Inhoud | Contents

Artikelen | Features

Rubrieken | Departments

Uitgelicht

Ionica Smeets neemt na vijf jaar afscheid van 'Het keerpunt' en wordt nu zelf geïnterviewd door haar opvolger Jan Beuving.

De Commissie Fokkema geeft een samenvatting van het Deltaplan voor de Nederlandse wiskunde.

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 een ieder 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

Hoofdredactie
Richard Boucherie (r.j.boucherie@utwente.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 (TU/e), Geertje Hek (UvA), Jan Hogendijk (UU), Teun Koetsier (VU), Sjoerd Rienstra (TU/e), Hans Sterk (TU/e), Mark Timmer (UT)

Problemenrubriek
Gabriele Dalla Torre (UL), Christophe Debry (KU Leuven), Jinbi Jin (UL), Marco Streng (UL), Wouter Zomervrucht (UL)

Bureauredactie
Gianni van den Braak
Lætitia Diemer

Illustraties
Ryu Tajiri, Amsterdam

Advertenties
advertenties@nieuwarchief.nl

Abonnementen/Subscriptions
Ledenadministratie KWG
Postbus 1064, 7940 KB Meppel
admin@wiskgenoot.nl, tel. 085-0160252
Abonnementsperiode/Subscription period
Een abonnement gaat in op de dag dat uw aanmelding ontvangen wordt. De opzegtermijn is twee maanden vóór het verstrijken van de abonnementsperiode.
The subscription starts on the day that your application is received. The term of notice is two months before the end of the current subscription period.
Exchange subscriptions
Koninklijk Wiskundig Genootschap Library
Centrum voor Wiskunde en Informatica
Postbus 94079, NL-1090 GB Amsterdam
KWG-exchange@cwi.nl, fax +31 20 5924026

Vormgeving
Kitty Molenaar (ontwerp)
Susanne Laws (omslag, begeleiding binnenwerk)
Programmatuur (CONTEXT)
Hans Hagen, PRAGMA-ADE, Hasselt (Overijssel)
Druk
Drukkerij Ten Brink, Meppel

Koninklijk Wiskundig Genootschap

Het Koninklijk Wiskundig Genootschap (KWG) is een landelijke vereniging van beoefenaren 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.
De contributie voor gewone leden bedraagt in 2016 € 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.
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.
Instituten in Nederland en buitenland kunnen zich abonneren op het Nieuw Archief voor Wiskunde voor € 100 (Nederland), € 110 (Europa) of € 112,50 (buiten Europa).
Website: www.wiskgenoot.nl
Ledenadministratie
Ledenadministratie KWG
Postbus 1064, 7940 KB Meppel
admin@wiskgenoot.nl, tel. 085-0160252
Bestuur (wiskgenoot@wiskgenoot.nl)
Geurt Jongbloed (TUD, voorzitter), Fetsje Bijma (VU, penningmeester), Joke Blom (CWI, secretaris), Erik van den Ban (UU), Theo van den Bogaart (HU), Sonja Cox (UvA), André Ran (VU), Herman te Riele (CWI), Mark Veraar (TUD)

Informatie voor auteurs

Kopij
Het Nieuw Archief voor Wiskunde is een gerediger tijdschrift. Kopij dient elektronisch te worden aangeboden aan de hoofdredactie. Uitgebreide informatie en auteursinstructies kunt u vinden op www.nieuwarchief.nl .
Reprints
Auteurs ontvangen een elektronische reprint in pdf-formaat.
Volgende nummers

nummer	*verschijningsdatum*	*deadline kopij*
1	07–03–2016	verstreken
2	06–06–2016	01–02–2016
3	05–09–2016	01–05–2016
4	05–12–2016	01–08–2016

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

Met een MRI-scan kunnen hersenen in beeld gebracht worden. Met 'Diffusion tensor imaging' (DTI) kunnen reconstructies gemaakt worden van verbindingen tussen verschillende delen van de hersenen. Lees het artikel van Neda Sepasian, Jan ten Thije Boonkkamp en Anna Vilanova op p. 259.

Bouwstenen

Onderzoek, onderwijs en maatschappij zijn drie hoekstenen van ons wiskundehuis dat we in het *Nieuw Archief voor Wiskunde* belichten. In de discussie over de nationale wetenschapsagenda stelt de maatschappij vragen aan onderzoekers. De vorm van deze discussie suggereert een afstand tussen maatschappij en onderzoek, zo ook tussen maatschappij en wiskunde.

Onze maatschappij leunt voor een aanzienlijk deel op de inzet van wiskundige methoden. Waar vinden we wiskundigen in Nederland? In onderzoek is dat duidelijk: overwegend aan universiteiten en hogescholen. Ook in onderwijs is het duidelijk: voor de klas op scholen van alle niveaus. In onze maatschappij lijken wiskundigen zich echter heel goed te verstoppen. Mijn definitie van een wiskundige in onze maatschappij is: "Iedereen met een diploma van een wiskundeopleiding, alsmede iedereen die zich wiskundige voelt." Wanneer ik u vraag drie bekende Nederlanders te noemen die wiskundige zijn, dan moet u mij zeer waarschijnlijk het antwoord schuldig blijven. Als lezer van dit blad heeft u bij de beantwoording van deze vraag wel een voorsprong. In 'Het keerpunt' heeft Ionica Smeets vijf jaar lang wiskundigen die midden in de maatschappij staan geïnterviewd, waarvoor het NAW haar hartelijk dankt. Bij haar afscheid wordt ze geïnterviewd door Jan Beuving die het stokje overneemt. Ze roept om een beroepsvereniging waar alle wiskundigen zich thuisvoelen. Hier ligt een mooie taak voor het NAW: alle wiskundigen samenbrengen in ons wiskundehuis.

Het nummer dat voor u ligt slaat een brug tussen wiskundig onderzoek en onze maatschappij. We kunnen met behulp van wiskundige methoden efficiënt zoeken in digitale databases. De stabiliteit van onze energievoorziening wordt mede gegarandeerd door inzet van wiskundige methoden. Voor de beschrijving van onze hersenen en ons DNA doen we een beroep op wiskunde. Om het geheel te completeren laat Christine van Vredendaal in een bijdrage naar aanleiding van de door haar gewonnen prijs voor beste masterscriptie van 2014 aan de Technische Universiteit Eindhoven zien dat wiskunde ook heel aantrekkelijke experimenten mogelijk maakt. Geniet u meer van de uitleg van een mooie theorie in historisch perspectief dan kunt u putten uit een bijdrage over het *abc*-vermoeden. Wanneer u wilt filosoferen over de hoekstenen van een wiskundige theorie biedt Arjeh Cohen uitkomst met zijn introductie van VROUW. Kortom, een mooie collage om in reactie op die lastige vraag op decemberborrels, "Wiskunde, wat kan je daar nu mee?", te laten zien dat wiskunde midden in de maatschappij staat.

Op verzoek van het Ministerie van Onderwijs, Cultuur en Wetenschap heeft een commissie onder leiding van Jacob Fokkema een Deltaplan opgesteld voor de Nederlandse wiskunde. De commissie beschrijft een wiskundehuis dat steunt op stevige wetenschappelijke fundamenten. In dit nummer geeft de commissie een samenvatting van het Deltaplan en stelt onder meer: "De Nederlandse wiskunde is bij uitstek in staat prominente vragen uit de Nationale Wetenschapsagenda te adopteren en een stevige rol te spelen in de publiek-private samenwerking."

Laten we samen wiskunde haar rol geven midden in onze maatschappij en bouwen aan een wiskundehuis waarin alle wiskundigen zich thuisvoelen! Wij wiskundigen vormen de bouwstenen van ons wiskundehuis.

Richard Boucherie, hoofdredacteur
Afdeling Toegepaste Wiskunde, Universiteit Twente

Servicepagina

Service page

Koninklijk Wiskundig Genootschap

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

❑ **Ledenadministratie KWG**
Met ingang van 1 december 2016 gaat het KWG over op lidmaatschapsgelden innen via automatische incasso. Mocht u een factuur willen blijven ontvangen dan zijn de kosten daarvoor 2,50 euro. De keuze voor automatische incasso kunt u aangeven via het webformulier op onze nieuwe website of op het formulier dat bijgesloten is bij dit nummer van het NAW. Als u uw gegevens opnieuw invoert via het webformulier dan weet u bovendien zeker dat alles weer correct in de ledenadministratie staat en dat u geen Mededelingen of NAW meer mist door adresseringsproblemen.

❑ **Wintersymposium 2016**
Op zaterdag 9 januari 2016 van 11.00–16.00 uur vindt in het Academiegebouw te Utrecht het Wintersymposium van het KWG plaats. Het symposium richt zich op wiskundedocenten in het voortgezet onderwijs en andere belangstellenden. Dit jaar is het thema 'Analytische meetkunde'. De sprekers zijn: Jeroen Spandaw (TU Delft), Jaap Top (RUG), Viktor Blåsjö (UU) en Remco Duits (TU/e). www.wiskgenoot.nl/wintersymposium

❑ **NMC 2016**
Het 52ste Nederlands Mathematisch Congres wordt onder auspiciën van het KWG en de Belgische en Luxemburgse zusterverenigingen op 22 en 23 maart 2016 op het Science Park in Amsterdam gehouden. Het programma omvat plenaire lezingen van Ellen Baake, James Maynard (Beegerlezing) en Stefaan Vaes; de N.G. de Bruijnprijs zal voor de eerste maal worden uitgereikt; en er is een avondprogramma in het Koninklijk Instituut voor de Tropen met lezingen van Marcus du Sautoy en Ionica Smeets. www.wiskgenoot.nl/bnlmc

❑ **L.E.J. Brouwer, vijftig jaar later**
Volgend jaar, in 2016, is het vijftig jaar geleden dat L.E.J. Brouwer overleden is. Er zullen verschillende activiteiten georganiseerd worden, waaronder op 9 december 2016 een symposium op Science Park in Amsterdam.

Platform Wiskunde Nederland

❑ **Bureau PWN**
Science Park 123, L013, 1098 XG Amsterdam
bureau@platformwiskunde.nl
pr@platformwiskunde.nl (publiciteit)
Tel: 020-5924006/06-51892525
www.platformwiskunde.nl

❑ **PWN Nieuwsbrief**
PWN publiceert eens in de drie maanden een PWN Nieuwsbrief, die op de website te vinden is. Actuele PWN-informatie wordt verder zowel via de website als via Twitter gegeven.

European Mathematical Society

❑ **EMS website**
www.euro-math-soc.eu

❑ **EMS Newsletter**
De elektronische versie van de meest recente uitgave van de EMS Newsletter is beschikbaar via www.ems-ph.org/publications.php.

IMU

❑ **IMU website**
www.mathunion.org

❑ **ICM 2018**
Het ICM 2018 vindt plaats van 1–9 augustus 2018 in het Rio Centro Convention Center, Rio de Janeiro, Brazilië. De *General Assembly* zal worden gehouden in São Paulo van 29–30 juli 2018. www.icm2018.org

❑ **Heidelberg Laureate Forum**
Het 4de Heidelberg Laureate Forum vindt plaats van 18–23 september 2016 in Heidelberg. www.heidelberg-laureate-forum.org

❑ **IMU Newsletter**
De twee-maandelijkse IMU Newsletter is beschikbaar via www.mathunion.org/imu-net.

Recente publicaties van het KWG

❑ **Epsilon Uitgaven** (www.epsilon-uitgaven.nl)
82. Een Variabele Constante, Martin Kindt, € 36, 2015.
81. Functionaalanalyse, Heinz Hanßmann, € 28, 2015.
80. Christiaan Huygens: Het Slingeruurwerk, Jan Aarts, € 32, 2015.
79. Galoistheorie, Frans Keune, € 24, 2015.

❑ **Zebra-reeks** (Epsilon Uitgaven)
46. Knopen in de wiskunde, Meike Akveld en Ab van der Roest, € 10, 2015.

Manuscripten voor Epsilon Uitgaven kunt u mailen naar redactie@epsilon-uitgaven.nl.

