + t_3 t_4, \\ \sigma_3 &= t_1 t_2 t_3 + t_1 t_2 t_4 + t_1 t_3 t_4 + t_2 t_3 t_4,\end{aligned}$$

the determinant can be written as

$$|1 \ (\sigma_1 - t_i) \ (\sigma_2 - t_i(\sigma_1 - t_i)) \ (\sigma_3 - t_i(\sigma_2 - t_i(\sigma_1 - t_i)))|.$$

But it is easy to see that this determinant is equal to the Vandermonde determinant

$$|1 \ t_i \ t_i^2 \ t_i^3| = \prod_{k>j} (t_k - t_j)$$

which indeed is non-zero if all t_i are distinct.

This completes the proof that any four lines, no two parallel, determine a unique cardioid for which they are osculant lines. But in a similar way a proof can be given for any $n \geq 2$, the only difference being an extra minus sign in the determinant for even n . $\square$

Curves B_n tangent to $n+1$ lines

Given three lines forming the extended sides of a triangle, there are four circles for which these lines are tangent lines. More generally, we have the following theorem:

Theorem 3. *Let $n \geq 2$ and suppose that $n+1$ lines, no two parallel, are given. Then there are at least one and at most n^n curves of type B_n tangent to each of the given lines.*

Proof. Let the given lines be $L_0, \dots, L_n$ and let the distinct turns $\tau_0, \dots, \tau_n$ be their clinants. If b_i is the image of the origin upon reflection in line L_i , then $b_i = -\tau_i \bar{b}_i$ and the line-equation of line L_i is

$$x - b_i - \tau_i \bar{x} = 0. \quad (22)$$

If these lines are tangent lines to a curve B_n there must be constants $c, a_1, \dots, a_{n-1}$ with $\bar{a}_k = a_{n-k}$ for all k , and parameter values $t_0, \dots, t_n$, such that line L_i is given by the line-equation

$$L_i: \ x - c + \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} a_k t_i^k + (\bar{x} - \bar{c}) (-1)^n t_i^n = 0$$

(cf. the line-equation (9) of a curve of type B_n). This should be the same as (22), so, in the first place, $-\tau_i = (-t_i)^n$ must hold, which leaves us with n choices for each t_i . Once we have chosen such parameter values t_i , the constants c and a_k should also satisfy the system of $n+1$ equations

$$b_i = c - \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} a_k t_i^k + \bar{c} (-1)^n t_i^n$$

with $a_{n-k} = \bar{a}_k$ for all k . In other words, the $(n+1)$ -tuple $(c, a_1, a_2, \dots, \bar{a}_2, \bar{a}_1, \bar{c})$ should be a solution of the system of $n+1$ equations in the unknown $z_0, \dots, z_n$,

$$b_i = z_0 - \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} t_i^k z_k + (-1)^n t_i^n z_n \quad (i = 0, \dots, n). \quad (23)$$

Up to a non-zero constant, the determinant of this system is the Vandermonde determinant

$$|1 \ t_i \ t_i^2 \ \dots \ t_i^n| = \prod_{k>j} (t_k - t_j)$$

which is non-zero since we have assumed that all clinants τ_i , and therefore also all parameter values t_i , are distinct. It follows that the system (23) has a unique solution $(z_0, \dots, z_n)$.

Conjugating the system (23), using $\bar{b}_i = -b_i/\tau_i = b_i/(-t_i)^n$, $\binom{n}{k} = \binom{n}{n-k}$ and multiplying equation i by $(-t_i)^n$ yields the equivalent system

$$b_i = (-t_i)^n \bar{z}_0 - \left(\sum_{k=1}^{n-1} \binom{n}{n-k} (-1)^{n-k} t_i^{n-k} \bar{z}_k \right) + \bar{z}_n$$

which, with the notation $n-k=j$, can be rewritten as

$$b_i = \bar{z}_n - \left(\sum_{j=1}^{n-1} (-1)^j \binom{n}{j} t_i^j \bar{z}_{n-j} \right) + (-1)^n t_i^n \bar{z}_0$$

showing that with each solution $(z_0, \dots, z_n)$ of the system (23) also $(\bar{z}_n, \dots, \bar{z}_0)$ is a solution. But since the solution of system (23) is unique, it must be of the form $(c, a_1, a_2, \dots, \bar{a}_2, \bar{a}_1, \bar{c})$, as desired.

In choosing the parameter values t_i satisfying $-\tau_i = (-t_i)^n$, we have n choices for each t_i . However, since multiplying each t_i by the same number ω^k , where $\omega = e^{2\pi i/n}$, yields the same curve B_n , there are not n^{n+1} , but at most n^n curves B_n tangent to $n+1$ given lines, no two parallel. Note that the actual number may be smaller than n^n , for example if the $n+1$ lines all pass through a common point, say the origin, so $b_i = 0$ for all i . Then $c = 0$ and $a_k = 0$ for all k and the only curve of type B_n that is tangent to all $n+1$ lines is the degenerated 'curve' given by the line-equation

$$x + \bar{x}(-1)^n t^n = 0$$

which, as a map-equation, is just the origin

itself. Since the system (23) is non-singular, there is always at least one curve of type B_n tangent to all $n+1$ lines. $\square$

The axes of a system of n lines

In the former section we have seen that there are at least one and at most n^n curves of type B_n tangent to each of $n+1$ given lines, no two parallel. However, if we remove one of these lines, there will be an infinitude of inscribed n -curves, i.e., curves of type B_n tangent to each of the n remaining lines.

Theorem 4. *Let $n \geq 2$ and suppose that n lines are given, no two parallel. Then the locus of the centers of the inscribed n -curves consists of a set of at most n^{n-1} lines, occurring in n equispaced directions, with the same number of lines in each direction.*

Morley [3, pp.468–469] and [4, chapter XXI] called these lines the *axes* of the given system of lines. For the main ideas in the following proof, I am indebted to my colleague Henk Pijls of the University of Amsterdam.

Proof. Let $L_1, \dots, L_n$ be the given lines. Add one more line L_0 to the system, not parallel to any of the given lines. Let τ_i be the clinant of L_i and let b_i be the image of the origin upon reflection in line L_i ($i = 0, 1, \dots, n$). For each i , let a turn t_i be chosen satisfying $(-t_i)^n = -\tau_i$. Then, on account of Theorem 3, there is a unique curve γ of type B_n given by a line-equation

$$\Gamma: \ x - c + \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} a_k t_i^k + (-1)^n (\bar{x} - \bar{c}) t_i^n = 0 \quad (24)$$

where $a_{n-k} = \bar{a}_k$ for all k , such that Γ is tangent to line L_i for parameter value t_i ($i = 0, 1, \dots, n$). Line L_i then is given by the line-equation

$$L_i: \ x - c + \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} a_k t_i^k + (-1)^n (\bar{x} - \bar{c}) t_i^n = 0. \quad (25)$$

Since the equation of line L_i is also given by $x - b_i - \tau_i \bar{x} = 0$, the $(n+1)$ -tuple $(c, a_1, \dots, \bar{a}_1, \bar{c})$ is the unique solution of the non-singular system (23) of $n+1$ linear equations in the $n+1$ unknown $z_0, \dots, z_n$, which we repeat here:

$$b_i = z_0 - \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} t_i^k z_k + (-1)^n t_i^n z_n.$$

Omitting line L_0 means omitting the first equation from this system, resulting in an $n \times (n+1)$ system of rank n . With the notation

$$p(t, z_0, \dots, z_n) = z_0 - \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} z_1 t^k + (-1)^n z_n t^n \quad (26)$$

this system may be written as

$$p(t_i, z_0, \dots, z_n) = b_i \quad (i = 1, \dots, n). \quad (27)$$

The solution set of this system is complex one-dimensional. The $(n+1)$ -tuple $(z_0, z_1, \dots, z_{n-1}, z_n) = (c, a_1, \dots, a_{n-1}, \bar{c})$ is a particular solution satisfying the additional condition

$$z_{n-k} = \bar{z}_k \quad (k = 0, \dots, n).$$

Let $(v_0, \dots, v_n)$ be any other solution of (27) satisfying

$$v_{n-k} = \bar{v}_k \quad (k = 0, \dots, n).$$

Then the difference

$$(w_0, \dots, w_n) = (c, a_1, \dots, \bar{a}_1, \bar{c}) - (v_0, v_1, \dots, v_{n-1}, v_n)$$

is a solution of the homogeneous system

$$p(t_i, w_0, \dots, w_n) = 0 \quad (i = 1, \dots, n) \quad (28)$$

satisfying the additional condition $w_{n-k} = \bar{w}_k$ for $k = 0, \dots, n$.

For any solution $(w_0, \dots, w_n)$ of (28), the function $p(t, w_0, \dots, w_n)$ is a polynomial in t of degree n with zeros $t = t_1, \dots, t = t_n$, so for some $\lambda \neq 0$,

$$\begin{aligned} p(t, w_0, \dots, w_n) &= \lambda(t - t_1) \cdots (t - t_n) \\ &= \lambda(t^n - \sigma_1 t^{n-1} \\ &\quad + \cdots + (-1)^n \sigma_n) \end{aligned}$$

must hold, where the σ_k are the symmetric functions defined by

$$\begin{aligned} \sigma_1 &= t_1 + t_2 + \cdots + t_n, \\ \sigma_2 &= t_1 t_2 + \cdots + t_{n-1} t_n, \\ \sigma_3 &= t_1 t_2 t_3 + \cdots + t_{n-2} t_{n-1} t_n, \\ &\vdots \\ \sigma_n &= t_1 t_2 \cdots t_n. \end{aligned}$$

Note that $\sigma_{n-k} = \bar{\sigma}_k \sigma_n$ for all $k = 1, \dots, n-1$. Equating coefficients yields

$$\begin{aligned} w_0 &= \lambda(-1)^n \sigma_n, \\ -(-1)^1 \binom{n}{1} w_1 &= \lambda(-1)^{n-1} \sigma_{n-1}, \\ -(-1)^2 \binom{n}{2} w_2 &= \lambda(-1)^{n-2} \sigma_{n-2}, \\ &\vdots \\ -(-1)^{n-1} \binom{n}{n-1} w_{n-1} &= \lambda(-1) \sigma_1, \\ (-1)^n w_n &= \lambda, \end{aligned}$$

so

$$\begin{aligned} w_0 &= \sigma_n w_n \text{ and} \\ w_k &= -(-1)^n \lambda \sigma_{n-k} \binom{n}{k}^{-1} \\ &= -(-1)^n \lambda \bar{\sigma}_k \sigma_n \binom{n}{k}^{-1} \end{aligned}$$

for $k = 1, \dots, n-1$. Hence, if and only if λ is chosen such that

$$\bar{\lambda} = \lambda \sigma_n \quad (29)$$

we have

$$\begin{aligned} \bar{w}_0 &= w_n \text{ and} \\ \bar{w}_k &= -(-1)^n \bar{\lambda} \sigma_k \bar{\sigma}_n \binom{n}{k}^{-1} \\ &= -(-1)^n \lambda \sigma_k \binom{n}{n-k}^{-1} \\ &= w_{n-k}. \end{aligned}$$

Equation (29) is a self-conjugate equation defining a line through the origin in the Argand plane. For any λ on this line, let

$$\begin{aligned} c' &= c + w_0 = c + \lambda(-1)^n \sigma_n, \\ a'_k &= a_k + w_k = a_k - \lambda(-1)^n \sigma_{n-k} \binom{n}{k}^{-1}. \end{aligned}$$

Then $a'_{n-k} = \bar{a}'_k$ for all $k = 1, \dots, n-1$, so the self-conjugate equation

$$\begin{aligned} x - c' + \sum_{k=1}^{n-1} (-1)^k \binom{n}{k} a'_k t^k \\ + (-1)^n (\bar{x} - \bar{c}') t^n = 0 \end{aligned}$$

defines an n -curve Γ_λ , which for the parameter values $t_1, \dots, t_n$ is tangent to the lines $L_1, \dots, L_n$, respectively.

The locus of the centers c' of these inscribed curves Γ_λ consists of the line through c with clinant

$$\frac{c' - c}{\bar{c}' - \bar{c}} = \frac{\lambda(-1)^n \sigma_n}{\bar{\lambda}(-1)^n \bar{\sigma}_n} = \sigma_n.$$

This line, with line-equation

$$L: x - \sigma_n \bar{x} = c - \sigma_n \bar{c}$$

is called an *axis* of the system $L_1, \dots, L_n$ of n lines. The right-hand side of this equation equals the image of the origin upon reflection in L (cf. equation (3)). Therefore, while c depends on the lines $L_0, L_1, \dots, L_n$, the expression $c - \sigma_n \bar{c}$ only depends on the lines $L_1, \dots, L_n$ (and the chosen tangent parameter values $t_1, \dots, t_n$). Morley ([3], p. 468) showed that

$$c - \sigma_n \bar{c} = (-1)^n \sum \frac{b_1 t_2 t_3 \cdots t_n}{(t_2 - t_1) \cdots (t_n - t_1)}$$

where in the summation the indices $1, \dots, n$ should be permuted cyclically. We leave it as a challenge to the reader to verify this result.

It follows that the axis L in terms of the images b_i of the origin upon reflection in the lines L_i and the chosen tangent parameter values t_i , is given by the self-conjugate equation

$$L: x - \sigma_n \bar{x} = (-1)^n \sum \frac{b_1 t_2 t_3 \cdots t_n}{(t_2 - t_1) \cdots (t_n - t_1)}. \quad (30)$$

There are n choices for each of the parameter values $t_i = -\sqrt[n]{-\tau_i}$ for $i = 1, \dots, n$, each yielding one axis, so at first sight there are n^n axes at most. However, multiplying each t_i by the same factor ω^k , where $\omega = e^{2\pi i/n}$, yields the same axis, so there are at most n^{n-1} axes. The axes constitute the locus of the centers of inscribed n -curves.

Since the clinant of an axis equals $\sigma_n = t_1 \cdots t_n = \sqrt[n]{\tau_1 \cdots \tau_n}$, the axes occur in n equispaced directions. Furthermore, multiplying one parameter value t_i by the factor ω^k for $k = 0, \dots, n-1$ while keeping the other parameter values t_j the same, yields n distinct clinants, so in each direction the number of axes is the same. $\square$

For $n \geq 3$ it is possible that one or more of the lines $L_1, \dots, L_n$ is a *double tangent line* of an inscribed n -curve Γ (cf. Figures 3 and 4). For instance, if L_1 is a double tangent line then there are two distinct parameter values t_1 and t'_1 for which L_1 is tangent to Γ . Then, necessarily, these parameter values must differ by some factor $\omega^k \neq 1$, since $(-t_1)^n = (-t'_1)^n = -\tau_1$. The center c of Γ then is the intersection of two axes with clinants $\sigma_n = t_1 t_2 \cdots t_n$ and $\sigma'_n = t'_1 t_2 \cdots t_n$. Therefore, in general, the center of an inscribed n -curve with multiple tangent lines among the given lines $L_1, \dots, L_n$ is the intersection of at least two axes.

Morley seems to claim that the converse is also true: if two axes intersect in a point c , then c is the center of an inscribed n -curve for which (at least) one of the lines $L_1, \dots, L_n$ is a multiple tangent line. He writes [3, p.469]: *The curves C^m which touch n lines fall then into n^{n-1} discrete systems. The transition from one system to another is when the center falls on two axes. One of the n lines then is a double line of C^m .* (Morley denotes n -curves by C^m .) For $n = 3$ this is true; it follows, e.g., from the results in Van de Craats and Brinkhuis [1]. But, as Henk Pijls pointed out, a proof for $n > 3$ still is lacking. Morley apparently assumes that the center determines the inscribed n -curve uniquely. Then, indeed, it

Morley, Miquel and Jiang Zemin

Theorem 1, which I termed ‘Morley’s five circles theorem’, is sometimes simply referred to as ‘The five circles theorem’. Although I didn’t make a deep search on its origin, I decided to name it after Morley, since it figures prominently on page 265 of his *Inversive Geometry* [4].

There are other related ‘five circles theorems’. Probably the oldest one was published in 1838 by Auguste Miquel (1816–1851) in his article ‘Théorèmes de géométrie’. In a slightly adapted notation, an English translation of his five circles theorem reads as follows:

Theorem III. *Let be given a pentagon $A_1A_2A_3A_4A_5$ with sides that are elongated to their mutual intersection points $B_1 = A_5A_1 \times A_2A_3$, $B_2 = A_1A_2 \times A_3A_4$, ..., Construct the circumscribed circles of the five triangles $A_1B_1A_2$, $A_2B_2A_3$, ..., formed by a side of the pentagon and the adjacent elongated sides. Then I say that the five other intersections of consecutive circles are concyclic. [2, p. 486].*

At first sight, this seems to be a kind of converse to Morley’s five circles theorem, but Miquel’s theorem doesn’t mention the centers of the five circles. Although some internet sources claim that these centers necessarily must lie on the Miquel circle, it is easily verified by counterexamples that this is not true in general. Recently, Miquel’s five circles theorem attracted attention when in 1999 Jiang Zemin, at that time the president of the People’s Republic of China, reportedly put it as a challenge to students during a visit to Macau, and also in 2002, when he attended the International Congress of Mathematicians in Beijing.

would follow that the intersection of two axes is the center of an inscribed n -curve with a multiple tangent line among the given lines. However, we doubt whether this assumption is true in general.

For $n = 2$ the inscribed 2-curves are circles touching two given intersecting lines

L_1 and L_2 , and the $2^{2-1} = 2$ axes are the two angle bisectors.

For $n = 3$ the inscribed 3-curves are cardioids touching three given lines, which leads us back to Morley’s famous trisector theorem. If the three lines are the extended sides of a triangle, then there are $3^{3-1} = 9$

axes, 3 axes in each of 3 equispaced directions, see [1] for more details on this case. If the three given lines are concurrent, then there are only 3 axes, with clinants that are the three geometric means of the clinants of the given lines, each axis passing through the common point of the given lines.

This brings to an end our introduction to Morley’s treatment of curves of type B_n . The interested reader is referred to Morley’s article [3] and his book [4] for other intriguing and challenging results in this field. Let us finish this paper by quoting the following lines from Morley’s 1929 article ([3], p. 469):

“If we apply the theory of this section to a triangle abc , we obtain as the locus of the centers of inscribed cardioids three sets of three parallel lines, forming equilateral triangles. The vertices of the triangles are the centers of the cardioids which touch a side (say bc) of the given triangle twice. If x_0 is such a center, then the angle x_0bc is a third of the angle abc . For x_0b is an axis of the 3 lines ab and bc twice. Thus if we take the interior trisectors of the angles, the points where those adjacent to a side meet form an equilateral triangle.”

In a footnote he adds: “This theorem, which I obtained in this way long ago, has excited much interest.” ☞

References

- 1 Jan van de Craats and Jan Brinkhuis, Cardioids and Morley’s trisector theorem, *Nieuw Archief voor Wiskunde* 5/18(1) (2017), 25–33.
- 2 A. Miquel, Théorèmes de géométrie, *J. de mathématiques pures et appliquées* 1 (1838), 485–487. <http://gallica.bnf.fr/ark:/12148/bpt6k1638m/f494.image>
- 3 F. Morley, Extensions of Clifford’s chain-theorem, *Amer. J. of Math.* 51 (1929), 465–472.
- 4 Frank Morley and F.V. Morley, *Inversive Geometry*, Ginn, Boston, 1933. (Reissued by Chelsea, 1954 and by Dover Books on Mathematics, 2013.)

Recommended internet page:

Alexander Bogomolny, *Morley’s Miracle* from Interactive Mathematics Miscellany and Puzzles, <http://www.cut-the-knot.org/triangle/Morley/index.shtml>.

Casper Albers

Psychometrie & Statistiek
Rijksuniversiteit Groningen
c.j.albers@rug.nl



Column Casper grijpt een kans

Column van het jaar

December is een speciale maand. Het is de maand van Sinterklaas, Kerst en de Oud van Oud & Nieuw. Het is ook de tijd van het jaar voor eindejaarslijstjes. Casper rekent u voor wat er mis kan gaan in dit soort lijstjes.

Wanneer stemmen?

Eindejaarslijstjes als ‘film van het jaar’, ‘ondernemer van het jaar’ en ‘speelgoed van het jaar’ zijn een jaarlijks terugkerend fenomeen. Het is een beetje raar dat dat soort lijstjes al verschijnen voordat het jaar voorbij is. Bij een voetbalwedstrijd wijs je ook niet in de tachtigste minuut al de winnaar aan.

Soms lijkt dat niet zo’n probleem te zijn. Zo las ik eind september een nieuwsbericht waarin stond dat Luis Fonsi’s ‘Desposito’ officieel verkozen was tot zomerhit van het jaar. (Welke instantie de bevoegdheid heeft zulke ‘officiële’ besluiten te maken, weet ik niet, maar dat is geen discussie voor een wiskundetijdschrift.) Het klinkt acceptabel om te bepalen wat de zomerhit was zodra het herfst is. Het wordt al iets discutabeler als je weet dat dat nummer al op 13 januari (in de winter dus) werd uitgebracht en in mei (in de lente dus) al op de nummer 1-positie kwam. Een ander voorbeeld waar het wel kan: op 23 oktober werd Lieke Martens door de FIFA tot beste voetbalster van de wereld gekozen. Aangezien de belangrijkste voetbalwedstrijden toen al achter de rug waren, was dat een veilige timing.

Vaak is het echter wel een probleem. Zo werd op 28 september Ingrid de Graaf van Aegon tot ‘topvrouw van het jaar’ gekozen [1]. Met nog 26% van het jaar te gaan een gewaagde timing, al zal het volgens de vakjury vast niet waarschijnlijk zijn geweest dat De Graaf in de laatste maanden iets doet waardoor ze opeens geen goed bestuurder meer is.

Anders is het bij verkiezingen waarbij je pas aan het eind van het jaar echt weet wat de kandidaten zijn. Zo organiseert de Van Dale jaarlijks de ‘Woord van het Jaar’-verkiezing [2]. Hier bepaalt de

bezoeker van de website van de Van Dale welk woord deze prestigieuze trofee op de schoorsteenmantel mag zetten. Klein probleem: de verkiezingsperiode begon begin november. Om in aanmerking te komen, moest een woord dus ergens tussen januari en oktober ontstaan zijn. Als we er, bij gebrek aan meer kennis, van uitgaan dat nieuwe woorden op willekeurige momenten in het jaar hun intrede doen, betekent dat dus dat de kans dat het beste woord van het jaar pas in november of december ontstaat ongeveer $\frac{2}{12} = \frac{1}{6}$ is. Gemiddeld eens in de zes jaar is het ‘Woord van het Jaar’ dus eigenlijk op z’n best het ‘Een-na-beste Woord van het Jaar’.

Omdat de Van Dale niet alleen de winnaar maar de gehele top 3 bekendmaakt, wordt de kans nog groter dat daar een woord uit november/december in mist. Die kans is $1 - (1 - \frac{1}{6})^3 = \frac{91}{216} \approx 42\%$. Zou de Van Dale de top 4 bekendmaken, dan is de kans groter dat dit verkeerd is dan dat dit juist is.

Meerdere verkiezingen

Schrale troost is dat ook het Genootschap Onze Taal (GOT) jaarlijks bekendmaakt wat het Woord van het Jaar is. Hoewel ook deze prijs voor het einde van het jaar wordt bepaald, is het interessant om te zien dat bij beide instanties vaak een ander woord de trofee in de wacht sleept. Vorig jaar won bij Van Dale het woord *treitervlogger* met 35% van de stemmen [3]. Bij het GOT won het woord *brexite* en kreeg *treitervlogger* slechts 8% van de stemmen [4]. Dit verschil kan door twee oorzaken komen: de populatie van potentiële Van Dale-stemmers wijkt af van de populatie van potentiële GOT-stemmers, of er was toeval. Omdat dit een statistiekcolumn is, gaan we dit toetsen en we beperken ons voor het gemak tot het woord *treitervlogger*. Formeel is de toetssituatie als volgt:

Laat π_1 en π_2 de proporties *treitervlogger*-liefhebbers zijn onder de populaties van potentiële stemmers bij Van Dale (1) en het Genootschap Onze Taal (2), en laat $p_1 = 0,35$ en $p_2 = 0,08$ de geobserveerde proporties zijn. De nulhypothese $H_0: \pi_1 = \pi_2$ versus

tweezijdig alternatief kan getoetst worden met een Z -toets met

$$Z = \frac{p_1 - p_2}{\sqrt{p^*(1-p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

waarbij $p^* = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$ een gewogen gemiddelde van beide proporties is. Als $|Z| > 1,96$, dan verwerpen we H_0 (indien we op een niveau van 5% toetsen). Een interessante complicatie is dat we hier niet weten hoe groot n_1 en n_2 zijn: het persbericht van Van Dale meldt alleen dat $n_1 > 100000$ en het bericht van het GOT zegt niets over het aantal stemmen. We weten alleen dat het percentage p_1 afgerond 8% is. Dit kan voor $n_2 \in \{12, 13, 24, 25, 26, 36, \dots\}$. Voor zowel n_1 als n_2 zijn wel ondergrenzen te vinden (waarbij het wat onrealistisch is om aan te nemen dat $n_2 = 12$) en hoe groter n_1 en n_2 , hoe groter Z . De p -waarde bij $n_1 = 100000$ en $n_2 = 12$ is gelijk aan $p = 0,049895$, nét kleiner dan de 5%-grens. Bij een $n_2 > 100$ geldt al dat $p < 10^{-7}$, dus we hebben hier wel een significant verschil te pakken: Van Dale-stemmers stemmen echt anders dan GOT-stemmers.

Hoe de stemmen te tellen

Maar zo'n verkiezing van X van het jaar is zo makkelijk nog niet. Hoe bepaal je de winnaar? Zowel Van Dale als GOT vragen stemmers om hun nummer 1-keuze, en kijken welk woord het vaakst gestemd wordt: dat woord wint. (En dan maar hopen dat november en december saai maanden zijn qua nieuw vocabulaire.) Het kan ook anders: je kan bijvoorbeeld iedereen vragen om een top 3 in te sturen en dan turven wat het meest genoemd wordt. In Tabel 1 staat een virtueel voorbeeld: Alice, Bob en Carol hebben een rangschikking gemaakt van vier woorden (de top 3 van Van Dale en de winnaar bij GOT). Zou je alleen de nummers 1 tellen, dan wint Brexit van Trumpisme, maar zou je top 3's tellen, dan zijn de verhoudingen omgedraaid: Trumpisme staat in elke top 3, Brexit maar in eentje. De gemiddelde ranking van Trumpisme, $2\frac{1}{3}$, is ook beter dan die van Brexit, 3.

Het maakt dus nogal uit of je alleen de kopposities meet of ook de nummers $2, 3, \dots, I$ voor een zekere I . Een omslachtige manier om dit op te lossen is om meerdere stemrondes te houden: als je N kandidaten hebt voor de hoofdprijs, kan je $N-1$ stemrondes houden. Bij elke ronde laat je dan de kandidaat met de minste steun wegvallen. Dit is vooral nuttig wanneer kandidaten in elkaars vaarwater zitten: als bijvoorbeeld bij presidentsverkiezingen een linkse kandidaat wegvalt, zal de steun voor een andere linkse kandidaat doorgaans toenemen. Een andere manier om via

Woord	Alice	Bob	Carol
Treitervlogger	1	3	3
Pokémonterreur	2	1	4
Trumpisme	3	2	2
Brexit	4	4	1

Tabel 1 Rankings gegeven aan vier woorden uit 2016.

meerdere stemrondes de winnaar te bepalen is via een knock-out-systeem waarbij telkens verkiezingen zijn tussen twee kandidaten en de winnaar doorgaat naar de volgende ronde. Ook dan heb je aan $N-1$ stemrondes voldoende.

Condorcet-paradox

Bij stemprocedures gebaseerd op rangschikkingen kan een interessante paradox optreden, vernoemd naar de markies van Condorcet die deze paradox al in de achttiende eeuw omschreef. Uit Tabel 1 is te af te lezen dat een (tweederde) meerderheid van de stemmers Treitervlogger een beter woord vindt dan Pokémonterreur. Tevens vindt een (tweederde) meerderheid dat Pokémonterreur een beter woord is dan Trumpisme. Intuïtief zou je verwachten dat er dus een meerderheid is die Treitervlogger verkiest boven Trumpisme. Dit is echter niet het geval, er zit een soort lus in de voorkeuren. Welk van deze drie woorden ook tot winnaar wordt uitgeroepen: er is een ander woord dat meer steun van de stemmers had!

Zo'n lus kan niet altijd optreden. Als 100% van de stemmers A prefereert boven B en B boven C, dan is het onmogelijk dat toch een meerderheid C boven A zet. Een voorwaarde voor het hebben van zo'n lus is het volgende: laat p_1 de proportie stemmers zijn die A prefereert boven B, p_2 de proportie met B boven C en p_3 de proportie met C boven A. Door de labels A, B en C te permueren, is het altijd mogelijk om $p_1 \geq \frac{1}{2}$ en $p_2 \geq \frac{1}{2}$ te veronderstellen. Er kan afgeleid worden [5] dat $p_3 \leq 2 - p_1 - p_2$. De paradox treedt op als $p_2 > \frac{1}{2}$ (en de ' $\geq$ '-relaties voor p_1 en p_2 strikte '>'-relaties zijn). Dit impliceert de voorwaarde $p_1 + p_2 < \frac{3}{2}$.

Conclusie

Alles samenvattend: zo makkelijk is het nog niet om een jaarverkiezing te houden. Hoe meer je er over puzzelt, hoe meer puzzels boven water komen. Een eenvoudige basisregel is echter wel om een verkiezing pas te houden nádat het jaar afgelopen is. Daar zit geen hogere wiskunde achter en daar is al winst te halen. ♦

Referenties

- 1 <https://www.nrc.nl/nieuws/2017/09/28/ingrid-de-graaf-is-topvrouw-van-het-jaar-a1575361>
- 2 <http://woordvanhetjaar.vandale.nl>
- 3 <http://www.vandale.nl/woord-van-het-jaar-2016-treitervlogger>
- 4 <https://onzetaal.nl/nieuws-en-dossiers/weblog/brexit-woord-van-het-jaar-bij-onze-taal/>
- 5 C.L. Silver, The voting paradox, *The Mathematical Gazette* 76(477) (1992), 387–388. <http://www.jstor.org/stable/pdf/3618386.pdf>

Koen Groenland

Centrum Wiskunde & Informatica
Amsterdam
k.l.groenland@cwi.nl

Tom Bannink

Centrum Wiskunde & Informatica
Amsterdam
tom.bannink@cwi.nl

Onderzoek

Sneller zoeken met quantum random walks

Het lijkt steeds meer een kwestie van tijd totdat werkende quantumcomputers een feit zijn. Vooruitlopend op dit scenario zijn onderzoekers aan het Centrum voor Wiskunde & Informatica al druk bezig met het ontwerpen van *kwantumalgoritmes*, waarmee we algoritmes bedoelen die op een quantumcomputer kunnen draaien. In dit artikel, dat geschreven is naar aanleiding van een Lorentz Center workshop, leggen Koen Groenland en Tom Bannink uit wat een *quantum random walk* is en laten zien hoe Grovers zoekalgoritme, historisch gezien een van de allereerste kwantumalgoritmes, beschreven kan worden met behulp van zo'n kwantumwandeling.

Kwantumrekenen

We zullen eerst laten zien hoe een quantumcomputer gebruikmaakt van de wonderbaarlijke wetten van de kwantummechanica. Het geheugen van een *klassieke* computer die beschikt over n bits kan $N = 2^n$ unieke configuraties of 'toestanden' aannemen. Een *toestand* kunnen we beschrijven als een reeks van n enen en nullen, zoals '0110011'.

Een *kwantumtoestand* $|\psi\rangle$ is een vector van lengte 1 in de (complexe) vectorruimte $\mathbb{C}^N$ die wordt opgespannen door alle N klassiek toegestane toestanden. Als we elke klassieke toestand een label j geven (een geheel getal tussen 1 en N), dan kunnen we elke kwantumtoestand schrijven als lineaire combinatie van deze klassieke toestanden. Iedere kwantumtoestand $|\psi\rangle$ laat zich dus eenduidig voorstellen als

$$|\psi\rangle = \sum_{j=1}^N \alpha_j |j\rangle \quad (1)$$

met normalisatie $\sum_{j=1}^N |\alpha_j|^2 = 1$.

De coëfficiënten α_j zijn complex. Hier zien we de eerste exotische eigenschap van kwantumtoestanden, namelijk dat ze een *superpositie* van klassieke toestanden zijn. Je zou kunnen zeggen dat een kwantumsysteem zich in meerdere klassieke toestanden tegelijkertijd kan bevinden! We gebruiken de notatie $|v\rangle$ om aan te duiden dat v een kwantumtoestand is. Het symbool $|\cdot\rangle$ noemen we een 'ket'. Om een complex geconjugeerde (en getransponeerde) vector weer te geven, gebruiken we de notatie $\langle v|$, die we een 'bra' noemen. De gedachte achter deze notatie wordt duidelijk wanneer we een inwendig product tussen vectoren v en w opschrijven als de zogenaamde bra(c)ket: $\langle v|w\rangle$. Evenzo kunnen we de matrix die de vector $|w\rangle$ afbeeldt op de vector $|v\rangle$ opschrijven als $|v\rangle\langle w|$. We zullen te maken krijgen met kwantumsystemen die interactie met elkaar hebben. In dat geval beschrijven we de gezamenlijke toestand als een vector in het tensorproduct van de twee afzonderlijke ruimtes. Als bijvoorbeeld de systemen A en B af-

zonderlijk beschreven worden met behulp van de vectorruimten $\mathbb{C}^{N_A}$ en $\mathbb{C}^{N_B}$, dan kunnen we de toestand van het gecombineerde systeem AB beschrijven met behulp van de vectorruimte $\mathbb{C}^{N_A} \otimes \mathbb{C}^{N_B}$. Als we de basisvectoren die $\mathbb{C}^{N_A}$ en $\mathbb{C}^{N_B}$ opspannen aangeven met $|j\rangle_A$ en $|k\rangle_B$, dan wordt $\mathbb{C}^{N_A} \otimes \mathbb{C}^{N_B}$ opgespannen door de (tensor) producten $|j\rangle_A |k\rangle_B$. Een toestandsvector in deze ruimte is dan te schrijven als

$$|\psi\rangle_{AB} = \sum_{j=1}^{N_A} \sum_{k=1}^{N_B} \alpha_{j,k} |j\rangle_A |k\rangle_B \quad (2)$$

met normalisatie $\sum_{j=1}^{N_A} \sum_{k=1}^{N_B} |\alpha_{j,k}|^2 = 1$.