Agenda

Upcoming Events

december 2015

7–11 december 2015

❑ **Quantum Random Walks and Quantum Algorithms**

Workshop georganiseerd door Harry Buhrman (CWI) en Frank den Hollander (LU).

plaats *Lorentz Center@Snellius, Leiden*

info *lorentzcenter.nl*

15 december 2015

❑ **Studiedag voor leraren wiskunde**

Lezingen en workshops over het vak wiskunde voor leraren in het voortgezet onderwijs, met als thema 'Nieuwe examenprogramma's in bedrijf'.

plaats *Rijksuniversiteit Groningen*

info *rug.nl/education/lerarenopleiding*

januari 2016

9 januari 2016

❑ **Wintersymposium KWG 2016**

Jaarlijks symposium van het KWG met dit jaar als thema 'Analytische meetkunde', hetgeen onderdeel is van de curriculumvernieuwing wiskunde op havo en vwo.

plaats *Academiegebouw, Utrecht*

info *wiskgenoot.nl*

12–14 januari 2016

❑ **41ste Lunteren Conferentie Mathematische Besliskunde**

Jaarlijkse conferentie van het Landelijk Netwerk Mathematische Besliskunde.

plaats *Congrescentrum De Werelt, Lunteren*

info *www.lnmb.nl/conferences/2016*

18–28 januari 2016

❑ **1ste ronde Nederlandse Wiskunde Olympiade**

Het begin van de jaarlijkse wiskundewedstrijd voor leerlingen van havo en vwo.

plaats *eigen school*

info *wiskundeolympiade.nl*

21–22 januari 2016

❑ **34ste Panama-conferentie**

Jaarlijkse conferentie die zich richt op alle aspecten van leren en onderwijzen van rekenen-wiskunde. Dit jaar is het thema 'Rekenen-wiskunde over( )denken'.

plaats *Congrescentrum Koningshof, Veldhoven*

info *www.fi.uu.nl/panama*

25–27 januari 2016

❑ **15th Winter School Mathematical Finance**

Winterschool over financiële wiskunde met onder andere minicursussen van Dirk Becherer (Berlijn) en Fred Espen Benth (Oslo).

plaats *Congrescentrum De Werelt, Lunteren*

info *www.mathfin.nl*

25–29 januari 2016

❑ **Studiegroep Wiskunde met de Industrie**

Jaarlijkse studiegroep waarin wiskunde en industrie samen problemen oplossen.

plaats *IMAPP, Nijmegen*

info *www.ru.nl/imapp*

28 januari 2016

❑ **Conferentie vmbo en onderbouw havo/vwo**

Conferentie voor docenten vmbo en onderbouw havo/vwo, georganiseerd door de Nederlandse Vereniging van Wiskundeleraren.

plaats *Aristo, Utrecht*

info *nvvw.nl*

29 januari 2016

❑ **Bijles geven in de bètavakken**

Masterclass voor leerlingen uit de bovenbouw van havo en vwo met interesse in onderwijs.

plaats *De Uithof, Utrecht*

info *uu.nl/u-talent*

februari 2016

3 februari 2016

❑ **OnderbouwWiskundeDag**

Leerlingen werken in teams van drie of vier leerlingen gedurende een dag aan een 'grote' wiskundige (denk)opdracht waarin probleemoplossen centraal staat.

plaats *eigen school*

info *www.fi.uu.nl/nl/onderbouwwiskundedag*

Suggesties voor deze agenda zijn welkom.
agenda@nieuwarchief.nl

5–6 februari 2016
❑ **Nationale Wiskunde Dagen**
22ste editie van de tweedaagse conferentie voor wiskundeleraren.
plaats *Noordwijkerhout*
info *uu.nl/nationale-wiskunde-dagen*

11 februari 2016 / 18 februari 2016
❑ **Bètadag**
Studenten van bètafaculteiten laten aan leerlingen van 3 havo/vwo zien waarom zij hun studie zo interessant, fascinerend en onmisbaar vinden.
plaats *bètafaculteiten in het hele land / TU/e*
info *betadagen.nl*

15 februari 2016
❑ **Logisch redeneren**
Workshop voor docenten wiskunde C (vwo) en docenten wiskunde D (havo), gegeven door Quintijn Puite en Theo van den Bogaart (Instituut Archimedes, HU).
plaats *Hogeschool Utrecht*
info *uu.nl/u-talent*

maart 2016

7–11 maart 2016
❑ **YEP XIII: Large Deviations for Interacting Particle Systems and Partial Differential Equations**
13de workshop voor Young European Probabilists (YEP). Workshop in het kader van de Stochastic Activity Month (SAM) 'Probability and Analysis'.
plaats *TU Eindhoven*
info *www.eurandom.nl*

11 maart 2016
❑ **2de ronde Nederlandse Wiskunde Olympiade**
De tweede ronde van de Nederlandse Wiskunde Olympiade. De beste leerlingen gaan door naar de finale in september.
plaats *Nederlandse universiteiten*
info *wiskundeolympiade.nl*

11–12 maart 2016
❑ **Finale Wiskunde A-lympiade**
Finale van competitie voor leerlingen uit de bovenbouw havo/vwo die wiskunde A in hun pakket hebben.
plaats *De Veluwe*
info *uu.nl/onderwijs/wiskunde-a-lympiade*

14–18 maart 2016
❑ **Variational methods in probability theory and statistical physics**
Workshop in het kader van de Stochastic Activity Month (SAM) 'Probability and Analysis'.
plaats *TU Eindhoven*
info *www.eurandom.nl*

17 maart 2016
❑ **W4Kangoeroe**
WereldWijde WiskundeWedstrijd voor basis- en middelbare scholieren.
plaats *eigen school*
info *w4kangoeroe.nl*

22–23 maart 2016
❑ **NMC 2016**
52ste Nederlands Mathematisch Congres, jaarlijkse bijeenkomst van wiskundig Nederland, ditmaal tezamen met de Belgische en Luxemburgse zustervereenigingen.
plaats *Science Park, Amsterdam*
info *nmc.wiskgenoot.nl*

23 maart 2016
❑ **Grote Rekendag**
14de editie van de Grote Rekendag, waarop het doel is om kinderen onderzoekend te laten rekenen. Dit jaar is het thema 'Kijkje achter de code'.
plaats *basisscholen in Nederland en Vlaanderen*
info *groterekendag.nl*

24 maart 2016
❑ **Vorm en ruimte**
Workshop voor docenten wiskunde C (vwo) en docenten wiskunde D (havo), gegeven door Quintijn Puite en Theo van den Bogaart (Instituut Archimedes, HU).
plaats *Hogeschool Utrecht*
info *uu.nl/u-talent*

29 maart – 11 april 2016
❑ **Simulation of rare events**
Workshop in het kader van de Stochastic Activity Month (SAM) 'Probability and Analysis'.
plaats *TU Eindhoven*
info *www.eurandom.nl*

31 maart 2016
❑ **Leren van raadsels**
Wiskunde leren zien in alledaagse problemen. Masterclass voor leerlingen uit de bovenbouw van havo en vwo.
plaats *De Uithof, Utrecht*
info *uu.nl/u-talent*

april 2016

8 april 2016
❑ **Leve de Wiskunde!**
Het jaarlijkse wiskundecongres voor inzicht in de actuele ontwikkelingen in de wiskunde, voor docenten en leerlingen.
plaats *UvA, Science Park, Amsterdam*
info *itsacademy.nl/leve-de-wiskunde*

10–16 april 2016
❑ **European Girls' Mathematical Olympiade**
Europese wiskundewedstrijd speciaal voor meisjes.
plaats *Buşteni, Roemenië*
info *www.egmo.org*

11–15 april 2016
❑ **Taking the Measure of One-Dimensional Dynamics**
Workshop georganiseerd door onder anderen Henk Broer (RUG), Robbert Fokkink (TUD), Frank den Hollander (UL) en Ale Jan Homburg (UvA).
plaats *Lorentz Center@Oort, Leiden*
info *lorentzcenter.nl*

18–22 april 2016
❑ **Game Theory and Evolutionary Biology: Exploring Novel Links**
Workshop georganiseerd door Johan Metz (UL), Nicole Mideo (Toronto), Kateřina Staňková (UM) en Frank Thuijsman (UM).
plaats *Lorentz Center@Oort, Leiden*
info *lorentzcenter.nl*

mei 2016

26–27 mei 2016
❑ **Diamant symposium**
Halfjaarlijks symposium van het DIAMANT-cluster met gastspreker Manjul Bhargava (Princeton University).
plaats *Congrescentrum De Werelt, Lunteren*
info *websites.math.leidenuniv.nl/diamant*

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 van ± 350 woorden, zo mogelijk met illustratie. De redactie behoudt zich het recht voor berichten te weigeren of in te korten.

nieuws@nieuwarchief.nl

Rekentoets grotendeels uitgesteld

Begin oktober werd bekend dat de rekentoets, die dit schooljaar voor het eerst mee zou gaan tellen in de zak-/slaagregelingen van het voortgezet onderwijs en het mbo, toch grotendeels nog verder wordt uitgesteld. Naar aanleiding van de resultaten uit 2015 besloot het ministerie van Onderwijs, Cultuur en Wetenschap zelf al dat de toets tot 2021 nog niet mee zal gaan tellen voor het behalen van een diploma in het mbo, aangezien "leerlingen niet de dupe mogen worden van slecht rekenonderwijs". De resultaten waren inderdaad niet best: 12 procent (mbo-4) tot 30 procent (mbo-2) van de leerlingen scoorde afgerond een 3 of lager en zou daarmee geen diploma kunnen halen, ongeacht hun prestaties op de andere vakken.

Het ministerie berekende ook de toename in procentpunten van het aantal leerlingen in het voortgezet onderwijs die gezakt zouden zijn voor hun eindexamen als de rekentoets in 2015 al was meegeteld. Op de havo betreft dat 1,4 procentpunt extra gezakten, op het vmbo-gt 2,3 procentpunt en op het vmbo-kb 3,8 procentpunt. Bij de havo en het vmbo-kb is hier bovendien al rekening gehouden met de zogeheten vangnetregeling: normaal gesproken is een 5 voor de rekentoets voldoende, maar als in een jaar blijkt dat meer dan 5 procentpunt van de leerlingen extra zou zakken vanwege de rekentoets, dan wordt het minimaal benodigde cijfer omlaag gebracht naar een 4. De bovengenoemde percentages voor havo en vmbo-kb betreffen dus leerlingen die lager dan een 3,5 hebben gehaald.

Naar aanleiding van onder andere deze bevindingen is een motie ingediend vanuit de PvdA, waarin gesteld werd dat de resultaten op de havo en het vmbo van dusdanige aard zijn dat ook hier de rekentoets nog niet mag worden ingevoerd. Deze motie werd aangenomen, zodat in 2016 alleen nog maar op het vwo gezakt kan worden op de rekentoets (bij het behalen van een cijfer lager dan 4,5). Voor hen geldt wel nog steeds de vangnetregeling, hoewel het onwaarschijnlijk lijkt dat deze ingezet zal worden — de resultaten uit 2015 duiden op slechts een extra 0,5 procentpunt gezakten vanwege de rekentoets.

Leerlingen van alle niveaus zijn niettemin verplicht om de toets te maken, en het resultaat daarvan zal ook op het diploma worden vermeld. De situatie voor het schooljaar 2016–2017 is nog onduidelijk. Wel is toegezegd dat het ministerie in overleg zal gaan met onder andere de Nederlandse Vereniging van Wiskundeleraren om te kijken op welke manier voldoende aandacht gegeven kan worden aan de verbetering van het rekenonderwijs.

Aanvullend op het bovenstaande is ook een motie aangenomen ter vermindering van het gebruik van de rekenmachine bij de rekentoets, evenals een motie die vraagt om een onderzoek naar de besteding van de financiële middelen die ter beschikking zijn gesteld ten behoeve van het rekenonderwijs.

Mark Timmer

Aantal wiskundestudenten stijgt tot recordhoogte

Het aantal wiskundestudenten in het hoger onderwijs neemt spectaculair toe. Dit jaar stijgt de instroom tot boven de 1000 nieuwe studenten. Dit is een vervijfvoudiging ten opzichte van het aantal studenten dat zich in 2002 voor een studie wiskunde inschreef. Het Platform Wiskunde Nederland (PWN) schrijft de hoge stijging toe aan de toenemende relevantie van wiskunde in de maatschappij en de hoge baankansen in het veld. Deze stijging is voor PWN een stimulans om het vakgebied onder de aandacht te blijven brengen.