We komen nu een nieuw exotisch fenomeen tegen, de zogenaamde *verstrengeling*. Stel dat kwantumsystemen A en B beide een basis hebben bestaande uit twee toestanden (dat wil zeggen we veronderstellen dat $N_A = N_B = 2$), die we met $|0\rangle_A$ en $|1\rangle_A$, respectievelijk $|0\rangle_B$ en $|1\rangle_B$ zullen aangeven. Zulke systemen noemen we *qubits* ('kwantumbits'). We kunnen een toestand van systeem A dus schrijven in de vorm

$$|\psi\rangle_A = \alpha_0 |0\rangle_A + \alpha_1 |1\rangle_A$$

en net zo voor de toestanden van systeem B . Beschouw nu de volgende toestand van het gecombineerde systeem AB :

$$|\chi\rangle = \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B). \quad (3)$$

Hoe zou je nu de toestand van systeem A omschrijven? En van systeem B ? Een (onjuiste) eerste gedachte zou kunnen zijn dat zij zich in toestand $\frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)$, respectievelijk $\frac{1}{\sqrt{2}}(|0\rangle_B + |1\rangle_B)$ bevinden. Echter, de combinatie van deze twee toestanden levert iets geheel anders: door simpelweg de haakjes uit te werken vinden we immers

$$\begin{aligned} & \frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A) \otimes \frac{1}{\sqrt{2}}(|0\rangle_B + |1\rangle_B) \\ &= \frac{1}{2}(|0\rangle_A |0\rangle_B + |0\rangle_A |1\rangle_B \\ & \quad + |1\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B). \end{aligned}$$

Het blijkt zelfs zo te zijn dat $|\chi\rangle$ onmogelijk te beschrijven is als combinatie $|\chi_1\rangle_A \otimes |\chi_2\rangle_B$ van twee losse qubit-toestanden $|\chi_1\rangle_A$ en $|\chi_2\rangle_B$. Om deze reden noemen we de toestand $|\chi\rangle$ een *verstrengelde toestand*: de systemen A en B zijn als het ware met elkaar ‘verstrengeld’ in deze toestand. Op grond van dit verschijnsel meende Einstein dat de kwantummechanica geen complete beschrijving van de werkelijkheid kan zijn. Immers, de qubits in een verstrengelde toestand zijn van elkaar afhankelijk, zelfs wanneer zij lichtjaren van elkaar verwijderd worden [1]. Inmiddels weten we dat het kwantumformalisme wel degelijk de juiste beschrijving geeft van onze wereld. Voor een kwantumberekening zullen we de kwantumtoestanden willen manipuleren. Dit kan op twee manieren: middels zogenaamde ‘gates’ (schakelingen) en door het verrichten van metingen. De fysisch toegestane gates worden beschreven door unitaire matrices. Dit zijn matrices met complexe coëfficiënten die voldoen aan $U^{-1} = (U^*)^T$ (de inverse is gelijk aan de complex geconjugeerde en getransponeerde van de matrix). Deze matrices hebben de eigenschap dat ze de norm van vectoren behouden. Een gate werkt op een kwantumtoestand volgens de gebruikelijke matrix-vectorvermenigvuldiging. Een voorbeeld van een gate is bijvoorbeeld de fase-gate R_θ die de relatieve fase tussen de coëfficiënten van een qubit-toestand kan veranderen:

$$R_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$

oftewel

$$R_\theta(\alpha_0 |0\rangle + \alpha_1 |1\rangle) = \alpha_0 |0\rangle + e^{i\theta} \alpha_1 |1\rangle.$$

Een ander voorbeeld is de controlled-NOT-gate (CNOT-gate). Deze werkt op twee

qubits als volgt. Wanneer qubit A zich in de toestand $|0\rangle$ bevindt laat CNOT de toestand van qubit B ongemoeid, en wanneer qubit A zich in de toestand $|1\rangle$ bevindt verwisselt CNOT de toestanden 0 en 1 van qubit B :

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

oftewel

$$\begin{aligned} \text{CNOT} |0\rangle_A |0\rangle_B &= |0\rangle_A |0\rangle_B, \\ \text{CNOT} |0\rangle_A |1\rangle_B &= |0\rangle_A |1\rangle_B, \\ \text{CNOT} |1\rangle_A |0\rangle_B &= |1\rangle_A |1\rangle_B, \\ \text{CNOT} |1\rangle_A |1\rangle_B &= |1\rangle_A |0\rangle_B. \end{aligned}$$

Hoewel de CNOT-gate volledig klassiek te beschrijven is, is deze in staat om kwantumverstrengeling te creëren en vormt deze de basis van veel kwantumalgoritmes. Net als bij de klassieke tegenhangers kan met een kleine set kwantumgates elke mogelijke functie f in de vorm van een ‘circuit’ van aaneengesloten kwantumgates worden gebouwd. Het bijzondere aan kwantumcircuits is dat de input kan bestaan uit een lineaire combinatie van alle mogelijke klassieke inputs. Stel bijvoorbeeld dat we een systeem A hebben met basisvectoren $|0\rangle_A, \dots, |N\rangle_A$ dat we koppelen met een systeem B met basisvectoren $|0\rangle_B, \dots, |M\rangle_B$. Zij voorts gegeven een functie f die $\{0, \dots, N-1\}$ afbeeldt naar $\{0, \dots, M-1\}$. Stel nu dat we in staat zijn een kwantumcircuit te maken dat het volgende effect heeft op de kwantumtoestand $|n\rangle_A |0\rangle_B$:

$$|n\rangle_A |0\rangle_B \mapsto |n\rangle_A |f(n)\rangle_B.$$

Met behulp van dit circuit kunnen we de functie f over *alle mogelijke inputs tegelijkertijd* uitrekenen, door simpelweg als input aan het circuit de volgende toestand te geven:

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} |n\rangle_A |0\rangle_B.$$

Het circuit voert deze input over in

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} |n\rangle_A |0\rangle_B = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} |n\rangle_A |f(n)\rangle_B. \quad (4)$$

In deze toestand treden alle waarden van f op. Met andere woorden: het circuit ‘berekent’ alle waarden $f(n)$ in één keer. Laat u echter niet misleiden: het is niet mogelijk om al deze informatie uit de resulterende kwantumtoestand te halen. Zoals we zo

dadelijk zullen beschrijven zal een meting van de kwantumtoestand (4) een van de waarden $f(n)$ geven. De enige manier om op niet-triviale wijze gebruik te maken van het ‘kwantumparallelisme’ inherent in (4) is door verder te rekenen met de gevonden superpositie — dit is precies wat we verderop in Grovers algoritme zullen doen.

Hoe werkt een meting dan precies? Wat er ‘werkelijk’ gebeurt bij een meting is nog steeds onderhevig aan meerdere interpretaties, maar voor de wiskundige beschrijving volstaat het om uit te gaan van de zogenaamde *Born-regel*. Deze zegt dat wanneer we een kwantumtoestand $|\psi\rangle = \sum_j \alpha_j |j\rangle$ meten, we met kans $|\alpha_j|^2$ de klassieke toestand $|j\rangle$ als uitkomst van onze meting zullen vinden. De normalisatie $\sum_j |\alpha_j|^2 = 1$ garandeert dat al deze waarschijnlijkheden optellen tot 1. Zo is de kans op het meten van een specifieke waarde $f(n)$ in de toestand van formule (4) precies $1/(N+1)$. Volgens de standaardinterpretatie van de kwantummechanica (de zogenaamde Kopenhagense interpretatie) verandert bij een meting de toestand van het systeem instantaan als volgt: als j de uitkomst van de meting is, dan gaat het systeem van toestand $|\psi\rangle = \sum_j \alpha_j |j\rangle$ over in toestand $|\psi'\rangle = |j\rangle$. Dit wordt ook wel de ‘ineenstorting’ (collapse) van de kwantumtoestand genoemd. Alle verdere informatie die was opgeslagen in de waarden α_j raakt daarmee verloren. Merk op dat het verrichten van een meting de enige manier is om informatie over de qubit te winnen — er zijn geen manieren om dit te omzeilen. De enige manier om alle waarden α_j precies te achterhalen is door een qubit telkens in dezelfde toestand te prepareren en opnieuw te meten.

Ondanks deze beperkingen blijkt dat we met de spelregels van de kwantummechanica nieuwe algoritmes kunnen maken die sneller zijn dan klassiek mogelijk is. Een voorbeeld hiervan zijn de zogenaamde kwantumwandelingen.

Kwantumwandelingen

De zogeheten *quantum random walk*, of kwantumwandeling, is een model voor een kwantumdeeltje dat (vrij) kan bewegen door een systeem, meestal een graaf. Het beschrijft de positie van het deeltje als functie van de tijd. De tijd t kan hierbij een continue of discrete parameter zijn. De kwantumwandeling wordt soms gezien als de kwantumversie van de klassieke

random walk. Als men bij elke tijdstap de positie van het vrije kwantumdeeltje *meet*, dan vindt men het gedrag van een klassieke random walk terug. We zullen dit illustreren aan de hand van een simpel voorbeeld, de zogeheten ‘Hadamard-wandeling’.

Hadamard-wandeling

De Hadamard-wandeling beschrijft de evolutie van een kwantumdeeltje op de discrete lijn $\mathbb{Z}$ als functie van een discrete tijd-parameter $t \in \mathbb{N}$. Het deeltje wordt op elk tijdstip beschreven door de positie en een extra interne variabele die de waarde $|\uparrow\rangle = |0\rangle$ of $|\downarrow\rangle = |1\rangle$ kan aannemen. De toestand $|x, \uparrow\rangle$ beschrijft een deeltje op positie x met interne toestand $\uparrow$. De volledige toestand van het kwantumdeeltje is een superpositie van zulke basistoestanden:

$$|\psi\rangle = \sum_{x \in \mathbb{Z}} (\alpha_{x\uparrow} |x, \uparrow\rangle + \alpha_{x\downarrow} |x, \downarrow\rangle)$$

met

$$\sum_{x \in \mathbb{Z}} (|\alpha_{x\uparrow}|^2 + |\alpha_{x\downarrow}|^2) = 1.$$

Het meten van deze toestand levert positie x op met kans $|\alpha_{x\uparrow}|^2 + |\alpha_{x\downarrow}|^2$. De evolutie van het deeltje verloopt als volgt. We passen eerst de zogeheten *Hadamard-operator* H toe op de interne toestand van het deeltje. Dit is de unitaire matrix

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

oftewel

$$\begin{aligned} H|\uparrow\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle), \\ H|\downarrow\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle). \end{aligned}$$

De operator H induceert een operator, die we met $1 \otimes H$ aangeven, op de algehele toestand van het deeltje: de identiteit 1 werkt op de positie en H op de interne toestand. Zo is bijvoorbeeld

$$(1 \otimes H)|3, \uparrow\rangle = \frac{1}{\sqrt{2}}(|3, \uparrow\rangle + |3, \downarrow\rangle).$$

Vervolgens passen we de ‘shift-operator’ S toe. Dit is de unitaire operator

$$\begin{aligned} S|x, \uparrow\rangle &= |x+1, \uparrow\rangle, \\ S|x, \downarrow\rangle &= |x-1, \downarrow\rangle. \end{aligned}$$

De tijdsevolutie van het deeltje bestaat uit het herhaaldelijk toepassen van de (unitaire) operator $S \circ (1 \otimes H)$. De operator H wordt in dit verband ook wel de *coin*-ope-

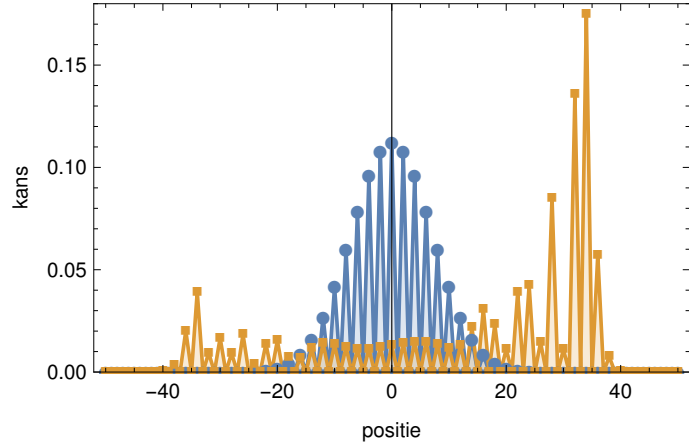
rator genoemd. De interne toestand speelt de rol van een muntje dat wordt ‘opgegooid’ door de coin-operator, en op basis van de uitkomst stuurt S het deeltje een stap naar links of naar rechts. We kunnen nu een begintoestand kiezen, bijvoorbeeld $|\psi\rangle_{(t=0)} = |0, \uparrow\rangle$, een deeltje dat begint in de oorsprong met interne toestand $\uparrow$. Als we de toestand van het systeem na vijftig stappen (vijftig maal de operator $S \circ (1 \otimes H)$ toepassen) bekijken vinden we een kansverdeling die is weergegeven in Figuur 1. Om dit resultaat te begrijpen gaan we stap voor stap kijken wat er met het systeem gebeurt. Neem aan dat het deeltje begint in de toestand $|\psi\rangle_{(t=0)} = |0, \uparrow\rangle$. We passen nu drie keer $S \circ (1 \otimes H)$ toe om $|\psi\rangle_{(t=3)}$ te vinden:

$$\begin{aligned} |\psi\rangle_{(t=0)} &= |0, \uparrow\rangle, \\ |\psi\rangle_{(t=1)} &= \frac{1}{\sqrt{2}}(|1, \uparrow\rangle + |-1, \downarrow\rangle), \\ |\psi\rangle_{(t=2)} &= \frac{1}{2}(|2, \uparrow\rangle + |0, \downarrow\rangle + |0, \uparrow\rangle - |-2, \downarrow\rangle), \\ |\psi\rangle_{(t=3)} &= \frac{1}{2\sqrt{2}}(|3, \uparrow\rangle + |1, \downarrow\rangle + 2|1, \uparrow\rangle \\ &\quad - |-1, \uparrow\rangle + |-3, \downarrow\rangle). \end{aligned}$$

Merk op dat op tijd $t=2$ de termen met positie 0 voorkomen in de specifieke combinatie $|0, \downarrow\rangle + |0, \uparrow\rangle$ (waarbij we de normalisatie even achterwege laten). Als we nu de Hadamard-operator toepassen vinden we

$$(1 \otimes H)(|0, \downarrow\rangle + |0, \uparrow\rangle) = \sqrt{2} |0, \uparrow\rangle,$$

dus de volledige operator $S \circ (1 \otimes H)$ stuurt de toestand $|0, \downarrow\rangle + |0, \uparrow\rangle$ naar $\sqrt{2} |1, \uparrow\rangle$, oftewel alleen maar naar rechts in plaats van een superpositie over links en rechts. De combinatie $|0, \downarrow\rangle - |0, \uparrow\rangle$ (met een minteken) zou echter naar $\sqrt{2} |-1, \downarrow\rangle$ zijn gegaan. Dit



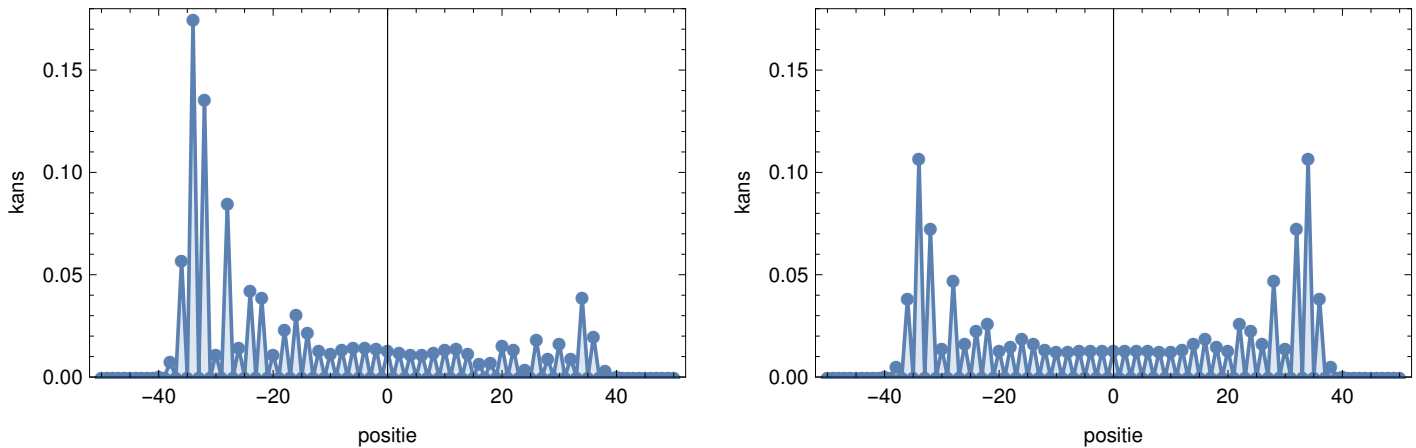
Figuur 1 De kansverdeling na vijftig stappen van een klassieke random wandeling in blauw met rondjes en de kansverdeling na vijftig stappen van de Hadamard-wandeling in oranje met vierkantjes. De begintoestand van de kwantumwandeling was $|0, \uparrow\rangle$.

laat zien dat het resultaat erg afhankelijk is van de interne toestand van het deeltje. Dit is de reden dat het gedrag van de Hadamard-wandeling heel anders is dan dat van de klassieke random walk op $\mathbb{Z}$.

Vergelijking: versus klassiek

We kunnen nu de verbinding leggen met de klassieke random walk. Als men op tijdstip $t=1$ een meting doet, wordt positie 1 of -1 gevonden met kans $1/2$. Na de meting is het systeem in de toestand $|1, \uparrow\rangle$ of $|-1, \downarrow\rangle$ en *niet* langer in een superpositie van die twee. Laten we aannemen dat het resultaat van de meting $|-1, \downarrow\rangle$ was. Als we vervolgens opnieuw een stap doen zal het systeem terecht komen in de toestand $\frac{1}{\sqrt{2}}(|0, \uparrow\rangle - |-2, \downarrow\rangle)$. Meten we nu weer dan zullen we 0 of -2 vinden, beide met kans $1/2$. We zien dat dit precies het gedrag is van de klassieke random walk. Het kwantumgedrag komt pas voor als we de superpositie intact laten door het systeem niet tussendoor te meten, waardoor de verschillende termen kunnen ‘interfereren’.

Een groot verschil is de snelheid waarmee het deeltje zich voortbeweegt, oftewel de snelheid waarmee de kansverdeling zich uitspreidt. De spreiding van de verdeling kunnen we wiskundig beschrijven met de verwachtingswaarde van het kwadraat van de positie, $\langle X_t^2 \rangle$. Hier staat X_t voor de positie op tijdstip t en $\langle \cdot \rangle$ voor de verwachtingswaarde. Het getal $\langle X_t^2 \rangle$ geeft intuïtief de breedte van de kansverdeling aan op tijdstip t . Men kan bewijzen dat dit getal zich bij een klassieke wandeling gedraagt als $\langle X_t^2 \rangle \sim t$, terwijl voor de kwantumwandeling geldt $\langle X_t^2 \rangle \sim t^2$. Het teken $\sim$ geeft



Figuur 2 De kansverdeling na vijftig stappen van de Hadamard-wandeling. Links was de begintoestand $|0, l\rangle$ en rechts was de begintoestand $\frac{1}{\sqrt{2}} |0\rangle (|l\rangle + i |l\rangle)$.

hier aan dat de linkerkant proportioneel is aan de rechterkant, met een niet nader gespecificeerde constante. In woorden betekent dit dat de Hadamard-wandeling zich kwadratisch sneller verspreidt dan de klassieke random walk.

Een ander groot verschil tussen het kwantumsysteem en het klassieke systeem is de afhankelijkheid van de tijdsevolutie van de begintoestand. In Figuur 2 wordt de kansverdeling van de Hadamard-wandeling weergegeven voor verschillende begintoestanden. Hoewel de operatoren hetzelfde blijven kan het verschil in begintoestand toch heel veel uitmaken. Als we in het voorbeeld hierboven de lijn $\mathbb{Z}$ vervangen door een eindig systeem zoals de discrete cirkel $\mathbb{Z}/N\mathbb{Z}$, dan heeft een klassieke wandeling een unieke *stationaire toestand*. Dit betekent dat hoe de begintoestand ook is, het systeem altijd zal uitkomen in deze stationaire toestand als men maar lang genoeg wacht. Op de cirkel zou dit de uniforme verdeling over alle punten zijn. De begintoestand kan bijvoorbeeld een deeltje op één punt zijn (oftewel een kansverdeling waarbij de kans 1 is op het beginpunt en 0 op alle andere punten), maar ook een kansverdeling over meerdere punten. In beide gevallen zal het systeem uiteindelijk terechtkomen in de uniforme kansverdeling. Voor de kwantumwandeling is dit echter niet het geval. Deze heeft geen stationaire toestand, en twee verschillende begintoestanden geven ook op lange termijn verschillende toestanden. Sterker nog, omdat de evolutie van een kwantumsysteem altijd unitair is zullen twee orthogonale toestanden altijd orthogonaal blijven. Een eigenschap van unitaire operatoren is namelijk dat deze

het inwendig product tussen vectoren behouden, $\langle Ua, Ub \rangle = \langle a, b \rangle$ in formules. Als $|a\rangle$ en $|b\rangle$ twee begintoestanden zijn die orthogonaal zijn, zoals $|0, l\rangle$ en $|l, l\rangle$ dan zijn de resulterende toestanden na t tijdstappen nog steeds orthogonaal. Deze zullen dus nooit in dezelfde stationaire toestand terecht kunnen komen.

De Hadamard-wandeling is maar een eenvoudig voorbeeld van een kwantumwandeling. We kunnen deze wat algemener maken door in plaats van de Hadamard-operator andere 2×2 unitaire matrices toe te laten als coin-operator. Nog algemener wordt het als we een algemene graaf toelaten in plaats van de discrete lijn. Bij een graaf waarbij alle knopen d burens hebben geven we het deeltje d interne toestanden en zal de coin-operator een unitaire $d \times d$ -matrix zijn. In het algemeen kan ook van elke klassieke Markov-keten (een algemeen willekeurig proces waar de random walk een voorbeeld van is) een kwantumversie worden gemaakt. Ook in dit algemene geval geldt dat de kwantumwandeling zich kwadratisch sneller uitspreidt. Er bestaat ook een versie van de kwantumwandeling met een continue tijdsparameter. In dit geval is de unitaire operator van de vorm e^{itH} waarbij H de zogeheten Hamiltoniaan van het systeem is (niet te verwarren met de Hadamard-operator H). De Hamiltoniaan beschrijft de energie van het deeltje. In dit artikel gaan we niet verder in op zulke systemen.

Kwantumwandelingen in algoritmen

Zoekalgoritmen. We zullen nu bekijken hoe kwantumwandelingen in kwantumalgoritmen kunnen worden gebruikt. Stel dat we

een lijst van N elementen hebben, bijvoorbeeld $1, 2, \dots, N$, en we zoeken een element in de lijst dat aan een bepaalde eigenschap voldoet. Het juiste element van de lijst noemen we ook wel het *gemarkeerde element* ('marked element') van de lijst. Een simpel klassiek algoritme is om de elementen van de lijst een voor een af te gaan en bij elk te controleren of het voldoet aan de gewenste eigenschap. Als we zo'n element vinden zijn we klaar. Dit algoritme kost in het slechtste geval N stappen, aannemend dat er in ieder geval één oplossing tussen zit. Een stap betekent hier het controleren van de eigenschap voor een gegeven element. De benodigde tijd voor dit algoritme is dus van de orde $O(N)$, dat wil zeggen de tijd schaalst lineair met N . Een ander algoritme is om steeds een *willekeurig* getal te kiezen en hiervan te controleren of het gemarkeerd is (oftewel of het voldoet aan de gewenste eigenschap). Dit herhalen we net zo lang tot we een oplossing vinden, aangenomen dat die er is. Dit lijkt een onhandige methode, maar gemiddeld zal dit ook $O(N)$ stappen kosten. Het kost soms meer stappen dan in het eerste algoritme, maar het verwachte aantal benodigde stappen is nog steeds lineair in N .

Het bovenstaande willekeurige algoritme kunnen we zien als een random walk op de *complete graaf*. De getallen vormen daarbij de knopen van de graaf, en tussen elke twee knopen zit een lijn. Het algoritme probeert een willekeurig getal, dus bevindt zich dan op een van de knopen. Vervolgens probeert het algoritme een ander willekeurig getal en dat betekent dat het naar een van de burens van de knoop beweegt. In dit geval zijn alle knopen burens en lijkt dit geen extra inzicht te geven.

We kunnen elk element van de lijst immers los van de andere bekijken om te controleren of het gemarkeerd is.

Er zijn echter veel problemen waarbij dit niet kan. In dat geval is er ook een lijst elementen waarvan er één of meer gemarkeerd zijn, maar is het niet mogelijk om direct elk element van de lijst te controleren. Om dit duidelijk te maken kijken we naar schaken. Het mogelijke spelverloop van een potje schaken kunnen we in een graaf weergeven. Elke knoop hoort bij een configuratie van de speelstukken en daarbij de kleur van de speler die aan zet is. Er zit een lijn tussen de knopen als het toegestaan is om van de ene naar de andere configuratie te gaan met een valide zet. De beginopstelling is bijvoorbeeld een witte knoop, en alle burens van deze knoop zijn zwarte knopen met opstellingen die je kan krijgen door precies één legale zet te doen met wit. Het verloop van een schaakspel is een pad door deze graaf.

Stel dat we nu de volgende vraag willen beantwoorden: kan ik in minder dan d beurten een paard op positie C4 krijgen en een loper op positie E5? Gegeven een knoop in de graaf die we zojuist omschreven kan men meteen controleren of er een paard op positie C4 staat en een loper op positie E5. Echter, of dit ook in minder dan d beurten kan is niet meteen duidelijk. Om dit te berekenen moeten we beginnen bij de huidige opstelling en alleen via de knopen van de graaf gaan om zo bij de knoop uit te komen die we zoeken.

In het algemeen is het probleem als volgt: gegeven is een graaf met N knopen en een beginknoop, en we zoeken één of meerdere knopen die gemarkeerd zijn. Het is hierbij alleen toegestaan om knopen te controleren die burens zijn van reeds gecontroleerde knopen. We mogen dus alleen ‘via de graaf’ naar andere knopen toe.

Bij een klassiek algoritme kan men aan de volgende procedure denken. Begin bij de beginknoop en kies een willekeurige buur. Controleer deze en als hij niet voldoet, kies dan opnieuw een willekeurige buur. Dit is hetzelfde algoritme als hiervoor, maar nu kiezen we alleen burens van de vorige knoop in plaats van een willekeurige knoop. Dit algoritme voert een random wandeling uit op de graaf, en met kansrekening kan men uitrekenen hoe lang het gemiddeld duurt tot we bij een gemarkeerd punt aankomen. De tijd die dit kost hangt af van de graaf.

Kwantumzoekalgoritmen. We kunnen nu een kwantumversie van dit algoritme maken. Een mogelijk idee zou zijn om een kwantumwandeling te beginnen op het beginpunt, analoog aan het klassieke algoritme. Dit werkt echter niet zomaar. Merk op dat we bij het klassieke algoritme meteen klaar zijn als we eenmaal bij een gemarkeerd punt zijn aangekomen. Bij de kwantumversie werkt zo’n idee echter niet. Als je namelijk meet op welke knoop je zit, dan ondergaat de kwantumtoestand een ‘collapse’. Zoals we al zagen bij de Hadamard-wandeling krijg je het klassieke gedrag terug als je na iedere stap een meting doet. Een kwantumzoekalgoritme werkt dus iets anders. Hoewel we niet steeds kunnen meten is het echter wel mogelijk om in superpositie te controleren of een knoop gemarkeerd is zonder een meting te doen, net zoals een functie $f(n)$ op meerdere inputs kan worden berekend zonder te meten. Dit betekent dat het is toegestaan om op de gemarkeerde knopen een andere unitaire operatie uit te voeren dan op de andere knopen. Bij de Hadamard-wandeling hierboven zou dit betekenen dat we overal de Hadamard-operator toepassen behalve op een gemarkeerde knoop, waar een andere operator naar keuze wordt toegepast. Stel bijvoorbeeld dat positie 3 gemarkeerd is. Dan zou het volgende een voorbeeld kunnen zijn van een stap uit het kwantumalgoritme:

$$\begin{aligned} &|2, \downarrow\rangle + |-4, \uparrow\rangle + |3, \uparrow\rangle + |0, \downarrow\rangle \\ &\rightarrow (1 \otimes H) |2, \downarrow\rangle + (1 \otimes H) |-4, \uparrow\rangle \\ &\quad + (1 \otimes A) |3, \uparrow\rangle + (1 \otimes H) |0, \downarrow\rangle \end{aligned}$$

waarbij A de gekozen operator is voor de gemarkeerde knoop. Dit betekent niet meteen dat na deze stap de kans heel hoog is om de gemarkeerde knoop te vinden, maar door A slim te kiezen en precies het goede aantal stappen te zetten kan het kwantumalgoritme toch werken.

Grovers algoritme. We zullen nu eerst Grovers algoritme [2] beschrijven. We moeten een gemarkeerd element vinden uit een lijst van lengte N , en we kunnen elk element direct los bekijken zoals op de complete graaf. Grovers algoritme kan ook zonder de technieken van kwantumwandelingen worden beschouwd, maar het is illustratief om het in dit formalisme te bekijken. We gaan ervan uit dat er een operator O is waarmee we kunnen controleren of een element gemarkeerd is. Deze operator

is gedefinieerd als

$$O|i\rangle = \begin{cases} |i\rangle & \text{als } i \text{ niet gemarkeerd is,} \\ -|i\rangle & \text{als } i \text{ wel gemarkeerd is,} \end{cases} \quad 1 \leq i \leq N.$$

De operator O wordt ook wel het *orakel* (‘oracle’) genoemd, en kan bijvoorbeeld een ander algoritme zijn dat controleert of het getal i aan een vergelijking voldoet. Wat dit andere algoritme ook is, het kan vrijwel altijd in deze vorm worden geschreven en dus zullen we nu aannemen dat zo’n operator bestaat. Merk op dat O precies overeenkomt met het toepassen van de identiteitsoperator 1 op de niet-gemarkeerde knopen en de operator -1 op de gemarkeerde knopen.

Om verder te gaan, bekijken we nu eerst de *uniforme superpositie*:

$$|U\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle.$$

Deze superpositie is over alle elementen uit de lijst waarbij elk element dezelfde amplitude $1/\sqrt{N}$ heeft. Vervolgens hebben we een operator R nodig die gedefinieerd is als

$$R = 2|U\rangle\langle U| - 1.$$

Dit is de *reflectie in $|U\rangle$* . Merk op dat $R|U\rangle = |U\rangle$, en als $|\phi\rangle$ een toestand is die loodrecht staat op $|U\rangle$, dan geldt $R|\phi\rangle = -|\phi\rangle$ omdat $\langle U|\phi\rangle = 0$. Deze operator R speelt de rol van ‘een stap zetten in de graaf’. Er geldt namelijk dat $R|i\rangle$ een superpositie over alle burens van knoop i is (inclusief i zelf).

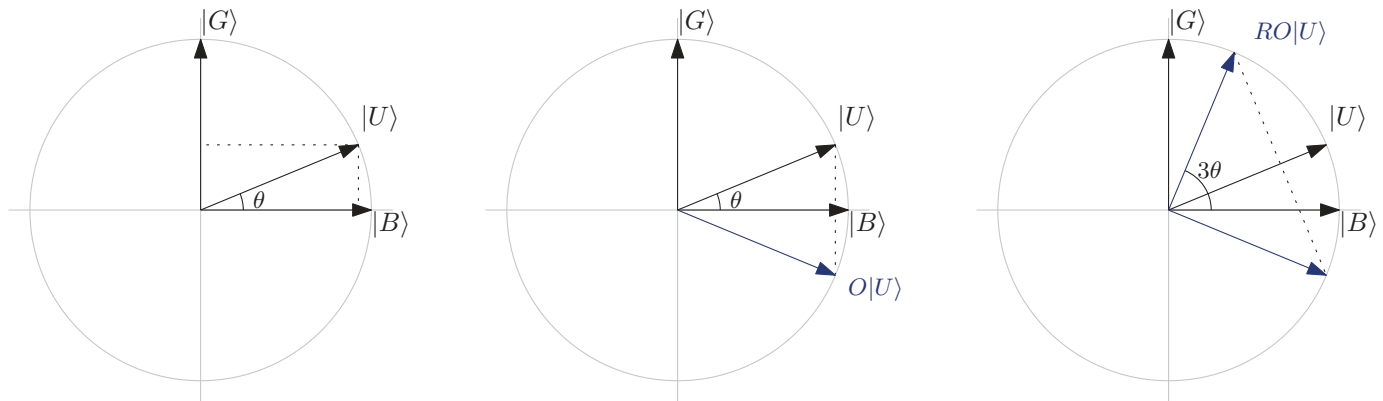
Het kwantumalgoritme van Grover is:

- maak de toestand $|U\rangle$ aan;
- pas T keer de operator $R \circ O$ toe;

waarbij we later zullen specificeren wat T is. We gaan nu kijken hoe en waarom dit werkt. Laten we aannemen dat er precies één gemarkeerd element is, dat we noteren met $|G\rangle$ (‘good’). In dat geval kunnen we de som van $|U\rangle$ splitsen in $|G\rangle$ en een som zonder $|G\rangle$ die we met $|B\rangle$ (‘bad’) aanduiden:

$$\begin{aligned} |U\rangle &= \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle + \frac{1}{\sqrt{N}} |G\rangle \\ &= \sqrt{\frac{N-1}{N}} |B\rangle + \sqrt{\frac{1}{N}} |G\rangle. \end{aligned}$$

De factor $\sqrt{(N-1)/N}$ is zo gekozen dat $|B\rangle$ genormaliseerd is. Merk op dat $|B\rangle$ en $|G\rangle$ loodrecht zijn omdat ze uit verschillende basisvectoren zijn gebouwd. Als we nu de



Figuur 3 Links: de begintoestand $|U\rangle$ van Grover's algoritme. Midden: de toestand na het toepassen van de operator O . Rechts: de toestand na het toepassen van O en R .

operator O toepassen vinden we $O|B\rangle = |B\rangle$ en $O|G\rangle = -|G\rangle$, direct uit de definitie van O . We kunnen dit weergeven door $|U\rangle$ te tekenen in de $|B\rangle$ -, $|G\rangle$ -basis, zie het linker diagram in Figuur 3. Het systeem begint in de toestand $|U\rangle$ en na het toepassen van de operator O wordt het gespiegeld in de horizontale as, zoals weergegeven in het middelste diagram. De operator R spiegelt per definitie in de begintoestand $|U\rangle$, en het resultaat daarvan is weergegeven in het rechter diagram van Figuur 3.

In de figuur is de hoek tussen $|U\rangle$ en $|B\rangle$ aangegeven met θ . We kunnen $|U\rangle$ schrijven als

$$|U\rangle = \cos\theta |B\rangle + \sin\theta |G\rangle,$$

dus geldt $\sin\theta = 1/\sqrt{N}$ en $\cos\theta = \sqrt{(N-1)/N}$. Als N voldoende groot is, dan is $1/\sqrt{N}$ voldoende klein zodat er geldt $\sin\theta \approx \theta$, dus $\theta \approx 1/\sqrt{N}$.

De figuur laat zien dat de hoek na één toepassing van $R \circ O$ gelijk is aan 3θ . Met elke iteratie van $R \circ O$ zal de hoek met 2θ verhogen, dus na T stappen zal de hoek gelijk zijn aan $\theta + 2T\theta$. We willen uitkomen in de toestand $|G\rangle$, dus we moeten het aantal stappen T zo kiezen dat de

hoek uiteindelijk gelijk is aan $\pi/2$, en dit is het geval als $T = \frac{\pi}{4\theta} - \frac{1}{2}$. De waarde van T moet een geheel getal zijn, dus voor veel waarden van θ komt dit niet meteen uit. Het heeft geen zin om langer door te gaan want dan draait het systeem te ver door en komen we uiteindelijk in $-|B\rangle$ uit. We kunnen de waarde $\frac{\pi}{4\theta} - \frac{1}{2}$ echter simpelweg afronden. Het systeem zal dan niet exact in $|G\rangle$ uitkomen maar wel erg dichtbij en is de kans om $|G\rangle$ te meten heel hoog. Naarmate N groter wordt, worden θ en de afrondfout kleiner.

Voor grote waarden van N geldt dus $T \approx \frac{\pi}{4}\sqrt{N}$. De tijd schaalst dus met $\sqrt{N}$ in plaats van N zoals bij een klassiek algoritme! Ongeacht hoe snel klassieke computers ook worden, een quantumcomputer is altijd sneller, als de lengte N van de lijst groot genoeg is.

Kwantumwandeling-algoritmes. Het kwantumwandeling-zoekalgoritme op een algemene graaf begint door eerst een aantal stappen te zetten om in de *uniforme* superpositie $|U\rangle$ uit te komen zoals bij Grover. Als de zoekgraaf de complete graaf is zal dit één stap kosten, en als de graaf

een lijn is zal dit N stappen kosten. In deze fase wordt nog niets speciaals met de gemarkeerde knoop gedaan. Vervolgens kan een kwantumwandeling worden uitgevoerd met verschillende operatoren op de gemarkeerde en ongemarkeerde knopen. Het hangt van de graaf af hoe snel het algoritme wordt. Voor sommige grafen is de kwantumversie niet sneller dan zijn klassieke tegenhanger.

Tot slot

We hebben laten zien hoe de kwantummechanica ons toestaat om andere algoritmes te maken die soms sneller zijn dan de klassieke algoritmes voor hetzelfde probleem. Er blijft nog veel werk te doen, want voor veel problemen is het nog niet duidelijk of een quantumcomputer ze sneller kan doen dan een klassieke computer en zo ja, hoeveel sneller het kan. Het onderzoek is vooralsnog enkel theoretisch omdat er op dit moment nog geen quantumcomputers zijn die goed genoeg zijn om berekeningen mee te doen. De hoop is dat dit over tien of twintig jaar wel het geval is en dat de benodigde algoritmes dan klaarliggen voor gebruik.

Referenties

- 1 A. Einstein, B. Podolsky en N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, *Phys. Rev.* 47 (1935), 777.
- 2 L. Grover, A fast quantum mechanical algorithm for database search, *Proceedings 28th Annual ACM Symposium on the Theory of Computing* (1996).

Remco van der Hofstad

Department of Mathematics and Computer Science
Eindhoven University of Technology
r.w.v.d.hofstad@tue.nl

Event Eurandom workshop honouring the 60th birthday of Frank den Hollander

Royal decoration for Frank den Hollander

On 15 December 2016, Frank den Hollander received a Royal Decoration in the Order of the Dutch Lion (Ridder in de Orde van de Nederlandse Leeuw). Mayor Emile Jaensch of Oegstgeest presented the decoration during a conference held at the research institute Eurandom located on the campus of Eindhoven University of Technology. Remco van der Hofstad reports on this special event.

This conference was devoted to the celebration of Frank den Hollander's sixtieth birthday. Frank was 'pleasantly surprised' by the honour awarded to him: "This was a very warm moment. It was wonderful that also my wife and children could attend. It is very attentive that the mayor of Oegstgeest came to Eindhoven especially for this event. I have always enjoyed working with young talents in science. Showcasing science is magnificent and you get to meet wonderful and beautiful people."

Eminent scientist and excellent manager

Frank den Hollander is a professor in probability and statistical mechanics. He is now

affiliated to the Department of Mathematics at Leiden University, and worked at Eindhoven University of Technology, Radboud University Nijmegen, Utrecht University and Delft University of Technology before that. He is an eminent Dutch scientist, who has made breakthrough international discoveries in probability theory, statistical mechanics and network theory. He acts as a role model for the many young scientists that he works with.

Frank den Hollander acted as the scientific director of the European institute in stochastics Eurandom from 2000 to 2005. He was one of the founding fathers of the Platform Wiskunde Nederland (PWN).