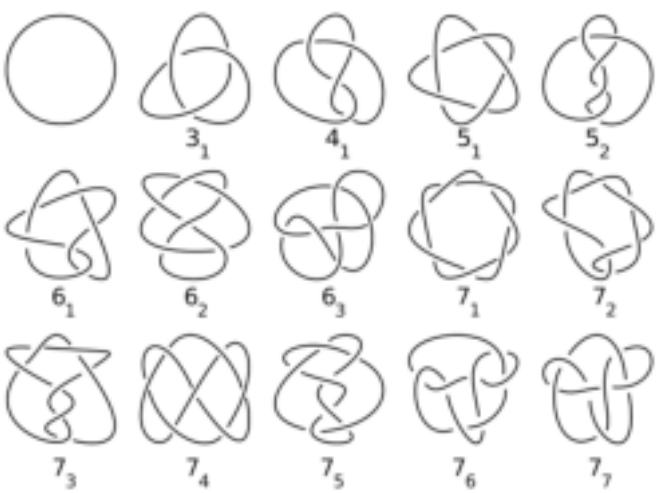
Remco van der Hofstad, hoogleraar wiskunde en woordvoerder PWN: "De vraag naar wiskundigen is op dit moment veel groter dan

het aanbod. Studenten beseffen dit steeds meer en kiezen voor een studie met een uitstekend toekomstperspectief. Bedrijven uit de meest uiteenlopende sectoren staan te springen om mensen die met getallen overweg kunnen. Geld, digitale communicatie, informatie, alles is in cijfers uit te drukken. Een wiskundige heeft inzicht in deze getallen en kan er patronen in ontdekken. Niet alleen in de techniek, maar bijvoorbeeld ook in de financiële sector, de gezondheidszorg, de logistiek, de landbouw, energie en verkeer." *platformwiskunde.nl*

Knopen in Bonita Avenue

De debuutroman *Bonita Avenue* van Peter Buwalda uit 2010 was zeer succesvol. Een van de hoofdpersonen is Siem Sigerius, rector van de Universiteit Twente die briljant wiskundig werk in Berkeley heeft gedaan en daarmee een Fieldsmedaille heeft gewonnen. Op pp. 112–113 van het boek beschrijft Siem zijn resultaten kort aan Aaron, de partner van zijn stiefdochter Joni. Hij geeft de definitie van een knoop, vertelt wanneer twee knopen als identiek moeten worden beschouwd, en vervolgt: "Hoe hou je ze uit elkaar? Zeker zestig jaar lang heeft dit onderzoek muurvast gezeten. En toen stelde iemand een polynoom op, een algebraïsche formule waarmee je knopen een eigen identiteit kunt geven. Die iemand was ik."

Dit geeft de wiskundige lezer associaties met het Jones-polynoom. Hoe kwam Peter Buwalda aan dit idee? Het antwoord lijkt, met dank aan Gerard Helminck, te vinden in een interview van Buwalda met Vaughan Jones dat op 15 oktober 1998 verscheen in *UT Nieuws*, weekblad van de Universiteit Twente, waarvan hij destijds redacteur was. Jones kwam naar Twente als spreker op het jaarlijks symposium van de TW-vakgroep Fundamentele Analyse. Een kopie van het interview is te vinden op https://staff.fnwi.uva.nl/t.h.koornwinder/fiction/1998_NieuwsUnivTwente.pdf. *Tom Koornwinder*



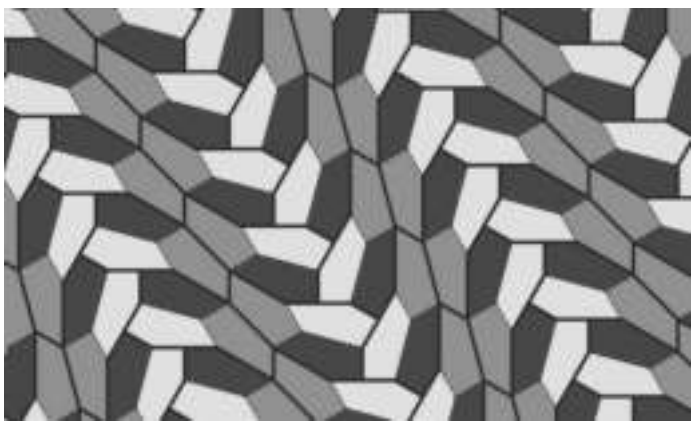
Wiskunde steeds populairder

Wegens verschrikkelijke ervaringen in haar jeugd had Lisbeth Salander een goede reden om de nationale veiligheidsdienst te wantrouwen. Na veel spannende avonturen kwamen zij en haar vriend Mikael Blomkvist erachter dat haar jeugdprobleem terugging tot een kleine groep binnen de dienst, een Sectie die de koude oorlog had overleefd.

Lezers van de Millennium-trilogie zullen zich afgevraagd hebben wat er verder met de hoofdpersonen zou gebeuren. De overleden auteur van de trilogie, Stieg Larsson, heeft nu een opvolger gekregen in David Lagercrantz. In Millennium deel 4, *Wat ons niet zal doden*, leest men over verdere verwikkelingen: Salander was altijd al een goede hacker, maar door de RSA-versleuteling werd het steeds moeilijker om computerbestanden binnen te dringen. Zij heeft zich daarom meer wiskunde eigen gemaakt, en gebruikt nu het liefst de factorisatiemethode van Lenstra met elliptische krommen. Daarmee komt zij een corrupte afdeling op het spoor van een bekende buitenlandse veiligheidsdienst. *Jaap Korevaar*

Kierloos tegelen met de vijfhoek

Met een regelmatige, vijfhoekige tegel kun je geen vloer betegelen zonder dat er gaten of kieren overblijven. Maar er bestaan scheve vijfhoeken waarmee dat wel mogelijk is. Tussen 1918 en 1985 werden veertien van zulke tegels gevonden, en toen bleef het heel lang stil. Casey Mann, Jennifer McLoud en David Von Derau van de Universiteit van Washington Bothell hebben nu een vijftiende exemplaar gevonden, met hulp van de computer. *kennislink.nl*



Illustratie: Casey Mann

Nieuwe veilingsite Wereldwiskunde Fonds

De veilingsite van het Wereldwiskunde Fonds is onlangs geheel vernieuwd. De NVvW heeft hiervoor het veilingprogramma PHP Pro Bid aangeschaft en de site hiermee ingericht. De biedmethode is nu te vergelijken met die op eBay.

Het Wereldwiskunde Fonds bestaat sinds 1993 en heeft als doel: "Ondersteuning van wiskundeonderwijs dat direct of indirect ten goede komt aan leerlingen op middelbare-schoolniveau in ontwikkelingslanden door middel van financiële bijdragen aan projecten." *wereldwiskundefonds.nl*

Wiskunde in de Nationale Wetenschapsagenda

In het rapport *Wetenschapsvisie 2025: keuzes voor de toekomst* heeft het kabinet de Kenniscoalitie gevraagd om het initiatief te nemen om een Nationale Wetenschapsagenda op te stellen. De Kenniscoalitie bestaat uit de universiteiten, hogescholen, Universitair Medische Centra, KNAW, NWO, VNO-NCW, MKB-Nederland en de instituten voor toegepast onderzoek. Iedereen in Nederland kon tot 1 mei vragen voor de Nationale Wetenschapsagenda insturen. Dit heeft geleid tot zo'n 11.700 reacties, met vragen van zeer verschillende diepgang.

Er zaten ook een honderdtal wiskundevragen bij, variërend van "Zijn de Langlandsvermoedens waar, en wat betekenen deze voor de fysi-

ca?" tot "Hoeveel boterhammen zijn er op de wereld?" van een zekere Aneirin (4 jaar) die zich dit momenteel afvraagt. Vele vragen die gesteld werden, zijn niet nieuw. Dit bracht wiskundige K.P. Hart van de TU Delft op het idee om alle gestelde wiskundevragen die al een antwoord hebben op een webpagina te verzamelen en van commentaar te voorzien. *wiskwa.tumblr.com*

Het wiskundehondje

Bij Uitgeverij Nieuwezijds is het kinderboek *Het wiskundehondje* verschenen van Margriet van der Heijden. De verhalen in deze bundel zijn een selectie uit de rubriek 'Vormen en getallen' van de kinderwetenschapspagina van *NRC Handelsblad*. Het zijn steeds korte stukjes van twee bladzijden met leuke weetjes over bijvoorbeeld het =-teken, een zielig getal, hoe je kunt goochelen met letters, hoe je van een cirkel een rechthoek maakt en wat de allermooiste formule is. *nieuwezijds.nl*

Illustraties: Iris van der Graaf

Het ultieme propellerblad ontwerpen

Er wordt steeds meer gevraagd van moderne vliegtuigmotoren, scheepspropellers, waterturbines en schroefmachines. Niet alleen op het gebied van efficiëntie, maar ook wat betreft slijtvastheid, gewicht, en concurrerende productiekosten. Dergelijke optimalisatieproblemen met meerdere doelen hebben niet een optimale oplossing. Met de huidige ontwerptools is het nagenoeg onmogelijk het beste compromis te vinden voor alle criteria. In september is in Delft het driejarige Europese MOTOR-programma gestart. Doel van dit project (4,3 miljoen euro, penvoerder TU Delft), is om met nieuwe wiskundige concepten en geavanceerde rekentools nieuwe technologieën te ontwikkelen waarmee in minder tijd dergelijke zogenaamde *fluid energy machines* optimaal kunnen worden ontworpen. *motor-project.eu*

Breng de schoonheid van wiskunde tot leven

Onder de titel 'Breng de schoonheid van wiskunde tot leven' worden alumni van de Universiteit Utrecht en anderen gevraagd om een financiële bijdrage voor de restauratie van ongeveer tachtig geometrische 3D-modellen van gips, draad of papier.

Wetenschappers brachten ruim honderd jaar geleden wiskundige formules tot leven in aansprekende 3D-modellen. Het is hoog tijd om deze prachtige stukken weer in hun volle glorie te herstellen en te openbaren aan nieuwe generaties. Sinds oktober staan de modellen permanent tentoongesteld in de bibliotheek van het Mathematisch Instituut in Utrecht (Budapestlaan 6, 7de verdieping). *uu.nl/alumni*

Koninklijke Bibliotheek gaat WisFaq archiveren

Sinds 2001 is de website www.wisfaq.nl een veelgebruikte vraagbaak op het gebied van wiskunde. De site is opgezet door Willem van Ravenstein en wordt nog steeds door hem geleid en onderhouden. Begin november heeft de Koninklijke Bibliotheek (KB) laten weten dat zij de inhoud van de website gaat archiveren als onderdeel van ons 'digitaal erfgoed'.

Websites bevatten vaak waardevolle informatie die niet analoog verschijnt en die ten gevolge van de grote omloopsnelheid het risico loopt voorgoed verloren te gaan. Dat websites als digitaal erfgoed het behouden waard zijn, is internationaal erkend in het *Unesco Charter on the Preservation of the Digital Heritage* uit 2003. Het signaleert dat digitaal erfgoed verloren dreigt te gaan en dat het bewaren daarvan voor gebruik door de huidige en toekomstige generatie onderzoekers zeer urgent is.

De KB archiveert websites die als verzameling een representatief beeld geven van de Nederlandse cultuur, geschiedenis en samenleving op het internet.De archiefversies worden beschikbaar gesteld aan een algemeen publiek via de website van de KB. *wisfaq.nl*

Rob Tijdeman
Mathematisch Instituut
Universiteit Leiden
tijdeman@math.leidenuniv.nl

Onderzoek

Het *abc*-vermoeden

Het abc-vermoeden is uitgegroeid tot een belangrijk gereedschap in de theorie over Diophantische vergelijkingen. Het wordt gebruikt om na te gaan wat je aan resultaat kunt verwachten, ook al kun je het niet bewijzen. Enige tijd geleden kwam het vermoeden in het nieuws vanwege de claim van de Japanner Shinichi Mochizuki (in ruim 500 bladzijden) het vermoeden bewezen te hebben. Of dat zo is, is nog niet duidelijk. Dit artikel is een inleiding met aandacht voor de historische ontwikkeling en enkele toepassingen.