PWN united the rather fragmented mathematical landscape in mathematics before that time. He was also one of the founding fathers of the cluster in Stochastics STAR. He is an excellent manager, who has gained the respect that is needed to act as a leader and the vision to decide what is the right course of action. He is one of the pillars of the Gravitation project NETWORKS, a consortium of scientists from the University of Amsterdam, Leiden University, Eindhoven University of Technology and the Center for Mathematics and Computer Science (CWI). In his extensive contributions shaping the mathematical landscape in the Netherlands, he has always put mathematics and the needs of the mathematical community in the Netherlands central.

Supervisor and networker

One of the distinguishing characteristics of Frank are his warm character and his enor-



Frank is being addressed by Mayor Jaensch of Oegstgeest



Frank receives his Royal Decoration



Michiel van den Berg and Erwin Bolthausen: two of Frank's long term collaborators

mous energy, his positive and constructive attitude and his personal relations to his colleagues — from their first steps into research until, sometimes, their professorships. Frank keeps on supporting them and is always ready for a chat when and where needed. In this way, he helps his colleagues in making the tough decisions, in science and beyond, that they are being confronted with. Having supervised some thirty plus postdocs and twenty PhD students, of whom now eight are full professors, he has had a marked influence on the development of probability and statistical physics in the Netherlands and internationally. He has an extensive network of collaborators, with a focus on Germany, England, France, Switzerland and Italy. Remarkably, he works with these colleagues in their own languages, showing that he combines mathematical strength with an unusual gift for languages.

A guided tour through random media

Den Hollander turned sixty on 1 December 2016. On this occasion and according to scientific tradition, a birthday conference

was organised on 12–16 December at Eurandom in Eindhoven, organised by Marek Biskup, Aernout van Enter, Frank Redig and the author, appropriately titled 'A guided tour through random media'.

His collaborators in the past forty years from all over the world were invited, and massively attended. This gave rise to a scientific feast, showcasing the beauty of mathematics and the strong personal ties that Frank forged with many colleagues across the world. During their talks, Frank's collaborators discussed research as well as how it was to work with him. These collaborations often spanned several decades, showing that working with Frank is a lasting and stimulating, maybe even invigorating, experience. Many anecdotes about these collaborations were shared, giving an extensive overview of Frank's work, as well as his personality.

The workshop mainly focussed on the areas that Frank made the most well-known contributions. These include the theory of large deviations, metastability and random polymers. In each of these topics, Frank has written an influential monograph summarizing the state of the art, as well as posing many open problems that have shaped the areas since. Other topics include the Parabolic Anderson Model and random walks in static and dynamic random environments. Several of the close collaborators gave panoramic talks on various topics that are dear to Frank's heart. Stu Whittington spoke about 'Self-avoiding walks and related polymer models', Andreas Greven about 'Populations, genealogies and fluctuating environments', Roberto Fernandez about 'Gibbs-non-Gibbs dynamical transitions: The large-deviation

paradigm and how we got there', Anton Bovier about 'Metastability according to Frank' and Wolfgang König about 'The spatially discrete parabolic Anderson model with time-dependent potential'. In each of these fields, there is an intricate relation between the objects studied and the (often random) environment in which they live, and in each of these topics Frank made profound contributions.

Frank tackles these problems with tools from across the mathematic fields, including variational problems, functional analysis, combinatorics, and perturbative techniques. These tools are used in ingenious and creative ways that not only tackle the problem, but also give insight into the how and why of the solutions found. The young people that he worked with in these areas have often flown out since, and have become research leaders in the field themselves. This shows the lasting influence that Frank has on the field at large, as exemplified in this workshop.

Piano recital

On Thursday 15 December the conference was complemented by a piano recital by Italian pianist Igor Roma, in 1996 winner of the prestigious Liszt Concours and a good friend of Den Hollander. Frank also played an important role in the careers of a number of Italian top-pianists, by offering them a home-away-from-home when they were in the Netherlands, for example for concerts. This shows Frank's great personality, as well as his generosity in sharing his love for piano music with his scientific friends. At the reception afterwards, he received his well-deserved Royal Decoration. ✦



Frank with pianist Igor Roma



Igor Roma's piano recital

Francesca Arici

IMAPP

Radboud University Nijmegen

f.arici@math.ru.nl, francesca.arici@gmail.com

Domenico Monaco

Fachbereich Mathematik, Eberhard Karls Universität Tübingen, and

Dipartimento di Matematica e Fisica, Università degli Studi di Roma Tre

domenico.monaco@uni-tuebingen.de, dmonaco@mat.uniroma3.it

Event Workshop Lorentz Center, 27 February–10 March 2017

KK-theory, gauge theory and topological phases

The history of science is full of moments when physics and mathematics both benefited from mutual interaction, with the early twentieth century providing us with two big examples. On one hand, Einstein's theory of gravitation could not have been developed without the work of Riemann on the geometry of manifolds. On the other hand, the advent of quantum mechanics fostered the development of new mathematics especially in operator algebras. In line with this spirit of interdisciplinarity, the school and workshop 'KK-theory, Gauge Theory and Topological Phases' took place from 27 February to 10 March 2017 at the Lorentz Center in Leiden. Francesca Arici and Domenico Monaco report about this event.

The program of 'KK-theory, Gauge Theory and Topological Phases', organized by Alan Carey (Canberra, AU), Steve Rosenberg (Boston, USA), and Walter van Suijlekom (Nijmegen, NL), aimed both at young and more senior scientists in mathematics and physics, focusing on the recent applications of Kasparov's bivariant K -theory in both high-energy and solid state physics.

This event was able to bring together experts from several diverse communities, creating an environment that stimulated fruitful interactions, fostering new interdisciplinary collaborations, and allowing participants to learn about cutting-edge research on the topics of the program from the leaders in the field.

The main subject of the program was *KK-theory*, an apparently abstract notion from operator algebra which has recently emerged as an important tool for applications of noncommutative geometry in mathematical physics, to describe gauge-theoretic phenomena originating both in high-energy physics and in condensed matter. Roughly speaking, *KK-theory* produces 'topological invariants' for C^* -algebras, which are a particular class of operator algebras often considered to be geometric: they include

the algebras of continuous functions over (Hausdorff, locally compact) topological spaces, and generalize topological notions to a noncommutative algebraic setting. Besides, C^* -algebras have repeatedly proved to be useful in modelling physical systems as well, most notably in quantum mechanics, where Max Born's commutation relations $[x, p] = i\hbar$ (which imply the paradigmatic result of quantum physics, namely Heisenberg's uncertainty principle) are considered to be the archetype of a noncommutative phase space.

Since its early days the field of noncommutative geometry and operator algebras has produced results of striking mathematical beauty, but also of profound relevance for mathematical physics. In the recent years, noncommutative index theory and *KK-theory* have provided insight in problems in particle physics [34], and in the mathematical formulation of the so-called 'bulk-edge correspondence' in solid state physics (see the subsection 'Topological phases' below). With these recent applications of *KK-theory* to gauge theories and topological quantum matter, Wigner's 'unreasonable effectiveness of mathematics' strikes once again.

The school

The first week of the program was devoted to the school. There were three main short courses, each of which was taught by two experts on the subject. The aim of these lectures was to get everyone 'up to speed' for the incoming workshop, and they served this purpose excellently, providing very clear exposition on the subjects of the program.

Each day of the school concluded with a discussion session, where the participants were invited to raise their doubts about the lectures, or to ask for clarifications regarding some of the points that were raised during them. This favoured a thorough understanding of the subjects taught, as well as the possibility to debate the physical motivations and applications for the mathematical tools that were illustrated and developed in the courses.

Index theory

The foundations of modern-day noncommutative geometry can be found in *index theory* [15]. Two introductory lectures on this subject were delivered by Matthias Lesch (Bonn, DE). The main aim was to state and discuss the celebrated *Atiyah–Singer index theorem* [3], one of the milestones of modern mathematics which draws a bridge between analysis (elliptic pseudo-differential operators on manifolds), topology (characteristic classes of vector bundles) and operator algebra (the theory of Fredholm operators). For this result Sir Michael F. Atiyah got both the Fields Medal (in 1966) and the

KK-theory

Kasparov's bivariate K -theory [21], or KK -theory, is a framework which incorporates and generalizes both K -theory and K -homology for C^* -algebras. In the spirit of noncommutative geometry, it can be thought of as an extension of index theory in which smooth closed manifolds are replaced by singular spaces, in the form of C^* -algebras.

Elements of the group $KK(A, B)$, also called *Kasparov modules*, are (homotopy classes of) triples (E, ϕ, T) , where

1. E is a $(\mathbb{Z}_2$ -graded) right Hilbert B -module,
2. ϕ is a representation of A on E as adjointable operators, and
3. T is an adjointable operator on E (of odd degree) such that the operators $[T, \phi(a)]$, $(T^* - T)\phi(a)$ and $(T^2 - 1)\phi(a)$ are compact for all $a \in A$.

One can show that $KK(\mathbb{C}, B)$ reproduces $K_0(B)$, the usual K -group of the C^* -algebra B , while $KK(A, \mathbb{C})$ gives back the K -homology group of the algebra A , that is, the group of K -cycles (or Fredholm modules) over A .

One of the deepest and most powerful features of Kasparov's bivariate K -theory is the fact that KK -cycles can be composed: one can define the internal *Kasparov product*

$$KK(A, B) \times KK(B, C) \rightarrow KK(A, C).$$

In the case $A = C = \mathbb{C}$, the internal Kasparov product reduces to the well-known *index pairing* of K -theory classes on B with Fredholm modules, landing in $KK(\mathbb{C}, \mathbb{C}) \simeq \mathbb{Z}$. Thus, one recovers classical index theory within the framework of KK -theory.

Motivated by the appearance of Dirac type operators in the index pairing, an unbounded version of KK -theory has also been formulated [4], where instead of an adjointable operator one considers a densely defined, self-adjoint and regular operator T on the Hilbert B -module E , satisfying the following two conditions:

1. $\phi(a)(T + i1)^{-1}$ is a compact operator on E for all $a \in A$, and
2. there exists a dense $*$ -subalgebra $\mathcal{A} \subset A$ such that for all $a \in \mathcal{A}$ the operator $\phi(a)$ leaves the domain D of T invariant, and moreover the (graded) commutator $[T, \phi(a)]$, initially defined on D , admits a bounded extension to the whole E .

In particular, unbounded representatives of K -cycles — i.e. elements in $KK(A, \mathbb{C})$ — are spectral triples $(\mathcal{A}, \mathcal{H}, D)$, where $\mathcal{A}$ is a dense subalgebra of A , whose elements have bounded commutators with D .

The two pictures are related via the bounded transform: given an unbounded Kasparov module, the formula

$$F_D := D(1 + D^2)^{-1/2}$$

gives a bounded (A, B) -Kasparov module. Note that the bounded transform is what allows one to move from Dirac operators to Fredholm operators.

While the unbounded theory presents more complication due to analytic subtleties, it also offers several advantages because it is more explicit and more geometric. Moreover, it allows to give, up to equivalence, an explicit realisation of the Kasparov product. This constructive approach to the Kasparov product, initiated by Mesland [26] and rigorously developed by Mesland, Kaad, Lesch, Rennie and others [19, 27, 28], has opened the door to a wide range of applications of geometric and physical nature. In the wake of these developments come new applications of Kasparov theory to index theory, with new proofs of many classical theorems by KK -methods.



The poster for the school and workshop depicts the complex Hopf fibration, an example of geometry that can be factorised using the Kasparov product.

Abel Prize, together with Isadore Singer (in 2004).

Fredholm operators are characterized by the property of having finite-dimensional kernel and cokernel. Many examples of such operators come from the field of *pseudo-differential operators*, a class of operators between vector bundles over a

manifold which, roughly speaking, act in the Fourier representation by multiplication times a function (the *symbol* of the operator) which is smooth and doesn't grow too fast at infinity. The Atiyah–Singer theorem computes the *index* of an *elliptic* pseudo-differential operator — a purely analytic concept which is defined, for any

Fredholm operator, as the difference between the dimensions of its kernel and its cokernel — in terms of certain characteristic classes of the manifold and of the bundles on which the operator acts (on the 'topological' side). We refer to the popular article [24] for an introduction. At a more abstract level, the index can be realized

Gauge theory

In the mathematical approach to gauge theories, gauge fields are realized as connections on principal bundles up to equivalence under the action of the gauge group.

The Atiyah–Singer index theorem has proven to be an effective tool in understanding the mathematical aspects of gauge theories: in the study of four-dimensional Yang–Mills gauge theories, this theorem was used to determine the dimension of the moduli space of self-dual connections [2]. With this in mind, it is not surprising that *KK*-theory, which sets its roots in index theory, has important and promising applications in the field of gauge theories.

In the framework of noncommutative geometry, gauge theories naturally arise from spectral triples [11]. Indeed, out of the ingredients of a spectral triple $(\mathcal{A}, \mathcal{H}, D)$, one can naturally obtain the defining elements of a gauge theory.

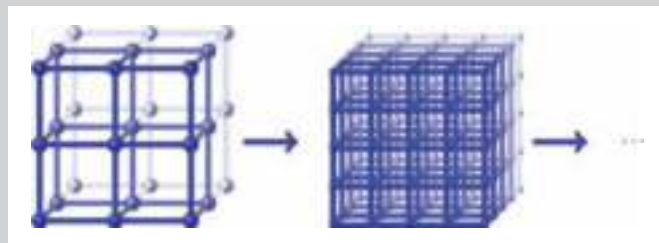
- Starting from the algebra $\mathcal{A}$ one constructs the gauge group $\mathcal{G}$, by looking at the group of inner automorphisms, which in the case of noncommutative algebras is non-trivial. In many situations, one can identify the group $\text{Inn}(\mathcal{A})$ with the gauge group of the theory.
- The gauge fields are obtained from the spectral data contained in the Dirac operator, with its spectrum playing a central role in the theory.
- The gauge group $\mathcal{G}$ acts on the Dirac operator (i.e., on fields) by conjugation with a unitary operator $D \mapsto UDU^*$, giving rise to a perturbation in terms of pure gauge fields. Since the spectrum of the Dirac operator is invariant under the action of unitary gauge transformations, one can define the *spectral action*, which is interpreted as an action functional for the theory, describing the dynamics and interactions of the gauge fields coming from D .

As described amply in [34], this notion is compatible with the classical definition of gauge theories in terms of connections

on a vector bundle. Moreover, one of the greatest achievements of the noncommutative approach to gauge theories is the derivation of the full Standard Model of particle physics starting from a spectral triple on a noncommutative manifold which is the product of a commutative geometry and a matrix algebra (cf. [33]).

Motivated by this product structure, one looks at the unbounded version of *KK*-theory, which allows for a bundle-theoretic description of gauge theories, both on commutative and noncommutative base spaces. In [9] the constructive Kasparov product is used to factorize a class of noncommutative differential geometries, in the form of spectral triples, into the product of two pieces: a commutative horizontal spectral triple on the base manifold, and a Kasparov module which describes the vertical noncommutative geometry.

More generally, it is interesting to understand whether a given spectral triple factorizes as the product of an unbounded Kasparov bimodule with another spectral triple, on a possibly noncommutative manifold, which would then represent the base manifold on which the gauge theory is defined. Many amongst the participants and speakers of the program have been dealing with this question in the past years, and their work has shed light on the problem, as well as on the technical aspects of Kasparov's theory.



In lattice gauge theories, ‘refining’ the lattice yields the thermodynamic limit. This procedure is best described in terms of limits of C^* -algebras.

Illustration: Walter van Suijlekom

as a pairing between *K*-theory (vector bundles) and *K*-homology (elliptic differential operators).

In the paradigm of noncommutative geometry due to Connes [11, 12] a central notion is that of a *spectral triple* $(\mathcal{A}, \mathcal{H}, D)$. This consists of an involutive algebra of operators acting on a Hilbert space $\mathcal{H}$ and a self-adjoint operator D on the same Hilbert space, subject to certain conditions. Such Dirac operators are ‘unbounded versions’ of Fredholm operators, as described in the box on *KK*-theory. Commutative examples of spectral triples are provided by the Hodge–de Rham operator on differential forms over a Riemannian manifold, and the spin Dirac operator that is naturally defined when the manifold has also a spin structure, which is a structure slightly finer than orientability [25] used in phys-

ics to model the behaviour of particles like electrons. Conversely, the notion of spectral triple, as introduced by Alain Connes, is modeled on these examples, and generalizes them to the noncommutative realm.

Coming back to the program of the school, more advanced topics in index theory, related to *Kontsevich formality theorem*, were discussed instead by Ryszard Nest (Copenhagen, DK). These shed some light on the connection between index theory and *Poisson geometry*, and on how this link naturally leads to the study of *Hochschild cohomology* of manifolds. This inspired Connes to formulate his *cyclic cohomology* for spectral triples, which heuristically plays the role for a noncommutative manifold of the usual de Rham cohomology of differential forms.

The presentation of index theory by Lesch and Nest served as an essential ‘motivating example’ for the other lectures taught during the school.

KK-theory

Siegfried Echterhoff (Münster, DE) gave an introductory overview on the main protagonist of the program, namely *KK*-theory (see box ‘*KK*-theory’).

In its complex, bounded variant, this theory associates to pairs of $(\mathbb{Z}_2$ -graded) C^* -algebras (A, B) an abelian group, denoted by $KK(A, B)$. In order to construct the *KK*-group, one needs to move beyond Hilbert spaces, the natural spaces on which C^* -algebras are represented as operators, and deal with more general modules, endowed with a scalar product taking values in a (possibly different)

Topological phases

The recent years have witnessed a booming expansion in the theoretical understanding and experimental realization of so-called *topological insulators* [10, 17], a class of materials which are insulating in the bulk but display robust edge modes, capable of conducting charge or spin without dissipation and virtually without being affected by the presence of impurities and defects in the system. Such topological materials, whose theoretical proposal has already lead to the assignment of several Nobel Prizes in Physics (including the latest one from 2016), could potentially be employed in the future in spintronic devices and in quantum computers.

At a theoretical level, topological insulators are classified according to the *symmetries* satisfied by the Hamiltonian H that models an effective one-particle picture of the quantum system. The fundamental symmetries come in three flavours:

- *charge-conjugation* (also called *particle-hole*): it is implemented by an antiunitary operator C over the one-particle Hilbert space which can square to ± 1 , and H is said to be charge-conjugation symmetric if

$$CHC^{-1} = -H;$$

- *time-reversal*: it is also implemented by an antiunitary operator $\mathcal{T}$ squaring to ± 1 , and H is said to be time-reversal symmetric if

$$\mathcal{T}H\mathcal{T}^{-1} = H;$$

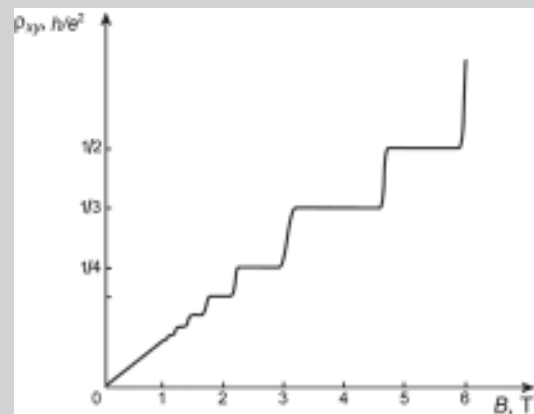
- *chiral*: this symmetry S is implemented unitarily, squares to 1, and H is said to be chiral symmetric if

$$SHS^{-1} = -H.$$

Owing to the fact that C and $\mathcal{T}$ can square either to $+1$ or -1 , and that the combination $C\mathcal{T}$ is a chiral symmetry S , one is lead to ten distinct symmetry classes: two of them, the one with absence of any symmetry and the one where only a chiral symmetry is present, are dubbed *complex*, while the other eight, which have at least one antiunitary symmetry and hence need the specification of a complex conjugation operator, are instead called *real*. When the underlying system is a crystal, other (spatial) symmetries may be imposed, like invariance under translations of the crystalline Bravais lattice and possibly under reflections or discrete rotations.

To each of these symmetry classes, and depending on the dimensionality of the system, one can attach certain topological labels, which essentially encode how many inequivalent Hamiltonian there exist obeying the prescribed symmetries: two Hamiltonians are considered equivalent if they can be adiabatically deformed one into the other, that is, if there is an homotopy of Hamiltonians between them for which the bulk gap stays open. Some of these topological labels have actual physical meaning as measurable quantities: the most celebrated of these instances is the quantization of the *quantum Hall transverse conductivity*, which occurs in a 2-dimensional gapped periodic system where all the other symmetries above are broken (the so-called Class A), and which is driven by the integer Chern number associated to the occupied Bloch states (see [16] and references therein). However, some of these classes, like the one corresponding to time-reversal symmetric systems in dimensions 2 and 3 under the assumption that $\mathcal{T}^2 = -1$ (Class AII), have more peculiar features, like the appearance of *torsion invariants*, with values in $\mathbb{Z}_2$ [13, 14, 20].

This classification scheme has lead to the compilation of ‘periodic tables’ of topological insulators [23, 31], depicted in Table 1. The realization that these tables are indeed periodic is the first success of mathematics, and more specifically of *K*-theory, in this field: *Bott periodicity* can be applied to show how to relate symmetry classes in different dimensions and how the classification repeats itself with an 8-fold periodicity in the dimension [32].



The transverse resistivity exhibits plateaus in the quantum Hall effect. Index pairings in *KK*-theory lurk behind the curtains of this and many other quantization phenomena in condensed matter physics.

C^* -algebra. These objects are known as *Hilbert C^* -modules*. Much of the theory of Hilbert spaces and operators thereon carries through to this more general setting, but with notable exceptions: not all bounded operators possesses an adjoint, and the class of *adjointable* operators plays a fundamental role in the theory. In particular, elements in the *KK*-group $KK(A, B)$ are Hilbert B -modules on which A acts (that is, is represented) as adjointable operators.

Starting from a brief reminder on *K*-theory for C^* -algebras, Echterhoff presented the theory of Hilbert C^* -modules, of the classes of operators acting on them (e.g. adjointable and compact operators), and of C^* -correspondences (in particular Morita equivalences), concluding with the definition of *KK*-theory in its bounded version and with the construction of a representative for the *Kasparov product*.

Echterhoff's lectures were followed by the ones given by Jens Kaad (Odense, DK),

which instead presented the *unbounded picture* of *KK*-theory. Much like in the case of spectral triples, in this framework the Fredholm operator which enters in the definition of a *KK*-class is replaced by a Dirac-like (unbounded) operator. This requires first of all to set up a theory for unbounded self-adjoint operators on Hilbert C^* -modules, which unfortunately combines the technical difficulties arising in the theory of unbounded operators on standard Hilbert spaces with those of operators on

C^* -modules, as for example the definition of the adjoint. Kaad gave an exposition of this theory, proceeding with the illustration of the main issues and possible solutions thereof for the definition of an unbounded representative of the Kasparov product (see box ‘*KK-theory*’). The general theory was corroborated with several instructive examples, both from the commutative and the noncommutative perspective.

Topological phases

Index theory and K -theory have found applications also in the thriving field of *topological quantum matter* (see box ‘*Topological phases*’). The seemingly trivial observation that the index of a Fredholm operator, being defined as a difference of dimensions of vector spaces, is an integer, has profound implications, underlying for example the quantization of the noncommutative topological invariants that label certain symmetry classes of topological insulators. These are a class of recently discovered materials, which offer promising applications in quantum devices due to their peculiar transport properties, induced mainly by their symmetry properties.

New advancements of the theory, which were discussed during the program, widely generalize the treatment of topological phases of matter to include *disordered systems*, *quasi-crystals* and *aperiodic solids* [29,30]. Inspired by the Bellissard–Connes approach to the quantum Hall effect [5], the correct framework to describe the symmetry classes of such systems was recently shown to be the KK -theory of certain C^* -algebras of observables, which are essentially constructed from the Hamiltonian of the system and the space of disorder or atomic configurations. Kasparov’s K -theory has indeed proven to be a flexible tool to incorporate all the relevant classes of symmetries, by considering in particular real and Real C^* -algebras (in the sense of Atiyah [1]: Real C^* -algebras are noncommutative spaces endowed with an involution). Furthermore, the use of the Kasparov product allowed to prove certain mathematical versions of the *bulk-edge correspondence*, which relates the topology of the bulk Hamiltonian to an effective description of the boundary of the system when this is terminated by an edge. These results prove rigorously the defining property of topological insulators, namely that their ‘topological’ character reflects in the stability

Symmetry				Dimension							
AZ	$\mathcal{T}$	C	$\mathcal{S}$	1	2	3	4	5	6	7	8
A	0	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$
AIII	0	0	1	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0
AI	1	0	0	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$
BDI	1	1	1	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$
D	0	1	0	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$
DIII	-1	1	1	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$	0
AII	-1	0	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$\mathbb{Z}$
CII	-1	-1	1	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
C	0	-1	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
CI	1	-1	1	0	0	$\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0

Table 1 The periodic table of topological insulators [23,31,32]. In the first column, ‘AZ’ stands for the Altland–Zirnbauer (sometimes called Cartan) label. The labels for the symmetries are: $\mathcal{T}$ (time-reversal), C (charge-conjugation), S (chirality). Time-reversal symmetry and charge conjugation are $\mathbb{Z}_2$ -symmetries implemented antiunitarily, and hence can square to plus or minus the identity: this is the sign appearing in the respective columns (0 stands for a broken symmetry). Chirality is instead implemented unitarily: 0 and 1 stand for absent or present chiral symmetry, respectively. Notice that the composition of a time-reversal and a charge conjugation symmetry is of chiral type. The table repeats periodically after dimension 8 (i.e. for example the column corresponding to $d = 9$ would be equal to the one corresponding to $d = 1$, and so on).

and robustness of their edge states even in the disordered regime [6,7,8,22].

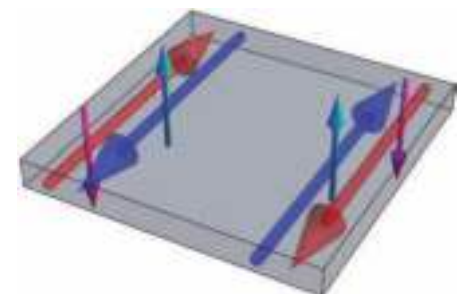
An introduction to the physics and mathematics of topological phases of matter was delivered by Guo Chuan Thiang (Adelaide, AU) and Chris Bourne (Sendai, JP).

Thiang introduced the periodic table of topological insulators (see Table 1), discussing symmetries of a quantum system from the point of view of K -theory and Clifford algebras. Indeed, the topology encoded in the bundle of occupied states of a gapped Hamiltonian determines a K -theory class of a C^* -algebra which essentially depends on the spatial symmetries (for example, a noncommutative torus in the case of periodic systems). Charge-conjugation, time-reversal and chiral symmetries form a representation of some Clifford algebra on the Hilbert space of the system, which turns the above-mentioned bundle into a Clifford module, and refines its geometry in terms of topological invariants. Thiang illustrated the general theory with two main examples from the periodic table, namely Class A in dimension 2 (the quantum Hall effect) and Class AII in dimensions 2 and 3 (the time-reversal symmetric topological insulators first proposed by Liang Fu, Charlie Kane and Eugene Mele [13,14]), which are paradigmatic for what concerns $\mathbb{Z}$ -valued and $\mathbb{Z}_2$ -valued invariants, respectively.

Bourne focused instead on the inclusion in this framework of disorder effects by means of C^* -algebraic techniques. This theory originated in the works of Jean Bellissard and collaborators, who gave a mathematical description of the quantum Hall effect based on the Connes–Chern character. The latter can be understood as a pairing between the K -theory class of the occupied states of the Hamiltonian (living in the K -theory of a *crossed product* C^* -algebra of the disorder configurations by the action of the group of translations by integer shifts) and a certain K -cycle (that is, a Fredholm module or a spectral triple), dictated essentially by the symmetries of the system. The fact that this pairing takes values in the group of integers is in turn a consequence of an index theorem.

This approach has been recently generalized to explain all the classes of the periodic table, as Bourne illustrated in his lectures. KK -theory arises as a description of the cycle needed to pair with the physical input from the Hamiltonian. Antiunitary symmetries like charge-conjugation and time-reversal require in particular certain ‘reality’ conditions, both in the algebra and in this KK -cycle, which in turn give rise to the torsion invariants, with values in $\mathbb{Z}_2$, via some index pairings in KK -theory.

Bourne’s lectures also covered the so-called *bulk-edge correspondence*, a principle which allows to ‘read’ the topology of the bulk physical system from the effects it induces on boundary states. It is this principle that underlies, for example, the quantization of the Hall current on the boundary of a sample of a quantum Hall system. This effect can be understood in terms of Toeplitz-like extensions of the ‘bulk’ algebra: the long-exact sequences in KK -theory then explain how to relate bulk and edge invariants.



Robust boundary currents and an insulating bulk are the defining properties of topological insulators. The bulk-edge correspondence has been established mathematically thanks to KK -theory methods.

The workshop

The five-days workshop followed the school and built on the topics introduced there. Elaborating on the material presented in the school, the newest uses of *KK*-theory in index theory, gauge theories and topological phases of quantum matter were illustrated by leading experts like Simon Brain (Nijmegen, NL), Maxim Braverman (Boston, USA), Motoko Kotani (Sendai, JP), Bram Mesland (Bonn, DE), and Emil Prodan (New York, USA). Together with several other international experts, the speakers reported on cutting-edge research, recently developed in the fields of interest to the program.

The more topical advancements and applications of *KK*-theory were debated both alongside the presentations and during dedicated discussion sessions, which served as well as a means to outline future lines of investigation for the communities which came together during this program.

The first discussion session was devoted to ‘*KK*-theory and gauge theory’ (see

box ‘Gauge theory’), chaired by Gianni Landi (Trieste, IT). The discussion concerned both open problems and applications of *KK*-theory, including the interpretation of *KK*-cycles as morphisms in the category of C^* -algebras, but also uniqueness of *KK*-theory and connections with the theory of extension of C^* -algebras. On the mathematical physics side, various applications were discussed, including the use of factorization in *KK*-theory to decompose space-time into hypersurfaces.

Ralph Meyer (Göttingen, DE) presided over a second discussion session focused on ‘*KK*-theory and topological phases’. The interdisciplinary atmosphere of the conference aroused the curiosity of the mathematicians in the audience for the physical aspects of topological quantum matter. For example, the discussion dealt with the possibility of measuring the invariants produced by *KK*-theory in the lab, in the spirit of the observation of Chern number as quantum Hall conductivity. The challenge for experimentalists and

theoreticians alike lies in particular in the measurement of bulk torsion invariants: the topology of $\mathbb{Z}_2$ insulators is usually probed through the edge states of the system.

The discussion sessions reflected several aspects of problems being currently investigated by participants in the workshop, and fostered both new and ongoing collaborations.

With its comfortable facilities and ample office space, the Lorentz Center proved to be a more than adequate accommodation for this kind of event, allowing and encouraging discussions and collaboration between the participants, and taking the ‘bureaucratic’ load off of the shoulders of the scientific organizers, under the motto “You focus on the science — we do the rest!” ☺

Acknowledgement

We would like to thank the organizers of the program and all participants. We are especially thankful to W. van Suijlekom for comments on an earlier version of this report.

References

- 1 M. F. Atiyah, *K-theory and Reality*, *Q. J. Math.* 17 (1966), 367–386.
- 2 M. F. Atiyah, N. J. Hitchin, I. M. Singer, Self-duality in four-dimensional Riemannian geometry, *Proc. Roy. Soc. London A* 362 (1978), 425–461.
- 3 M. F. Atiyah, I. M. Singer, The index of elliptic operators on compact manifolds, *Bull. Amer. Math. Soc.* 69 (1963), 422–433.
- 4 S. Baaj, P. Julg, Théorie bivariante de Kasparov et opérateurs non bornés dans les C^* modules hilbertiens, *CR Acad. Sci. Paris Sér. I Math* 296 (1983), 875–878.
- 5 J. Bellissard, H. Schulz-Baldes and A. van Elst, The non commutative geometry of the quantum Hall effect, *J. Math. Phys.* 35 (1994), 5373–5471.
- 6 C. Bourne, A. L. Carey and A. Rennie, The bulk-edge correspondence for the quantum Hall effect in Kasparov theory, *Lett. Math. Phys.* 105 (2015), 1253–1273.
- 7 C. Bourne, A. L. Carey and A. Rennie, A non-commutative framework for topological insulators, *Rev. Math. Phys.* 28 (2016), 1650004.
- 8 C. Bourne, J. Kellendonk and A. Rennie, The K -theoretic bulk-edge correspondence for topological insulators, *Ann. Henri Poincaré* 18 (2017), 1833–1866.
- 9 S. Brain, B. Mesland and W. van Suijlekom, Gauge theory for spectral triples and the unbounded Kasparov product, *J. Noncommut. Geom.* 10 (2016), 135–206.
- 10 Ch.-K. Chiu, J. C. Y. Teo, A. P. Schnyder and Sh. Ryu, Classification of topological quantum matter with symmetries, *Rev. Mod. Phys.* 88 (2016), 035005.
- 11 A. Connes, Gravity coupled with matter and the foundation of Noncommutative Geometry, *Commun. Math. Phys.* 182 (1996), 155–176.
- 12 A. Connes, *Noncommutative Geometry*, Academic Press, 1994.
- 13 L. Fu and C. L. Kane, Time reversal polarization and a $\mathbb{Z}_2$ adiabatic spin pump, *Phys. Rev. B* 74 (2006), 195312.
- 14 L. Fu, C. L. Kane and E. J. Mele, Topological insulators in three dimensions, *Phys. Rev. Lett.* 98 (2007), 106803.
- 15 P. B. Gilkey, *Invariance Theory, the Heat Equation, and the Atiyah-Singer Index Theorem*, CRC Press, 1995, 2nd edition.
- 16 G. M. Graf, Aspects of the integer quantum Hall effect, in: *Spectral Theory and Mathematical Physics: A Festschrift in Honor of Barry Simon's 60th Birthday*, F. Gesztesy, P. Deift, C. Galvez, P. Perry and W. Schlag, eds., Proceedings of Symposia in Pure Mathematics, No. 76, American Mathematical Society, 2007, pp. 429–442.
- 17 M. Z. Hasan, S.-Y. Xu and M. Neupane, Topological insulators, topological Dirac semimetals, topological crystalline insulators, and topological Kondo insulators, in: *Topological Insulators, Fundamentals and Perspectives*, F. Ortmann, S. Roche and S. O. Valenzuela, eds. Wiley-VCH Verlag, 2015.
- 18 N. Higson, A primer on KK -theory, in: *Operator Theory: Operator Algebras and Applications, Part 1*, AMS, 1990, pp. 239–283.
- 19 J. Kaad and M. Lesch, Spectral Flow and the unbounded Kasparov product, *Adv. Math.*, 248 (2013), 495–530.
- 20 C. L. Kane and E. J. Mele, $\mathbb{Z}_2$ topological order and quantum spin Hall effect, *Phys. Rev. Lett.* 95 (2005), 146802.
- 21 G. G. Kasparov, The operator K -functor and extensions of C^* algebras, *Math. USSR Izv.* 16 (1980), 513–572.
- 22 J. Kellendonk, On the C^* -algebraic approach to topological phases for insulators, *Ann. Henri Poincaré* 18 (2017), 2251–2300.
- 23 A. Kitaev, Periodic table for topological insulators and superconductors, *AIP Conference Proceedings* 1134 (2009), 22.
- 24 N. P. Landsman, De indexstelling van Atiyah en Singer, *Nieuw Archief voor Wiskunde* 5/5 (2004), 207–211.
- 25 H. B. Lawson and M. L. Michelsohn, *Spin Geometry*, Princeton University Press, 1989.
- 26 B. Mesland, *Bivariant K -theory of Groupoids and Noncommutative Geometry of Limit Sets*, Ph.D. Thesis.
- 27 B. Mesland, Unbounded bivariant K -theory and correspondences in noncommutative geometry, *J. Reine Angew. Math.* 691 (2014), 101–172.
- 28 B. Mesland, A. Rennie, Nonunitary spectral triples and metric completeness in unbounded KK -theory, *J. Funct. Anal.* 271 (2016), 2460–2538.
- 29 E. Prodan, *A Computational Non-Commutative Geometry Program for Disordered Topological Insulators*, SpringerBriefs in Mathematical Physics, No. 23, Springer, 2017.
- 30 E. Prodan and H. Schulz-Baldes, Bulk and boundary invariants for complex topological insulators, *Mathematical Physics Studies*, Springer, 2017.
- 31 Sh. Ryu, A. P. Schnyder, A. Furusaki and A. W. W. Ludwig, Topological insulators and superconductors: tenfold way and dimensional hierarchy, *New J. Phys.* 12 (2010), 065010.
- 32 G. Ch. Thiang, On the K -theoretic classification of topological phases of matter, *Ann. Henri Poincaré* 17 (2016), 757–794.
- 33 W. D. van Suijlekom, The spectral model of particle physics, *Nieuw Archief voor Wiskunde* 5/15 (2014), 240–250.
- 34 W. D. van Suijlekom, *Noncommutative Geometry and Particle Physics*, Springer, 2015.

Jan Beuving

cabaretier, Zeist
janbeuving@gmail.com



Het keerpunt van Marc van Loo

Wiskunde geeft je een onmetelijk zelfvertrouwen

Na zijn studie wiskunde in Nijmegen ging Marc van Loo een jaar backpacken, om vervolgens te promoveren aan de Universiteit van Amsterdam. Maar toch ging hij zijn droom achterna: hij bouwde Loola Adventure Resort op het Indonesische eiland Bintan, vlakbij Sumatra, met een adembenemend uitzicht op de Zuid-Chinese Zee.

Een adventure resort, dat is niet het eerste waar ik aan denk bij een wiskundige. Hoe is het zo gekomen?

“Tijdens het backpacken werd ik verliefd op Zuidoost-Azië en vooral op Sumatra. Toen al wist ik dat ik daar ooit een resort wilde bouwen. De mensen hebben er een natuurlijke gastvrijheid. Die gastvrijheid, in een iets moderner jasje, wilde ik graag aan de wereld laten zien. Bovendien waren er daar oude opaatjes die je voor 200 gulden de schitterendste lappen grond aanboden. Ik voelde gelijk dat er een goede onderneming in zat.”

Toch ging je eerst nog promoveren.

“Ik zat op de Filipijnen toen ik hoorde dat ik een baan als aio in Amsterdam kon krijgen. Ik moest halsoverkop terug. Ik had helemaal geen kamer, en ben toen nachtportier geworden van een hotel op de Warmoesstraat, midden op de Wallen. In ruil voor dat werk kreeg ik een kamer in het hotel.”

Wat voor onderzoek deed je?

“Ik promoveerde bij Eduard Looijenga, en die gaf me heel erg de vrije hand. Ik wilde heel graag het Poincaré-vermoeden bewijzen, en daar heb ik me vol overgave op gestort. Maar na anderhalf jaar kwam ik erachter dat ik niks was opgeschoten.

Bovendien wilde ik ook terug naar Sumatra, want wiskunde kun je overal doen, dacht ik. Toen heb ik de Universiteit van Singapore aangeschreven, en daar heb ik mijn onderzoek voortgezet.

Singapore was in die tijd nog wildwest. Alles kon. Om in mijn levensonderhoud te voorzien gaf ik onder andere wiskundeles op internationale scholen, waar je fenomenale salarissen kreeg voor weinig werk. Dat geld spaarde ik om het resort te gaan bouwen. Ik zat de helft van de tijd overigens op het eiland Bintan,

anderhalf uur varen van Singapore. Ik had daar een eigen hangmatje in een van de guesthouses, en in die hangmat heb ik het grootste deel van mijn promotieonderzoek gedaan: een axiomatische onderbouwing van de relativiteitstheorie.”

Kun je die grondslagen axiomatisch benaderen?

“Veel mensen vertelden mij dat axiomatische grondslagen van de natuurkunde een hopeloos terrein was. Maar als mensen tegen me zeggen dat iets niet kan, dan wil ik juist laten zien dat het wel kan.”

Je zou een goede trainer van het Nederlands voetbalelftal zijn.

“Haha! Er moet wel een eerlijke kans van slagen zijn.”

Maar die eerlijke kans was er dus wel voor die axiomatische grondslagen?

“Ja. Het leuke van wiskunde is, zeker in de complexe meetkundige hoek: als je het goed doet, krijg je er een onmetelijk zelfvertrouwen van. Alles wat met nadenken te maken heeft, dat kan ik, dacht ik toen. Dus als de natuurkundige grondslagen niet duidelijk zijn, dan kon ik die in orde brengen. Die arrogantie had ik wel.”

Is het echt gelukt?

“Ik vind van wel. Ik heb wel heel ongebruikelijke conclusies in mijn thesis. Maar Looijenga heeft het doorgelezen, dus wiskundig klopt het in elk geval, en Jayant Narlikar, een beroemde kosmoloog, vond



Marc van Loo

het ook heel goed. Ik ben er toen wel achter gekomen dat het toch uitmaakt wáár je je wiskunde doet. Want toen ik op het eind van mijn promotie wilde publiceren was ik die jongen van een universiteit in Singapore, en Singapore was, zeker in die tijd, een wiskundig achtergebleven gebied. Ik werd niet serieus genomen. Dat kwam ook door mijn eigenwijsheid. Narli- kar had me heel goed advies gegeven. Hij zei: publiceer nu eerst je puur meetkun- dige resultaten, die zijn heel sterk, dan bouw je een naam op, en dan volgen de kosmologische implicaties. Maar ik wilde meteen het meest spectaculaire resultaat laten zien. Dat staat echter lijnrecht te- genover de gebruikelijke relativiteitsthe- orie, dus de reviewers zeiden: deze man heeft de relativiteitstheorie gewoon niet begrepen.”

Is dat het moment dat je afgeknapt bent op de wetenschap?

“Nee, helemaal niet! Tot op de dag van vandaag houd ik van die grondslagen. Ik heb laten zien dat je speciale relati- viteitstheorie kunt afleiden uit twee heel basale aannames. Je begint met een echt universum. Dus niet zoals Einstein — tra- ditionele verklaringen beginnen altijd met een ‘leeg’ universum, maar wat dan wel ineens vierdimensionaal en lineair zou moeten zijn. Dus zo leeg is het nou ook weer niet. Dat klopt niet. Ik vind dat je moet beginnen met het echte univer- sum, dat globaal homogeen en isotroop is, dus dat er in alle richtingen hetzelfde uitziet. Als je dát aanneemt, en je neemt aan dat je oneindig blokken kunt stape- len, dus kubussen op elkaar kunt zetten, dan komt daar wiskundig uit dat je maar twee mogelijke universa hebt: of je hebt een statisch newtoniaans universum, óf je hebt speciale relativiteitstheorie. Dat vond ik de mooiste conclusie. Je krijgt het cadeau uit twee simpele beginselen! Maar je hebt dan wel ook meteen een universum dat met constante snelheid uitdijt, en dat rijmt niet met de resultaten van de algemene relativiteitstheorie. En dan kom je er bij de reviewers niet in.”

En dus heeft niemand jouw spectaculaire resultaten ooit kunnen lezen.

“Nee. Vooral door mijn eigenwijsheid hoor. Ik bleef maar proberen om gelijk

de meest controversiële conclusies te publiceren. Ik heb ook geprobeerd mijn wiskundig bewijs te slijten dat de licht- snelheid niet constant kan zijn. Niet in een echt universum. En op een draaiende aarde is het onmogelijk om alle klokken gelijk te zetten als je de lichtsnelheid in verschillende richtingen constant wilt houden; je kunt wiskundig laten zien dat je op onze aarde geen natuurlijke tijd kunt creëren. Inmiddels ben ik natuurlijk wiskundig veel minder scherp dan toen, maar ik zou het nu wel veel beter aan de man kunnen brengen. Wiskunde is ook sales, in zeker opzicht. Maar wiskunde is vooral zo mooi. Ik ben er nog steeds van overtuigd dat het ook ten grondslag ligt aan de natuurkunde.”

Heb je iets van die wiskunde grondigheid ook meegenomen in je huidige werk?

“Zeker. Wiskunde is de beste manier om je voor te bereiden op een carrière in de business. Ten eerste leert wiskunde je te concentreren op de onderliggende reali- teit. Een concreet probleem is vaak het uiteinde van iets veel fundamentele. Wis- kunde leert je door de bomen het bos te zien. Ten tweede geeft wiskunde een diep gevoel van waarheid van getallen. Ieder project zit vol met getallen. Tachtig pro- cent van alle business om me heen loopt stuk omdat mensen geen goed gevoel voor getallen hebben. Ze gooien wat data in een of ander achterlijk Excel-template en dan geloven ze het verder wel.

Verder leert wiskunde je dat een idee dat niet op de achterkant van een en- velop past, geen goed idee is. De beste ideeën zijn de kortste ideeën. En ten vier- de maakt de wiskunde je eerlijk. In deze regio heeft men de neiging om fouten te camoufleren, maar een wiskundige weet dat fouten maken geweldig leerzaam is. Als je namelijk gedegen werkt maak je nooit een domme fout, en van ‘slimme’ fouten leer je een boel.

Mensen zitten trouwens helemaal niet wiskundig in elkaar. Als je met mensen werkt, moet je op je ervaring afgaan. En toch kun je ook bij mensen gemene de- lers vinden. De belangrijkste is denk ik dat iedereen gewaardeerd wil worden. Iets anders is dat je mensen moet stimu- leren: als dingen niet veranderen, gaan mensen zich vervelen.”

Je resort is ook constant in ontwikkeling, toch?

“Ja. Onze filosofie is dat we met alleen maar locals werken. Toen we begonnen, heb ik iedereen die wilde werken aange- nomen. Slechts één van hen kon een bé- tje Engels. In de loop van de jaren is het resort organisch gegroeid. We hebben ons vanaf het begin toegelegd op ecologisch verantwoord ondernemen. Inmiddels zijn we gekroond tot het beste ecoresort van de wereld.

De samenwerking met de lokale bevol- king hebben we ook geïntensiveerd. We bieden gasten die bij ons logeren aan om, voor 20 euro extra per dag aan materiaal- kosten, mee te helpen om bijvoorbeeld scholen op te knappen of wegen aan te leggen. Tijdens hun vakantie gaan ze dan overdag aan de slag om die projecten te realiseren. Dat is een win-winsituatie: de toeristen vinden het geweldig, en de lo- kale community ook.”

Wat zijn je eigen taken binnen het resort?

“Het resort is inmiddels bijna helemaal zelfstandig. Ik woon ook niet op Bintan, maar in Singapore. Ik heb mijn bakens in- middels verzet. Gebruikmakend van onze awards is mijn doel voor de komende vijf jaar om Indonesië van schoon drinkwater te voorzien. In Indonesië sterven dagelijks 370 mensen door gebrek aan goede sani- taire voorzieningen. In samenwerking met de TU Eindhoven en drie universiteiten uit deze regio zijn we bezig om de oplossing van een zogeheten Safe Water Garden aan te bieden voor een betaalbaar bedrag. Mensen kunnen dan in hun eigen tuin voor 200 euro een installatie aanleggen die hun afvalwater veilig recyclet. Samen met die universiteiten willen we het zo simpel ma- ken dat iedereen het zelf aan kan leggen.”

En denk je nog wel eens aan je axioma- tische grondslagen?

“Ik lees nog ieder jaar één keer mijn thesis door, om te kijken of ik het nog begrijp. Ooit wil ik de resultaten echt pu- bliceren.”

Kijk voor meer informatie over Loola Adventure Resort op www.loola.net.

Goede suggesties voor een Nederlandse wis- kundige met een keerpunt in zijn of haar carri- ère zijn welkom via keerpunt@nieuwarchief.nl.

Frans Schurer

Faculteit Wiskunde en Informatica
Technische Universiteit Eindhoven
f.schurer@tue.nl

Klaas van Harn

Afdeling Wiskunde
Faculteit der Exacte Wetenschappen
Vrije Universiteit Amsterdam
k.van.harn@vu.nl

In Memoriam Fred Steutel (1931–2017)

Wiskundige, columnist, ingezonden brieven-schrijver en sportman

Op 1 juni 2017 overleed Fred Steutel, emeritus hoogleraar van de Technische Universiteit Eindhoven, in de leeftijd van 85 jaar. Frans Schurer en Klaas van Harn beschrijven zijn wetenschappelijke carrière, zijn schrijverschap, zijn vriendschap en zijn sportactiviteiten. Dit levensbericht is eerder in een wat andere vorm verschenen in *STATOR* 18(3) (2017), 23–31.

Proloog

Op 1 juni 2017 is Frederik Willem (Fred) Steutel, volstrekt onverwacht, vredig in zijn slaap overleden — *media vita in morte sumus*, zoals de eerste regel van een middeleeuwse antifoon luidt. Freds wiskundige interesse — hij was emeritus hoogleraar van de Faculteit Wiskunde en Informatica van de Technische Universiteit Eindhoven (TU/e) — betrof vooral de kansrekening, maar ook binnen de analyse was hij actief; met name de *approximatietheorie* mocht zich in zijn belangstelling verheugen. Binnen de kansrekening raakte Fred gaandeweg steeds meer geïnteresseerd in de zogenoemde *oneindige deelbaarheid* van kansverdelingen, een in feite heel eenvoudig concept, dat echter tot vele uitdagende vragen aanleiding geeft en vaak onverwacht in andere delen van die discipline opduikt. In het navolgende levensbericht beschrijven we zijn wetenschappelijke carrière en, heel globaal, het soort problemen in de kansrekening en de approximatietheorie waarmee hij zich bezighield; in een aantal stukjes aan het eind van dit artikel wordt daarop meer in detail ingegaan. Daarnaast geven we ook een schets

van zijn schrijverschap als columnist, en zeggen we iets over de aard van zijn vriendschap en zijn sportactiviteiten. Als achtergrond moge dienen dat we beiden jarenlang met Fred hebben samengewerkt en het voorrecht hadden tot zijn intieme

vriendenkring te behoren: Schurer was vanaf 1964 — eerst op de Technische Hogeschool Twente (THT) en daarna op de Technische Hogeschool Eindhoven (THE), later de TU/e — een directe collega van Fred, en Van Harn verrichtte onderzoek op het gebied van de oneindige deelbaarheid, van 1973 tot 1978 als Freds eerste promovendus op de THE, daarna als collega ‘op afstand’ op de Vrije Universiteit Amsterdam.



Fred Steutel

Wetenschappelijke carrière

Fred werd geboren in Tubbergen op 17 november 1931. Hij bezocht de lagere school in Haaksbergen en ‘genoot’ vervolgens — dat was gedurende het laatste oorlogsjaar; van het onderwijs kwam weinig terecht — één jaar Uitgebreid Lager Onderwijs aldaar. Daarna bezocht hij het gymnasium in Hengelo. Vervolgens, na het behalen van het einddiploma, ging hij twee jaar in militaire dienst. Zijn initiële studiekeuze aan de Universiteit van Amsterdam (UvA), scheikunde en natuurkunde met bijvak wiskunde, was geen succes: van de scheikundepractica bracht Fred weinig terecht en ook de natuurkundecolleges interesseerden hem maar matig. Na vier maanden zei hij de scheikundestudie vaarwel en kwam het accent op wiskunde en natuurkunde te liggen, met als bijvak sterrenkunde. Na het behalen van het kandidaatsexamen in 1956 moest gekozen worden tussen wiskunde en natuurkunde. Het advies van J. de Groot, hoogleraar topologie, gaf de doorslag; hij beoordeelde Freds aanleg voor wiskunde als ‘zeer goed, maar niet briljant’, een kwalificatie waarin Fred zich, ook later, goed kon vinden. Het doctoraal werd behaald in 1961, maar reeds van 1956 af was hij werkzaam bij de Statistische Afdeling van het Mathematisch Centrum, het huidige Centrum Wiskunde & Informatica.

In 1964 deed zich de gelegenheid voor terug te keren naar zijn geboortestreek. Fred werd, met een zevental anderen — onder wie D.W. Bresters, de latere rector magnificus van de UvA gedurende 1981–1987, en Schurer —, op voorstel van I.W. van Spiegel, benoemd tot wetenschappelijk medewerker aan de in november 1961 opgerichte THT; de eerste colleges, oefeningen en practica werden gegeven in september 1964. Na Van Spiegel werden, in de beginjaren, achtereenvolgens benoemd de hoogleraren A.J.W. Duijvestijn (numerieke wiskunde), T.J. Terpstra (stochastiek), E.M. de Jager (mathematische fysica) en P.J. Zandbergen (toegepaste wiskunde). Allengs werden de medewerkers over de leerstoelen verdeeld en Fred — dat lag voor de hand — werd toegevoegd aan die van Terpstra. Het is wellicht interessant te vermelden dat de wiskundigen destijds bestuurlijk waren ondergebracht bij één der afdelingen Elektrotechniek, Chemische Technologie



Fred tijdens zijn promotie op 20 januari 1971, met links zijn promotor Theo Runnenburg

en Werktuigbouwkunde, en niet, zoals in de beginjaren in Delft en Eindhoven, een onderafdeling waren van de afdeling Algemene Wetenschappen.

Gedurende het studiejaar 1969–1970 doceerde en studeerde Fred aan de University of Texas at Austin. In 1973 werd hij, mede op voorspraak van de hoogleraren R. Doornbos en J. Wessels, benoemd aan de THE tot lector in de wiskunde, in het bijzonder kansrekening en statistiek; in 1980 werd Fred hoogleraar. Intussen (1971) was hij gepromoveerd aan de UvA op een proefschrift [17] over oneindige deelbaarheid van kansverdelingen; J.Th. (Theo) Runnenburg was zijn promotor. Gaandeweg ontwikkelde Fred zich tot een voor- aanstaand wetenschapper; zo was hij gedurende de zomers van 1970 (aansluitend op het verblijf in Austin) en 1974 ‘research associate’ in Rochester en in 1979 gasthoogleraar aan de Johns Hopkins University in Baltimore.

In 1996 ging Fred met emeritaat. In zijn afscheidscollege *Laatste kansen* [18], een terugblik op de afgelegde weg, geeft hij zich onder andere rekenschap van zijn bijdragen aan onderwijs, onderzoek en bestuur. Hoewel Fred in veel commissies binnen en buiten de faculteit heeft gezeten, stelt hij, kennelijk zonder spijt, vast dat hij zich niet tot een bestuurder heeft

ontwikkeld. Over zijn functioneren als docent is hij niet bijzonder tevreden (cf. [18, p.17] en dat geldt dan met name het geven van grootschalig onderwijs; inderdaad, door zijn niet vêrdragende en soms wat haperende stem, door zijn bescheidenheid en zich wars te betonen van toneel en theater, was hij minder geschikt voor colleges voor grote aantallen studenten. Toch gaf hij graag onderwijs. Maar zonder twijfel lag Freds grootste kracht in het onderzoek en hij heeft, ook in samenwerking met anderen (voor wie hij vaak een stimulans was), daaraan veel vreugde en voldoening ontleend. Zijn wetenschappelijk werk omvat, behalve het ‘oneindig deelbaar boek’ [20], ongeveer zeventig artikelen, congresverslagen en rapporten; verder heeft hij een groot aantal voordrachten gehouden, in Oberwolfach natuurlijk, maar ook in exotische plaatsen als Hanoi in Vietnam, Isfahan in Iran en Tashkent in Oezbekistan. Na Van Harn heeft Fred nog drie andere promovendi gehad en hij was vier keer tweede promotor.

Fred hield van ‘sometjes maken’; hij was lid van de roemrijke Eindhovense groep O.P. Lossers, twaalf jaar redacteur van de Opgavenrubriek in *Statistica Neerlandica*, het tijdschrift van de Vereniging voor Statistiek en Operationele Research — in 2008 werd hij benoemd tot erelid —, en nadien

redacteur van de soortgelijke rubriek in het *Nieuw Archief voor Wiskunde*.

Bijdragen aan de approximatietheorie

Zoals eerder is aangegeven, hebben Fred en de eerste auteur van dit levensbericht elkaar medio 1964 voor het eerst ontmoet. Fred, kansrekenaar van professie, met een grondige kennis van de analyse, ging zich, in het voetspoor van Schurers lopende promotieonderzoek, ook interesseren voor problemen in de *approximatietheorie*, hetgeen voor beiden leidde tot een jarenlange vruchtbare samenwerking: samen hebben ze, naast een aantal rapporten, dertien artikelen geschreven. Centraal daarin staat de bepaling van de *graad van approximatie* van continue functies met welomschreven gladheidseigenschappen door algebraïsche polynomen, waaronder die van de Russische wiskundige Bernstein (1880–1968); cf. [15]. Voor continue 2π -periodieke functies met analoge eigenschappen werd soortgelijk onderzoek verricht en dan met name voor singuliere integralen, genoemd naar de Amerikaanse wiskundige Jackson (1888–1946); cf. [14]. In een tweetal stukjes aan het eind van dit artikel worden de belangrijkste resultaten gegeven uit de genoemde artikelen.

De operatoren die in de genoemde approximatiemethoden als intermediair fungeren, dat wil zeggen die aan de beschouwde functie de bijbehorende benadering toevoegen, zijn, behalve lineair, *positief*: het beeld van een niet-negatieve functie is ook weer niet-negatief op het beschouwde interval. Dergelijke operatoren zijn, dankzij het baanbrekende werk van de Russische wiskundige Korovkin [8], vanwege hun elegante eigenschappen sinds de jaren vijftig van de vorige eeuw onderwerp geweest van talrijke publicaties; Schurers dissertatie [13] is daar een voorbeeld van.

Expert in oneindige deelbaarheid

Fred kon zijn kennis van de analyse goed gebruiken in zijn eigenlijke vakgebied, de (analytische) kansrekening. Daarbinnen richtte hij zich met name op *oneindige deelbaarheid* van kansverdelingen, een begrip dat in de jaren dertig van de vorige eeuw door bekende kansrekenaars als De Finetti, Kolmogorov, Lévy en Khintchine ontwikkeld was in de context van (continue-tijd) *processen met stationaire en onafhankelijke aangroeiingen* en

het *Centrale Limietprobleem*. Overigens is er, naast dit theoretisch aspect, ook een meer praktisch belang van oneindige deelbaarheid in (stochastische) modelering binnen de verzekeringswiskunde (jaarlijkse schadeclaim) en de biologie (regenval, landopbrengst), zoals duidelijk zal zijn uit de volgende definitie. Een stochastische variabele X wordt *oneindig deelbaar* (eigenlijk beter: *onbeperkt deelbaar*) genoemd indien X voor ieder natuurlijk getal n dezelfde verdeling heeft als een som van n onafhankelijke, gelijkverdeelde grootheden; de kansverdeling van X wordt dan ook *oneindig deelbaar* genoemd. Van vele bekende verdelingen, zoals de normale, Cauchy-, Laplace-, gamma-, negatief-binomiale en Poisson-verdeling, is de oneindige deelbaarheid gemakkelijk aan te tonen; maar dat ligt geheel anders voor bijvoorbeeld de in de financiële wiskunde populaire log-normale verdeling.

In de jaren dertig waren ook canonieke voorstellingen voor de karakteristieke functies van de oneindig deelbare verdelingen afgeleid. Maar bij het vaststellen of een gegeven verdeling oneindig deelbaar is, helpen deze voorstellingen doorgaans niet. Fred was, samen met Charles Goldie, één van de eersten die, in de jaren zestig en zeventig, nuttige criteria voor oneindige deelbaarheid ontwikkelde in termen van de kansverdelingen zelf. Daarmee toonde hij bijvoorbeeld aan dat alle *mengsels van exponentiële dichtheden* en, algemener, alle *log-convexe* dichtheden op de positieve halfrechte oneindig deelbaar zijn. Freds aanpak gaf aanleiding tot vele aanvullende resultaten en generalisaties, zoals die te vinden zijn in [17] en [20].

Fred ontwikkelde zich tot een deskundige op het gebied van oneindige deelbaarheid. Zijn brede kennis stelde hem bovendien in staat ongedachte verbanden te leggen. We willen daar twee voorbeelden van geven, die we hier kort aanstippen en in twee stukjes aan het eind van dit artikel nader uitwerken. In het promotieonderzoek van de tweede auteur van deze bijdrage kwamen functionaalvergelijkingen voor die Fred op het lumineuze — en, naar later bleek, zeer vruchtbare — idee brachten om op zoek te gaan naar *geheelwaardige fracties* van niet-negatieve, geheelwaardige stochastische variabelen X . Er bleken uit-

eindelijk vele mogelijkheden te zijn, maar de eenvoudigste variant is die waarbij bijvoorbeeld de helft van 1 niet gelijk is aan $\frac{1}{2}$ (en ook niet langer constant), maar een stochastische variabele is met een alternatieve ($\frac{1}{2}$) verdeling op $\{0,1\}$: met kans $\frac{1}{2}$ gelijk aan 0 en met kans $\frac{1}{2}$ gelijk aan 1. Met behulp van deze vermenigvuldiging konden vervolgens discrete analogons gedefinieerd en bestudeerd worden van de bekende (oneindig deelbare, absoluut-continue) *stabiele en zelf-ontbindbare* verdelingen. Een en ander is, behalve in [3] en [20], te vinden in het basisartikel [19], dat zeer veel vervolgonderzoek heeft gegenereerd. Fred hield graag de citatiescore van het artikel bij; volgens Google Scholar is deze score nu bijna vijfhonderd. In deze fase werkte Fred, behalve met Van Harn, soms samen met Wim Vervaat en met Jim Wolfe, die overigens beiden een tragisch levenseinde hebben gekend, en later met Lennart Bondesson en Ken-iti Sato; ook waren er stimulerende contacten met Theo Runnenburg, Piet Holewijn, Guus Balkema en Laurens de Haan.

Een tweede ongedacht verband dat we willen noemen, is een mooi voorbeeld van wederzijdse bevruchting van onderwijs en onderzoek, waar Fred naar eigen zeggen veel plezier aan heeft beleefd. Het betreft hier de bekende *inspectieparadox*, ook wel *wachttijdparadox* genoemd, in een zogenoemde vernieuwingscontext. Door de daarin vigerende *generieke levensduur* X geschikt *oneindig deelbaar* te kiezen kan men laten zien (cf. [4]) dat de paradox een ‘onbeperkt’, en daarmee *extreem*, aspect heeft.

In de proefschriften van Fred zelf en zijn promovendi Klaas van Harn, Björn Hansen en Roel Wilms komen slechts bepaalde aspecten van oneindige deelbaarheid aan bod — Aegle Hoekstra promoveerde bij Fred op een ander onderwerp. Maar in het begin van de jaren negentig, ongeveer vijf jaar vóór hij met emeritaat ging, werd Fred door de redactie van uitgever Marcel Dekker (New York) benaderd om in de bekende serie *Pure and Applied Mathematics* een monografie te schrijven over zijn onderzoek. Fred stemde met enige aarzeling toe en haalde Van Harn over aan dit project mee te doen met de woorden: “We schrijven wat leuke dingen op waar we iets van af weten.” Maar al snel stelden zij hun doelen bij en streefden ze naar een zo volledig mogelijke behande-

ling van de oneindig deelbare kansverdelingen op de reële rechte. Het was echter een gigantisch werk om de ontwikkelingen van vooral de laatste 25 jaar te ordenen en te verwerken in een toegankelijke tekst. Uiteindelijk heeft dit project zo'n twaalf jaar geduurd, eigenlijk veel te lang. Maar in 2004 verscheen dan toch de monografie [20] van 550 pagina's, met ruim honderd concrete voorbeelden, verspreid over zes van de zeven hoofdstukken. Het boek wordt veelvuldig aangehaald — volgens Google Scholar nadert de citatiescore de vierhonderd — en lijkt intussen te zijn uitgegroeid tot een standaardwerk over dit onderdeel van de kansrekening. Fred heeft zo zijn wetenschappelijke carrière op een indrukwekkende wijze kunnen afsluiten.

Affiniteit met taal

Bij het samenstellen van dit herdenkingsartikel, tevens bedoeld als eerbetoon aan Fred, hebben we zijn afscheidscollege nog weer eens geraadpleegd. Het verging ons daarbij als bij het herlezen van sommige boeken van Willem Elsschot, *Lijmen* en *Het been* bijvoorbeeld. De rede heeft nog niets aan frisheid ingeboet, is, door de finzinnige humor en de puntige formuleringen, een genot om te lezen, bovendien een feest der herkenning omdat we in Eindhoven gedurende lange tijd tot dezelfde faculteit behoorden. Hier openbaarde zich een gave van Fred: de kunst om lenig proza te schrijven, ontdaan van opsmuk, beknopt, boeiend, met verrassende wendingen. Dat hij goed kon schrijven, dat hij daar plezier in had, was eerder al gebleken: regelmatig kwam men bijdragen van hem tegen in *de Volkskrant* en in *NRC Handelsblad*. Maar de drempel om een ingezonden brief in die kranten geplaatst te krijgen is hoog; hij had behoefte aan een vast podium. Zo nu en dan schreef hij ook stukjes in *Cursor*, de TU/e-periodiek, en rond de eeuwwende kreeg hij een eigen rubriek.

Fred behandelde — zijn belezenheid hielp hem daarbij — een breed scala aan onderwerpen: politiek, taal en literatuur, onderwijs en onderzoek, de opmars van het Engels in het wetenschappelijk onderwijs, waarvan hij een fervent tegenstander was, en 'Effe zeuren' was jarenlang een sieraad voor het blad. Van één zijner columns willen we hier iets zeggen, ook al omdat het Freds bedrevenheid in het maken van

limericks illustreert; het rijmpje heeft betrekking op een vermaard wiskundige. Ter inleiding het volgende. In het begin van de jaren tachtig werd door W.J. Deetman, minister van Onderwijs en Wetenschappen, de pensioengerechtigde leeftijd van hoogleraren verlaagd van 70 naar 65 jaar. Ongeveer twintig jaar later, in 2005, komt Fred daarop terug. In de bewuste column (cf. [7]) hekelt hij het benoemen tot hoogleraar aan onze universiteiten van Beken-de Nederlanders die niet of nauwelijks op wetenschappelijke prestaties kunnen bogen — 'de pronkprofessor verkleutert het hoogleraarsambt' — en plaatst daartegenover het gedwongen terugtreden van prominente wetenschappers die nog lang niet uitgedoofd zijn en graag hun onderwijs en onderzoek aan de universiteit willen voortzetten. Fred had hier met name — hoewel, niet letterlijk: in het stukje is sprake van ene professor De Wit — N.G. (Dick) de Bruijn (1918–2012) op het oog, die als gevolg van Deetmans oekaze met tegenzin in 1984 met emeritaat ging. Nu de limerick:

*Geen vis in de wiskundevijver
betwistte zijn inzicht en ijver.
Nu, door Deetman genept,
op het droge geschept,
moet 'ie weg, maar het is toch een blijver.*

Een blijver, inderdaad: De Bruijn was nog jarenlang na zijn terugtreden, samen met Fred en anderen, een trouw bezoeker van 'Huize Avondrood', de ruimte die de faculteit genereus ter beschikking stelde aan haar oud-medewerkers. Enkele jaren nadat hij zijn rubriek bij *Cursor* kreeg, werd Fred medewerker van *STATOR*, een uitgave van de Vereniging voor Statistiek en Operationele Research; zijn eerste column verscheen in december 2002.

Talent voor vriendschap

Naast een grote aanleg voor wiskunde en affiniteit met taal, had Fred 'le talent de l'amitié' en volgens Adriaan Roland Holst (1888–1976), in een stukje [12, p.84] over de dichter en essayist Jan Greshoff, is dat een gave die zeldzamer is dan meestal wordt verondersteld; spottenderwijs noemt hij vriendschap de centrale verwarming van de ziel. Wat ons bij Fred opviel was dat zijn vriendschap zo volstrekt *onbaatzuchtig* was, zo gespeend van mogelijk eigenbelang. Jan Hendrik Leopold (1865–1925), wel eens Neerlands 'meest verstilde dichter'

genoemd, heeft een vriend — het vers is treffend van toepassing op Fred — als volgt gekarakteriseerd [9, p.206]:

*Een vriend is niet, die u aan 't hart wil sluiten
in uw geluksuur en zich niet genoeg doen kan,
maar die den balling bij zich binnen roept en dan
de deur toeslaat tegen de wolven buiten.*

Wij memoreren in dit verband hier ook de bezoeken die hij bracht aan Runnenburg tijdens diens laatste moeilijke en eenzame levensjaren; hart- en zielverwarmend en een illustratie van de 'milk of human kindness' waarvan Fred zo veel in zich had.

Passie voor sport

We hebben samen aan wiskunde gedaan, maar ook veel aan sport; medewerkers van universiteiten staan daartoe uitstekende faciliteiten ter beschikking. Fred hield van schaatsen op de kunstijsbaan in Eindhoven en, als er natuurijs was, van tochten maken met Van Harn op de Ankeveense Plassen en de Gouwzee; ook van tennissen (dat heeft hij tot op hoge leeftijd gedaan), maar joggen deed hij toch het liefst. Met een vast groepje van de TU/e heeft hij duizenden kilometers gelopen, in de middagpauze vanuit het Studentensportcentrum Eindhoven rondjes om de Karpendonkse Plas, halve marathons, trainen voor de eerste marathon in Eindhoven in oktober 1982 (Fred was toen bijna 51 jaar). Die liep hij in de uitstekende tijd van 3.16.01 uur; nadien heeft hij er nog vier gelopen. Aan dat joggen met vrienden, ruim twee decennia lang — hij was erelid van de Karpendonk Road Runners —, heeft hij ongelooflijk veel plezier beleefd.

Epiloog

Uit het bovenstaande moge duidelijk zijn dat Frederik Willem Steutel — wiskundige, columnist, ingezonden brieven-schrijver en sportman — begiftigd was met grote gaven, iemand die zich verheugde in de samenwerking met anderen en zijn kennis en inzichten — al dan niet gevraagd — ruimhartig met hen deelde. Zijn heengaan is een onherstelbaar verlies voor Vita, naar eigen zeggen dé vrouw in zijn leven, en hun twee kinderen, maar ook vele anderen zullen hem node missen. Wij verloren een dierbare vriend.

Approximatie met singuliere Jackson-integralen

Zij $C_{2\pi}$ de verzameling van continue functies op $\mathbb{R}$ met periode 2π . In [14] wordt de graad van approximatie onderzocht van functies f in $C_{2\pi}$ met behulp van de singuliere integralen

$$(L_{n,p}f)(x) := A^{-1}(n,p) \int_{-\pi}^{\pi} f(x+t) \left(\frac{\sin \frac{1}{2}nt}{\sin \frac{1}{2}t} \right)^{2p} dt, \quad n \in \mathbb{N}, \quad (1)$$

waarbij $p \geq 3$ een vast natuurlijk getal is en

$$A(n,p) := \int_{-\pi}^{\pi} \left(\frac{\sin \frac{1}{2}nt}{\sin \frac{1}{2}t} \right)^{2p} dt. \quad (2)$$

Uit de gedaante van (1) volgt dat $L_{n,p}$ een *positieve* lineaire operator is. De keuzes $p=1$ en $p=2$ zijn in (1) uiteraard ook geoorloofd; de naam van Fejér (1880–1959) is verbonden met $L_{n,1}$, terwijl $L_{n,2}$ de klassieke Jackson-operator is [6].

Voor de bepaling van $A(n,p)$ en voor het onderzoek van de approximatie-eigenschappen van $L_{n,p}$ is het navolgende resultaat uit [14], dat ook geldt voor $p=1$ en $p=2$, van cruciaal belang.

Stelling 1. De coëfficiënten $\mu_k^{(n,p)}$ in de ontwikkeling

$$\left(\frac{\sin \frac{1}{2}nt}{\sin \frac{1}{2}t} \right)^{2p} = \mu_0^{(n,p)} + 2 \sum_{k=1}^{np-p} \mu_k^{(n,p)} \cos kt$$

worden gegeven door

$$\mu_k^{(n,p)} = \sum_{j=0}^{2p} (-1)^j \binom{2p}{j} \binom{np+p-k-nj-1}{2p-1}, \quad (3)$$

met de gebruikelijke conventie dat $\binom{a}{b} = 0$ als $a < b$.

Uit (2) en (3) volgt na enig rekenwerk dat

$$\begin{aligned} A(n,p) &= 2\pi \mu_0^{(n,p)} \\ &= \frac{2\pi n}{(2p-1)!} \sum_{j=0}^{p-1} (-1)^j \binom{2p}{j} (p-j) \prod_{k=1}^{p-1} ((p-j)^2 n^2 - k^2). \end{aligned} \quad (4)$$

Daarmee is de expliciete vorm van (1) bepaald. Uit (4) volgt bijvoorbeeld dat $A(n,1) = 2\pi n$ en dat $A(n,2) = 2\pi n(2n^2 + 1)/3$. Opgemerkt zij nog dat $A(n,p)$ een veelterm is in n van oneven graad. Door toepassing van een aantal resultaten van Korovkin (cf. [14, p. 156–157]) over positieve lineaire operatoren van een bepaald type

waartoe ook de $L_{n,p}$ behoren, kunnen uitspraken worden gedaan over de graad van approximatie van functies in $C_{2\pi}$ met behulp van (1). De volgende ongelijkheid, waarin de continuïteitsmodulus $\omega(f; n^{-1}) := \sup\{|f(x) - f(y)| : x, y \in [-\pi, \pi] \text{ met } |x - y| \leq n^{-1}\}$ optreedt, is daarvan een voorbeeld; deze is van *globale* aard en vertoont gelijkenis met de resultaten van Sikkema [16] en van Schurer en Steutel [15], zoals gegeven voor Bernstein-polynomen. Echter, hier is geen sprake van een *beste* approximatieconstante, maar slechts van een bovengrens daarvoor.

Stelling 2. Voor alle $f \in C_{2\pi}$ en alle $p \geq 3$ geldt:

$$\max_{-\pi \leq x \leq \pi} |(L_{n,p}f)(x) - f(x)| \leq \left(1 + \frac{1}{2}\pi\sqrt{2}\right) \omega(f; n^{-1}), \quad n \in \mathbb{N}. \quad (5)$$

Opmerking. Voor $p=2$ dient in het rechterlid van (5) de constante $\sqrt{2}$ te worden vervangen door $\sqrt{3}$ (cf. [11, p. 76]). De graad van approximatie in het geval $p=1$ is essentieel slechter: in het rechterlid van (5) komt dan te staan $\omega(f; n^{-1/2})$ in plaats van $\omega(f; n^{-1})$.

De volgende stelling bevat een resultaat van *lokale* aard.