In 1981 bewees Stothers [20]:

Als $A[x], B[x], C[x] \in \mathbb{C}[x]$ polynomen zijn zonder gemeenschappelijk nulpunt, niet alle constant, zó dat $A(x) + B(x) = C(x)$, dan is

$$\max(\deg(A), \deg(B), \deg(C)) < N_1(ABC)$$

waarbij $N_1(f)$ het aantal verschillende nulpunten van $f(x)$ is.

Kies bijvoorbeeld $A(x) = (x+1)^5, B(x) = -(x-1)^5, C(x) = 10x^4 + 20x^2 + 2$. Dan is het linkerlid 5 en volgt dat de vier nulpunten van $C(x)$ enkelvoudig zijn. Een ander gevolg van de stelling is dat voor alle n de veelterm $(x+1)^n - x^n$ van graad $n-1$ geen meervoudige nulpunten heeft.

De stelling was een hulpresultaat en bleef onopgemerkt tot Mason in 1984 het resultaat herontdekte en in zijn proefschrift publiceerde [10]. Ik verwijs naar de stelling als de *stelling van Stothers–Mason*. Een bewijs geef ik in de volgende paragraaf.

Sommige wiskundigen vroegen zich af of er een analogon voor gehele getallen zou zijn. Het voor de hand liggende analogon gaat niet op, want als a, b, c onderling ondeelbare positieve gehele getallen zijn met $a+b=c$, dan hoeft niet te gelden dat $c < N(abc)$ waarbij $N(abc)$ het product van de priemfactoren van abc is. Neem bijvoorbeeld $a = 49, b = 32, c = 81$. Dan is $c = 81, N = 42$.

Oesterlé vermoedde in 1985 dat *er een constante C is zodat als a, b, c onderling ondeelbare positieve gehele getallen zijn met $a+b=c$, dat dan $c < N(abc)^C$.*

Masser dacht dat C zelfs asymptotisch gelijk aan 1 zou kunnen zijn in de volgende zin:

Voor elke $\varepsilon > 0$ bestaan slechts eindig veel onderling ondeelbare positieve gehele drietallen a, b, c met $a+b=c$ en $c > (N(abc))^{1+\varepsilon}$.

Dit staat nu bekend als *het abc-vermoeden* en is nog niet bewezen of weerlegd.

De Japanner Mochizuki [13] heeft meer dan twee jaar geleden op zijn homepage vier artikelen gepubliceerd. Hij beweert dat hij daarin het abc-vermoeden bewijst. Verschillende wiskundigen proberen het bewijs te begrijpen, maar de theorie is zo ingewikkeld dat nog niemand het heeft geverifieerd of weerlegd. Zoals we zullen zien heeft het abc-vermoeden op Diophantische vergelijkingen van drie termen ongeveer hetzelfde effect als kwantumcomputing op factorisatie van grote getallen. Een bewijs van het abc-vermoeden zou heel wat te weeg brengen. Het zijn dus spannende tijden.

Waarschuwing. Vergeet niet te eisen dat a, b, c onderling deelbaar zijn. Beschouw bijvoorbeeld de vergelijking $2^{k-1} + 2^{k-1} = 2^k$ voor een willekeurig positief geheel getal k. Dan is $c = 2^k = (N(abc))^k$ en hierbij kan de exponent k willekeurig groot gemaakt worden.

Bewijs van de stelling van Stothers–Mason

We schrijven (f, g) voor de grootste gemene deler van de polynomen $f(x)$ and $g(x)$, $\deg(f)$ voor de graad van $f(x)$, en $N_1(f)$ voor het aantal verschillende nulpunten van $f(x)$. We geven een schets van het bewijs van de stelling van Stothers–Mason waarbij de lezer zelf de details kan invullen. We nemen in deze sectie de voorwaarden stilzwijgend aan. We beginnen met enkele lemma's. Het bewijzen van de eerste twee wordt aan de lezer overgelaten.

Lemma 1. *Als $f \in \mathbb{C}[x]$ een wortel x_0 van orde $k > 0$ heeft, dan is x_0 een wortel van f' van orde $k-1$.*

Lemma 2. $\deg((f, f')) = \deg(f) - N_1(f)$.

Lemma 3. $A'B - AB' = BC' - B'C \neq 0$.

Bewijs. Als B constant is, is A dat niet en is

$A'B - AB' \neq 0$. Als A en B niet constant zijn, volgt uit $A'B - AB' = 0$ dat B een deler is van B', een tegenspraak. Dus $A'B - AB' \neq 0$. □

Lemma 4. $(A, A') \times (B, B') \times (C, C') \mid (BC' - B'C)$.

Bewijsschets. Merk op dat (A, A') en (B, B') en (C, C') geen nulpunt gemeen hebben en elk nulpunt van het product deler is van $A'B - AB' = BC' - B'C$. Dus is $(A, A') \times (B, B') \times (C, C')$ een deler van $BC' - B'C$. □

Bewijs van de stelling van Stothers–Mason. Uit Lemma's 4 en 3 volgt

$$\begin{aligned}\deg(A, A') + \deg(B, B') + \deg(C, C') \\ \leq \deg(BC' - B'C) = \deg(BC) - 1.\end{aligned}$$

Door toepassing van Lemma 2 vinden we

$$\begin{aligned}\deg(A) - N_1(A) + \deg(B) - N_1(B) \\ + \deg(C) - N_1(C) \\ < \deg(B) + \deg(C).\end{aligned}$$

Dus $\deg(A) < N_1(ABC)$. Vanwege de symmetrie in A, B, C volgt

$$\max(\deg(A), \deg(B), \deg(C)) < N_1(ABC). \quad \square$$

Het $abc2$-vermoeden

Veel mensen hebben onderzocht hoe groot de optimale constante in het vermoeden van Oesterlé zou moeten zijn. Laat a, b, c onderling ondeelbare positieve gehele getallen zijn met $a + b = c$. Schrijf $q = q(abc) = \log c / \log(N(abc))$. De huidige records zijn gegeven in Tabel 1, zie [14].

Genoemde waarden van a, b, c waren al vóór 1995 bekend. In 2000 waren dertien drietallen met $q > 1{,}5$ gevonden en sindsdien geen enkele meer. Ook heuristisch is er goede reden om aan te nemen dat er niet meer zijn. Het bovenstaande maakt aannemelijk dat in het vermoeden van Oesterlé $q = 1{,}63$ voldoet en dat dit nauwelijks te verbeteren is. Baker [1] heeft in 2004 een verscherping van het abc-vermoeden geformuleerd waaruit volgt dat $q = 1{,}75$ een correcte waarde in het vermoeden van Oesterlé zou moeten zijn. In feite vermoedt hij [1] dat voor alle onderling ondeelbare positieve gehele getallen a, b, c met $a + b = c$ geldt dat

$$c < \frac{6}{5} \frac{N(abc)(\log N(abc))^{\omega(abc)}}{(\omega(abc))!} \quad (1)$$

waarbij $\omega(abc)$ het aantal priemdelers van abc is. Laten we een beetje voorzichtiger zijn:

Het $abc2$-vermoeden. *Er bestaan geen onderling ondeelbare positieve gehele getallen a, b, c met $a + b = c$ zó dat $c < (N(abc))^2$.*

Het ABC@home-project

In 2005 startte Bart de Smit een project om drietallen onderling ondeelbare positieve gehele getallen a, b, c te vergaren met $a + b = c$ en $c > N(abc)$. Zulke drietallen worden *ABC-drietallen* genoemd. Hij deed dit met een tweeledig doel, enerzijds om een grote groep mensen bij wiskundig onderzoek te betrekken, anderzijds om gebruik te maken van anders ongebruikte rekenkracht. Alle deelnemers kregen de software toegestuurd om een bepaald traject van drietallen a, b, c te onderzoeken zodat in totaal een groot gebied onderzocht werd. De resultaten tot nog toe zijn te vinden op [17] waar men zich ook nog kan aanmelden. Inmiddels zijn 23.827.716 ABC-drietallen gevonden.

Het begon volgens de statistieken op 31 augustus 2009 met 147.317 drietallen. Het grootste aantal dat op één dag gevonden werd was 822.036 en wel op 15 oktober 2010. Het vinden van nieuwe drietallen wordt steeds lastiger: op 12 februari 2014 werden 197 nieuwe ABC-drietallen in veertien dagen gerapporteerd. Het zoeken naar drietallen met $c < 10^{18}$ heeft in totaal 14.482.065 van die ABC-drietallen opgeleverd. Het streven is nu om alle ABC-drietallen met $c < 2^{63}$ te bepalen.

Interessant is natuurlijk ook om te zien welke kwaliteit q de gevonden ABC-drietallen hebben. Van de 14.482.065 gevonden drietallen met $c < 10^{18}, q > 1$ geldt voor 0,15 procent dat $c < 10^9$. De overeenkomstige cijfers zijn voor de drietallen met

$q > 1{,}05$,	2.352.105,	0,37%,
$q > 1{,}1$,	449.194,	0,82%,
$q > 1{,}2$,	24.013,	2,94%,
$q > 1{,}3$,	1.843,	7,81%,
$q > 1{,}4$,	160,	21,25%.

Het is duidelijk dat de ABC-drietallen met hoge kwaliteit steeds zeldzamer worden. In het onderzochte hogere segment van de c's groeit het aantal ABC-drietallen ruwweg exponentieel met q, maar het aantal drietallen met $q > 1{,}4$ groeit ongeveer lineair met q.

Vergelijkingen van Fermat, Catalan, Pillai

De laatste stelling van Fermat zegt dat er geen positieve gehele getallen $n > 2, x, y, z$ bestaan met $x^n + y^n = z^n$. Hier mogen we zonder verlies van algemeenheid aannemen dat x, y, z onderling ondeelbaar zijn. Nadat Faltings in 1983 bewezen had dat er voor elke $n > 2$ maar eindig veel onderling ondeelbare oplossingen x, y, z zijn, werd de stelling van Fermat in 1995 volledig door Wiles en Taylor bewezen [23].

Met het $abc2$-vermoeden wordt de Laatste Stelling van Fermat voor $n > 5$ een fluitje van een cent. Pas het $abc2$-vermoeden toe op $x^n + y^n = z^n$. We mogen aannemen dat de termen onderling ondeelbaar zijn, want anders delen we eerst de gemene deler uit. We krijgen $z^n < (xyz)^2 < z^6$. Dus $n \leq 5$. (Het is al bijna 200 jaar bekend dat er voor $n \leq 5$ geen oplossingen zijn.)

Het vermoeden van Catalan zegt dat 8 en 9 de enige opeenvolgende positieve gehele getallen zijn die beide een pure macht zijn, met andere woorden dat de enige oplossing van de vergelijking $x^m - y^n = 1$ met $m > 1, n > 1, x > 1, y > 1$ gegeven wordt door $(m, n, x, y) = (2, 3, 3, 2)$. Nadat Tijdeman in 1976 bewezen had dat er een (heel groot) getal is waarboven geen oplossingen voorkomen, werd het volledige vermoeden in 2004 door Mihailescu bevestigd [11].

$q = 1.6299$	$a = 2$, $b = 3^{10} \times 109$, $c = 23^5$	Reyssat
$q = 1.6260$	$a = 11^2$, $b = 3^2 \times 5^6 \times 7^3$, $c = 2^{21} \times 23$	De Weger
$q = 1.6235$	$a = 19 \times 1307$, $b = 7 \times 29^2 \times 31^8$, $c = 2^8 \times 3^{22} \times 5^4$	Browkin and Brzezinski
$q = 1.5808$	$a = 283$, $b = 5^{11} \times 13^2$, $c = 2^8 \times 3^8 \times 17^3$	Browkin and Brzezinski, Nitaj
$q = 1.5679$	$a = 1$, $b = 2 \times 3^7$, $c = 5^4 \times 7$	De Weger

Tabel 1 De huidige records van de optimale constante q.

Ook de oplossing van de vergelijking van Catalan wordt eenvoudig, als we het abc2-vermoeden aannemen. Uit $x^m - y^n = 1$ volgt dan dat $x^m < (xy)^2 < (x^m)^{2(\frac{1}{m}+\frac{1}{n})}$. Dus is $1 < 2(\frac{1}{m}+\frac{1}{n})$. De ongelijkheid $\frac{1}{m}+\frac{1}{n} > \frac{1}{2}$ geldt alleen als $\min(m,n) \le 3$. Maar het is sinds 1964 bekend dat er voor deze waarden geen oplossingen zijn.

Een vergelijking waarvan de vergelijkingen van Fermat en Catalan speciale gevallen zijn werd in 1945 door Pillai bestudeerd:

$$x^m + y^n = z^k$$

in onderlinge ondeelbare gehele getallen $x > 1$, $y > 1$, $z > 1$ met $m > 1$, $n > 1$, $k > 1$. Beukers [2] en Edwards [5] bewezen dat voor elk drietal m,n,k met $1/m + 1/n + 1/k > 1$ er oneindig veel oplossingen zijn die tot eindig veel parameterklassen zijn terug te brengen. Dat betreft de exponenten $(2,2,*)$, $(2,3,3)$, $(2,3,4)$ en $(2,3,5)$. Het is bekend dat er geen oplossing is voor drietallen m,n,k met $1/m+1/n+1/k = 1$, dat wil zeggen voor exponenten $(3,3,3)$, $(2,4,4)$, $(2,3,6)$. Darmon en Granville [4] bewezen in 1995 dat er voor elk drietal m,n,k met $1/m+1/n+1/k < 1$ maar eindig veel oplossingen x,y,z zijn. De methode is ineffectief, dat wil zeggen dat ze het niet mogelijk maakt voor vaste m,n,k de oplossingen daadwerkelijk te bepalen. Tot 1993 waren de volgende oplossingen van de vergelijking van Pillai bekend:

$$1^m + 2^3 = 3^2,$$

$$2^5 + 7^2 = 3^4,$$

$$13^2 + 7^3 = 2^9,$$

$$2^7 + 17^3 = 71^2,$$

$$3^5 + 11^4 = 122^2.$$

Bij de voorbereiding van de Fermatdag in Utrecht in november 1993 vonden Beukers en Zagier tot ieders verrassing nog vijf grote oplossingen:

$$33^8 + 1549034^2 = 15613^3,$$

$$1414^3 + 2213459^2 = 65^7,$$

$$9262^3 + 15312283^2 = 113^7,$$

$$17^7 + 76271^3 = 21063928^2,$$

$$43^8 + 96222^3 = 30042907^2.$$

Sindsdien heeft niemand meer een oplossing gevonden. Merk op dat er in alle tien gelijkheden een kwadraat voorkomt. Dit leidde mij ertoe om tijdens die Fermatdag het volgende vermoeden uit te spreken:

De vergelijking $x^m + y^n = z^k$ heeft geen oplossingen $x > 1$, $y > 1$, $z > 1$ met $m > 2$, $n > 2$, $k > 2$ en x,y,z onderling ondeelbaar.

Dit is een generalisatie van zowel de Laatste Stelling van Fermat als de Vergelijking van Catalan. De bankier Beal heeft een geldbedrag uitgeloofd voor de oplossing van dit vermoeden en daarom staat dit vermoeden wel bekend als het Beal-vermoeden en ook als het vermoeden van Tijdeman–Zagier.

Wat zegt het abc-vermoeden over de vergelijking van Pillai? Het is een leuke opgave, ook voor een klas leerlingen, om te bewijzen dat uit $1/m + 1/n + 1/k < 1$ volgt dat $1/m + 1/n + 1/k \le 1 - 1/42$. Passen we het abc-vermoeden met $\varepsilon = 0,01$ toe, dan volgt dat $z^k < (xyz)^{1,01}$ met slechts eindig veel uitzonderingen. Dus bestaat een constante $C > 1$ zo dat $z^k < C(xyz)^{1,01}$ voor alle toegelaten zestallen k,m,n,x,y,z. Omdat $x < z^{k/m}$ en $y < z^{l/m}$, volgt dat voor alle k,m,n met $1/m + 1/n + 1/k < 1$ geldt

$$\begin{aligned} z^k &< Cz^{1,01k(1/m+1/n+1/k)} \\ &\le Cz^{1,01k(1-1/42)} \\ &< Cz^{0,99k}. \end{aligned}$$

Dat impliceert $z^k < C^{100}$. Maar dan zijn ook x^m en y^n kleiner dan C^{100}. Als het abc-vermoeden waar is, heeft de vergelijking van Pillai dus maar eindig veel toegelaten oplossingen m,n,k,x,y,z *in totaal*. Waarschijnlijk zijn bovenstaande tien oplossingen de enige met $x > 1$, $y > 1$, $z > 1$, $m > 1$, $n > 1$, $k > 1$ en x,y,z onderling ondeelbaar en $1/m + 1/n + 1/k < 1$.

Ik ga in het vervolg nog op een meer recente toepassing in. Er zijn echter nog vele andere toepassingen gepubliceerd. Voor een lijst van zulke toepassingen verwijs ik naar [14].

Kwadraatvrije deel van blok getallen

Het *kwadraatvrije deel* van een positief getal n is het kleinste getal a zodat n/a een kwadraat is. We schrijven $a = Q(n)$. Dus $Q(12) = 3$ en $Q(600) = 6$.

Een blok getallen is een rij opeenvolgende natuurlijke getallen. Onder het kwadraatvrije deel van een blok getallen verstaan we gemakshalve het kwadraatvrije deel van het product van die rij opeenvolgende getallen. Het begrip speelt een rol bij een nog onopgelost probleem van Erdős [8], namelijk of het product van twee blokken getallen een kwadraat kan zijn. Dat is equivalent met de vraag of twee blokken hetzelfde kwadraatvrije deel kunnen hebben. Het probleem van Erdős ligt in het verlengde van het resultaat dat een enkel blok opeenvolgende getallen geen kwadraat als product kan hebben, in 1939 onafhankelijk van elkaar bewezen door Erdős [7] en Rigge [15].

Als je een blok van twee getallen beschouwt, kan het kwadraatvrije deel zo klein zijn als 2. Neem bijvoorbeeld 8 en 9. Het komt zelfs oneindig vaak voor: Als $Q((n-1)n) = 2$, dan is n een kwadraat en $n-1$ twee keer een kwadraat of andersom. Welnu, de Pell-vergelijkingen $x^2 - 2y^2 = \pm 1$ hebben elk oneindig veel oplossingen. (Start met $(x_1,y_1) = (3,2)$ en definieer inductief $x_n = x_{n-1} + 2y_{n-1}$, $y_n = x_{n-1} + y_{n-1}$ voor $n = 2,3,\ldots$.) Zo vinden we de blokken $(49,50)$, $(288,289)$, $(1681,1682)$, ... met kwadraatvrij deel 2.

Wat weten we van het kwadraatvrije deel van een blok van drie getallen, zeg $n-1, n, n+1$? Weinig. Nog niet eens dat het groter is dan $\log n$. Het abc-vermoeden geeft wel een mooie ondergrens.

Stel $n - 1 = by^2, n = ax^2, n + 1 = cz^2$ met a,b,c kwadraatvrij (dat wil zeggen niet deelbaar door een kwadraat > 1). Dan is

$$a^2x^4 - bcy^2z^2 = n^2 - (n^2 - 1) = 1.$$

Het abc-vermoeden impliceert dat, voor $\varepsilon > 0$ en bijbehorende C,

$$(abc)^{2/3}(xyz)^{4/3} < a^2x^4 < C(abcxyz)^{1+\varepsilon}.$$

Hieruit volgt

$$\begin{aligned} C(abc)^{\frac{1}{2}+\frac{1}{2}\varepsilon} &> (abc)^{\frac{1}{6}-\frac{1}{2}\varepsilon}(xyz)^{\frac{1}{3}-\varepsilon} \\ &= (n(n-1)(n+1))^{\frac{1}{6}-\frac{1}{2}\varepsilon}. \end{aligned}$$

Dus is $abc > C^{-3}n^{1-4\varepsilon}$ voor voldoend grote n. Door voor het paar $(n,n+1)$ twee getallen te kiezen waarvan het product een kwadraatvrij deel 2 heeft, zien we in dat het kwadraatvrije deel van abc oneindig vaak kleiner dan $2n$ is. Het abc-vermoeden geeft dus de op $\varepsilon > 0$ na best mogelijke exponent van n, namelijk 1.

Wat weten we over het abc-vermoeden?

Nog voordat het abc-vermoeden zijn naam gekregen had, bewezen Stewart en Tijdeman

[18] de volgende twee stellingen. We schrijven in het vervolg N voor $N(abc)$.

Stelling 1. *Alle onderling ondeelbare positieve gehele getallen a, b, c met $a + b = c$ voldoen aan*

$$\log c < CN^{15}, \tag{2}$$

waarbij C een uit te rekenen constante is.

Stelling 2. *Zij $\varepsilon > 0$. Dan bestaan er oneindig veel onderling ondeelbare positieve gehele drietallen a, b, c met $a + b = c$ en*

$$c > N \exp\left((4-\varepsilon)\frac{\sqrt{\log N}}{\log\log N}\right). \tag{3}$$

Sindsdien zijn alleen de constantes in de exponenten verbeterd. Stewart and Yu [19] hebben het rechterlid van (1) verbeterd tot $CN^{1/3}(\log N)^3$. Frankenhuysen [9] heeft de constante 4 in (3) verbeterd tot een constante tussen 6 and 7. Er zit een orde verschil tussen de bovengrens (2) en de ondergrens (3). Vermoed wordt dat (3) heel dicht bij de asymptotisch juiste grens zit. Schatting (2) berust namelijk op een ongelijkheid uit de theorie van lineaire vormen in logaritmen van algebraïsche getallen waarvan vermoed wordt dat deze een orde van grootte te zwak is. Baker [1] heeft een vermoede scherpe schatting voor lineaire vormen in logaritmen geïntroduceerd en laat zien dat deze equivalent is met zijn vermoede schatting (1) voor het abc-vermoeden.

Een recent artikel van Robert, Stewart and Tenenbaum [16] bevat een heuristisch argument dat de juiste orde gegeven wordt door

$$N \exp\left(4\sqrt{3}\sqrt{\frac{\log N}{\log\log N}}\right).$$

Zij geven daarbij een hoofdterm, een tweede-orde-term en een restterm. De groei van $\exp(\sqrt{\log N})$ is te traag om numerieke evidentie voor deze formule te vinden. Maar de eerdere numerieke resultaten ondersteunen anders dan de heuristiek op overtuigende wijze dat het abc-vermoeden waar is. Dit is uitgewerkt in Baker [1] en heeft ook geleid tot de keuze van de constante 6/5 in (1).

Meer over abc

Wie meer over het abc-vermoeden zelf wil lezen, wordt verwezen naar [12], [14] en [22]. Er bestaan generalisaties van de stelling van Stothers–Mason en het abc-vermoeden in verschillende richtingen. Voor generalisaties van de stelling van Stothers–Mason voor meer dan drie variabelen verwijs ik naar [3] en [21]. Voor generalisaties van het abc-vermoeden, consequenties, numerieke resultaten en referenties naar verwante resultaten, zie [6], [14] en verwijzingen daar.

Referenties

1 A. Baker, Experiments on the abc-conjecture, *Publ. Math. Debrecen* 65 (2004), 253–260.

2 F. Beukers, The Diophantine equation $Ax^p + By^q = Cz^r$, *Duke Math. J.* 91 (1988), 61–88.

3 P. Corvaja en U. Zannier, An $abcd$ theorem over function fields and applications, *Bull. Soc. Math. France* 139 (2011), 437–454.

4 H. Darmon en A. Granville, On the equation $z^m = F(x, y)$ and $Ax^p + By^q = Cz^r$, *Bull. London Math. Soc.* 27 (1995), 513–543.