Stelling 3. Zij gegeven dat $f \in C_{2\pi}$ in een vast punt $x \in (-\pi, \pi)$ een tweede afgeleide heeft, dan geldt voor $p \geq 2$:

$$(L_{n,p}f)(x) - f(x) = a_p f''(x) / n^2 + o(1/n^2), \quad n \rightarrow \infty, \quad (6)$$

waarbij

$$a_p = (2p-1)(p-1) \frac{\sum_{j=0}^{p-1} (-1)^{j+1} \binom{2p}{j} (p-j)^{2p-3}}{\sum_{j=0}^{p-1} (-1)^j \binom{2p}{j} (p-j)^{2p-1}}. \quad (7)$$

Dit zijn zogenaamde asymptotische approximatieformules, vergelijkbaar met het door Voronovskaya [21] gegeven resultaat voor Bernstein-polynomen. Bij de bewijzen van (5), (6) en (7) spelen de coëfficiënten $\mu_k^{(n,p)}$ uit (3) voor $k=0, 1, 2$ een essentiële rol. Uit Stelling 3 volgt in het bijzonder dat geldt (cf. [11, p. 34]):

$$(L_{n,2}f)(x) - f(x) = \frac{3}{2} f''(x) / n^2 + o(1/n^2), \quad n \rightarrow \infty.$$

Frans Schurer

Bernstein-polynomen en beste approximatieconstanten

Zij $C_1[0,1]$ de verzameling van functies die gedefinieerd en continu-differentieerbaar zijn op het interval $[0,1]$. In [15] wordt de graad van approximatie onderzocht van functies in $C_1[0,1]$ met behulp van Bernstein-polynomen, gedefinieerd door

$$(B_n f)(x) := \sum_{k=0}^n f(k/n) \binom{n}{k} x^k (1-x)^{n-k}, \quad x \in [0,1], \quad n \in \mathbb{N}. \quad (1)$$

Uit het rechterlid van (1) blijkt dat B_n een *positieve* lineaire operator is. De polynomen (1) zijn door Bernstein [1] in 1912 gebruikt om een elegant en constructief bewijs te geven van de beroemde

approximatiestelling (1885) van Weierstrass (1815–1897). De resultaten in [15] — de bepaling van beste approximatieconstanten en het asymptotisch gedrag daarvan — zijn vergelijkbaar met die van Sikkema [16] en Esseen [2], die rond 1960 soortgelijk onderzoek hebben verricht voor continue functies. Bij de introductie van de approximatieconstanten hebben we de continuïteitsmodulus ω_1 van de afgeleide van f nodig, gedefinieerd als volgt:

$$\omega_1(f; \delta) := \sup_{|x-y| \leq \delta} |f'(x) - f'(y)|, \quad x, y \in [0,1], \quad \delta > 0.$$

Zij nu

$$\begin{cases} c_n := \sup_{f \in C_1[0,1]} \max_{0 \leq x \leq 1} \frac{n^{1/2} |(B_n f)(x) - f(x)|}{\omega_1(f; n^{-1/2})}, \\ c^{(j)} := \sup_{n \geq j} c_n, \quad j = 1, 2, \\ c := \lim_{n \rightarrow \infty} c_n. \end{cases}$$

Stelling. De constanten $c^{(1)}$, $c^{(2)}$ en c hebben de volgende waarden:

$$c^{(1)} = \frac{1}{4}, \quad (2)$$

$$c^{(2)} = c_5 = \frac{2 \cdot 5^{1/2} - 1}{16} \doteq 0,217008, \quad (3)$$

$$c = (2\pi)^{-1/2} \left(\frac{1}{2} + \sum_{j=1}^{\infty} e^{-2j^2} \right) - 2 \sum_{j=1}^{\infty} j(1 - \Phi(2j)) \doteq 0,207969, \quad (4)$$

$$\text{waarbij } \Phi(x) := (2\pi)^{-1/2} \int_{-\infty}^x e^{-\frac{1}{2}u^2} du \text{ voor } x \in \mathbb{R}.$$

Opmerkingen bij de stelling. Eerder is door Lorentz [10, p.21] bewezen dat $c^{(1)} \leq 3/4$. Het is wellicht instructief (2) te herformuleren in de volgende vorm: voor alle $f \in C_1[0,1]$ geldt

$$\max_{0 \leq x \leq 1} |(B_n f)(x) - f(x)| \leq \frac{1}{4} \omega_1(f; n^{-1/2}) n^{-1/2}, \quad n \in \mathbb{N}, \quad (5)$$

waarbij de constante $\frac{1}{4}$ in het rechterlid van (5) best mogelijk is, dat wil zeggen niet verder verscherpt kan worden. Bij het bewijs van (2) speelt de positiviteit van de operator B_n een essentiële rol. De afleiding van (3) is uiterst complex, vergelijkbaar met die in [16]. Bij het bewijs van (4) wordt gebruik gemaakt van de Berry-Esseen-versie van de Centrale Limietstelling.

Frans Schurer

Discreet zelf-ontbindbare en stabiele kansverdelingen

In het *Centrale Limietprobleem* wordt men op een natuurlijke wijze geleid naar een belangrijke klasse van *oneindig deelbare* kansverdelingen, namelijk die van de zogenoemde *zelf-ontbindbare* verdelingen; voor een stochastische variabele X met een dergelijke verdeling geldt per definitie dat

$$X \stackrel{d}{=} \alpha X + X_\alpha \quad (0 < \alpha < 1; \stackrel{d}{=} \text{betekent gelijkheid in verdeling}), \quad (1)$$

met in het rechterlid X_α onafhankelijk van X . De zelf-ontbindbare verdelingen zijn in allerlei modelleringssituaties heel aantrekkelijk, onder andere vanwege het feit dat ze alle *unimodaal* zijn; deze eigenschap, jarenlang slechts een vermoeden, werd uiteindelijk in 1978 door Yamazato bewezen.

Zij X een $\mathbb{Z}_+$ -waardige (niet-ontaarde) stochastische variabele. Dan kan X natuurlijk niet aan (1) voldoen; het is zelfs zó dat alle zelf-ontbindbare verdelingen noodzakelijk absoluut-continu zijn. We proberen daarom de fractie αX in (1) te vervangen door een *geheelwaardige fractie* $\alpha \odot X$: beschouw daartoe αX als $\alpha + \dots + \alpha$ (X maal) en vervang hierin de α 's door onafhankelijke, alternatief (α) op $\{0,1\}$ verdeelde variabelen $Z_1, Z_2, \dots$, ook onafhankelijk van X ; dan krijgen we

$$\alpha \odot X := Z_1 + \dots + Z_X. \quad (2)$$

Merk op dat voor $n \in \mathbb{Z}_+$ de fractie $\alpha \odot n$ geen constante is, maar een binomiaal (n, α) verdeelde stochast. De operatie $\odot$, die daarom wel *binomial thinning* genoemd wordt, heeft allerlei eigenschappen van een vermenigvuldiging, en $\alpha \odot X$ heeft een verwachting zoals gewenst, namelijk gelijk aan $\alpha \mathbb{E}X$.

Een $\mathbb{Z}_+$ -waardige stochast X wordt nu *discreet zelf-ontbindbaar* genoemd indien, analoog aan (1):

$$X \stackrel{d}{=} \alpha \odot X + X_\alpha, \quad 0 < \alpha < 1, \quad (3)$$

met in het rechterlid X_α onafhankelijk van $\alpha \odot X$. De resulterende discreet zelf-ontbindbare verdelingen blijken alle *oneindig deelbaar* te zijn en *unimodaal*. De *negatief-binomiale* en de *Poisson*-verdeling zijn speciale gevallen. De laatste verdeling is ook *discreet stabiel* in de (analoge) zin dat een bijbehorende variabele X voor zekere $\gamma \in (0,1]$ voldoet aan

$$X \stackrel{d}{=} \frac{1}{n^{1/\gamma}} \odot (X_1 + \dots + X_n), \quad n \in \mathbb{N}, \quad (4)$$

met $X_1, X_2, \dots$ onafhankelijk en $\stackrel{d}{=} X$; neem maar $\gamma = 1$.

Meer informatie over discrete zelf-ontbindbaarheid en stabiliteit is, behalve in [3] en [20], te vinden in het basisartikel [19]. In [5] wordt aangetoond dat alle redelijke geheelwaardige fracties in verdeling gekarakteriseerd worden door compositie-halfgroepen $\mathcal{F} = (F_t)_{t \geq 0}$ van kansgenererende functies (kgf's) zoals die voorkomen bij sub-kritische *vertakkingsprocessen*; voor de kgf van de bij $\mathcal{F}$ horende fractie $\alpha \odot_{\mathcal{F}} X$ geldt dan:

$$P_{\alpha \odot_{\mathcal{F}} X}(z) = P_X(F_t(z)) \quad (0 < \alpha < 1; t := -\log \alpha). \quad (5)$$

Merk op dat het eerst beschouwde (standaard) geval verkregen wordt door gebruik van de halfgroep $\mathcal{F}$ met $F_t(z) = 1 - e^{-t} + e^{-t}z$ voor $t \geq 0$. Door in (3) en (4) $\odot$ te vervangen door $\odot_{\mathcal{F}}$ krijgen we $\mathcal{F}$ -zelf-ontbindbare en $\mathcal{F}$ -stabiele verdelingen op $\mathbb{Z}_+$ die alle weer *oneindig deelbaar* blijken te zijn; de unimodaliteit gaat echter wel verloren.

Klaas van Harn

Een extreem geval van de inspectieparadox

Een bekend deelgebied van de kansrekening is *vernieuwings-theorie*. Men gaat daarbij uit van onafhankelijke, niet-negatieve stochastische variabelen $X_1, X_2, \dots$, de *levensduren* (van achtereenvolgens ingeschakelde onderdelen), die alle dezelfde verdeling hebben, zeg $\stackrel{d}{=} X$; we beperken ons hier tot het geval waarin X

absoluut-continu verdeeld is met dichtheid f en een eindige verwachting μ heeft. Zij $S_n := X_1 + \dots + X_n$, het *n-de vernieuwings-tijdstip*, en beschouw voor $t > 0$ het *aantal vernieuwingen* voor of op tijdstip t (eindig met kans 1 wegens de sterke wet van grote aantallen):

$$N(t) := \#(\{n \in \mathbb{N} : S_n \leq t\}).$$

Inspecteer nu, voor $t > 0$, het onderdeel dat ten tijde t in bedrijf is; voor de levensduur hiervan, de *momentane levensduur* $L_X(t)$ ten tijde t , geldt: $L_X(t) = X_{N(t)+1}$. Men zou kunnen vermoeden dat deze (weliswaar speciale) levensduur dezelfde verdeling heeft als alle andere (gewone) levensduren: $L_X(t) \stackrel{d}{=} X$, maar dat is *niet* het geval; dit fenomeen staat bekend als de *inspectieparadox*. Een en ander wordt nog concreter wanneer $t \rightarrow \infty$; men kan bewijzen dat dan $L_X(t) \stackrel{d}{\rightarrow} L_X$, waarbij L_X absoluut-continu verdeeld is met dichtheid g gegeven door

$$g(x) = \frac{1}{\mu} x f(x), \quad x > 0. \quad (1)$$

Hiermee volgt dat L_X stochastisch groter dan (of gelijk aan) X is in de zin dat $\mathbb{P}(L_X > a) \geq \mathbb{P}(X > a)$ voor alle a , en men zou zich kunnen afvragen voor welke levensduren X geldt dat

$$L_X \stackrel{d}{=} X + Y \quad \text{met } X \text{ en } Y \geq 0 \text{ onafhankelijk.} \quad (2)$$

Door de gelijkheid hierin in termen van dichtheden weer te geven krijgen we een functionaalvergelijking waarvan bekend is dat deze oneindige deelbaarheid karakteriseert; voor details verwijzen we naar [4]. We concluderen dat (2) geldt als en alleen als X *oneindig deelbaar* is.

Maar dit is nog niet alles. Neem voor de generieke levensduur X in het voorgaande de variabele $X(t)$ met $t > 0$, waarbij $(X(t))_{t \geq 0}$ een proces is met stationaire en onafhankelijke aangroeiingen; behalve de generator $X(1)$ is ook $X(t)$ dan oneindig deelbaar. Dat betekent dat voor de momentane levensduur $L_{X(t)}$ de volgende decompositie geldt:

$$L_{X(t)} \stackrel{d}{=} X(t) + Y(t) \quad \text{met } X(t) \text{ en } Y(t) \geq 0 \text{ onafhankelijk.} \quad (3)$$

Nu blijkt echter dat de ‘additionele component’ $Y(t)$ *niet afhangt van t* : $Y(t) = Y$ voor alle t . Laat nu $t \downarrow 0$, dan gaat de (gewone) levensduur $X(t)$ naar 0: $X(t) \stackrel{d}{\rightarrow} 0$, en men zou verwachten dat de momentane levensduur, hoewel stochastisch groter, dan ook naar 0 gaat. Dat is echter *niet* het geval:

$$L_{X(t)} \stackrel{d}{\rightarrow} Y, \quad t \downarrow 0. \quad (4)$$

Ingeval $X(1)$ exponentieel verdeeld is, blijkt Y dat ook te zijn; door dan $t = 1/(n-1)$ te nemen wordt duidelijk dat de inspectieparadox in dit geval een ‘onbeperkt’, en daarmee extreem, aspect krijgt:

Voor iedere $n \in \mathbb{N}$ is er een levensduur X zó dat de corresponderende momentane levensduur L_X als onafhankelijke componenten n gewone levensduren heeft.

Klaas van Harn

Referenties

- 1 S.N. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, *Comm. Soc. Math. Kharkov* 2 Series XIII (1912), 1–2.
- 2 C.G. Esseen, Über die asymptotisch beste Approximation stetiger Funktionen mit Hilfe von Bernstein-Polynomen, *Numer. Math.* 2 (1960), 206–213.
- 3 K. van Harn, *Classifying Infinitely Divisible Distributions by Functional Equations*, Mathematical Centre Tracts 103, Mathematisch Centrum, Amsterdam, 1978.
- 4 K. van Harn en F.W. Steutel, Infinite divisibility and the waiting-time paradox, *Stoch. Models* 11 (1995), 527–540.
- 5 K. van Harn, F.W. Steutel en W. Vervaat, Self-decomposable discrete distributions and branching processes, *Z. Wahrsch. verw. Gebiete* 61 (1982), 97–118.
- 6 Dunham Jackson, *The Theory of Approximation*, Amer. Math. Soc., Providence, 1930.
- 7 J.L. Konings en B.E.M. Span (red.), *Alle 80 goed! 17 november 2011 – Fred Steutel 80 jaar*, Technische Universiteit Eindhoven, 2011.
- 8 P.P. Korovkin, *Linear Operators and Approximation Theory*, Hindustan Publishing Corporation, 1960.
- 9 J.H. Leopold, *Verzen*, G.A. van Oorschot, 1977.
- 10 G.G. Lorentz, *Bernstein Polynomials*, University of Toronto Press, 1953.
- 11 I.P. Natanson, *Konstruktive Funktionentheorie*, Akademie-Verlag, 1955.
- 12 A. Roland Holst, *In den verleden tijd*, Boelen, 1975.
- 13 F. Schurer, *On Linear Positive Operators in Approximation Theory*, Waltman, 1965.
- 14 F. Schurer en F.W. Steutel, On linear positive operators of the Jackson type, *Mathematica (Cluj)* 9(32) (1967), 155–184.
- 15 F. Schurer en F.W. Steutel, The degree of local approximation of functions in $C_1[0,1]$ by Bernstein polynomials, *J. Approx. Theory* 19 (1977), 69–82.
- 16 P.C. Sikkema, Der Wert einiger Konstanten in der Theorie der Approximation mit Bernstein-Polynomen, *Numer. Math.* 3 (1961), 107–116.
- 17 F.W. Steutel, *Preservation of Infinite Divisibility under Mixing, and Related Topics*, Mathematical Centre Tracts 33, Mathematisch Centrum, Amsterdam, 1970.
- 18 F.W. Steutel, *Laatste kansen*, Technische Universiteit Eindhoven, 1996.
- 19 F.W. Steutel en K. van Harn, Discrete analogues of self-decomposability and stability, *Ann. Probab.* 7 (1979), 893–899.
- 20 Fred W. Steutel en Klaas van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, Pure and Applied Mathematics 259, Marcel Dekker, 2004.
- 21 E. Voronovskaya, Détermination de la forme asymptotique de l’approximation des fonctions par les polynômes de M. Bernstein (Russisch), *Dokl. Akad. Nauk SSSR* A 4 (1932), 79–85.

Julia Kuhn

Korteweg-de Vries Instituut
Universiteit van Amsterdam
j.kuhn@uva.nl

Interview Simon Singh

It would be great if politicians understood what science is

Simon Singh is a British science writer and author of the bestseller *Fermat's Last Theorem*. Julia Kuhn and Sonja Cox met him in May at the Amsterdam Science Park, where he had just presented his book *The Simpsons and Their Mathematical Secrets*.

When Simon Singh proposed the idea of a book about The Simpsons and mathematics, the response was one of shock and surprise. However, as Singh explains, many of the writers of the popular TV show are deeply fond of mathematics, some have advanced maths degrees. To keep themselves entertained, they hide references to mathematics in The Simpsons episodes to see whether their audience would pick up on them.

Evidently, Singh was one of those viewers who did. He draws our attention to many little hints to mathematical formulas that my colleague and I had missed in spite of being mathematicians and having watched a non-negligible number of Simpsons episodes. In his book, Singh uses these references to take his reader on a fun and informative journey from Euler's equation via perfect numbers to the unsolved problem of P versus NP and many other topics from number theory and other areas of maths. My personal favourite are the 'tests' he provides at the end of each chapter: a series of mathematical jokes that the reader can use to test her understanding of the subject that has been explained in the chapter. For each joke she can laugh about, she may give herself a

point. If you know who Mandelbrot is, here is one for you:

Question: What does the B stand for in Benoit B Mandelbrot?

Answer: Benoit B Mandelbrot.

Singh hopes that due to the connection to the famous TV show, his book is able to

draw some young readers towards maths that have not discovered their passion for it yet. On his website (simonsingh.net), he provides powerpoint slides for maths teachers who may wish to use them in class.

Indeed, it is easy to imagine how the book could be used as an inspiration to design an entertaining maths lecture. And it is a book Singh must certainly have enjoyed writing: "My wife was also writing a book at the same time, spending hours and hours each day at the library doing her research. When she came home in the evening, I'd still be doing my research lying on the couch watching The Simpsons."

Singh seems to enjoy telling such little anecdotes. Interviewing him is easy; it is evident that he is a great communicator. After having completed a PhD in particle physics, he decided he would use that communication skill for the promotion of science and took on a job at the BBC science department.

During his time at the British broadcaster, Singh directed the emotionally capturing documentary *Fermat's Last Theorem*, which won a BAFTA award for Best Documentary in 1997. But Singh saw more to the story: "A picture is worth a thousand words as we know, but still, there was just so much that we had to leave out. This was the last thing I did for the BBC. I could have done this job forever,



Simon Singh

but I was interested in trying something else.”

He decided to write a book about the subject. “I just thought it would be an interesting exercise. I suppose I learned how to write from TV, which needs drama and clarity. I started writing the book the same way I would start writing a TV program. The publisher seemed to think it would be OK.”

Fermat’s Last Theorem — in The Netherlands published by De Arbeiderspers — turned out to be a big success. “I thought it would be great if the book would get into the science top 10. And then it got into the UK top 10, and then it reached No.1, and then it got translated, by now in thirty languages. It’s amazing!” Never before had a book about maths made it to the top of the UK bestseller list.

By now Singh has published seven books in which he covers a range of topics from different areas of science, making them accessible to a broad audience. He hopes to engage, to push more young people to pursue a scientific career. And is maths becoming sexier? Singh thinks that, luckily, it is. He praises free online resources such as XKCD, Assassin, Morning Breakfast Cereal, and Numberphile that have helped this development. Recent movies he has seen with his sons (he mentions *Mr. Peabody and Sherman*, *Big Hero 6*, *Cloudy with a Chance of Meatballs*, *Queen of Katwe*) teach children that nerds are cool.

The problem is not to lose those kids along the way. “In England,” Singh explains, “if you are good at maths, you don’t get stretched at school, it’s just easy. You get really confident, you get excellent grades but you never get challenged.” He tells us about a school experiment carried out in the UK: Instead of calling upon those pupils that raised their hand to answer the teacher’s questions, the pupils’ names were drawn at random from a bowl. “Interestingly, it was the really bright students who started skipping class because for the first time they were risking to get questions they would not know the answer to. It was embarrassing for them because they had defined themselves as the ones who get it right.” Singh envisions a top-of-the-top group in which exceptionally bright students can learn to feel comfortable about being confused. In which they learn that not having succeeded yet is not equal to failure. “We don’t need people who *think* they are good because they

passed every test, we want people who *are* good because they have been thoroughly challenged.”

People with the potential to become good scientists are an increasingly precious resource. “Science is important for society”, Singh explains. “For a very long time, politicians did whatever they wanted to do according to ideology. They would tell scientists: ‘Go find me the evidence that proves that I’m right.’ In the UK over the last twenty years this has changed in a positive way. Now every department has its own chief scientist. And the government is slowly moving towards evidence-based policy.”

While he acknowledges this development, Singh also emphasises that there is still a long way to go. “Sometimes people misunderstand what ‘evidence-based policy’ means. They think it means that the scientist tells the politician what to do but it is not that at all.” The scientist provides the evidence but it is up to the politicians

“We don’t need people who think they are good because they passed every test, we want people who are good because they have been thoroughly challenged.”

to act upon it. And, of course, many other considerations are important apart from the scientific facts. “It would be great if politicians understood what science is, what it does and what it can provide for society.”

The day before our interview, Singh tells us, he heard a radio show about the late David MacKay, author of the book *Sustainable Energy—Without the Hot Air*. “Somebody said to MacKay, ‘you are pro nuclear’, and he answered, ‘I’m not pro nuclear, I’m pro maths, and that’s it. It’s about the numbers and this is what the numbers say, you can’t argue with them’.” Singh seems to agree. Real science is not a matter of opinion.

The battle against pseudo-science and quackery has taken up much of Singh’s time in recent years. In 2012 he founded the Good Thinking Society, a non-profit organisation advocating scientific scepticism. Similar to the pioneering Dutch ‘Vereniging tegen de kwakzalverij’, the Good Thinking

Society focusses most of its scepticism on alternative medicine. When asked how this particular subject had caught his interest, Singh tells us about a program he had watched on the BBC. “The presenter was a professor, backed by the Open University, so I thought it should be very scientific, very credible.” The show talked about a woman undergoing major heart surgery in China that was carried out without a general anaesthetic, only using acupuncture. This could have changed the history of medicine, had it been true. But the story as presented was misleading: as a report of the Royal College of Anaesthetists pointed out afterwards, the woman in surgery may not have had a general anaesthetic, but she had received so many sedatives that she did not need it. Singh complained to the BBC, and eventually the broadcaster admitted to having made a mistake.

Together with Edzard Ernst, a professor of complementary medicine, Singh decided to investigate available evidence for the effectiveness of alternative medicine. The book *Trick or Treatment? Alternative Medicine on Trial* is the compilation of their conclusions. “All my other books were promoting science,” says Singh, “but this was about the other side, challenging the ‘bad stuff’. We tried to be as fair as possible. We just thought a parent spending €100 a year on alternative medicine should invest €10 in this book to find out what works and what doesn’t. However, it turned out to be my least successful book.”

Worse still, it is a book that brought Singh an enormous amount of stress and trauma, mainly due to a related column he wrote in April 2008 for *The Guardian* in which he accused the British Chiropractic Association of ‘happily promoting bogus treatment’. The chiropractors sued for libel. It took two years, a second ruling and cost more than €100,000 in defence costs before he was cleared on the grounds that his statements were to be treated as ‘fair comment’.

The scientific community rallied and supported Singh, calling for a reform of the English libel law which they felt was stifling scientific debate. In the aftermath of the case the ‘Defamation Act 2013’ became law, which includes the defence of ‘responsible publication on matters of public interest’ and a clause that is designed to protect free speech in academic journals.

Madelon de Kemp

Korteweg-de Vries Instituut voor Wiskunde
Universiteit van Amsterdam
m.a.dekemp@uva.nl

Onderzoek

Inplannen van afspraken voor een groot aantal patiënten

Voor haar masterscriptie heeft Madelon de Kemp onderzoek gedaan naar het optimaal inplannen van afspraken. Hiervoor heeft zij vorig jaar november de ASML afstudeerprijs ontvangen, uitgereikt door de Koninklijke Hollandse Maatschappij der Wetenschappen (KHMW). In dit artikel vertelt zij iets over haar onderzoek.

Stel we willen graag afspraken inplannen bij bijvoorbeeld een dokter. Aan de ene kant willen we dat patiënten niet te lang hoeven wachten. Aan de andere kant willen we dat de dokter niet teveel tijd onbenut laat tussen de behandeling van twee patiënten in. Hiervoor moeten we twee zaken van tevoren bepalen: hoeveel tijd er voor iedere afspraak wordt ingepland, en in welke volgorde de dokter de verschillende patiënten ziet.

De tijd die een behandeling duurt is niet van tevoren bekend. In plaats daarvan zullen we aannemen dat de behandeltijd een stochast is. We nemen aan dat de behandeltijden van verschillende patiënten onafhankelijk zijn. We zullen ook aannemen dat we voor iedere patiënt de verwachting en variantie van de behandeltijd kennen, bijvoorbeeld door vorige metingen, maar niet de volledige verdeling. Onze planning moet dan goed presteren voor alle mogelijke verdelingen met deze verwachting en variantie.

Doelfunctie

Voordat we op zoek kunnen gaan naar een optimale planning, zullen we eerst precies

moeten maken wat we bedoelen met optimaal. Noteer de wachttijd van patiënt i met W_i . We willen graag dat $\mathbb{E}W_i$, de verwachte wachttijd van patiënt i , klein is. Noteer met I_i de tijd die de dokter onbenut laat tussen patiënt $i-1$ en patiënt i . We willen dan ook dat $\mathbb{E}I_i$, de verwachte onbenutte tijd voor patiënt i , klein is. De tijd van de dokter is niet altijd evenveel waard als de tijd die een patiënt moet wachten. Daarom introduceren we een parameter $\omega \in (0,1)$ die de relatieve waarde van de tijd van de dokter representeert. Een mogelijke doelfunctie is dan

$$\omega \sum \mathbb{E}I_i + (1-\omega) \sum \mathbb{E}W_i,$$

waarbij we sommeren over alle patiënten.

Bij deze doelfunctie doet zich nog een probleem voor: $\mathbb{E}I_i$ en $\mathbb{E}W_i$ hangen, naast de planning, ook af van de verdeling van de behandeltijden. Echter, deze verdeling weten we niet, alleen de verwachting en variantie. Omdat we willen dat de planning goed werkt voor alle verdelingen met deze verwachting en variantie, nemen we als doelfunctie

$$\sup \{ \omega \sum \mathbb{E}I_i + (1-\omega) \sum \mathbb{E}W_i \},$$

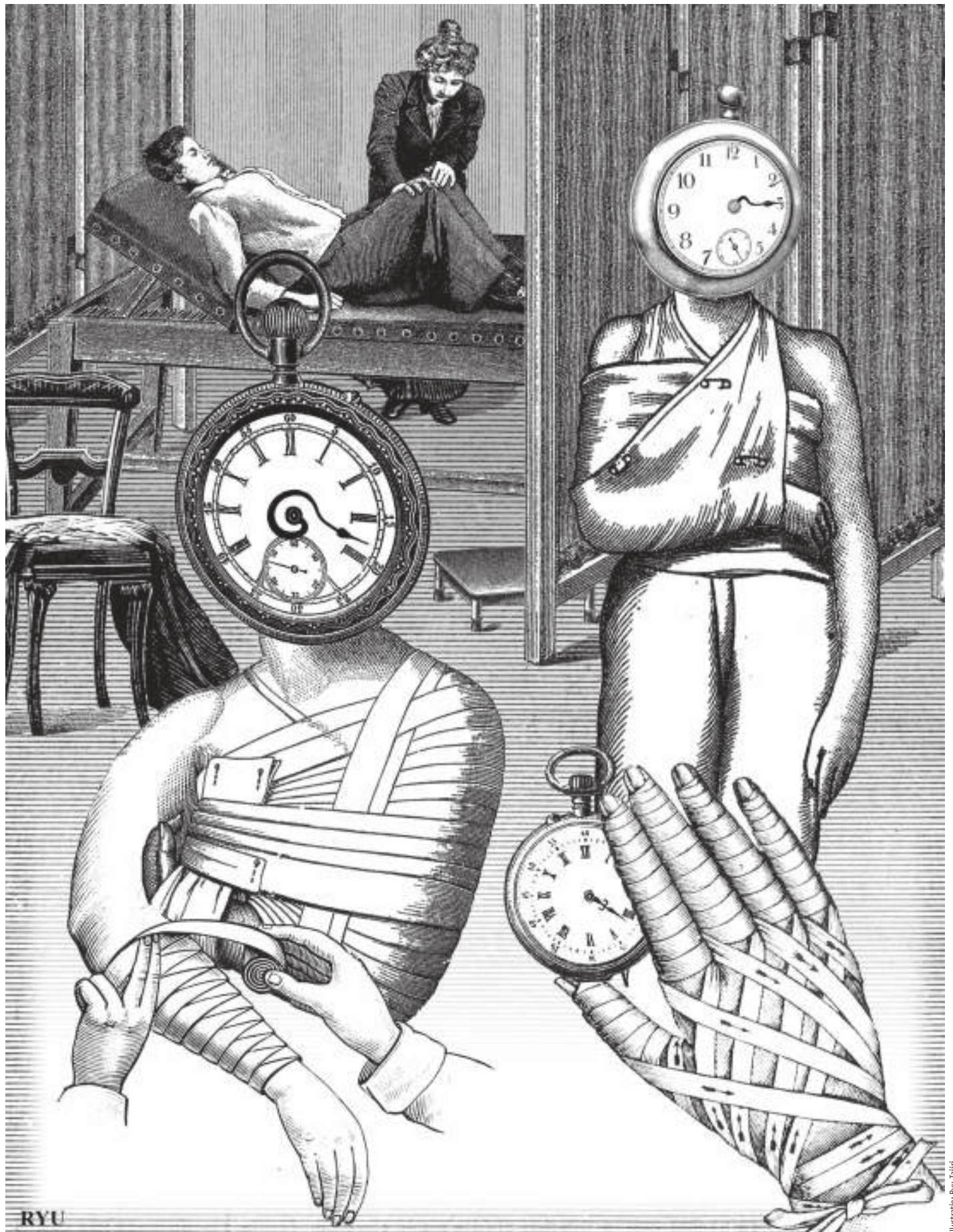
waar we het supremum nemen over alle verdelingen voor de behandeltijden, ieder met de gegeven verwachting en variantie. Deze doelfunctie geeft dan een garantie voor alle mogelijke behandeltijden.

Deze doelfunctie hangt nu nog af van twee zaken: de tijd die per afspraak wordt ingepland, en de volgorde waarin de patiënten behandeld worden. Deze zullen zo gekozen moeten worden dat de doelfunctie minimaal is. In mijn masterscriptie heb ik naar beide deelproblemen onderzoek gedaan, maar niet in precies dezelfde situatie. Het gezamenlijke probleem is dus nog niet opgelost.

Geplande tijd per afspraak

Eerst bekijken we het probleem om te bepalen hoeveel tijd we per afspraak moeten inplannen. Hiervoor nemen we aan dat de behandeltijden van alle patiënten dezelfde verdeling hebben, met gegeven verwachting μ en variantie σ^2 . We zullen kijken naar de limiet als het aantal patiënten naar oneindig gaat. Dit geeft ons dan een benadering voor situaties met een eindig maar groot aantal patiënten.

Als er oneindig veel patiënten zijn, zal de situatie van een patiënt niet verschillen van zijn voorganger. We nemen daarom aan dat dezelfde hoeveelheid tijd, die we x zullen noemen, gepland is voor alle afspraken. De verwachte wachttijd, $\mathbb{E}W$, van een



patiënt is nu hetzelfde voor iedere patiënt. Hetzelfde geldt voor de verwachte ongebruikte tijd tussen twee patiënten, $\mathbb{E}I$. Als doelfunctie krijgen we dan

$$\sup\{\omega\mathbb{E}I + (1-\omega)\mathbb{E}W\}.$$

Voor de verwachte ongebruikte tijd is het bekend dat $\mathbb{E}I = x - \mu$, voor iedere verdeling voor de behandeltijd met verwachting μ . Om de doelfunctie te kunnen bepalen moeten we dus nog $\sup\mathbb{E}W$ bepalen. We kunnen nu een bovengrens van Kingman [4] voor algemene wachtrijen toepassen. Dit geeft de volgende bovengrens voor de verwachte wachttijd:

$$\mathbb{E}W \leq \frac{\sigma^2}{2(x-\mu)}.$$

Deze bovengrens is in deze situatie ook strikt (zoals bewezen in [2]), dat wil zeggen er zijn verdelingen voor de behandeltijd met verwachting μ en variantie σ^2 zodanig dat $\mathbb{E}W$ deze bovengrens willekeurig dicht benaderd.

Nu weten we dus dat $\mathbb{E}I = x - \mu$ en $\sup\mathbb{E}W = \frac{\sigma^2}{2(x-\mu)}$. De doelfunctie wordt nu

$$\begin{aligned} &\sup\{\omega\mathbb{E}I + (1-\omega)\mathbb{E}W\} \\ &= \omega(x-\mu) + (1-\omega)\frac{\sigma^2}{2(x-\mu)}. \end{aligned}$$

We zoeken nu de x waarvoor deze doelfunctie minimaal is. Standaard calculus geeft ons dat deze x gegeven wordt door

$$x = \mu + \sigma\sqrt{\frac{1-\omega}{2\omega}}.$$

Volgorde bepalen

Het volgordeprobleem is het probleem om te bepalen in welke volgorde patiënten behandeld moeten worden als er verschillende soorten patiënten zijn. We nemen aan dat we de patiënten kunnen indelen in m groepen van soortgelijke patiënten. Patiënten in dezelfde groep worden achter elkaar in een tijdsblok behandeld. We willen nu

weten in welke volgorde de m groepen moeten worden ingedeeld.

We nemen aan dat de behandeltijden van patiënten in dezelfde groep dezelfde verdeling hebben. In groep k heeft de behandeltijd verwachting μ_k en variantie σ_k^2 . We plannen μ_k tijd in voor iedere patiënt in groep k , dus de tijd die we inplannen per afspraak is gelijk aan de verwachte tijdsduur van de afspraak (wat niet optimaal hoeft te zijn).

We zullen de limiet bekijken als het aantal patiënten naar oneindig gaat, terwijl de fractie patiënten in groep k , die we T_k noemen, gelijk blijft voor alle k .

We kunnen de wachttijd van een patiënt als volgt uitrekenen: tel bij de wachttijd van de vorige patiënt op hoeveel vertraging de vorige patiënt veroorzaakt of goedmaakt. Als dit positief is, is dit de wachttijd van de huidige patiënt, en als dit negatief is, is de wachttijd van de huidige patiënt nul.

Binnen een groep zijn de vertragingen (de behandeltijd min de geplande tijd) steeds gelijk verdeeld. Wegens de centrale limietstelling zullen sommen van vertragingen van een groot aantal patiënten zich steeds meer volgens een normale verdeling gaan gedragen. Een generalisatie hiervan is de stelling van Donsker, die ons vertelt dat het hele proces van vertragingen voor een groot aantal patiënten zich steeds meer zal gedragen als een zogenaamde 'Brownse beweging'.

Voor de wachttijden zal het proces nu steeds meer gaan lijken op een omhoog geduwde versie van een Brownse beweging. Voor deze omhoog geduwde Brownse beweging kunnen we nu op ieder moment de verwachting berekenen. Deze blijkt voor ieder moment het kleinst te zijn als we de groepen behandelen in volgorde van hun variantie σ_k^2 , met de kleinste variantie eerst.

De onbenutte tijd kan ook met behulp van de Brownse beweging benaderd wor-

den. De verwachting van de onbenutte tijd blijkt dan niet van de volgorde af te hangen. In totaal is de volgorde met de kleinste variantie eerst het beste.

Conclusie

Als de behandeltijden van alle patiënten dezelfde verdeling hebben, vonden we dat voor een groot aantal patiënten de optimale behandeltijd gegeven wordt door

$$x = \mu + \sigma\sqrt{\frac{1-\omega}{2\omega}}.$$

We plannen dus voor een afspraak tijd in gelijk aan de verwachting plus een extra tijd lineair aan de standaardafwijking om eerdere vertragingen te kunnen opvangen. Vooral voor ω dicht bij nul, als vooral de wachttijd belangrijk is, is dit opvangen van eerdere vertragingen belangrijk. De formule lijkt dus met die intuïtie overeen te komen.

We keken ook naar het volgordeprobleem, waarbij we te maken hebben met een aantal grote groepen patiënten. Als binnen een groep de behandeltijden dezelfde verdeling hebben, en we plannen voor iedere afspraak tijd in gelijk aan de verwachte behandeltijd, dan is het het beste om de groepen te behandelen in volgorde van de variantie van hun behandeltijd. Ook dit is intuïtief aannemelijk, aangezien groepen met een grote variantie een groot risico vormen om vertragingen te veroorzaken, en dus het beste als laatste behandeld kunnen worden.