5 J. Edwards, A complete solution to $X^2+Y^3+Z^5$, *J. Reine Angew. Math.* 571 (2004), 213–236.

6 N.D. Elkies, The ABC's of number theory, *The Harvard College Mathematics Review* 1(1) (2007), 57–76.

7 P. Erdős, Notes on the products of consecutive integers, I and II, *J. London Math. Soc.* 14 (1939), 194–198 en 245–249.

8 P. Erdős en R.L. Graham, On products of factorials, *Bull. Inst. Math. Acad. Sinica* 4 (1976), 337–355. Zie ook: *Old and new problems and results in combinatorial number theory*, Monograph Enseign. Math. 28, Geneva, 1980.

9 M. van Frankenhuijsen, A lower bound in the ABC conjecture, *J. Number Th.* 82 (2000), 91–95.

10 R.C. Mason, *Diophantine Equations over Function Fields*, LMS LNS 96, Cambridge Univ. Press, 1984.

11 P. Mihăilescu, Primary cyclotomic units and a proof of Catalan's conjecture, *J. Reine Angew. Math.* 572 (2004), 167–195.

12 P. Mihăilescu, Around ABC, *EMS Newsletter* September 2014, 29–34, http://www.ems-ph.org/journals/newsletter/pdf/2013-09-89.pdf.

13 S. Mochizuki, http://www.kurims.kyoto-u.ac.jp/~motizuki/Inter-universal%20Teichmuller%20Theory%20IV.pdf.

14 A. Nitaj, www.math.unicaen.fr/~nitaj/abc.html

15 O. Rigge, Ueber ein diophantisches Problem, *9th Congress Math. Scand.*, pp. 155–160.

16 O. Robert, C.L. Stewart en G. Ténenbaum, A refinement of the abc conjecture, *Bull. London Math. Soc.* 46 (2014), 1156–1166.

17 B. de Smit, http://www.abcathome.com/data.

18 C.L. Stewart, R. Tijdeman, On the Oesterlé-Masser conjecture, *Monatsh. Math.* 102 (1986), 251–257.

19 C.L. Stewart en K.R. Yu, On the abc-conjecture II, *Duke Math. J.* 108 (2001), 169–181.

20 W.W. Stothers, Polynomial identities and Hauptmoduln, *Quart. J. Math. Oxford* (2) 32 (1981), 349–370.

21 L.N. Vaserstein en E.R. Wheland, Vanishing polynomial sums, *Comm. Algebra* 31 (2003), 751–772.

22 M. Waldschmidt, On the abc-conjecture and some of its consequences, https://www.imj-prg.fr/~michel.waldschmidt/articles/pdf/abcLahore2013VI.pdf.

23 A. Wiles, Modular elliptic curves and Fermat's last theorem, *Ann. Math. (2)* 141 (1995), 443–551.

Jan Beuving
cabaretier, Zeist
janbeuving@gmail.com

Het keerpunt van Ionica Smeets

Er valt nog zoveel te winnen voor wiskundig Nederland

Vijf jaar lang interviewde Ionica Smeets voor de rubriek 'Het keerpunt' mensen die na hun studie de gebaande paden van de wiskunde verlieten. Nu neemt ze afscheid van deze serie. Toen ze begon was ze vooral bekend als Wiskundemeisje, onlangs is ze benoemd tot hoogleraar wetenschapscommunicatie aan de Universiteit Leiden. Hoogste tijd om de interviewer nu zelf eens aan het woord te laten.

Is één van al die keerpunten je in het bijzonder bijgebleven?
"Als ik één iemand moet noemen, dan Steven Engelsman. Hij was toen ik hem interviewde directeur van het Museum Volkenkunde in Leiden. Het was zo'n leuke man, normaal zou ik hem nooit gesproken hebben. Hoewel hij iets heel anders deed, was hij zo blij dat hij wiskunde gestudeerd had. Dat was geweldig.

Ik liet de interviews altijd aan mijn moeder lezen voordat ik ze publiceerde. Na het lezen van het keerpunt met Engelsman zei ze: 'Wat moet het leuk zijn om voor die man te werken.' Ze had tussen de regels door feilloos gelezen dat het een bijzonder gesprek was."

Is er een gemene deler te ontdekken in de interviews?
"Er was heel veel diversiteit in hoe de mensen het roer omgegooid hadden; rigoureus of stapje-voor-stapje. Maar wat opviel is dat iedereen, ver afgedwaald van de wiskunde of niet, zich nog steeds wiskundige voelde."

Is het typisch iets voor wiskundigen, denk je, dat ze met zoveel plezier terugkijken op hun studie?
"Dat weet ik niet. Het is natuurlijk ook een kwestie van selectie. Ik heb alleen maar mensen gesproken die min of meer geslaagd zijn in hun carrière. Dan is het misschien ook makkelijker om blij te zijn met je studie. De mensen die werkloos thuis zitten heb ik niet opgezocht. "

Wat heeft jou zelf naar de wiskunde gedreven?
"Ik dacht dat ik als kind niet zo heel erg gefocust was op wiskunde en cijfers. Maar toen ik laatst een dagboek vond van toen ik een jaar of negen was, las ik: 'Ik was vandaag in het zwembad. Buiten was het plusminus 25 graden Celsius. Mama kwam ons om 12.45 ophalen.' Het zat er dus toch al een tikje in. Mijn lievelingsspelletje was bovendien een spel waar je een boter-kaas-en-eierenrekje moest vullen met de getallen die je met een dobbelsteen gooide, om ze vervolgens op te tellen tot zo dicht mogelijk bij duizend. Dat kon ik uren doen in mijn eentje. Maar mijn studiekeuze was toch heel moeilijk. Mijn lievelingsvak op school was Nederlands, en ik deed ook toneel. Uiteindelijk gaf het de doorslag dat je in een bètastudie wel alfahobby's kunt doen, terwijl het andersom vrijwel onmogelijk is. Daarnaast voelde het

Foto: Ype Driessen

Ionica Smeets

ook een beetje als verspilling van talent om niet iets met mijn bètakant te doen."

Toch ben je niet meteen wiskunde gaan studeren.
"Dat kwam door een advies op de middelbare school. De decaan raadde me aan om informatica te gaan doen. Want dat was wiskunde én programmeren. Mijn vader, zelf natuurkundige, vond dat ook een goed idee. Dus dat ging ik toen maar doen. Ik was heel autoriteitsgevoelig. Dat maakte het ook moeilijk om te stoppen. Want ik haalde mijn vakken, maar ondertussen was ik jaloers op de wiskundestudenten die ik bij mijn studievereniging tegenkwam. Toen ik bij mijn propedeuse-uitreiking hoorde dat de rest van de studie nu een formaliteit was, dacht ik: dan kan ik met een gerust hart stoppen. Een week later studeerde ik wiskunde."

Heb je nog iets aan die informatica gehad?
"Ik heb uiteindelijk nog best veel keuzevakken gedaan. Fundamentele informatica vond ik heel interessant; ik vind algoritmes en procedures uitdenken heel leuk. Maar, ik ben slordig in de implementatie en daar liep ik steeds tegenaan. Ik vind het belangrijker om een goed idee te hebben, dan een vlekkeloze uitwerking. Ook nu als ik columns schrijf: ik heb liever een column met een goed idee en een d/t-fout erin, dan een foutloos geschreven tekst die je niet verheft."

Je noemt de columns — heb je altijd voor ogen gehad dat je die kant op wilde?
"Wel dat ik wilde schrijven. Toen ik zeven was maakte ik al boekjes, en later zat ik bij de school- en universiteitskrant. Na mijn studie wist ik dat ik journalist wilde worden, maar ik kreeg het advies om eerst te promoveren. In het sollicitatiegesprek daarvoor werd mij door Hendrik Lenstra gevraagd: 'Wat wil je nu worden? Journalist of wiskundige?' Toen antwoordde ik 'journalist'. Ik dacht eerst: 'shit, verkeerde antwoord', maar later bleek dat juist een aanbeveling. Ze dichtten me niet het talent toe om een onderzoekswiskundige te worden, maar vonden het wel heel erg de moeite waard om een toekomstige journalist een gedegen promotieonderzoek te laten doen. Het was het begin van alles, want daar kwam ik Jeanine Daems tegen, met wie ik de Wiskundemeisjes ben gaan vormen."

Wat was de aanleiding voor die blog?
"Frustratie. Ik vond een geweldige blog over natuurkunde (qulog.wordpress.com), maar er was nergens een goede blog over wiskunde, ook niet in het Engels. Toen zei Jeanine: moeten wij dat dan niet beginnen? Dat hebben we gedaan."

Heb je het gevoel dat er ergens tussen dat meisje dat de zwembadtemperatuur opmerkte en de vrouw die nu hoogleraar is geworden, een keerpunt heeft plaatsgevonden?
"Niet echt. Ik heb een rare vorm van carrièreplanning; voor mijn gevoel doe ik altijd maar wat. Maar ik heb altijd in mijn achterhoofd een paar dingen gehad die ik het liefst zou willen. Ik wilde graag een column over wiskunde — zodat ik geen nieuwsberichten hoefde te maken over Chinese pubers die ingewikkelde problemen oplossen. Daar kwam ik achter door die blog. En achteraf kan ik dingen aanwijzen die later van belang zijn geweest voor mijn werk. Tijdens mijn studie schreef ik brochureteksten voor een theater. De vaardigheid om mensen de zaal in te schrijven, bleek later heel handig om wetenschap te verkopen."

Mis je de Wiskundemeisjes nog wel eens?
"Ja. Het gevoel dat je met z'n tweeën tegen de wereld was, en vooral ook de gemeenschap die de blog volgde. Bij een krantencolumn krijg je vaak vragen die je zelf moet beantwoorden. Maar als er een vraag op ons blog kwam, was er vaak al iemand anders die een vriendelijk antwoord gegeven had voor we er zelf over na konden denken. Nu nog steeds kom ik in mijn werk mensen tegen waarvan blijkt dat ze jaren geleden al onderdeel van die gemoedelijke bloggemeenschap waren. Het lijkt me best leuk om het nog eens voor een afgebakende periode, een jaar bijvoorbeeld, te doen."

Jullie motief was destijds om mensen voor de wiskunde te motiveren. Is dat in grote lijnen ook je doel als hoogleraar?
"Ik ben natuurlijk geen cheerleader van de wetenschap. En wetenschapscommunicatie als leerstoel is iets anders dan wetenschapscommunicatie doen. Dat maakt het vakgebied soms wel een beetje raar. Het doel is vooral om onderwijs te geven en onderzoek te doen naar hoe wetenschapscommunicatie werkt. Veel wetenschapscommunicatie is nu ad hoc. Dat is tof, maar daardoor gaan ook dingen mis. Ik wil mensen op die fouten wijzen, vanuit mijn eigen ervaring."

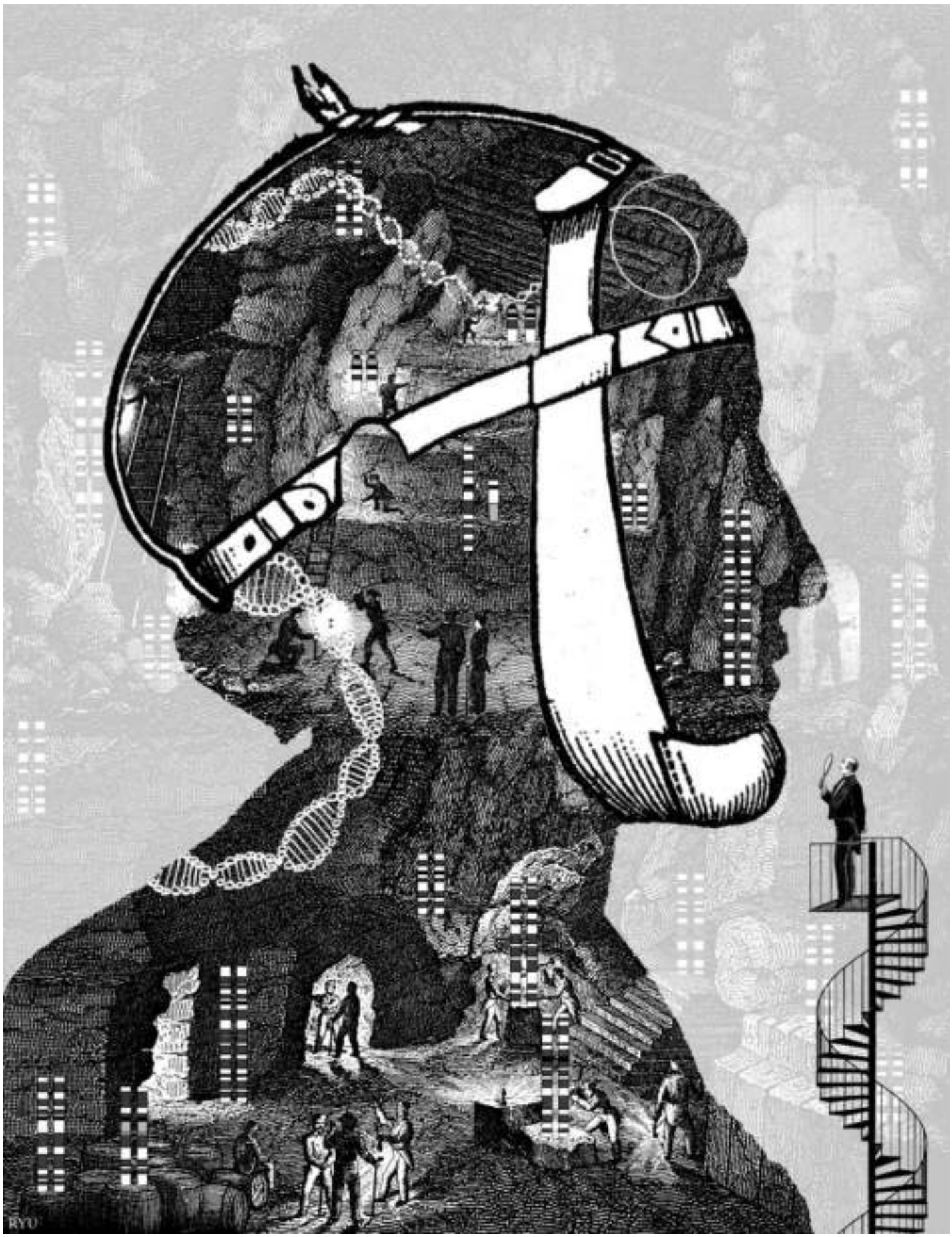
Je eigen promotieonderzoek betrof keiharde wiskunde over kettingbreuken. Nu verdiep je je in wetenschap over wetenschap. Verlang je nooit terug naar die wiskunde?
"Nee. Het rare is: juist sinds ik de harde wiskunde verlaten heb, ben ik veel meer bezig met wiskunde. Ik vind mijn werk nu duizend keer leuker. En vaak ook moeilijker. Hier is trouwens toch echt sprake van een keerpunt: wetenschapscommunicatie is iets wezenlijk anders dan wiskunde.

Wat me altijd opvalt als ik op congressen kom, is dat ik van bijna alle vakgebieden de praatjes kan volgen, maar dat ik bij een praatje over wiskunde vaak afhaak omdat het verhaal te specialistisch is. Er valt nog zoveel te winnen voor wiskundig Nederland. Neem nu het *Nieuw Archief voor Wiskunde*, waarin dit interview staat. Dat is als blad heel erg gericht op onderzoekswiskundigen, net als het Nederlands Mathematisch Congres. Maar het is toch heel tragisch dat al die mensen die ik voor deze rubriek gesproken heb, zich nog wel wiskundige voelen, maar er niet meer bij horen. Het zou toch veel leuker zijn als er een beroepsvereniging was waar alle wiskundigen zich thuis voelden? Dan kun je ook veel meer bereiken, omdat je mensen op alle posities van de samenleving hebt, die op willen komen voor wiskunde."

Herkennen we daar een speerpunt van de verse hoogleraar?
"Ik zou graag de instituten bij elkaar willen brengen. Met name ook om af te komen van het imago dat communicatie iets is wat je gaat doen als je niet goed genoeg bent voor onderzoek. Als ik vroeger zelf had kunnen kiezen voor een communicatiemaster, had ik dat niet gedaan, want dat was voor 'de sukkels'. Maar dat is een belachelijke gedachte. Het is echt een denkfout aan de universiteit dat alleen wetenschappelijk onderzoek intellectueel bevredigend zou zijn."

Denk je dat je ooit nog iets heel anders gaat doen? Zien we je ooit nog terug in deze keerpuntrubriek met een echt rigoureuze ommezwaai?
"Hoogleraar was altijd mijn droombaan. Dat ben ik nu, dus eerst wil ik dit goed doen. Voorlopig zit ik hier goed dus."

Deze rubriek wordt voortgezet door Jan Beuving. Goede suggesties voor een Nederlandse wiskundige met een keerpunt in zijn of haar carrière zijn welkom via *keerpunt@nieuwarchief.nl*.



Illustratie: Ryu Tajiri

Mark van de Wiel
Afdeling Wiskunde, Vrije Universiteit Amsterdam, en
Afdeling Epidemiologie en Biostatistiek, VUmc, Amsterdam
mark.vdwiel@vumc.nl

Onderzoek

Statistiek op het genoom: 'Big Data', maar dan anders

In het voorjaar van 2014 werd Mark van de Wiel benoemd tot hoogleraar statistiek voor genoomanalyse aan de Faculteit Exacte Wetenschappen van de Vrije Universiteit en aan het VU Medisch Centrum. Dit artikel is gedeeltelijk gebaseerd op zijn oratie gehouden op 23 mei 2014. Hij beoogt duidelijk te maken waar de statistiek bijdraagt aan de analyse van Big Data, en dan in het bijzonder genomische data.

Big Data: Big p or Big n?

Big Data is overal tegenwoordig. De Googles en Facebooks van onze maatschappij verzamelen enorme hoeveelheden data op een automatische manier, onze vele mobiele telefoons leveren locatie-informatie voor bijvoorbeeld file-voorspellingen, en in een medische setting geven MRI-beelden en metingen aan het genoom gigantische aantallen datapunten per individu. Maar wat is 'Big'? Hiertoe moeten we eerst onderscheid maken tussen n: het aantal individuen en p: het aantal variabelen. En het is juist dat onderscheid dat zo essentieel is voor de statistische analyse van Big Data: als n groot is en veel groter dan p ($n >> p$) dan werkt traditionele statistiek goed, zelfs heel goed, vooral voor de vele asymptotische benaderingen die de statistiek rijk is. Wellicht de grootste uitdaging voor dit geval is het bestuderen van complexe, (niet-lineaire) modellen, omdat de grote n daartoe de ruimte geeft. Als echter $p >> n$ spreken we van hoog-dimensionale data, en stapelen de problemen zich op, omdat de traditionele statistiek niet meer werkt. Ik geef hier verderop enkele voorbeelden van. Genomische data zijn vrijwel altijd hoog-dimensionaal: niet zelden is p van de orde 10^5 terwijl $n \approx 100$. De oorzaak is simpel: nieuwe technologieën stellen moleculair biologen in staat met het grootste gemak veel stukjes van het genoom parallel te meten. Echter, zeker in klinische omgevingen, is het lastig en duur om weefsel te verkrijgen van grote aantallen (zieke) individuen. Ik vervolg met een voorbeeld van (al dan niet opzettelijke) verwarring tussen p en n in genomisch onderzoek.

Big p, Big Errors (p is geen n)

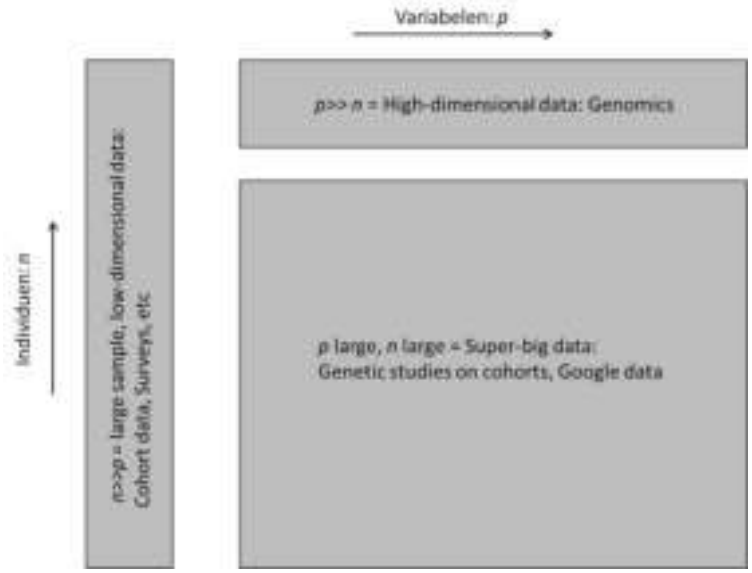
Er wordt wel gezegd: "There are lies, damned lies and statistics" [8]. Ik zou willen zeggen "wrong statistics". Met goede statistiek lieg je niet, sterker nog: met goede statistiek achterhaal je leugens, of op zijn minst foute conclusies. Een beroemd historisch voorbeeld in de genetica is het experiment van Mendel met de erwtenplanten [4]. Op dit experiment heeft hij zijn erfelijkheidswetten gebaseerd. Statisticus én geneticus Fisher gebruikte Pearsons chi-kwadraattoets om aan te tonen dat Mendels resultaten te goed om waar te zijn waren [2]. Wetenschappers zijn dus net mensen, en zelfs briljante wetenschappers als Mendel gaan weleens de mist in. Wanneer het genomisch onderzoek betreft, denk ik dat wetenschappelijke nalaatbaarheid een veel negatievere invloed heeft op de kwaliteit dan moedwillige fraude. Te vaak publiceren absolute toptijdschriften genomisch onderzoek waarvan de statistiek gewoon slecht is.

Een voorbeeld. Methylatie is een belangrijk proces op het DNA waarmee genen kunnen worden uitgeschakeld. Een artikel in *Nature* uit 2011 [3] beschrijft methylatie van gebieden op het genoom in stamcellen en normale cellen. De auteurs rapporteren vele gebieden die verschillend gemethyleerd zijn tussen deze twee types cellen. Prachtige plaatjes, krachtige conclusies. Het artikel is dan ook al meer dan 600 keer geciteerd. Ik was geïnteresseerd, want ik dacht dat men maar liefst tien biologische herhalingen had per conditie. Statistici zijn dol op herhalingen en tien is veel voor de dure techniek die werd gebruikt, 'bisulphite sequencing'.

Echter wat bleek: de herhalingen waren fictief. Men had het genoom in stukjes opgehakt en tien achtereenvolgende stukjes als onafhankelijke herhalingen beschouwd. Dus gedaan alsof dát de echte biologische variatie representeert. Dat is zoiets als zeggen dat de variatie tussen zoogdieren bestudeerd kan worden door de variatie tussen mensen te beschouwen. Voor het artikel leidde de sterk onderschatte variatie natuurlijk tot prachtige p-waardes. Deze waren blijkbaar klein genoeg om de referenten te overtuigen. Ik heb het aangekaart bij *Nature*; men vond het niet urgent genoeg voor publicatie van het commentaar, maar ik mocht het commentaar wel online plaatsen [6].

Big p, Big Problems

Waarin wijkt de statistiek voor $p >> n$ nu af van de traditionele statistiek? Simpel gezegd: in alle gevallen moet men een oplossing vinden voor een slecht-geconditioneerd probleem vanwege de 'curse of dimensionality'. Vergelijk het met het fitten van een



Figuur 1 Verschillende soorten 'Big Data'.

100ste-graads polynoom aan tien datapunten: daar zijn oneindig veel oplossingen voor. Ik noem een aantal gebieden waarop de laatste tien jaar veel onderzoek is gedaan.