De twee deelresultaten sluiten helaas nog niet helemaal op elkaar aan. Bij het volgordeprobleem namen we aan dat de geplande tijd per afspraak gelijk is aan de verwachte behandeltijd, terwijl we eerder al zagen dat we eigenlijk meer tijd zouden moeten inplannen. Het gezamenlijke probleem van zowel de optimale behandeltijden en de optimale volgorde is dus nog steeds onopgelost. $\square$

Referenties

- 1 A. Ahmadi-Javid, Z. Jalali en K. Klassen, Out-patient appointment systems in healthcare: a review of optimization studies, *European Journal of Operational Research* 258(1) (2017), 3–34.
- 2 D. Daley, A. Kreinin en C. Trengove, Inequalities concerning the waiting-time in single-server queues: a survey, in: *Queueing and Related Models*, Clarendon Press, 1992, pp. 177–223.
- 3 M.A. de Kemp, *Robust Appointment Scheduling and Sequencing for Many Patients*, M.Sc. Thesis, VU Amsterdam, 2016.
- 4 J.F.C. Kingman, Some inequalities for the GI/G/1 queue, *Biometrika* 49(3-4) (1962), 315–324.

Nicolaos Starreveld

Korteweg-de Vries Instituut voor Wiskunde
Universiteit van Amsterdam
n.j.starreveld@uva.nl

Boekbespreking D'Arcy Thompson: Over groei en vorm

Over groei en vorm

In 2017 vierten we het honderdjarig bestaan van de eerste editie van het boek *On Growth and Form* geschreven door D'Arcy Wentworth Thompson. Uitgeverij Meromorf Press heeft, in samenwerking met Begonia Translations, het boek voor het eerst in het Nederlands vertaald onder het titel *Over Groei en Vorm*. Het oorspronkelijke boek werd voor het eerst in 1917 gedrukt en telt negentien hoofdstukken en ongeveer duizend pagina's. Het boek dat we hier bespreken is een vertaling van een verkorte versie die in 1961 verscheen. Daarnaast heeft er een aantal lezingen plaatsgevonden, de eerste was in maart in de Balie in Amsterdam en de tweede in oktober op het Lorentz Center in Leiden. Nicolaos Starreveld bezocht de lezing in maart en bespreekt het boek.

On Growth and Form heeft een substantiële bijdrage geleverd aan de natuurwetenschappen. Het wordt beschouwd als het eerste boek over wiskundige biologie. D'Arcy Thompson (1860–1948) was een Schotse bioloog en wiskundige gespecialiseerd in ongewervelde dieren. In die tijd waren het de ideeën van Darwin die dat vakgebied domineerden en de vorm en structuur van organismen werden voornamelijk gezien als een resultaat van natuurlijke selectie. D'Arcy Thompson postuleerde de stelling dat biologen, en vooral morfologen, meer aandacht moesten besteden aan de natuurlijke wetten en hun invloed op de ontwikkeling van organismen. De morfologie is het onderdeel van de biologie dat de bouw en vorm bestudeert van organismen. In de epilooft verklaart D'Arcy Thompson zijn bedoeling met dit boek: "Te laten zien dat een bepaald wiskundig aspect van de morfologie essentieel is voor het goed bestuderen en begrijpen van groei en vorm."

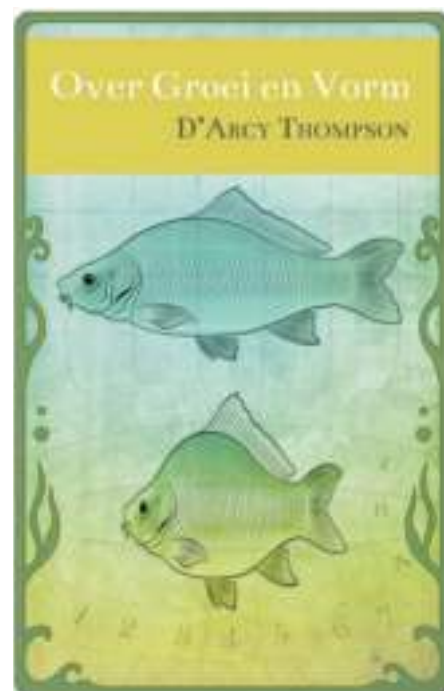
D'Arcy Thompson had een klassieke opleiding, zijn vader was hoogleraar Grieks, en dat is evident in zijn boek. In zijn onderzoek gebruikte hij methoden

die Aristoteles introduceerde. We moeten de wereld waarnemen en meten, daarna gaan we vergelijken en potentiële correlaties ontdekken. Hij durfde een stap verder te gaan dan de deskundigen van zijn tijd en hij probeerde de empirische kennis van de bioloog en de dierkundige uit te drukken door middel van het getal. Hij gebruikte vergelijkingen om relaties te bestuderen en symmetrieën om patronen te identificeren. Zoals Stephen J. Gould van Harvard University het samenvat: "De kruising van de theorieën van Pythagoras en Newton stelt dat natuurwetten direct de vormgeving van levende wezens bepalen, waarbij de 'interne' en genetische invloeden het bouw materiaal verzorgen en de natuurkundige wetten de opbouw bepalen."

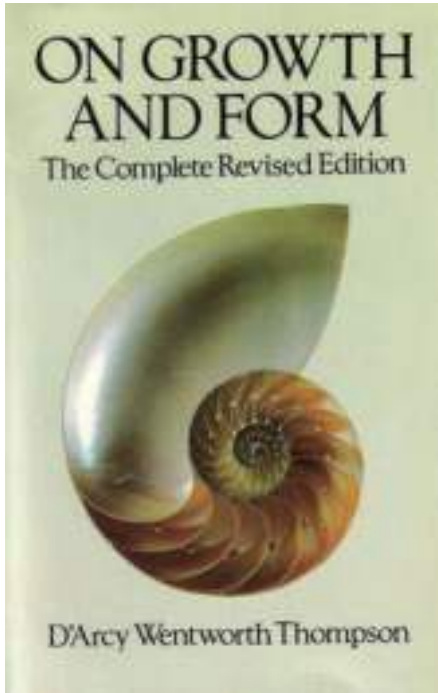
Het boek

Deze verkorte versie telt tien hoofdstukken en ongeveer driehonderd pagina's. Beide edities beginnen met het hoofdstuk 'Over afmeting' dat cruciaal is voor het begrijpen van de hoofdstukken die volgen. In de oorspronkelijke editie volgt een uitgebreide analyse van de maat van groei; die is weg-

gelaten in de verkorte versie. Daarna volgen hoofdstukken over de vorm van cellen, weefselvorm en celaggregaatvorm en drie hoofdstukken over sponsnaalden en spiculaire skeletten, de gelijkhoekige spiraal en de vorm van hoorns en tanden. De laatste twee hoofdstukken gaan over vorm en mechanische efficiëntie en de transformatietheorie. D'Arcy Thompson probeerde zijn wiskundige opleiding volledig te gebruiken ten gunste van *Over Groei en Vorm*, maar zijn betoog is toch hoofdzakelijk een biologische theorie.



D'Arcy Thompson: *Over groei en vorm*, Meromorf Press, 2016, 332 p., ISBN 9780992900809, prijs € 24,95.



Omslag van de verkorte Engelstalige versie uit 1961

Over afmeting

In de elementaire meetkunde leren we dat het volume van een bol gelijk is aan $\frac{4}{3}\pi r^3$ en het oppervlak gegeven wordt door $4\pi r^2$, waarbij r de straal van de bol is. Het was Archimedes die dit voor het eerst bewees maar dit resultaat kan ook bewezen worden door analyse. Beschouw een infinitesimale bol rond een punt x en integreer van $-r$ tot r . Een vergelijkbare methode kan toegepast worden om het oppervlak te bepalen. Hieruit volgt dat de verhouding van het volume tot het oppervlak gelijk aan $\frac{1}{3}r$ is; en r is dus evenredig met deze verhouding.

D'Arcy Thompson was bekend met het werk van Newton en de infinitesimaalanalyse. Dat begrijp ik tenminste uit zijn redeneringen, hij gebruikte deze methode van opsplitsen in infinitesimaal kleine bollen en dan op te tellen om de vorm van verschillende wezens te bestuderen. Een toepassing van deze wiskundige methoden vinden we als D'Arcy Thompson begint te spreken over afmeting en warmte-uitwisseling. Het warmteverlies door uitstraling is evenredig aan het huidoppervlak van het dier, maar de door verbranding geproduceerde warmte is evenredig aan de massa van het dier. Dus de verhouding van het oppervlak tot het volume is evenredig aan $\frac{1}{r}$, waar r de lineaire dimensie is. Daarom hoort de verhouding tussen warmteverlies en warmteproductie, evenals die tussen

oppervlakte en volume, toe te nemen naarmate het dier kleiner wordt. Hij kwam tot de conclusie dat kleine dieren meer warmte per kilo produceren dan grote dieren, om gelijke tred te kunnen houden met het warmteverlies door oppervlakte-uitstraling. Met zulke prachtige wiskundige methoden kon D'Arcy Thompson verschillende restricties op de grootte, de vorm en de functies van wezens beschouwen, bijvoorbeeld op het springvermogen, het vliegen en de opbouw van het skelet van dieren.

Maat van groei

In de originele tekst bestaat er een hoofdstuk (Hoofdstuk III) met de titel 'The rate of growth'. D'Arcy Thompson verzamelde in dit hoofdstuk uitkomsten van experimenten die hij had uitgevoerd en probeerde daarna, door het gebruik van statistische analyse, groei te analyseren. We zien onder andere ook toepassingen van differentiaalvergelijkingen, van de normale verdeling en van de stelling van Stokes. Dit hoofdstuk is volgens een aantal critici zwak en daarom is het weggelaten in de verkorte versie van 1961. Vandaag de dag is er ook een duidelijker beeld over dit onderwerp en er zijn boeken die een betere uitleg bieden. De belangstelling voor dit hoofdstuk blijft in ieder geval nog steeds bestaan. Historisch gezien stijgt het afleiden van wiskundige vergelijkingen gebaseerd op waargenomen biologische processen uit boven het wetenschappelijke niveau van zijn tijd. Daarom kan dit hoofdstuk nog nuttig zijn voor hedendaagse wetenschap-

pers, zelfs als het niet in overeenstemming is met de wiskundige strengheid en de recente wetenschappelijke ontdekkingen.

Celvorm, weefselvorm en celaggregaatvorm

Om celvorm te kunnen bestuderen is het vooral belangrijk alle krachten die een rol spelen in aanmerking te nemen. Het gaat voornamelijk over de zwaartekracht en de oppervlaktespanning. D'Arcy Thompson beperkte zich in eerste instantie tot die weefsels en die omstandigheden waarin de zwaartekracht relatief onbelangrijk is en oppervlakte-energie dus een belangrijk deel van de totale energie in het systeem betreft. Dit is in het algemeen het geval waar we te maken hebben met heel kleine hoeveelheden vloeistof en afmetingen die de moleculaire afmetingen benaderen. In die kleine cellulaire organismen wordt de vorm grotendeels door oppervlaktespanning bepaald, in tegenstelling tot grote organismen, van wie vorm tot stand komt onder invloed van andere krachten zoals de zwaartekracht. De manier waarop celvorm wordt beïnvloed door oppervlaktespanning is bestudeerd door het gebruik van het natuurkundige concept van evenwicht en het wiskundige concept van een minimaaloppervlak. Een evenwicht wordt bereikt nadat het vlak tot het kleinste mogelijke oppervlak is gereduceerd zodat de inwaartse en uitwaartse spanningen gelijk aan elkaar zijn. In het werk van D'Arcy Thompson wordt een minimaaloppervlak beschouwd als een oppervlak dat op elk punt dezelfde gemiddelde kromming kent, niet perse gelijk aan nul, en het heeft het minimale oppervlak voor het ingesloten volume.

Na eenvoudige alleenstaande cellen belande Thompson uiteindelijk bij het bestuderen van complexe weefsels. Deze transitie wordt in stappen gezet, van eenvoudige cellen gaan we naar cellen die met elkaar in contact staan, of anders gezegd celaggregaten. Technieken van de vlakke meetkunde leiden tot prachtige formules die de opbouw van verschillende complexe systemen kunnen beschouwen. Een voorbeeld dat in het boek te vinden is, bevat het berekenen van de hoeken waaronder de bloedvaten afsplitsen. De algemene stelling die Thompson voorstelde, is de volgende: "de vorm, ordening en de hoeken waaronder de bloedvaten afsplitsen is zodanig, dat de bloedsomloop met een minimale hoeveelheid arbeid en een minimum van wandoppervlak plaatsvindt".



D'Arcy Wentworth Thompson (1860–1948)

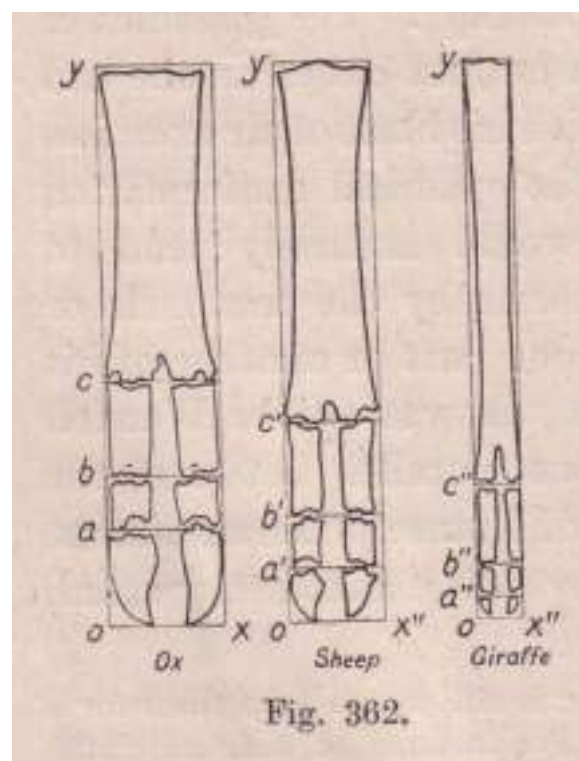
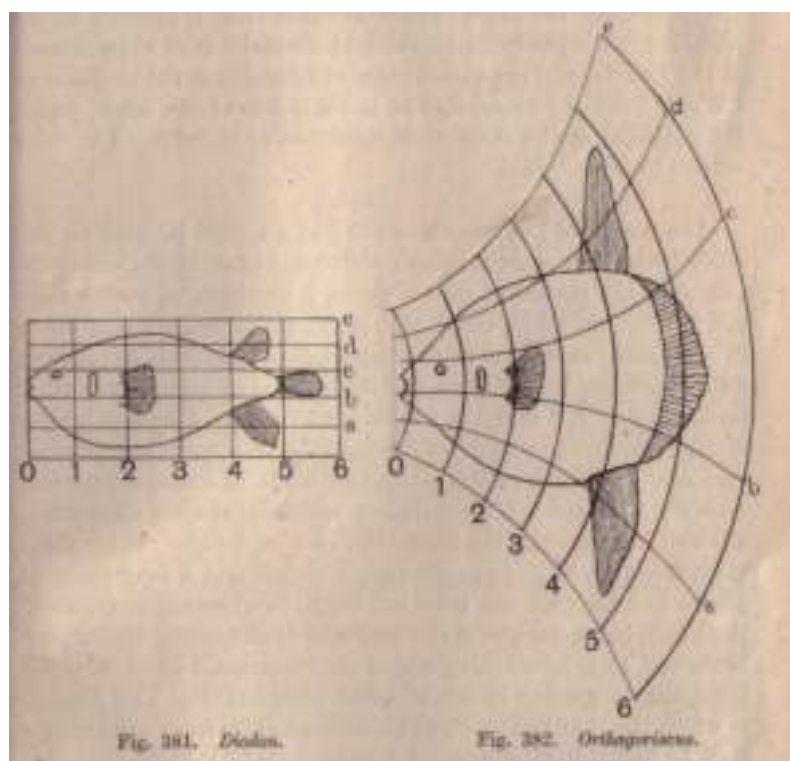
Transformatietheorie

De transformatietheorie is waarschijnlijk een van de grootste bijdragen aan de morfologie van D'Arcy Thompsons werk. Om groei en vorm op een uniforme schaal te kunnen bestuderen, gebruikte hij de wiskundige methode van de 'coördinatenmethode'. Het herkennen van een vorm als een van de vele mogelijke permutaties of vervormingen van een andere vorm is de kern van de transformatietheorie. Een ander significant voordeel van de coördinatenmethode ligt in het feit dat hij de eenheid van het organisme onder de aandacht brengt. De onderliggende intuïtie is precies hetzelfde als in de functietheorie en het verschil tussen puntsgewijze en uniforme eigenschappen. Als we juist diverse en ongelijke organismen en vormen in hun geheel als een identieke 'functie' binnen ver-

schillende coördinatenstelsels kunnen zien, vormt dit een sterke aanwijzing dat variatie volgens bepaalde lijnen is voortgeschreden en dat hier een 'groeiwet', in de vorm van een krachtstelsel, verantwoordelijk voor is. D'Arcy Thompson heeft over drie verschillende coördinatenstelsels gesproken: het cartesische, het radiale en het logaritmische. Hij heeft een uitgebreid experiment gepresenteerd om te laten zien dat zijn transformatietheorie nuttig is in de morfologie. Transformaties in coördinaten zijn in het algemeen wiskundig uitdagend om te analyseren, maar zijn doel is alsnog in voldoende mate bereikt. Met behulp van dergelijke transformaties, die vaak "niet aan een strikt wiskundige analyse onderworpen kunnen worden", kunnen we grafisch nog steeds de relatie tussen het nieuwe en het oude coördinatenstelsel aantonen.

Tot slot

In het algemeen vond ik het een goed verzorgde vertaling die een prachtig werk hopelijk in de schijnwerpers zal zetten. Als wiskundige raak ik gefascineerd door de eenvoudige manier waarop wiskundige principes zo natuurlijk voorkomen in verschillende biologische processen en in de morfologie. Het boek is niet altijd makkelijk om te lezen en de lezer moet alert zijn om belangrijke details niet te missen. Hoewel er geen voorkennis wordt geëist om dit boek te kunnen lezen, zou het toch nuttig zijn om wat basiskennis van analyse en mechanica te hebben. Het is prettig dat de hoofdstukken onafhankelijk van elkaar gelezen kunnen worden. Ik zou iedereen aanraden om het boek, of ten minste een paar hoofdstukken van het boek, te lezen. ☼



Twee afbeeldingen uit de oorspronkelijke uitgave van *On Growth and Form* uit 1917.

Jakob Krarup

Department of Computer Science
University of Copenhagen, Denmark
krarup@di.ku.dk

Kees Roos

Faculty of EEMCS
Technical University Delft
c.roos@tudelft.nl

Research

On the Fermat point of a triangle

For a given triangle $\triangle ABC$, Pierre de Fermat posed around 1640 the problem of finding a point P minimizing the sum s_P of the Euclidean distances from P to the vertices A, B, C . Based on geometrical arguments this problem was first solved by Torricelli shortly after, by Simpson in 1750, and by several others. Steeped in modern optimization techniques, notably duality, however, Jakob Krarup and Kees Roos show that the problem admits a straightforward solution. Using Simpson's construction they furthermore derive a formula expressing s_P in terms of the given triangle. This formula appears to reveal a simple relationship between the area of $\triangle ABC$ and the areas of the two equilateral triangles that occur in the so-called Napoleon's Theorem.

We deal with two problems in planar geometry. Both problems presuppose that three noncollinear points A, B and C are given:

- (F) Find a point P minimizing the sum s_P of the Euclidean distances from P to the vertices of $\triangle ABC$.
- (D) Find an equilateral triangle T of maximum height h_T , with the points A, B and C on different sides of T .

Fermat's name is associated with various problems and theorems. Fermat posed problem (F) around 1640.¹ In other literature (F) is often referred to as the 3-point Fermat Problem. For its long history we refer to, e.g., [2, 3, 11].

Using a result of his student Viviani, Torricelli presented a geometrical construction of the Fermat point P , repeated below in the section 'Solution of Fermat's Problem' and accompanied by his correctness proof in the section 'Torricelli's triangle'. Viviani's result will be discussed in between those sections.

Problem (D) was posed in 1755 by T. Moss in the *Ladies Diary* [13], as shown in Figure 1. As this figure indicates, the prob-

lem has been solved but it is not known by whom.² As will be made clear in the section 'Duality of (F) and (D)', Torricelli's approach already reveals that the problems (F) and (D) are dual to each other.

Our first aim is to show that application of duality in *conic optimization* yields the solution of (F), straightforwardly. Developed in the 1990s, the field of conic optimization (CO) is a generalization of the more well-known field of linear optimization (LO). Its main source is [1]. No prior knowledge of LO or CO is needed for reading the paper.



Pierre de Fermat (1601?–1665)



Evangelista Torricelli (1608–1647)

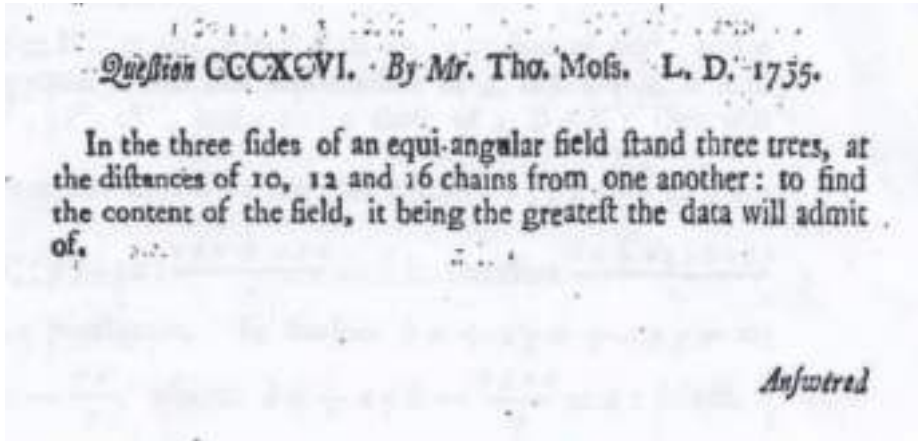


Figure 1 Original formulation of problem (D) in the *Ladies Diary*.

It turns out that two cases need be distinguished depending on the largest angle in the triangle ABC :

- Case F1: the largest angle in $\triangle ABC$ is less than 120° ;
 Case F2: the largest angle in $\triangle ABC$ is at least 120° .

Our second aim is to show that in case F1 the problems (F) and (D) are each others dual problem; the meaning of this sentence will be made clear below. The corresponding result can be stated as follows.

Theorem 1. *Let X be any point of a triangle ABC of type F1 and T an equilateral triangle such that A, B and C lie on different sides of T . Then one has*

$$h_T \leq s_X. \quad (1)$$

Equality holds if and only if X solves problem (F) and T solves (D).

The inequality (1) expresses so-called *weak duality* for problems (F) and (D), whereas the last statement in Theorem 1 says that also *strong duality* holds. The proof of Theorem 1 will be given later.

As far as we know there does not exist a simple formula expressing s_P in terms of the given points A, B and C . Such a formula will be derived. It uses a simple geometric construction of Simpson that yields a line segment with length s_P . We show that the sides of the so-called Napoleon's outward triangle of $\triangle ABC$ have length $s_P/\sqrt{3}$.

In the next section we reformulate (F) as a conic optimization problem and introduce its dual problem. As will be shown in the section thereafter, the optimality conditions for this pair of problems imme-

diately yield the correctness of Torricelli's construction of the Fermat point.

Analytic approach to Fermat's problem

Fermat's problem as stated in (F) is a geometric problem. To begin with we put it in analytic form. To each of the relevant points we associate a vector as follows:

$$x = OP, \quad a_1 = OA, \quad a_2 = OB, \quad a_3 = OC.$$

Here we assume that the plane is equipped with a two-dimensional coordinate system, with origin O . Moreover, OP denotes the vector that goes from O to P . The distance from P to A is the length of the vector PA . Since $OP + PA = OA$ we have $PA = a_1 - x$. Hence, the distance from P to A is just the Euclidean norm of the vector $a_1 - x$. As a consequence, the sum of the distances from P to A, B and C is equal to $\sum_{i=1}^3 \|a_i - x\|$. This function will be called our *objective function*. We need to minimize this function when x runs through all vectors in $\mathbb{R}^2$. So (F) can be reformulated as the following optimization problem:

$$\min_x \left\{ \sum_{i=1}^3 \|a_i - x\| \mid x \in \mathbb{R}^2 \right\}. \quad (2)$$

Each term in the above sum is nonnegative, and since A, B and C are noncollinear, at most one of them can be zero. Hence, the objective function is positive for every $x \in \mathbb{R}^2$. Furthermore, each term is strictly convex in x , as follows from the triangle inequality, and so will be their sum. We conclude that the objective function is strictly convex and positive. This implies that problem (2) has a unique solution, which necessarily is the Fermat point of $\triangle ABC$.

Since the objective function is a sum of Euclidean norms of linear functions of x , and the domain is $\mathbb{R}^2$, (2) is a special case

of a convex optimization problem, namely a so-called *conic optimization problem* or a CO problem for short.

Every convex optimization problem has a dual problem. This is a maximization problem with the property that the objective value of any dual feasible solution is less than or equal to the objective value of any 'primal' feasible solution. This is the property called *weak duality*.

Finding a problem with this property can be a tedious task. For the class of CO problems there exists a simple recipe to get a dual problem. For the current purpose it is not necessary to go into further detail. It suffices to mention that the theory of CO yields the following dual problem of (2):

$$\max_{y_i} \left\{ \sum_{i=1}^3 a_i^T y_i \mid \sum_{i=1}^3 y_i = 0, \|y_i\| \leq 1, \right. \\ \left. i \in \{1, 2, 3\} \right\}. \quad (3)$$

To show that (2) and (3) considered together exhibit the desired property of weak duality, assume that x and the triple (y_1, y_2, y_3) are feasible for (2) and (3), respectively. We may then write

$$\sum_{i=1}^3 \|a_i - x\| \\ \geq \sum_{i=1}^3 \|a_i - x\| \|y_i\| \\ \geq \sum_{i=1}^3 (a_i - x)^T y_i = \sum_{i=1}^3 a_i^T y_i.$$

The first inequality in (4) is due to the fact that $\|y_i\| \leq 1$ for each i and the second inequality follows from the Cauchy-Schwarz inequality; the equality in (4) follows since $\sum_{i=1}^3 y_i = 0$, whence also $x^T \sum_{i=1}^3 y_i = 0$. This shows that we have weak duality.

This is all we need to reveal a property of the Fermat point that enabled Torricelli to geometrically construct this point in a very simple way. As will become clear in the next section the property in question follows by elementary means from (4).

Since (3) maximizes a linear function over a bounded and closed convex domain, an optimal solution certainly exists. This, by the way, is predicted by the duality theory for CO, which also guarantees that the optimal values of (2) and (3) are equal and attained.

Solution of Fermat's problem

In this section we show that we have also *strong duality*, i.e., the optimal values of

(2) and (3) are equal. Using this we derive the main result of this section, namely that in case F1 the Fermat point P satisfies

$$\angle APB = \angle BPC = \angle CPA = 120^\circ. \quad (5)$$

As was established in the previous section, (2) has a (unique) optimal solution, denoted by x_P . We have strong duality if and only if there exists a dual feasible y such that we have equalities throughout in (4), with $x = x_P$. This happens if and only if, for each i ,

$$\|a_i - x_P\| = \|a_i - x_P\| \|y_i\| \\ = (a_i - x_P)^T y_i. \quad (6)$$

Let us first consider the case where $x_P \neq a_i$, for each i . Then we have $a_i - x_P \neq 0$ for each i . The first equality in (6) then implies $\|y_i\| = 1$, and then the second equality implies

$$y_i = \frac{a_i - x_P}{\|a_i - x_P\|}. \quad (7)$$

Thus, if x_P is not one of the vertices of $\triangle ABC$ we can conclude that (6) determines each of the vectors y_i uniquely. We need still to verify whether these vectors are dual feasible. This holds if and only if $\sum_{i=1}^3 y_i = 0$. Since $y_1 + y_2 = -y_3$ it follows that $\|y_1 + y_2\| = \|y_3\| = 1$. Since

$$1 = \|y_1 + y_2\|^2 \\ = \|y_1\|^2 + 2y_1^T y_2 + \|y_2\|^2 \\ = 2y_1^T y_2 + 2,$$

it follows that $y_1^T y_2 = -\frac{1}{2}$. Hence, y_1 and y_2 must make an angle of 120° . According to (7) the angle between the vectors y_1 and y_2 is the same as the angle between the vectors $a_1 - x$ and $a_2 - x$, which is the angle between PA and PB . Thus, $\angle APB = 120^\circ$. The same argument applies to any two of the vectors y_i . So we may conclude that the Fermat point P satisfies (5).

The Fermat point must lie in the interior of $\triangle ABC$. This can now be shown by using $\sum_{i=1}^3 y_i = 0$ once more. This relation is equivalent to

$$\frac{a_1 - x_P}{\|a_1 - x_P\|} + \frac{a_2 - x_P}{\|a_2 - x_P\|} + \frac{a_3 - x_P}{\|a_3 - x_P\|} = 0,$$

which can be rewritten as

$$x_P = \lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3, \quad (8)$$

where

$$\lambda_i = \frac{\frac{1}{\|a_i - x_P\|}}{\frac{1}{\|a_3 - x_P\|} + \frac{1}{\|a_3 - x_P\|} + \frac{1}{\|a_3 - x_P\|}}, \quad (9) \\ i = 1, 2, 3.$$

Note that $\lambda_i > 0$ for each i . The sum of these coefficients being 1, this means that x_P is a convex combination of the vectors a_i . Since each λ_i is positive, it follows that P lies in the interior of $\triangle ABC$.

Since the sum of the angles in the $\triangle APC$ is 180° , we have $\angle APC + \angle ACP + \angle PAC = 180^\circ$. Since $\angle APC = 120^\circ$ and $\angle PAC > 0$ we get $\angle ACP < 60^\circ$. In the same way we get $\angle BCP < 60^\circ$. Hence, $\angle C = \angle ACP + \angle BCP < 120^\circ$. The same holds for $\angle A$ and $\angle B$. We conclude that if $x_P \neq a_i$ for each i , we are in case F1.

It remains to deal with the case where $x_P = a_i$ for some i . Without loss of generality we assume that this happens for $i = 3$. In that case the optimal value is simply given by $\|a_3 - a_1\| + \|a_3 - a_2\|$. We next show that this happens only in case F2.

Since the points A , B and C are not collinear, we have $x_P \neq a_i$ for $i \in \{1, 2\}$. Hence, y_1 and y_2 are given by (7) and $y_3 = -y_1 - y_2$. Dual feasibility requires $\|y_3\| \leq 1$, which holds if and only if $\|y_1\|^2 + 2y_1^T y_2 + \|y_2\|^2 \leq 1$. Since $\|y_1\| = \|y_2\| = 1$, this holds if and only if $y_1^T y_2 \leq -\frac{1}{2} = \cos 120^\circ$. Therefore, $\angle APB \geq 120^\circ$. Since $x_P = a_3$ we have $P = C$ and therefore also $\angle ACB \geq 120^\circ$, proving that we are in case F2.

At this stage we may conclude that we have completely solved (F) in case F2: then the Fermat point is the vertex with the largest angle, and s_P equals the sum of the distances of that vertex to the other two vertices. As is pointed out in [2] this case has either been overlooked or weakly treated in the early literature. It was explicitly stated for the first time by Bonaventura Cavalieri (1598–1647) [4, pp. 504–510]. A simple geometric proof of its solution had to wait until 1976 [17].

It is worthwhile to show Sokolowsky's geometric proof for case F2. Without loss of

generality we suppose that $\angle C = \gamma > 120^\circ$. Now let Q be an arbitrary point in $\mathbb{R}^2$ different from C . In Figure 2 Q lies inside $\triangle ABC$, but the argument below is also valid when Q lies outside the triangle or on its boundary. With $\beta = 180^\circ - \gamma$ we turn the triangle CQA clockwise around the vertex C over the angle β . This yields a triangle $CQ'A'$ such that the vectors $A'C$ and CB point in the same direction and

$$|CQ| = |CQ'|, |AQ| = |A'Q'|, |CA| = |CA'|.$$

Since $\angle Q'CQ = \beta < 60^\circ$, the equal base angles in the isosceles triangle $Q'CQ$ exceed β , which implies $|CQ| > |QQ'|$. By using this and $|AQ| = |A'Q'|$ we get

$$s_Q = |QA| + |QB| + |QC| \\ > |Q'A'| + |QB| + |QQ'|.$$

The last expression is the length of the polygonal path $A'Q'QB$, which is certainly larger than the length of the straight path $A'CB$. Since $|A'C| = |AC|$ we obtain

$$s_Q > |AC| + |CB| = s_C.$$

Since this holds for any point Q in $\mathbb{R}^2$ different from C , it follows that C is the Fermat point.

From now on we restrict ourselves to case F1. In that case neither x_P nor s_P is yet known; we only have proved (5). But, as observed by Torricelli³, this property already enables us to construct P . His surprisingly simple construction is demonstrated in Figure 3. The circumcircles of the outward equilateral triangles on the sides AC and BC intersect at the Fermat point, because the three angles between the dashed lines PA , PB and PC all equal 120° .

Torricelli also found a geometric proof that (5) must hold. We present this later. But first Viviani's Theorem is presented.

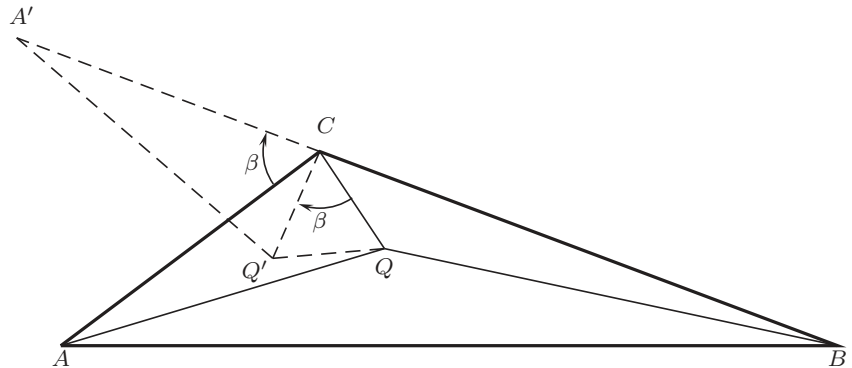


Figure 2 Sokolowsky's proof for case F2.

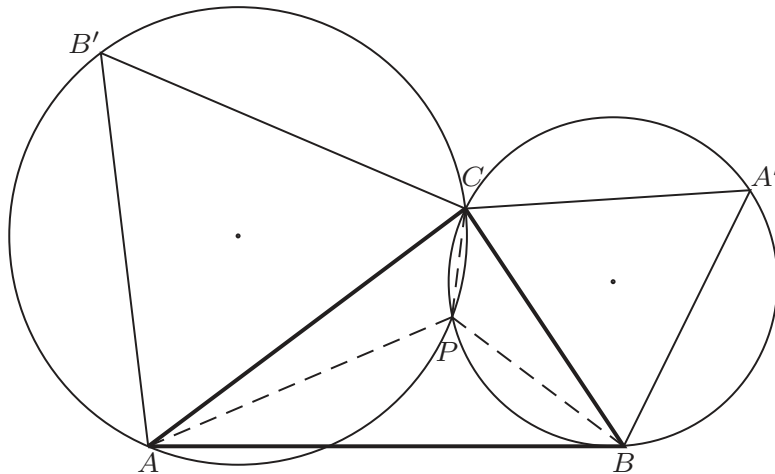


Figure 3 Torricelli's construction of the Fermat point in case F1.

Viviani's Theorem

In this section we deal with a lemma due to Torricelli's student Vincenzo Viviani (1622–1703) [18] now known as Viviani's Theorem.

Lemma 1 (Viviani's Theorem). *If T is an equilateral triangle then for each point X in T the sum of its distances to the sides of T equals the altitude h_T of T .*

Proof. The proof is easy and demonstrated in Figure 4. The point X divides $\triangle ABC$ into three triangles: $\triangle AXB$, $\triangle BXC$ and $\triangle CXA$. Since the sum of their areas equals the area of $\triangle ABC$ we get

$$\frac{1}{2}|AB|h_1 + \frac{1}{2}|BC|h_2 + \frac{1}{2}|CA|h_3 = \frac{1}{2}|AB|h.$$

Since $|AB|=|BC|=|CA|$, this implies $h = h_1 + h_2 + h_3$, proving the lemma. $\square$

It is worth noting that if X is on the boundary of T then one or two of the three triangles in the above proof reduce to a single point. But the argument used in the above proof remains valid. This also follows by considering a problem that is

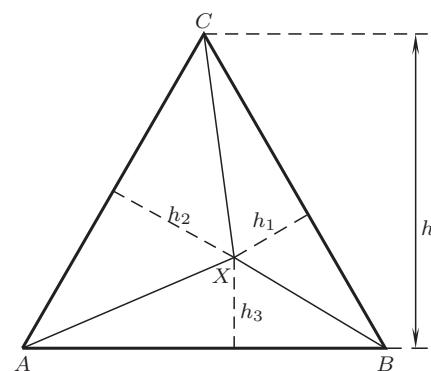


Figure 4 Viviani's Theorem: $h = h_1 + h_2 + h_3$.

quite similar to Fermat's problem but much easier to solve, namely

(V) Find the point V that minimizes (or maximizes) the sum τ_V of the distances of V to the sides of $\triangle ABC$.

It will be convenient below to use the notations

$$a := OA, \quad b := OB, \quad c := OC.$$

Lemma 2. *The minimal (or maximal) value of τ_V is attained at a vertex of $\triangle ABC$, and it equals the height of $\triangle ABC$ seen from that vertex.*

Proof. Let n_A be a unit vector that is orthogonal to BC such that $n_A^T(b-a) > 0$. Then the distance to BC from any point X in $\triangle ABC$ is equal to $n_A^T(b-x)$. Defining n_B and n_C in a similar way, the sum of the distances of X to the sides of $\triangle ABC$ is given by

$$f(x) := n_A^T(b-x) + n_B^T(c-x) + n_C^T(a-x).$$

This function depends linearly on X . Hence, its minimal (or maximal) value is attained at a vertex of $\triangle ABC$ and equals the distance of that vertex to the opposite side of $\triangle ABC$. $\square$

The height value of an equilateral triangle seen from one of its vertices is independent of that vertex. Hence, $f(x)$ is constant on an equilateral triangle, which yields a second proof of Viviani's Theorem.

Torricelli's triangle

We now show how Torricelli used Viviani's Theorem to solve Fermat's problem. We

have already seen how Torricelli constructed a point P in $\triangle ABC$ that satisfies (5). How did he know that the Fermat point has this property?⁴

For that purpose he introduced another triangle whose sides contain the vertices A , B and C and are perpendicular to AP , BP and CP , respectively, with P such that the three angles at P are 120° . In this way he obtained the dashed triangle $A_1B_1C_1$ in Figure 5, now called *Torricelli's triangle*. Since $\angle CPB$ is 120° and $\angle PCA_1 = \angle PBA_1 = 90^\circ$ we have $\angle A_1 = 360^\circ - 120^\circ - 2 \times 90^\circ = 60^\circ$. For similar reasons also $\angle B_1 = \angle C_1 = 60^\circ$. Hence, $\triangle A_1B_1C_1$ is equilateral.

By the construction of Torricelli's triangle the sum of the distances of P to its sides is equal to $|PA| + |PB| + |PC|$, which is s_P . By Viviani's Theorem this sum is equal to the height of Torricelli's triangle, which we denote by h . Thus, $s_P = h$.

Now let Q be any other point in $\triangle ABC$. Viviani's Theorem tells us that the sum of the distances from Q to the sides of Torricelli's triangle also equals h . The grey triangles in Figure 5 make clear that this sum is less than or equal to $|QA| + |QB| + |QC|$, which is s_Q . So we obtain $s_Q \geq h = s_P$, proving that P minimizes the sum of the distances to A , B and C . It means that P solves (F), which implies that it is the Fermat point.

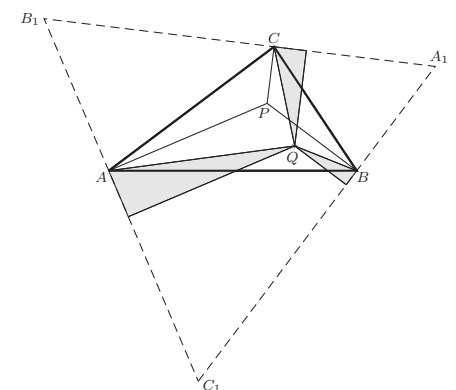
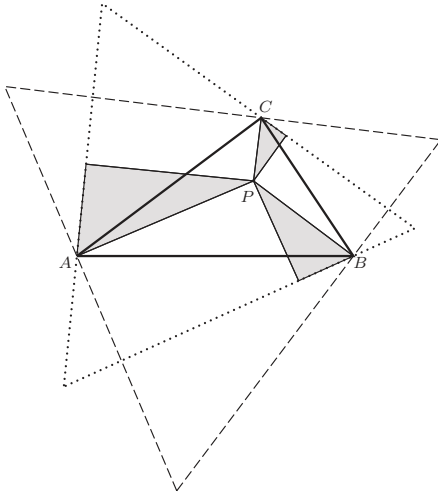


Figure 5 Torricelli's proof of (5).

Duality of (F) and (D)

Torricelli's triangle can also be used to prove the duality relation between the problems (F) and (D), as described in Theorem 1. In order to prove this theorem we have also drawn another equilateral (dotted) triangle in Figure 6 such that the points A , B and C lie on different sides, as well as line segments from P orthogonal to the sides of this triangle.

Figure 6 Torricelli's triangle maximizes h_T .

Considering the grey triangles in the figure it becomes clear that the sum of the distances of P to the sides of the dotted triangle is less than the sum of the distances of P to the sides of Torricelli's triangle. By Viviani's Theorem these sums are equal to the heights of the respective triangles. The dotted triangle's height is therefore less than the height of the Torricelli triangle. Since the height of Torricelli's triangle equals s_P , this proves Theorem 1.

Note that the above proof also makes clear that Torricelli's triangle for $\triangle ABC$ solves problem (D). So, when Torricelli was solving Fermat's problem, i.e., problem (F'), he also solved problem (D), although most likely being unaware of this. This is typical for problems that are each others dual. At first sight they seem to be unrelated—though both are based on the same input data, but when solving one of the two problems usually one gets enough information to also solve the other problem. Harold W. Kuhn [12] states: "Until further evidence is discovered this must stand as the first instance of duality in nonlinear programming" as was established in 1811–1812 by M. Vecten, professor of *mathématiques spéciales* at the Lycée de Nîmes.

In the next section we discuss another method to construct the Fermat point. This method will enable us to find an expression for s_P in terms of the given $\triangle ABC$. This will be done in the section thereafter.

Simpson lines

To introduce Simpson's method we use Figure 7, which shows a triangle ABC satisfy-

ing F1 and its Fermat point P . So the three angles at P are all equal to 120° .

We turn $\triangle BPC$ over 60 degrees clockwise around B , yielding $\triangle BP'A'$, as shown in Figure 7. Then $\triangle BCA'$ is isosceles. Since $\angle CBA'$ is 60 degrees it is even equilateral. The same argument yields that also $\triangle BPP'$ is equilateral. Hence, all corners in both triangles are 60° . Since $\angle APB = 120^\circ$ it follows that $\angle APP' = 180^\circ$. This proves that APP' is a straight line segment. The same argument yields that also $PP'A'$ is a straight line segment. It follows that $APP'A'$ is a straight line segment. The length of this segment equals

$$|AP| + |PP'| + |P'A'| = |AP| + |BP| + |CP|,$$

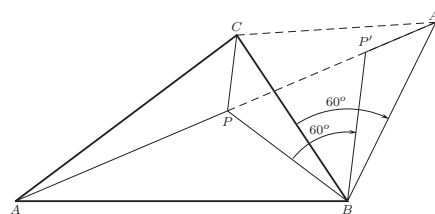
which is exactly s_P , the sum of the distances of the Fermat point P to the points A , B and C .

The line AA' is named a *Simpson line*. It can be easily constructed by ruler and compass by first drawing the outward equilateral triangle BCA' . In the same way one may construct Simpson lines BB' and CC' . The Simpson lines intersect at P . This construction of the Fermat point is due to Simpson⁵ (1710–1761) [16]. Strangely enough Simpson did not observe that the lengths of the three Simpson lines are the same and equal to s_P . This was first shown by Heinen [8].

As a byproduct of the fact that $\triangle BPP'$ is equilateral we get $\angle BPP' = 60^\circ$, which means that the Simpson line PA' is the bisector of $\angle BPC$ as was observed in [11]. This in turn implies that P is also the Fermat point of $\triangle A'B'C'$, because of (5). Note that $|AA'| + |BB'| + |CC'| = 3s_P$ implies

$$|AP| + |PA'| + |BP| + |PB'| + |CP| + |PC'| = 3s_P.$$

Hence, $|PA'| + |PB'| + |PC'| = 2s_P$, which is the sum of the distances from the Fermat point of $\triangle A'B'C'$ to its vertices.

Figure 7 Simpson line AA' .

Formula for s_P

Using the property of the Simpson lines just found, an expression of s_P in terms of $\triangle ABC$ is immediately at hand, simply by computing the length of the line segment AA' in Figure 7. To compute the point A' we use that it connects the mid point of BC by a line perpendicular to BC . Defining $a' := OA'$, we then have

$$a' = \frac{1}{2}(b+c) + d,$$

where d is a vector perpendicular to BC whose length equals the distance of A' to BC . Since $\triangle CBA'$ is equilateral, this distance equals $\frac{1}{2}\sqrt{3}\|b-c\|$. Moreover, the vector $[b_2 - c_2; c_1 - b_1]$ is orthogonal to BC and has length $\|b-c\|$. Hence,

$$d = \pm \frac{1}{2}\sqrt{3} \begin{bmatrix} b_2 - c_2 \\ c_1 - b_1 \end{bmatrix}. \quad (10)$$

One easily understands that d is an outward direction at $\frac{1}{2}(b+c)$ if $d^T(b-a) > 0$ and an inward direction if $d^T(b-a) < 0$. It is worth to have a closer look at the quantity $d^T(b-a)$. We may write

$$\begin{aligned} & \begin{bmatrix} b_2 - c_2 \\ c_1 - b_1 \end{bmatrix}^T \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \end{bmatrix} \\ &= (b_2 - c_2)(b_1 - a_1) + (c_1 - b_1)(b_2 - a_2) \\ &= -b_2a_1 - c_2b_1 + c_2a_1 + c_1b_2 - c_1a_2 + b_1a_2 \\ &= \det \begin{bmatrix} c & b \end{bmatrix} - \det \begin{bmatrix} c & a \end{bmatrix} + \det \begin{bmatrix} b & a \end{bmatrix} \\ &= \det \begin{bmatrix} 1 & 1 & 1 \\ c & b & a \end{bmatrix}, \end{aligned}$$

where the last equality sign follows by evaluating the determinant $\det \begin{bmatrix} 1 & 1 & 1 \\ c & b & a \end{bmatrix}$ to its first row. The absolute value of this determinant equals two times the area of $\triangle ABC$. In the sequel we assume that the determinant is positive; if this is not so, interchanging the names of two vertices will make it positive. As a consequence d will be an outward direction at $\frac{1}{2}(b+c)$. So we must use the plus sign in (10). This gives

$$a' = \frac{1}{2}(b+c) + \frac{1}{2}\sqrt{3} \begin{bmatrix} b_2 - c_2 \\ c_1 - b_1 \end{bmatrix}. \quad (11)$$

As before, we use $|AB|$ to denote the length of line segment AB . So $|AB| = \|a-b\|$.

We also use $|ABC|$ to denote the area of $\triangle ABC$.

Lemma 3. In case F1 one has

$$s_P = \sqrt{\frac{1}{2}(|AB|^2 + |BC|^2 + |CA|^2) + 2\sqrt{3}|ABC|}. \quad (12)$$

Proof. From Section 7 we know that $s_P = \|a - a'\|$. Therefore, $s_P^2 = \|a\|^2 + \|a'\|^2 - 2a^T a'$. Since

$$\begin{aligned} |a'|^2 &= \frac{1}{4}\|b+c\|^2 + \frac{3}{4}\|b-c\|^2 + \frac{1}{2}\sqrt{3}(b+c)^T \begin{bmatrix} b_2-c_2 \\ c_1-b_1 \end{bmatrix} \\ &= \|b\|^2 + \|c\|^2 - b^T c + \frac{1}{2}\sqrt{3}((b_1+c_1)(b_2-c_2) + (b_2+c_2)(c_1-b_1)) \\ &= \|b\|^2 + \|c\|^2 - b^T c + \sqrt{3}(b_2 c_1 - b_1 c_2) \\ &= \|b\|^2 + \|c\|^2 - b^T c + \sqrt{3} \det[c \ b]. \end{aligned}$$

and

$$\begin{aligned} 2a^T a' &= a^T \left(b+c + \sqrt{3} \begin{bmatrix} b_2-c_2 \\ c_1-b_1 \end{bmatrix} \right) \\ &= a^T b + a^T c + \sqrt{3}(a_1 b_2 - a_2 b_1 - a_1 c_2 + a_2 c_1) \\ &= a^T b + a^T c + \sqrt{3}(\det[a \ b] - \det[a \ c]), \end{aligned}$$

it follows that

$$\begin{aligned} s_P^2 &= \|a\|^2 + \|b\|^2 + \|c\|^2 - a^T b - b^T c - a^T c \\ &\quad + \sqrt{3}(\det[c \ b] - \det[a \ b] + \det[a \ c]) \\ &= \frac{1}{2}(\|a-b\|^2 + \|b-c\|^2 + \|c-a\|^2) + \sqrt{3} \det \begin{bmatrix} 1 & 1 & 1 \\ c & b & a \end{bmatrix}. \end{aligned}$$

This gives the lemma. $\square$

Surprisingly enough the formula for s_P just found also appears when dealing with so-called Napoleon triangles. This will be the subject of the next section, where we find a line segment of length $s_P/\sqrt{3}$.

Napoleon's triangles

Figure 8 shows a triangle ABC and the three outward equilateral triangles on its sides. The centres of the triangles BCA' , CAB' and ABC' are denoted by X , Y and Z , respectively. Napoleon's Theorem⁶ states that the triangle XYZ is equilateral. Moreover, the same holds for the triangle $X'Y'Z'$, where X' , Y' and Z' are the centres of the inward equilateral triangles on the sides of ABC . The triangle XYZ is called the *outward Napoleon triangle* and $X'Y'Z'$ the *inward Napoleon triangle*.

The first part of the next theorem is just Napoleon's Theorem. Part of the second statement is also known, namely that XYZ and $X'Y'Z'$ have the same centre [9].

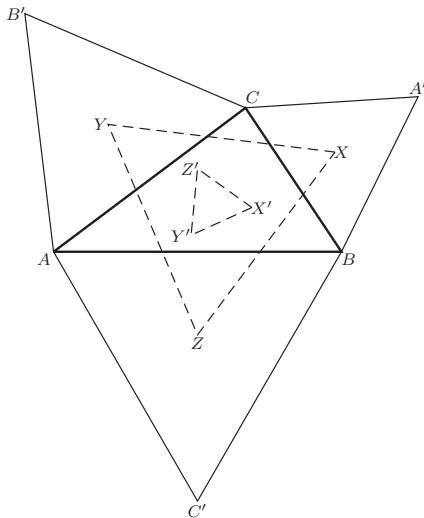


Figure 8 Napoleon's triangles.

Theorem 2. The triangles XYZ and $X'Y'Z'$ are equilateral. The centres of these triangles coincide with the centre of ABC . Moreover,

$$s_P = \sqrt{3} |XY|$$

and

$$|XY|^2 + |X'Y'|^2 = \frac{1}{3}(|AB|^2 + |BC|^2 + |CA|^2), \quad (13)$$

$$|XY|^2 - |X'Y'|^2 = \frac{4}{\sqrt{3}} |ABC|. \quad (14)$$

Proof. Let us define $a = OA$, $x = OX$, et cetera. Then $a' = OA'$ is given by (11). Since $x = \frac{1}{3}(b+c+a')$ it therefore follows that

$$x = \frac{1}{3} \left(b+c + \frac{1}{2}(b+c) + \frac{1}{2}\sqrt{3} \begin{bmatrix} b_2-c_2 \\ c_1-b_1 \end{bmatrix} \right) = \frac{1}{2}(b+c) + \frac{1}{6}\sqrt{3} \begin{bmatrix} b_2-c_2 \\ c_1-b_1 \end{bmatrix}.$$

The centre X' of the inward equilateral triangle on side BC is obtained by mirroring X in the line BC , which gives

$$x' = \frac{1}{2}(b+c) - \frac{1}{6}\sqrt{3} \begin{bmatrix} b_2-c_2 \\ c_1-b_1 \end{bmatrix}.$$

In the same way we get the following expressions for y , y' , z and z' :

$$\begin{aligned} y &= \frac{1}{2}(c+a) + \frac{1}{6}\sqrt{3} \begin{bmatrix} c_2-a_2 \\ a_1-c_1 \end{bmatrix}, & y' &= \frac{1}{2}(c+a) - \frac{1}{6}\sqrt{3} \begin{bmatrix} c_2-a_2 \\ a_1-c_1 \end{bmatrix}, \\ z &= \frac{1}{2}(a+b) + \frac{1}{6}\sqrt{3} \begin{bmatrix} a_2-b_2 \\ b_1-a_1 \end{bmatrix}, & z' &= \frac{1}{2}(a+b) - \frac{1}{6}\sqrt{3} \begin{bmatrix} a_2-b_2 \\ b_1-a_1 \end{bmatrix}. \end{aligned}$$

From the above expressions one readily obtains $x+y+z = x'+y'+z' = a+b+c$, which implies that the centres of XYZ , $X'Y'Z'$ and ABC coincide.

One has $|XY| = \|x-y\|$ and $|X'Y'| = \|x'-y'\|$. We start by computing $\|x-y\|$. We may write

$$\begin{aligned} x-y &= \frac{1}{2}(b-a) + \frac{1}{2\sqrt{3}} \left(\begin{bmatrix} b_2-c_2 \\ c_1-b_1 \end{bmatrix} - \begin{bmatrix} c_2-a_2 \\ a_1-c_1 \end{bmatrix} \right) \\ &= \frac{1}{2}(b-a) + \frac{1}{2\sqrt{3}} \begin{bmatrix} a_2+b_2-2c_2 \\ 2c_1-a_1-b_1 \end{bmatrix}. \end{aligned}$$

As a consequence we have

$$\begin{aligned} \|x-y\|^2 &= \left(\frac{1}{2}(b_1-a_1) + \frac{\sqrt{3}}{6}(a_2+b_2-2c_2) \right)^2 + \left(\frac{1}{2}(b_2-a_2) + \frac{\sqrt{3}}{6}(2c_1-a_1-b_1) \right)^2 \\ &= \frac{1}{4}(b_1-a_1)^2 + \frac{1}{4}(b_2-a_2)^2 + \frac{1}{12}(a_2+b_2-2c_2)^2 + \frac{1}{12}(2c_1-a_1-b_1)^2 \\ &\quad + \frac{\sqrt{3}}{6}(b_1-a_1)(a_2+b_2-2c_2) + \frac{\sqrt{3}}{6}(b_2-a_2)(2c_1-a_1-b_1) \\ &= \frac{1}{12}(3(b_1-a_1)^2 + 3(b_2-a_2)^2 + (a_2+b_2-2c_2)^2 + (2c_1-a_1-b_1)^2) \\ &\quad + \frac{\sqrt{3}}{6}((b_1-a_1)(a_2+b_2-2c_2) + (b_2-a_2)(2c_1-a_1-b_1)) \\ &= \frac{1}{3}(a_1^2 + a_2^2 + b_1^2 + b_2^2 + c_1^2 + c_2^2 - a_1 c_1 - a_2 c_2 - a_1 b_1 - b_1 c_1 - a_2 b_2 - b_2 c_2) \\ &\quad + \frac{\sqrt{3}}{3}(a_2(b_1-c_1) + b_2 c_1 - b_1 c_2 + a_1(-b_2+c_2)) \\ &= \frac{1}{3}(\|a\|^2 + \|b\|^2 + \|c\|^2 - c^T b - b^T a - a^T c) \\ &\quad + \frac{1}{\sqrt{3}}(\det[b \ a] - \det[c \ a] + \det[c \ b]) \\ &= \frac{1}{6}(\|a-b\|^2 + \|b-c\|^2 + \|c-a\|^2) + \frac{1}{\sqrt{3}} \det \begin{bmatrix} 1 & 1 & 1 \\ c & b & a \end{bmatrix}. \end{aligned}$$

Comparison of the last expression with the expression for s_P^2 in (12) yields $s_P^2 = 3\|x-y\|^2$, which gives the value for $|XY|$ in the theorem. Since the expression is invariant under cyclic permutations of

a , b and c it follows that $\|x - y\| = \|y - z\| = \|z - x\|$, which implies that $\triangle XYZ$ is equilateral.

Now consider the triangle $X'Y'Z'$. We have

$$\begin{aligned} x' - y' &= \frac{1}{2}(b - a) - \frac{1}{2\sqrt{3}} \left(\begin{bmatrix} b_2 - c_2 \\ c_1 - b_1 \end{bmatrix} + \begin{bmatrix} c_2 - a_2 \\ a_1 - c_1 \end{bmatrix} \right) \\ &= \frac{1}{2}(b - a) - \frac{1}{2\sqrt{3}} \begin{bmatrix} a_2 + b_2 - 2c_2 \\ 2c_1 - a_1 - b_1 \end{bmatrix}. \end{aligned}$$

The only difference with the expression for $x - y$ is the sign of the second term. It is easy to check that a similar derivation as for $\|x - y\|^2$ leads to

$$\begin{aligned} \|x' - y'\|^2 &= \frac{1}{6}(\|a - b\|^2 + \|b - c\|^2 + \|c - a\|^2) \\ &\quad - \frac{1}{\sqrt{3}} \det \begin{bmatrix} 1 & 1 & 1 \\ c & b & a \end{bmatrix}. \end{aligned}$$

From this and the expression for $\|x - y\|^2$ we readily obtain (13) and (14). $\square$

The area of an equilateral triangle with side length s equals $\frac{\sqrt{3}}{4}s^2$. Moreover, the area $|ABC|$ of $\triangle ABC$ is equal to $\frac{1}{2} \det \begin{bmatrix} 1 & 1 & 1 \\ c & b & a \end{bmatrix}$. As a consequence of (14) we therefore get a surprisingly simple relation between the areas of Napoleon's triangles and the area of $\triangle ABC$, namely (cf. [5] and [9]) $|ABC| = |XYZ| - |X'Y'Z'|$.

Concluding remarks

As was made clear the key to the solution of Fermat's problem is property (5), which we obtained straightforwardly by using the conic dual of Fermat's problem. It may be worth pointing out that at first (and second!) sight there is no obvious relation between the 'analytic' dual (3) and the 'geometric' dual (D) of Fermat's problem. Whereas (2) easily is identified as a mathematical model for Fermat's problem (F), it seems harder to recognize (3) as a mathe-

matical model for (D). This is remarkable and seems to deserve further investigation.

Different variants of Fermat's original problem have been investigated by several authors. Among the more obvious extensions are weighted versions where the distances are multiplied by positive or negative weights associated with the vertices of $\triangle ABC$ as in e.g. [7] and [9]. Other lines of research have focused upon the number of given points and/or the dimension of the space. For an overview the interested reader is referred to [3]. $\diamond$

Acknowledgement

Thanks are due Professor Marijana Zekić-Sušac, the chief organizer of the Croatian Operations Research Society's conference held in Osijek, Croatia, September 2016. Without her invitation to both of us to attend this conference, the present paper, let alone its potential successors, would never have materialized.

Notes

- 1 There seems to be uncertainty about the birth year of Fermat, see e.g. [11].
- 2 *The Ladies' Diary: or, Woman's Almanack* appeared annually in London from 1704 to 1841 after which it was succeeded by *The Lady's and Gentleman's Diary*. It was a respectable place to pose mathematical problems (brain teasers!) and sustain debate.
- 3 Torricelli was a famous French number theorist, student of Galileo and discoverer of the barometer.
- 4 Some authors who ask this question leave it unanswered, suggesting that it was based on Torricelli's intuition [3, p. 88; 7, p. 444].
- 5 Thomas Simpson (1710–1761) was an English mathematician. He taught mathematics at the Royal Military Academy and became known for the Simpson rule for numerical integration [6].
- 6 According to [11] and others, any connection between Napoleon's Theorem and the French emperor is highly doubtful. It may

amuse the reader to know that Napoleon's Theorem is claimed to be one of the most often rediscovered results in triangular geometry. Thus, no less than 32 sources related to it are mentioned in [19].

References

- 1 A. Ben-Tal and A. Nemirovski, Lectures on modern convex optimization, *MPS/SIAM Series on Optimization*, SIAM, 2001.
- 2 V. Boltyanski, H. Martini and V. Soltan, *Geometric Methods and Optimization Problems*, Combinatorial Optimization, Vol. 4, Kluwer, 1999.
- 3 G. Bruno, A. Genovese and G. Improta, A historical perspective on location problems, *Journal of the British Society for the History of Mathematics* 29(2) (2014), 83–97.
- 4 B. Cavalieri, *Exercitationes geometricae*, 1657.
- 5 H.S.M. Coxeter and S.L. Greitzer, *Geometry Revisited*, New Mathematical Library, Vol. 19, Random House, 1967.
- 6 S.I. Gass and A.A. Assad, *An annotated timeline of operations research: an informal history*, International Series in Operations Research & Management Science, Vol. 75, Kluwer, 2005.
- 7 S. Gueron and R. Tessler, The Fermat–Steiner problem, *Amer. Math. Monthly* 109(5) (2002), 443–451.
- 8 F.A. Heinen, Über systeme von Kräften, Gymnasium zu Cleve, 1834.
- 9 R. Honsberger, *Mathematical Gems from Elementary Combinatorics, Number Theory, and Geometry*, The Mathematical Association of America, 1973.
- 10 G. Jalal and J. Krarup, Geometrical solution to the Fermat problem with arbitrary weights, *Ann. Oper. Res.* 123 (2003), 67–104.
- 11 J. Krarup and S. Vajda, On Torricelli's geometrical solution to a problem of Fermat, *IMA J. Math. Appl. Bus. Indust.* 8(3) (1997), 215–224.
- 12 H.W. Kuhn, Nonlinear programming: A historical note, in: J.K. Lenstra, A.H.G. Rinnooy Kan and A. Schrijver, eds., *History of Mathematical Programming. A Collection of Personal Reminiscences*, CWI, 1991, pp. 82–96.
- 13 T. Moss, *Ladies Diary* 620, Question CCCXCVI, 1755.
- 14 R.T. Rockafellar, *Convex Analysis*, Prentice-Hall, 1970.
- 15 C. Roos, M. Balvert, B.L. Gorissen and D. den Hertog, A universal and structured way to derive dual optimization problem formulations. Submitted.
- 16 T. Simpson, *The Doctrine and Application of Fluxions*, 1750.
- 17 D. Sokolowsky, A note on the Fermat problem, *Amer. Math. Monthly* 83(4) (1976), 276.
- 18 V. Viviani, *De maximis et minimis divinatione in quintum Conicorum Apollonii Pergaei. Part II*, Firenze, 1659.
- 19 J.E. Wetzel, Converses of Napoleon's theorem, *Amer. Math. Monthly* 99(4) (1992), 339–351.

In de verdediging

| In defence

Pas gepromoveerden brengen hun werk onder de aandacht. Heeft u tips voor deze rubriek of bent u zelf pas gepromoveerd? Laat het weten aan onze redacteur.

Redacteur: Geertje Hek
la Voie-du-Coin 7
1218 Grand-Saconnex
Zwitserland

verdediging@nieuwarchief.nl

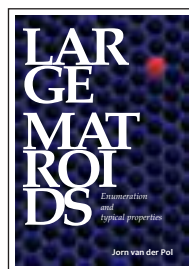


Foto: Elisa van Hout

Large Matroids, Enumeration and Typical Properties

Jorn van der Pol

‘Promovendus blaast nieuw leven in wiskundig vakgebied’, kopte *Cursor*, de nieuwssite van de Technische Universiteit Eindhoven, nadat Jorn van der Pol op 20 september *cum laude* was gepromoveerd. De combinatie van promotor prof. dr. Remco van der Hofstad en co-promotor dr. Rudi Pendavingh had de nieuwsgierigheid al gewekt: een kansrekenaar en een discreet wiskundige. En de titel van zijn proefschrift, *Large Matroids, Enumeration and Typical Properties*, vroeg om uitleg: “Wat zijn matroïden eigenlijk?” De *Cursor*-kop bevestigde de keuze om Van der Pol te interviewen nog eens extra.

Matroïden

Matroïden zijn abstracte meetkundige objecten die bestaan uit punten, lijnen en hoger-dimensionale tegenhangers. Deze objecten zijn sterk verbonden met verschillende gebieden binnen de wiskunde, zoals lineaire algebra en grafentheorie. Veel interessante problemen, zoals het vinden van minimale opspannende bomen, kunnen worden geformuleerd in termen van matroïden, en veel algoritmen om dergelijke problemen op te lossen zijn in feite matroïde-algoritmen.

De term matroïde werd door Hassler Whitney geïntroduceerd in 1935 in zijn artikel ‘On the abstract properties of linear dependence’. Hij kwam als volgt tot zijn definitie: neem een verzameling vectoren E die leeft in een of andere eindig-dimensionale vectorruimte V (je kunt dus zeggen dat E de verzameling kolommen van een of andere matrix is) en veronderstel dat I een onafhankelijke verzameling vectoren in E is. Wat betekent dit? Ten eerste kun je stellen dat als J een deelverzameling van I is, ook J linear onafhankelijk is. Ten tweede geldt een soort uitbreidingsprincipe: als een deelverzameling J van E ook lineair onafhankelijk is en J bevat strikt meer elementen dan I , dan moet J een vector bevatten die niet in het opspansel van I zit.

In zijn artikel bestudeerde Whitney structuren die aan precies deze eigenschappen voldoen: hij definieerde een matroïde als een verzameling E en een collectie $\mathcal{I}$ van deelverzamelingen van E , met de volgende eigenschappen:

1. $\mathcal{I}$ is niet-leeg;
2. Als $I \in \mathcal{I}$ is, en $J \subset I$, dan is ook $J \in \mathcal{I}$;
3. Als $I \in \mathcal{I}$ en $J \in \mathcal{I}$, en $|J| > |I|$, dan is er een element $j \in J$ zodat de verzameling $I \cup \{j\} \in \mathcal{I}$.

Eigenschap 1 sluit een triviaal object uit, terwijl eigenschappen 2 en 3 precies corresponderen met de genoemde eigenschappen van vectorruimten. Whitney noemde elementen in $\mathcal{I}$ ‘onafhankelijk’.

Je kunt laten zien dat de inclusiegewijs-maximale elementen van $\mathcal{I}$ allemaal even groot zijn: dit noem je de *rang* van de matroïde. In het geval van vectoren is dit precies de dimensie van $\langle E \rangle$. Inclusiegewijs-maximale onafhankelijke verzamelingen zelf noem je om voor de hand liggende reden *bases*.

Voorbeeld: grafische matroïden

Een belangrijk voorbeeld is dat van grafische matroïden, matroïden die afkomstig zijn van een graaf. Zij $G = (V, E)$ een graaf met V de verzameling punten (vertices) en E de verzameling kanten (edges), en bekijk de verzameling $\mathcal{I}$ van deelverzamelingen van E die geen cykel in G bevatten (zo'n verzameling kanten heet een 'bos'). Met een beetje moeite kun je laten zien dat $(E, \mathcal{I})$ aan de eigenschappen 1, 2, en 3 hierboven voldoet, en dus een matroïde is. Als G samenhangend is, dan zijn de bases van deze matroïde precies de opspannende bomen van G , en de rang van de matroïde is $|V| - 1$.

Een matroïde heet lineair als hij kan worden gerealiseerd als de matroïde behorende bij een verzameling vectoren, zoals hierboven beschreven. Je kunt laten zien dat iedere grafische matroïde lineair is. Andersom kun je je afvragen, en Whitney deed dat al, of iedere matroïde lineair is. (Het antwoord is nee.)

Grote matroïden

Met de opkomst van *data science* en grote netwerken van informatie kwamen grote matroïden steeds meer in de belangstelling. In zijn proefschrift behandelt Van der Pol twee vragen over zulke grote matroïden. Ten eerste: Hoe hard groeit het aantal matroïden op grondverzameling $\{1, 2, \dots, n\}$ wanneer $n \rightarrow \infty$? Of algemener, hoe hard groeit het aantal matroïden op E als functie van $|E|$? Hierop slaat het woord 'Enumeration' uit de ondertitel. Ten tweede: Wat voor eigenschappen kunnen we verwachten als n groot wordt?

Een exact antwoord op de eerste vraag is lastig te vinden, maar je kunt wel onder- en bovengrenzen maken. Van der Pol: "Zowel een onder- als een bovengrens bestond al, maar er zat nog een groot gat tussen. Wij vermoedden dat de bovengrens verbeterd zou kunnen worden, en dit hebben we gedaan."

Korte omschrijvingen en telefoonnummers

Een probleem bij het vinden van goede bovengrenzen is dat de gebruikelijke omschrijvingen van matroïden lastig te analyseren zijn. Het belangrijkste ingrediënt van Van der Pols argument bestaat daarom uit een 'korte' omschrijving van een matroïde. Hiervoor heeft hij het verband tussen enumeratie en complexiteit uitgebuit: als je voor iedere matroïde M op E een unieke 'korte' omschrijving $f(M)$ kunt vinden, en je kunt het aantal omschrijvingen begrenzen, dan kun je ook het aantal mogelijke matroïden begrenzen. Met dit argument heeft Van der Pol de vorige bovengrens, die ruim veertig jaar oud was, verbeterd. Een echte doorbraak.

Van der Pol illustreert zijn idee voor een korte omschrijving als volgt: stel dat je in een land bent, en je wilt weten hoeveel mensen daar wonen. Het is lastig om die vraag exact te beantwoorden. Maar als je weet dat iedereen een 4-cijferig telefoonnummer heeft, dan weet je wel dat er nooit meer dan 10000 mensen in dat land kunnen wonen. Hier zijn twee dingen van belang: (1) iedereen heeft een uniek telefoonnummer (anders tel je sommige mensen niet mee, of juist dubbel), en (2) je telefoonnummer is niet langer dan noodzakelijk (als iedereen een 10-cijferig telefoonnummer heeft, dan weet je wel dat er hooguit 10^{10} mensen in het land wonen, maar dat is een beroerde bovengrens).

Typische eigenschappen

De tweede vraag in Van der Pols proefschrift was de vraag hoe een 'typische' matroïde eruitziet. Ofwel: als je alle matroïden in

een zak zou stoppen en er willekeurig één uit zou halen, wat voor eigenschappen kun je dan verwachten?

Bewezen stellingen zijn van de vorm "bijna alle matroïden hebben eigenschap X ", wat wil zeggen dat de fractie van matroïden met eigenschap X naar 1 gaat als de grootte van de grondverzameling E naar oneindig gaat. Het lastige bij het bewijzen van dit soort stellingen is dat we niet goed weten hoe we een 'random' matroïde kunnen maken. Van der Pol buite daarom zijn korte omschrijvingsmethode verder uit. Soms heeft de afwezigheid van een bepaalde eigenschap X tot gevolg dat zijn omschrijving nog veel korter is. Het aantal matroïden zonder die eigenschap is dan veel kleiner dan het totaal aantal matroïden en dus moet het aantal matroïden met eigenschap X juist groot zijn. De eigenschap X is daarmee typisch.

Van der Pol gebruikt wederom de telefoonanalogie: Veronderstel dat er 5000 mensen in het land wonen, en neem verder aan dat mensen in de hoofdstad allemaal een nummer hebben dat met 0 begint. Er wonen dan hooguit 1000 mensen in de hoofdstad. Je weet dan ook dat minstens 80 procent van de mensen juist buiten de hoofdstad woont.

Een van de belangrijkste resultaten uit Van der Pols proefschrift is een afschatting van het typische aantal bases in een matroïde. Dit resultaat is een gevolg van de bovengenoemde korte omschrijving. Een aantal van zijn beste resultaten over typische eigenschappen van matroïden volgt relatief makkelijk uit dit algemene resultaat.

Afwisseling

Van der Pol vond vooral de afwisseling tijdens zijn promotietijd prettig: de afwisseling tussen onderwijs geven, zelf een college volgen, en het doen van onderzoek, maar ook de afwisseling die het werken aan verschillende projecten bracht. Die afwisseling hielp ook om over de moeilijker momenten heen te komen. Lang worstelen met een probleem kan demotiverend werken, aldus Van der Pol. Als het onderzoek lastig was hielp het hem om meerdere ijzers in het vuur te hebben. En dan was het uiteindelijk extra motiverend als hij na lang worstelen een probleem toch wist te temmen.

Zijn project was multidisciplinair, met een kansrekeningcomponent en een discrete-wiskundecomponent, wat ook uitdagingen met zich meebracht. Maar zijn begeleiders roemen Van der Pols talent om de verschillende disciplines in de wiskunde te kunnen verbinden.

Met die begeleiders was het heel prettig samenwerken. Het kantoor van Rudi Pendavingh bevond zich naast dat van Van der Pol, en hij kon altijd bij hem binnenlopen. Het dagelijks samen lunchen en af en toe ondernemen van gezamenlijke activiteiten met de andere promovendi en postdocs uit de gang droeg ook bij aan de goede sfeer. Van der Pol was ook betrokken bij de oprichting van de PhD/PDEng Council van de faculteit, een club die workshops en sociale activiteiten organiseert.

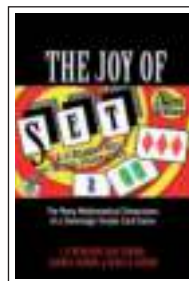
Als mooie momenten noemt Van der Pol nog een aantal mooie reizen die hij gemaakt heeft om andere universiteiten te bezoeken en conferenties bij te wonen: vaak uitputtend, maar hij heeft er veel leuke contacten aan overgehouden. En, hoewel hij daar tijdens het schrijven zelf wel eens anders over dacht, vond hij ook het schrijven van zijn proefschrift heel leuk. Een groot aantal van de resultaten was al opgeschreven, maar het schrijven van het proefschrift gaf hem de kans de teksten nog eens extra te polijsten en met elkaar te verbinden. Met succes: bij zijn promotie zei James Oxley, een van de autoriteiten binnen het vakgebied, dat hij in zijn 35 jaar als onderzoeker nog nooit zo'n goed proefschrift had gezien. ☺

Boekbesprekingen

| Book Reviews

Redactie: Hans Cuypers en Hans Sterk

Review Editors NAW - MF 7.092
 Faculteit Wiskunde & Informatica
 Technische Universiteit Eindhoven
 Postbus 513
 5600 MB Eindhoven
reviews@nieuwarchief.nl
www.win.tue.nl/wgreview



Liz McMahon, Gary Gordon,
 Hannah Gordon, Rebecca Gordon

The Joy of SET
The Many Mathematical Dimensions
of a Seemingly Simple Card Game

Princeton University Press, 2016

320 p., prijs \$29.95

ISBN 9780691166148

Het is geen wonder dat er artikelen en blogs zijn volgeschreven over de wiskunde achter SET. Het is een simpel en kleurrijk 'real-time' kaartspel gebaseerd op patroonherkenning. Tegelijkertijd heeft het iets abstracts en geeft het op een natuurlijke manier aanleiding tot tal van (combinatorische) vragen die soms zelfs raken aan wiskundig onderzoek (zie bijvoorbeeld de nummers van december 2002 en maart 2017 van dit tijdschrift).

The Joy of SET is het eerste boek gewijd aan de wiskunde die samenhangt met het spel SET. Het is bedoeld voor 'anyone with a deep curiosity about games' en vereist geen wiskundige voorkennis. Je zult dus geen formele stellingen of bewijzen vinden in dit boek, maar wel puzzels, vragen, inzichten en trucs. *The Joy of SET* houdt het midden tussen een boek over recreatieve wiskunde in de stijl van Gardner en een schoolboek over kansrekening en combinatoriek.

In het eerste hoofdstuk worden de SET-kaarten en de regels van het spel geïntroduceerd door middel van een dialoog tussen drie fictieve studenten, Stefan, Emily en Tanya, die zelf SET aan het spelen zijn. Meteen komen er natuurlijke vragen naar voren, zoals: "Hoeveel SETs zijn er?", "Wat is het verwachte aantal SETs in twaalf kaarten?", "Wat is het maximum aantal kaarten zonder SET?", enzovoort.

In de volgende vier hoofdstukken worden deze en andere vragen onderzocht. Zo wordt in het hoofdstuk over modulorekenen duidelijk dat de structuur van SET te maken heeft met rekenen modulo 3 en wordt in het hoofdstuk over meetkunde een verband gelegd met (eindige) meetkunde. In het hoofdstuk over tellen wordt uitgebreid stilgestaan bij combinaties, permutaties en binomiaalcoëfficiënten (inclusief een uitstapje naar de driehoek van Pascal).

In een intermezzo worden aardige variaties op het normale spel gegeven, waaronder manieren om een handicap te implementeren. De favoriete variatie van de auteurs is 'the Endgame'. Aan het begin wordt één kaart (dicht) opzij gelegd. Wanneer de laatste kaarten op tafel zijn gelegd mag ook de dichte kaart in een SET gebruikt worden. De kunst is om die kaart snel te bepalen uit de andere kaarten op tafel.

De laatste vijf hoofdstukken gaan wiskundig iets dieper dan de eerste vijf hoofdstukken. Zo wordt het spel gegeneraliseerd van 4 naar n eigenschappen en wordt bijvoorbeeld het aantal 2-dimensionale deelruimten (magische SET-vierkanten) geteld. Interessant is ook het laatste hoofdstuk waarin computereperimenten worden gedaan om bijvoorbeeld de kans te bepalen dat er aan het eind geen kaarten overblijven.

Het hele boek is luchtig en toegankelijk geschreven. Ook in latere hoofdstukken komen de dialogen terug, zoals in hoofdstuk 5 waarin Socrates, Euclides en Theano de postulaten van euclidische meetkunde onderzoeken. De auteurs maken scheutig gebruik van humor en voetnoten (die soms wat flauw zijn, maar eenvoudig

kunnen worden genegeerd). Ook de vele voorbeelden en (kleurige) figuren vergroten de toegankelijkheid. Ten slotte wordt elk hoofdstuk afgesloten met een groot aantal opgaven (met antwoorden achterin) waarmee de lezer zelf aan de slag kan. Het is duidelijk dat de auteurs (die als gezin al twintig jaar SET spelen) het boek met veel liefde hebben geschreven.

The Joy of SET is gericht aan een heel algemeen publiek. Onderwerpen uit de wetenschappelijke literatuur (denk aan Fourier-analyse en de polynomiale methode voor het begrenzen van cap sets, en relaties met andere onderwerpen uit de combinatoriek en theoretische informatica) komen niet of nauwelijks aan bod. Aan de andere kant is veel materiaal uit het boek juist prima bruikbaar op school: bij wiskunde D of buiten een reguliere les om. De grotere opgaven ('Projecten') bieden heel wat uitdaging en kunnen een goed startpunt zijn voor een praktische opdracht. *Dion Gijswijt*



Keith Devlin

Finding Fibonacci
The Quest to Rediscover the Forgotten
Mathematical Genius Who Changed
the World

Princeton University Press, 2017

256 p., prijs \$29.95

ISBN 9780691174860

De ondertitel zegt heel bondig direct al waar dit boek in essentie over gaat: 'The Quest to Rediscover the Forgotten Mathematical Genius Who Changed the World'. Inderdaad het verhaal van een 'zoektocht'. Maar wat valt er dan te zoeken? Iedereen weet toch waar het bij Fibonacci over gaat? Het woord 'konijnenrijtje' zegt toch al genoeg? Desnoods komen er dan nog wat andere punten naar boven, zoals de formules voor de algemene term van de rij en zogenaamde Lucas-rijen, maar dan hebben we het vaak wel gehad. Bovendien speelt dat alles zich af in de dertiende eeuw, dus vreselijk lang geleden in de nog 'donkere middeleeuwen'. Wat valt er dan allemaal te zoeken? En welk belang wordt met zo'n zoektocht gediend?

Op de een of andere manier was Keith Devlin gegrepen door de vraag wie Leonardo van Pisa, bekend staande als Fibonacci, nu eigenlijk was. En we herkennen allemaal wat er dan met je gebeurt. Als een onderwerp je eenmaal in de greep krijgt, dan wil je het op een bepaald moment niet meer loslaten. En zo liet Keith Devlin zijn onderwerp 'Fibonacci' niet meer los: wie was deze Leonardo, hoe zag zijn leven eruit, wat heeft hij voor ons vak betekend, waar zijn zijn geschriften gebleven, wat was het belang dat hij vertegenwoordigd heeft, enzovoort: enkele van de vragen die Devlin op zijn zoektocht bezighielden. Een letterlijke zoektocht die meer dan tien jaar heeft geduurd.

In de eerste plaats bleek dat, naast de bekende gemeenplaatsen, slechts bitter weinig bekend was rond de figuur van Fibonacci. Ook in de standaard historische werken was/is maar weinig specifieke te vinden. Een letterlijke tocht om de gegevens te zoeken en mogelijk te vinden, bracht Devlin naar archieven, met name die in Italië. Smakelijk en vermakelijk zijn zijn verhalen over de moeite die hij moest doen om toegang tot archieven te krijgen.

Natuurlijk op zoek naar de oudste nog bestaande kopieën van het hoofdwerk van Fibonacci, zijn *Liber Abbaci*, verschenen in 1202. Het origineel is, voor zover bekend, verloren gegaan, maar er zijn nog wel enkele kopieën die gemaakt zijn in diezelfde dertiende eeuw. Zelfs nog een kopie die geschreven is tijdens het leven van Fibonacci. Wellicht staat deze kopie het dichtst bij het origineel. 'Uit de werken herkent men de schrijver' en het is deze opvatting die Devlin dreef de oudst bekende kopieën te kunnen raadplegen om zodoende zicht te krijgen op de persoon Fibonacci. Het was de tijd waarin het positiestelsel van ons getallensysteem in Europa geïntroduceerd werd. Maar dat was al gebeurd vóór de tijd van Fibonacci. Dit positiestelsel was al geruime tijd bekend in kringen van geleerden, maar deze beschouwden dat stelsel puur als een studieobject voor geleerden. Praktische consequenties werden er niet aan verbonden. Op de markt werd van oudsher gerekend door middel van een systeem met de vingers, of met behulp van een telraam. Vliegenvlug, ja vingervlug, werden de uitkomsten op die manier berekend. We zien dat nog steeds gebeuren op die grote telramen op markten in het verre oosten. De uitkomsten werden/worden verkregen met een ongelooflijke snelheid die de snelheid van het tegenwoordige schermpje benadert. Maar één groot nadeel: als je wilde controleren of de handelaar je niet oplichtte, moest jijzelf die berekening volledig opnieuw maken. We kunnen ons het gekrakeel voorstellen dat ontstaat als er een fout(je) gemaakt was. Dan nog maar een keer vingervlug gerekend. Hét grote praktische voordeel van het rekenen met het positiestelsel was dat de rekenarij op schrift gesteld kon worden en zo objectief door ieder beoordeeld kon worden.

De middeleeuwse geleerden waren niet op het idee gekomen om 'het volk' deze mogelijkheid te bieden. Niet uit onwil, maar het paste gewoon niet in hun denkraam. Zo'n democratische gedachte was ver van hun bed. We zien dat op allerlei terreinen, in de wetenschap, in de kerk en eigenlijk in alle menselijke verbanden. Zo was de Bijbel er alleen voor de priesters, het gewone volk hoefde het niet te lezen, het hoorde wel van de priesters wat er in stond. En zo was het ook met de nieuwe rekenmethode: alleen voor geleerden, niets voor gewone mensen. Het is de onschatbare verdienste van Fibonacci dat hij de nieuwe rekenmethode toegankelijk gemaakt heeft voor alle mensen op alle terreinen van het leven. Met dit in het achterhoofd schreef hij zijn *Liber Abbaci*: boek van berekeningen. En letterlijk op alle terreinen van het leven geeft Fibonacci voorbeelden van hoe je een bepaald probleem zou kunnen/moeten aanpakken. Zijn boek is daarmee een lijvig werk geworden waar je maar moeilijk helemaal doorheen kunt komen. Juist omdat wij 21ste-eeuwers met andere ogen naar die problemen kijken. Problemen die in de dertiende eeuw wezenlijk van elkaar verschilden, vatten wij nu samen onder hetzelfde kopje. Daarmee wordt Fibonacci's boek voor ons een beetje saai, maar dat zegt niets over de enorme betekenis die de ideeën uit dit boek in de cultuurgeschiedenis hebben gespeeld. Zes eeuwen lang is het boek niet aan de oppervlakte gekomen, tot het in de negentiende eeuw weer werd ontdekt. Maar het heeft tot de zestiger jaren van de vorige eeuw geduurd voor er meer bekendheid aan gegeven werd.

Intmiddels is er een volledige Engelse vertaling van dit boek verschenen, verzorgd door Laurence E. Sigler. Devlin was een van de eersten die bij Amazon een exemplaar van dit werk wist te bemachtigen. Het verging hem zoals hiervoor beschreven: spannend

om te lezen wat er zeven eeuwen geleden geschreven is, maar een beetje saai in 21ste-eeuwse ogen. Maar daarmee niet minder belangrijk voor de ontwikkeling van de wereld. Prof. W.N. Goetzman (econoom van Yale) behoorde met Devlin ook tot de eersten die de vertaling van Sigler in handen kreeg. Hij noemt Leonardo van Pisa degene die aan de wieg heeft gestaan van de 'Birth of Modern Finance'. En dit alweer vanwege de vele praktische financiële problemen die Fibonacci in zijn boek rekenonderwijs tot een oplossing brengt. Inderdaad: *Liber Abbaci*, een boek dat de wereld veranderd heeft, niet alleen op het gebied van de wiskunde en de economie, maar op alle terreinen van het leven. Het was de zoektocht van Keith Devlin meer dan waard.

Een schitterend boek, mooi uitgevoerd: harde koft, leeslint met een uitvoerige bibliografie en index. Mooi Engels, erudiet opgezet: een genot om te lezen. *Wim Kleijne*



Ghislaine Gueudet, Marianna Bosch,
Andrea A. diSessa, Oh Nam Kwon,
Lieven Verschaffel

Transitions in Mathematics Education

Springer, 2016

x + 34 p., prijs €9,34

ISBN 9783319316215

Dit boekje bevat een aantal artikelen van diverse auteurs over een interessant en veelbelovend thema. Transitie in de wiskunde en in het wiskundeonderwijs is een breed onderwerp. Na een overzicht van de literatuur, wordt ingegaan op de pedagogische en didactische transitie: is er nu sprake van veel discontinue stappen en het daaraan verbonden begrip van misconceptie, of is er toch sprake van een continue leerlijn bij het leren van wiskunde? Een beknopte uiteenzetting van de diverse standpunten en empirische resultaten geeft helaas wel voorbeelden uit de natuurwetenschappen maar nauwelijks uit de wiskunde.

De zogenaamde dubbele discontinuïteit die Felix Klein in 1908 formuleerde is voor de lezers van dit blad wellicht het meest interessant: de soms forse discrepantie in het wiskundeonderwijs enerzijds als studenten van het voortgezet onderwijs naar de universiteit gaan en anderzijds als men als docent van de universiteit weer naar het voortgezet onderwijs gaat. Wellicht door het geringe aantal pagina's blijft de analyse van wat er de afgelopen honderd jaar hier verder over bekend is geworden, zeer oppervlakkig en beperkt zich inhoudelijk tot de rol van de theorie en formele bewijzen en contextueel tot aantallen uren en groepsomvang. Er wordt tamelijk negatief gerapporteerd over programma's die mogelijk een brug kunnen slaan tussen schoolwiskunde en universitaire wiskunde terwijl er toch wel degelijk positieve ervaringen zijn zoals de zomerkampen in Nederland of de bridging programs in Zuidelijk Afrika. Deze laatste functioneerden vele jaren uitstekend om meer studenten van betere kwaliteit naar exacte wetenschappen te loodsen. Hoewel het niet duidelijk is welke auteur verantwoordelijk is voor welk hoofdstuk, had de vermoedelijke auteur (Marianna Bosch, Barcelona) hier wel iets verder mogen uitweiden. Waarschijnlijk is het Lieven Verschaffel (Leuven) die in het laatste

artikel een goed gebalanceerd overzicht geeft van de discussie over schoolwiskunde en buitenschoolse wiskunde-ervaringen: kunnen die elkaar nu versterken of belemmeren ze elkaar wellicht op sommige momenten?

Al met al een aardig en snel te lezen boekje voor een plezierige prijs. Het is ook via open access toegankelijk. *Kees Smit*



Willem Bouman

De kunst van het hoofdrekenen

Epsilon Uitgaven, deel 89, 2017

xiv + 246 p., prijs €25,00

ISBN 9789050411646

Wat zou het heerlijk zijn als je zomaar de derdemacht van 2017 zou kunnen berekenen, of de vierkantswortel uit 4068289 zou kunnen trekken. Je zou als een zwaluw door de getallenwereld scheren, overal cijferpatronen herkennen, getallen in nieuwe subcategorien indelen, bewerken en steeds weer andere verbanden ontdekken. Je zou een elastische geest hebben met eindeloze geheugenruimte en acrobatische rekenlenigheid. Willem Bouman heeft zo'n geest. Wat zouden we graag zijn hoofd een dagje lenen. Dat kan nu in overdrachtelijke zin, want dit rekentalent dat drie wereldreCORDS hoofdrekenen op zak heeft en in 2014 de 'Prix d'Excellence' ontving heeft een boek geschreven over zijn kunst van het hoofdrekenen, waarin hij uitvoerig, helder en gedetailleerd beschrijft wat hij doet en hoe hij het doet, maar de lezer krijgt het niet cadeau.

Concentratie en doorzettingsvermogen zijn nodig om in detail te begrijpen hoe hij de dertiende wortel uit een getal van honderd cijfers trekt, of een getal van acht cijfers ontbindt in priemfactoren. Het boek wemelt van tabellen waarin bijvoorbeeld te vinden is op welke drie cijfers de 21ste macht van een gegeven grondgetal eindigt. Modulorekenen is de basis van veel van zijn rekenmethodes. Het is geen boek om dyscalculie te genezen, of alledaagse rekenproblemen op te lossen. Stuntelaars met wisselgeld en de prijs van treinkaartjes kunnen weinig compassie van Bouman verwachten. Het is wel een boek om je brein te scherpen en waardering te krijgen voor het prachtige getalsysteem met zijn vele regels, verbanden en patronen. Wat een rijkdom! En wat benutten wij — zelfs als we niet onmiddellijk voor elke berekening de rekenmachine grijpen — schrikbarend weinig van deze rijkdom!

Kunnen we nu allemaal rekenwonder worden? Met dit boek als bron in theorie wel. Bouman houdt geen enkele techniek achter, maar wie het bestudeert, ontdekt ook dat het bepaald niet om een kunstje gaat dat je blind kunt toepassen. Wie rekenwonder wil worden moet een Mount Everest aan getallenkennis in zijn hoofd pompen, en dan nog de scherpzinnige vaardigheid verwerven om op het juiste moment de juiste kennis vlot, vaardig en foutloos in te zetten. Men zegt wel dat voor het bereiken van de wereldtop 10.000 gerichte trainingsuren nodig zijn. Maak daar voor de hoofdrekenkunst maar 100.000 uren van. Wanhoop niet. Je bent nooit te laat om te beginnen. Bouman is zelf het bewijs, want hoewel hij zijn hele leven al een opvallend rekentalent was is hij pas na

zijn pensioen er doelgericht mee aan de slag gegaan. Op 67-jarige leeftijd nam hij onvoorbereid deel aan het wereldkampioenschap hoofdrekken en werd hij meteen zevende. Nu hij richting de 80 gaat, heeft hij een levenswerk geschreven dat respect afdwingt, door zijn gedrevenheid, denkvermogen, discipline en de bereidheid dit alles met ons te delen.

Marjolein Kool



Daniel Huybrechts

Lectures on K3 Surfaces

Cambridge University Press, 2016

xi + 485 p., prijs £49.99

ISBN 9781107153042

De mysterieuze benaming ‘K3-oppervlak’ is afkomstig van André Weil en slaat terug op de namen Kummer, Kähler en Kodaira enerzijds en anderzijds op de moeilijk te beklimmen berg K2 in Kasjmir — de K3-oppervlakken zouden een grotere moeilijkheidsgraad hebben. Tot mijn verbazing bestaat de K3 ook als berg; een oude foto ervan vindt men in het voorwoord van het boek.

De wiskundige definitie van een K3-oppervlak is even kort als weinig zeggend: het is een ‘oppervlak’ zonder reguliere 1-vormen en met een reguliere 2-vorm die nergens verdwijnt. Hierbij dient te worden aangetekend dat het woord ‘oppervlak’ gebruikt wordt in twee betekenissen. Men kan er een tweedimensionale complexe variëteit onder verstaan, maar ook een tweedimensionale complete en gladde *algebraïsche* variëteit over een willekeurig lichaam.

K3-oppervlakken vormen een belangrijke klasse van oppervlakken en duiken op in vele deelgebieden van de wiskunde, zoals differentiaalmeetkunde van complexe variëteiten, getaltheorie, aritmetische algebraïsche meetkunde, aftellende meetkunde en mathematische fysica. De meeste van die raakvlakken komen in dit boek aan de orde en maken het lezen ervan extra boeiend.

Maar eerst enige voorbeelden. Het eerste basisvoorbeeld is een glad oppervlak van graad 4 in de projectieve ruimte. Het tweede is het zogenaamde *Kummer-oppervlak*, een (gedesingulariseerd) quotiënt van een 2-dimensionale torus onder de standaardinvolutie. Dit voorbeeld toont al aan dat er onder de complexe K3’s niet-algebraïsche oppervlakken voorkomen. In zekere zin zijn de meeste complexe K3’s inderdaad niet-algebraïsch.

Dit boek is de weerslag van reeksen voordrachten voor wisselend publiek en is daarom heel divers van inhoud. Desondanks is de presentatie verbazingwekkend homogeen en is het boek zorgvuldig geschreven met vele verwijzingen binnen het boek zodat de lezer op zoek kan gaan naar verwante onderwerpen en snel iets kan vinden wat hij niet (meer) weet. Een ander fenomeen dat ik nog niet eerder zag, ziet men in de literatuurlijst. Van elk item wordt aangegeven waar het in de tekst geciteerd wordt. Heel handig als je bijvoorbeeld weet dat persoon X iets over onderwerp Y gedaan heeft. Al met al is dit behalve een originele en uitvoerige tekst over K3’s ook een handig naslagwerk, en niet alleen voor onderwerpen betreffende K3’s want er wordt nogal wat behandeld!

Hier volgt een overzicht van de inhoud. De basiskennis over K3’s vindt men in de eerste twee hoofdstukken. In de volgende hoofdstukken wordt enige basale Hodge-theorie ‘herhaald’. Hieronder valt volgens de schrijver ook de theorie van de Mumford–Tate-groepen en de Kuga–Satake-constructie! Die laatste constructie is van belang: zij maakt het mogelijk aritmetische eigenschappen van Hodge-structuren van gewicht 2 en van ‘K3-type’ terug te voeren tot bekende eigenschappen van (gepolariseerde) abelse variëteiten. Dit alles wordt kort aangestipt en helder uiteengezet met als toepassing het bewijs van Deligne van de Weil vermoedens voor K3’s. Dit bewijs ging vooraf aan het — veel ingewikkeldere — bewijs dat voor een willekeurige gladde projectieve variëteit geldt.

Het volgende onderwerp is de modulistheorie van K3’s. In het complexe geval hangen K3’s van 20 parameters af, maar algebraïsche hangen van maar 19 parameters af, conform de bovenstaande bewering dat de meeste K3’s niet algebraïsch zijn. De algebraïsche aanpak via Hilbert-schema’s wordt in hoofdstuk 5 uiteengezet; dit vooronderstelt een inbedding in een vaste projectieve ruimte, of, zo men wil, de keuze van een polarisatie. De modulistheorie in het complexe geval is veel lastiger en gebruikt de periodenafbeelding. Dit wordt in de hoofdstukken 6 en 7 uit de doeken gedaan. De aanpak hier wijkt af van wat men in de literatuur vindt en neemt niet de Kummer-oppervlakken als startpunt, maar juist de meest niet-algebraïsche K3’s (dat wil zeggen die met triviale Picard-groep). Hier speelt de differentiaalmeetkunde van twistorruimtes een doorslaggevende rol. Deze aanpak is succesvol gebleken voor de aanpak van het moduliprobleem (door D. Huybrechts, M. Verbitsky en E. Markman) van hyperkähler-variëteiten, hoger-dimensionale equivalenten van K3-oppervlakken. Wat men onder andere voor het bewijs van de Torelli-stelling nodig heeft over de Kähler-kegels staat in een handzaam apart hoofdstuk 8.

Meetkundige objecten op K3’s zoals vectorbundels en schoven vormen het volgende thema, dat in de hoofdstukken 9 en 10 behandeld wordt. Hierbij staat centraal de door Mukai gestarte studie van moduliruimten van stabiele bundels en schoven. Sommige van deze ruimtes blijken voorbeelden te geven van complex symplectische variëteiten, een betrekkelijk zeldzame en moeilijk te vinden klasse van variëteiten. Andere objecten, zoals rationale krommen, de Picard-groep en de Brauer-groep komen in latere hoofdstukken (13, 17, respectievelijk 18) aan de orde. Deze hoofdstukken bevatten een schat aan informatie. Om maar wat te noemen, in hoofdstuk 13 wordt de Yau–Zaslow-formule besproken, een expliciete door de mathematische fysica gemotiveerde formule voor het aantal nodale rationale krommen van verschillende graad op een vaste ‘algemene’ K3, voorts de Bogomolov–Tschinkel-resultaten over dichtheid van de vereniging van rationale krommen, het Tate-vermoeden voor K3-oppervlakken, en de formele groepstructuur op de Brauer-groep van K3’s in positieve karakteristiek.

Alsof dit niet genoeg is, is er nog een hoofdstuk dat de rijke wereld van elliptische K3-oppervlakken betreedt, en hoofdstukken over Chow-groepen van nulcykels op K3-oppervlakken, de mogelijke automorfismen groepen van K3’s, en een uitgebreid hoofdstuk over afgeleide categorieën waarin de zogenaamde getwiste K3’s een belangrijke rol vervullen. Samen met de eigenschappen van de Brauer-groep vormt dit laatste onderwerp een belangrijk ingrediënt in het bewijs van het Hodge-vermoeden voor het zelfproduct van een grote klasse van K3’s in het recente proefschrift van Buskin (zie arXiv:1510.02852 [math.AG]). Dit is het onderwerp dat uitge-

breid aan bod is gekomen in het Nederlandse 'Intercity seminarium meetkunde en algebra' van 2016, met Huybrechts als vaste deelnemer (en spreker).

Het zal duidelijk zijn dat ik dit boek van harte aanbeveel aan mijn collega's algebraïsch meetkundigen. Natuurlijk behoort ook elke wiskundebibliotheek een exemplaar in zijn bezit te hebben: dit boek is een uniek naslagwerk over K_3 's en zal vast een klassieker worden.

Chris Peters



Jean-Pierre Serre

Finite Groups: An Introduction

International Press, 2016

190 p., prijs \$79.00

ISBN 9781571463272

In 1978–1979 gaf Serre een college aan de École Normale Supérieure de Jeune Filles, Paris. “Its aim is to give an introduction to the main theorems of finite group theory.” Dit boek bevat de tekst van dat college, met twee nieuwe hoofdstukken, en 150 vraagstukken. Vergelijking met de originele Franse tekst (2008, zie <https://arxiv.org/pdf/math/0503154.pdf>) laat zien dat deze versie meer dan twee keer zo lang is, met prachtig nieuw materiaal.

We kennen boeken over eindige groepen (bijvoorbeeld Huppert, Suzuki, Gorenstein, Kurzweil–Stellmacher, Aschbacher–Smith), omvangrijke artikelen en uitvoerige tabellen. Compleet zijn is haast niet mogelijk in dit onderwerp. Ook is het in sommige gevallen nog niet duidelijk wat er bewezen is (bijvoorbeeld in de classificatie van simpele groepen). We zien hoe Serre tussen al deze moeilijkheden door laveert met precisie en inzicht.

In 1832 gaf Galois de definitie van een normale ondergroep, en hij liet zien dat A_n (een groep van orde $(n!)/2$) voor $n \geq 5$ en $\text{PSL}_2(\mathbb{F}_p)$ voor een priemgetal $p \geq 5$ (een groep van orde $p(p^2 - 1)/2$) simpele groepen zijn. Dat was het begin van een lange zoektocht naar alle simpele groepen. In een college algebra geven we graag de opgave, in navolging van Galois, dat een sim-

pele groep van orde kleiner dan 60 een cyclische groep van priem orde is.

We zien hoe deze problemen benaderd worden in dit boek: in hoofdstuk 7 worden simpele groepen van oneven orde hoogstens 2000 en simpele groepen van orde hoogstens 200 geïdentificeerd. (Kent de lezer van deze bespreking het resultaat?)

Het is mooi om te zien hoe Serre kort, en ook heel precies zijn tekst redigeert. Bijvoorbeeld (Definition 1.7 op pagina 4): “A group G is *simple* if the number of its normal subgroups is 2.” (En dan daarna een uitleg.)

Is de classificatie van eindige simpele groepen (CFSG) nu compleet? Keer op keer krijgen we te horen dat die classificatie compleet is, tot er dan in die aaneenschakeling van 500 artikelen weer een lacune gevonden wordt. Van Serre is bekend dat hij sceptisch is over deze claims (althans dat was hij in 2003), zie het interview in de *European Mathematical Society Newsletter*, September 2003, pp. 18–20, zie ook *Notices AMS* 51, (2004) pp. 210–214. Voor niet-specialisten (ondanks het werk van Michael Aschbacher en Stephen D. Smith, 2004) is het moeilijk uit te maken of we dit nu als stelling mogen aanmerken. In dit boek (uit 2016) is Serre voorzichtig en zorgvuldig: of CFSG waar is in de huidige vorm wordt in het midden gelaten.

Behalve de grote nauwkeurigheid van Serre kennen we ook zijn inzicht in voorbeelden naast het begrijpen van diepe stellingen als kenmerk van zijn werk. Dit zien we ook in dit boek.

De flaptekst vermeldt: “Finite group theory is a topic remarkable for the simplicity of its statements and the difficulty of their proofs. It is used in an essential way in several branches of mathematics — for instance, in number theory. This book is a short introduction to the subject, written both for beginners and for mathematicians at large. There are ten chapters: Preliminaries, Sylow theory, Solvable groups and nilpotent groups, Group extensions, Hall subgroups, Frobenius groups, Transfer, Characters, Finite subgroups of GL_n , and Small Groups.”

Elke bladzijde van dit boek laat parels van wiskundig denken zien. Zie hoe moeilijke stellingen schijnbaar eenvoudig bewezen worden, zie hoe motiverende vraagstukken veel informatie geven. Iedereen met nauwelijks kennis van algebra kan aan de slag (zoals de ‘Jeune Filles’). Zeer aanbevolen, voor ons allemaal, bij een college algebra en ook om te zien hoe je compact, precies en met een grote reikwijdte over wiskunde schrijft.

Frans Oort

Recent verschenen publicaties. Als u een van deze boeken wilt bespreken of als u suggesties heeft voor andere boeken voor deze rubriek, laat dit dan per e-mail weten aan reviews@nieuwarchief.nl.



Lesley B. Cormack, Steven A. Walton,
John A. Schuster (Eds.)

Mathematical Practitioners and the Transformation of Natural Knowledge in Early Modern Europe

Springer, 2017

ISBN 9783319494296

springer.com/9783319494296



Peter Cameron

Notes on Counting: An introduction to enumerative combinatorics

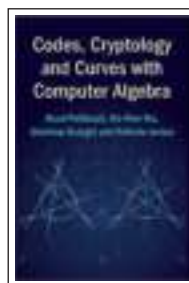
Cambridge University Press, 2017

ISBN 9781108404952

cambridge.org/9781108404952



Lizhen Ji, Eduard J.N. Looijenga
Introduction to Moduli Spaces of Riemann Surfaces and Tropical Curves
International Press, 2017
 ISBN 9781571463531
intlpress.com/site/pub/pages/books/items/00000498



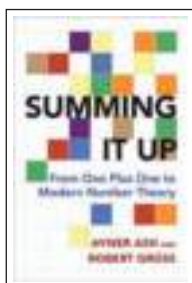
Ruud Pellikaan, Xin-Wen Wu, Stanislav Bulygin, Relinde Jurrius
Codes, Cryptology and Curves with Computer Algebra
Cambridge University Press, 2017
 ISBN 9780521520362
cambridge.org/9780521520362



Mircea Pitici
The Best Writing on Mathematics 2017
Princeton University Press, 2017
 ISBN 9781400888559
press.princeton.edu/titles/11102.html



Øystein Linnebo
Philosophy of Mathematics
Princeton University Press, 2017
 ISBN 9781400885244
press.princeton.edu/titles/11024.html



Avner Ash, Robert Gross
Summing It Up
Princeton University Press, 2016
 ISBN 9781400880539
press.princeton.edu/titles/10692.html



Persi Diaconis, Brian Skyrms
Ten Great ideas about Chance
Princeton University Press, 2017
 ISBN 9781400888283
press.princeton.edu/titles/11082.html



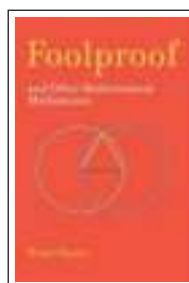
J.S. Milne
Algebraic Groups: The theory of group schemes of finite type over field
Cambridge University Press, 2017
 ISBN 9781107167483
cambridge.org/9781107167483



Russell Lyons, Yuval Peres
Probability on Trees and Networks
Cambridge University Press, 2017
 ISBN 9781107160156
cambridge.org/9781107160156



Reviel Netz (Ed. and translator)
The works of Archimedes Translation and Commentary, Vol 2: On spirals
Cambridge University Press, 2017
 ISBN 9780521661454
cambridge.org/9780521661454



Brian Hayes
Foolproof, and Other Mathematical Meditations
MIT Press, 2017
 ISBN 9780262036863
mitpress.mit.edu/9780262036863

Problemen

| Problem Section

This Problem Section is open to everyone; everybody is encouraged to send in solutions and propose problems. Group contributions are welcome. For each problem, the most elegant correct solution will be rewarded with a book token worth €20. (To compete for the book token you should have a postal address in The Netherlands.)

Please send your submission by e-mail (LaTeX is preferred), including your name and address to problems@nieuwarchief.nl.

The deadline for solutions to the problems in this edition is 1 March 2018.

Problem A

Consider two identical bags of stones, all having integral weights. No two weights in a bag are the same. Suppose that the stones from these two bags are divided into proper subsets of equal cardinality and equal total weight. No two weights in a subset are the same. Is it then possible to readjust this division such that each subset contains stones from the same bag?

Problem B

Let $H_1, \dots, H_k$ be k planes in $\mathbb{R}^3$. Prove that the unit ball contains an open ball of radius $\frac{1}{k+1}$ which does not intersect any of the planes.

Problem C (proposed by Hendrik Lenstra)

Suppose that a, b, n, m are integers, and that m, n are positive, such that $a^n \equiv b^n \equiv -1 \pmod{m}$. Prove there exists an integer c such that $ab \equiv c^2 \pmod{m}$. Also provide a fast method to compute such a c which works even if the prime factors of m are unknown.

Edition 2017-2 We received solutions from Rik Bos (Bunschoten), Souvik Dey (Kolkata, India), Math Dicker (Hoensbroek), Alex Heinis (Amsterdam), Thijmen Krebs (Nootdorp), Hans van Luipen (Zaltbommel), José Polo Gómez (Sevilla, Spain), Ludo Pulles (Zeist), Hendrik Reuvers (Maastricht), Hans Samuels Brusse (Den Haag), Toshihiro Shimizu (Kawasaki, Japan), Djurre Tijsma (Zeist) and Robert van der Waall (Huizen). The book tokens for problems A, B and C go to Ludo Pulles, Thijmen Krebs, respectively Rick Bos.

Problem 2017-2/A

Show that there are no infinite antichains for the partial order $\leq$ on $\mathbb{N}^k$ defined by $(x_1, x_2, \dots, x_k) \leq (y_1, y_2, \dots, y_k)$ iff $x_i \leq y_i$ for all $1 \leq i \leq k$.

Solution This problem was solved by Alex Heinis, Thijmen Krebs, Hans van Luipen, Ludo Pulles, Hendrik Reuvers, Hans Samuels Brusse, Toshihiro Shimizu and Djurre Tijsma. Several solvers noticed that this is Dickson's Lemma. All solutions are similar and Thijmen Krebs is very concise. Below is his solution.

By induction. The statement is obviously true if $k = 0$. Suppose on the contrary that such an antichain C exists, and let $x \in C$. For every other $y \in C$ there is some $1 \leq i \leq k$ such that $0 \leq y_i \leq x_i$. By the pigeonhole principle we find some infinite antichain $A \subseteq C$ and some $1 \leq i \leq k$ such that $y_i = z_i$ for all $y, z \in A$. Deleting the i -th entry from each $y \in A$ then gives an infinite antichain in $\mathbb{N}^{k-1}$, which is impossible by induction.

Problem 2017-2/B

A sector is a portion of a disk enclosed by two radii and an arc. For each pair of radii, there are two complementary sectors. If the two complementary sectors have unequal area, then

we say that the larger sector is the major sector. Let S be a subset of the plane. We say that $x \in S$ is virtually isolated if x is the only element of S in a major sector of a disk centered on x . Suppose that all elements of S are virtually isolated. Prove that S is countable.

Solution The problem was suggested by Jan Aarts, with a solution. The problem was solved by Thijmen Krebs, José Polo Gómez and Toshihiro Shimizu. Below is the solution of Thijmen Krebs. Again very concise. Each $x \in S$ is the center of some closed semi-disk with rational radius and rational angle that contains no point from S other than x . Distinct $x, y \in S$ with up to translation the same semi-disk must be more than its radius apart, so sending $x \in S$ to a translation of its semi-disk to a rational center within half the radius from x is an injection to a countable set.



Problem 2017-2/C (proposed by Hendrik Lenstra)

Let n be a natural number > 1 . Suppose that for every prime $p < n$ we have that $p^n = (p-1)^n + 1 \pmod{n^2}$. Prove that $n = 2$.

Solution Solved by Rik Bos, Souvik Dey, Math Dicker, José Polo Gómez, Thijmen Krebs, Hans van Luipen, Toshihiro Shimizu, Djurre Tijsma and Robert van der Waall. All solutions are similar. The solution by Rik Bos is below.

We first argue by induction that $a^n = a \pmod{n^2}$ for all $a < n$. If a is prime then $a^n = (a-1)^n + 1 \pmod{n^2}$, which by induction is equal to $(a-1) + 1 \pmod{n^2}$. If a is composite, then $a = bc$ and by induction both $b^n = b \pmod{n^2}$ and $c^n = c \pmod{n^2}$. Therefore, $a^n = a \pmod{n^2}$. By the binomial theorem $(n-1)^n = (-1)^n \pmod{n^2}$. Taking $a = n-1$ produces $(-1)^n = n-1 \pmod{n^2}$. It follows that $n = 2$.