Multiple testing

Hoe beperk je het aantal fout-positieven wanneer je meerdere hypotheses toetst? Bijvoorbeeld wanneer we willen weten welk gen zich anders gedraagt in een zieke groep dan in een gezonde groep individuen. Feitelijk een 'oud probleem', omdat ook al voor het $p >> n$-tijdperk men vaak meerdere statistische hypotheses toetste. De Bonferroni-regel [1] schrijft dan voor dat wanneer je de kans op minimaal één fout-positieve bevinding lager wilt hebben dan α, je simpelweg alleen variabelen (bijvoorbeeld genen) met een p-waarde kleiner dan α/p positief verklaart. Zie hier het dimensie-probleem: hoe groter p, des te kleiner de drempelwaarde. Veel onderzoek op het gebied van multiple testing richt zich óf op (a) andere, liberalere fout-positief criteria die meer meeschalen met p, zoals False Discovery Rate; óf op (b) het formuleren van geneste hypotheses en het dan gebruikmaken van de ontstane verbanden tussen de hypotheses om krachtigere procedures te ontwikkelen.

Geregulariseerde multiple regressie

Een klassiek probleem in de statistiek is het fitten van een lineaire multiple regressie met behulp van de kleinste-kwadratenmethode:

$$b = \text{argmin}_\beta \|Y - \mathbf{X}\beta\|_2,$$

waarbij Y de $n \times 1$-responsvector is (voor elk individu één respons), X de $n \times p$-matrix met verklarende variabelen (p voor elk individu) en β de $p \times 1$-vector met coëfficiënten. In een medische setting kan men bijvoorbeeld hiermee proberen de 'body mass index' (Y) te verklaren vanuit genetische factoren ($\mathbf{X}$). De oplossing is eenvoudig:

$$b = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Y.$$

Echter, als $p > n$ dan is $\mathbf{X}^T\mathbf{X}$ rang-deficiënt en dus niet inverteerbaar. Dit wordt opgelost door het minimalisatie probleem te regulariseren:

$$b = \text{argmin}_\beta \{\|Y - \mathbf{X}\beta\|_2 + \lambda\|\beta\|_q\},$$

met tuning parameter $\lambda > 0$ en veelgebruikte normen $q = 1$ (lasso-regressie) en $q = 2$ (ridge-regressie). Voor $q = 2$ kunnen we de oplossing direct opschrijven:

$$b = (\mathbf{X}^T\mathbf{X} + \lambda I)^{-1}\mathbf{X}^T Y,$$

waarbij I de $p \times p$-identiteitsmatrix is. Het inverteren van de grote $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ matrix kunnen we efficiënt doen door singuliere-waardenontbinding van $\mathbf{X}$. Hoewel ridge-regressie vaak goede predictie-eigenschappen heeft (voorspel Y_{new} met $\mathbf{X}_{\text{new}}b$), selecteert het geen variabelen. Lasso-regressie doet dat wel: wanneer $q = 1$ zijn maximaal $k \leq n$ componenten van de p-dimensionale vector b ongelijk aan 0. Dit is prettig in de praktijk: een volgende keer hoeft de bioloog slechts k genen te meten. Onder een zogenaamde spaarzaamheids-aanname (het aantal echte niet-nulcoëfficiënten is klein), heeft de lasso goede asymptotische eigenschappen. Veel mathematisch statistisch werk richt zich op het verfijnen van deze resultaten voor (variaties op de) lasso. De lasso voorspelt echter niet zo goed als de ridge. Recentelijk is er veel aandacht voor een tweetal oplossingen: (i) hybriden tussen lasso en ridge, zoals de combinatie van de twee: het elastische net; of (ii) vanuit een ander perspectief: na ridge-regressie 'thresholding'-methodes toepassen om kleine coëfficiënten op nul te zetten zonder veel verlies van voorspellende waarde.

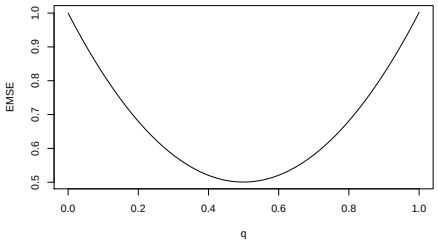
Moleculaire netwerken

Genen werken niet alleen, maar interacteren. Met elkaar, maar ook met andere moleculen zoals eiwitten. Wanneer we data hebben, kunnen we correlaties tussen genen berekenen om een idee te krijgen over welke interacties er mogelijk zijn. Dit is echter te naïef, omdat veel correlaties tussen twee genen veroorzaakt worden door interacties van beide genen met een of meerdere andere genen. Daarom is het beter om partiële correlaties te bestuderen, welke in het populaire Gaussische grafische model direct relateren aan de elementen van Σ^{-1} : de inverse correlatiematrix. Veel statistisch onderzoek op dit vlak richt zich dan ook op het schatten van Σ , een gruwelijk probleem, omdat het aantal onbekenden hierbij van de orde p^2 is. Weer spelen regularisatie-technieken zoals hierboven besproken een belangrijke rol. Het schatten van het netwerk is nu echter niet het eindpunt. Het geschatte netwerk bevat niet zelden honderden of duizenden verbindingen, wat tot een typisch 'haarbal'-plaatje leidt dat een bioloog natuurlijk onmogelijk kan interpreteren. Daarom richt men zich nu vaak op technieken om zulke netwerken samen te vatten, interessante kleine sub-modules te onttrekken, en ook om verschillen tussen netwerken (bijvoorbeeld gezond versus ziekte) te vinden.

Big p, Big Blessing

De grote p is niet altijd een vloek. Vaak is er veel structuur bekend over de p variabelen, in het bijzonder in genomische studies. Genen organiseren zich in padwegen, hetgeen modules zijn met één bepaalde biologische functie (bijvoorbeeld celdeling). Vaak zijn biologen geïnteresseerd of zo'n verzameling genen zich verschillend gedraagt onder twee of meer condities. Hiervoor zijn genset-toetsen ontwikkeld die in de regel meer onderscheidingsvermogen hebben dan toetsen op afzonderlijke genen. Zo kunnen we dus profiteren van de grote p. Een ander principe dat gebruik maakt van de grote p is 'Empirical Bayes', dat zich tevens goed leent voor het systematisch gebruikmaken van voorkennis.Het mooie van genomisch onderzoek is dat veel voorkennis opgedaan kan worden uit publiek beschikbare databases.

In het echte leven maken we continu gebruik van voorkennis om inschattingen te



Figuur 2 Expected mean square error (EMSE) als functie van q voor de schatter $m_i(q) = (1-q)X_i + q \cdot m$, wanneer $p = 1{,}000$.

doen. Een voorbeeld uit het casino, een plek waar een statisticus zich natuurlijk thuis voelt. Als iemand veel geld heeft verloren met een nieuw spel in het casino, en ik schat in dat die persoon redelijk intelligent is, dan speel ik liever het spel helemaal niet, dan dat ik eerst duizenden euro's uitgeef aan het spel om vervolgens waarschijnlijk te concluderen dat het voor mij ook niet werkt. Bayesiaanse methoden zijn bij uitstek geschikt hiervoor: de externe informatie kan verpakt worden in de zogenaamde prior, of à priori verdeling.

Kritiek uit de niet-Bayesiaanse hoek is dat in kleine, laag-dimensionale studies de keuze van de à priori verdeling subjectief is. Welnu, dat is het mooie aan hoog-dimensionale data: we hebben zoveel data van soortgelijke entiteiten, bijvoorbeeld genen, dat het mogelijk is om de à priori verdeling te schatten. Dat noemen we 'Empirical Bayes'. De te gebruiken bronnen van externe informatie zijn goeddeels keuzes, maar de relevantie van die bronnen wordt bepaald door de data zelf.

Laat ik even teruggaan naar het casinovoorbeeld, en ik neem aan dat het nieuwe spel een behendigheidsspel is, zoals bijvoorbeeld Black Jack. Stel nu dat niet alleen ik voor de keuze sta of en hoe vaak ik het spel ga spelen, maar u, de lezers, ook allemaal. Bovendien hebben we allemaal een ander, soortgelijk spel al heel vaak gespeeld. En helaas: we verloren daar gemiddeld gezien wat.

Nu zouden wij gelijk kunnen besluiten dit spel ook maar niet te spelen. Maar een slimmere strategie is de volgende: wij spelen allemaal het spel een paar keer met geringe inzet, waarna ieder voor zich gaat beslissen om door te gaan of niet. Waren we toen we dat andere spel speelden allemaal dronken, dan zullen we het nu met dit spel een stuk beter doen, hoop ik. En als 'dronken zijn' redelijkerwijs impliceert dat wij met dat vorige spel zomaar wat deden, dan zullen de data uitwijzen dat de correlatie tussen onze prestaties bij het nieuwe en oude spel zo goed als afwezig is. In dit geval vertellen de data ons dat we weinig gewicht moeten geven aan de externe informatie, hier onze prestaties bij het oude spel. Dan moet ieder voor zich een beslissing nemen op basis van de kleine hoeveelheid gegevens voor het nieuwe spel.

Het wordt echter interessanter als wij bij het vorige spel nuchter waren. En helemaal als de data van beide spellen ons dan vertellen dat er wél een behoorlijke correlatie is tussen de prestaties. Blijkbaar betreft het dan soortgelijke vaardigheden die nodig zijn om beide spellen relatief goed te kunnen spelen. Let wel: die correlatie, of algemener de à priori verdeling van een parameter die beschrijft in hoeverre de prestaties op elkaar lijken, kunnen we behoorlijk accuraat schatten (met Empirical Bayes technieken), want u bent (hopelijk) met velen. We kunnen dan veel meer gewicht toekennen aan de externe informatie. Juist omdat voor ieder van ons de data van het nieuwe spel van zeer beperkte omvang zijn, kan dit enorm helpen om betere beslissingen te nemen dan wanneer we deze informatie negeren. Niet voor iedereen zal de beslissing beter zijn, maar gemiddeld gezien vaak wel. En wellicht maken we als groep zelfs winst. Als we nu vooraf afspreken dat we die winst delen, dan is iedereen blij.

Een klein, iets meer wiskundig voorbeeld uit de genomics is het volgende. We veronderstellen een simpel model waarbij de (log)-genexpressie (= maat voor aantal kopieën van dat gen) voor gen i, X_i , een normale verdeling volgt met verwachtingswaarde μ_i en standaarddeviatie (sd) = 1. We meten echter vele genen en we veronderstellen dat μ_j een à priori normale verdeling volgt met verwachtingswaarde 0 en sd = 1 voor alle genen $j = 1, \ldots, p$. We willen μ_i schatten. De kwaliteit van een schatter m_i kwantificeren we met de zogenaamde Expected Mean Square Error (EMSE), de verwachtingswaarde onder de à priori verdeling van de MSE gedefinieerd als:

$$\text{MSE}(m_i) = [\text{Bias}(m_i)]^2 + \text{Variantie}(m_i).$$

Hier is Bias de verwachte afwijking van de schatter, m_i, van de waarheid, μ_i. Een simpele, zuivere (unbiased) schatter is $m_{i,1} = X_i$, met $\text{EMSE}(m_{i,1}) = 1$. Een mogelijke Empirical Bayes schatter (of krimp-schatter) is $m_{i,2} = X_i/2 + m/2$, waarbij m het gemiddelde van alle andere $p-1$ genexpressiemetingen is (dus we proberen te 'leren' van de andere genen). De schatter $m_{i,2}$ is niet zuiver, maar heeft wel een veel kleinere variantie. We kunnen de EMSE berekenen: $\text{EMSE}(m_{i,2}) = [1 + 1/(p-1)]/2$, welke dus bijna een factor 2 kleiner is dan die van $m_{i,1}$. Figuur 2 laat zien dat dit ook (nagenoeg) het minimum is. Dus ja, we kunnen echt leren van de andere genen. Let wel: als μ_i sterk verschilt van de à priori verwachtingswaarde van μ_j, $j \neq i$, dan kan de krimpschatter ook slechter zijn dan de simpele schatter. Voorzichtigheid is dus geboden als je vermoedt dat gen i een heel afwijkend gen is.

Big p, Big Business

Tot slot: Big Data, en dus ook hoog-dimensionale data, is Big Business. Dat is op zich goed nieuws voor de statistiek, welke niet voor niets door een hooggeplaatste Google-econoom de "sexiest job of the 21st century" [7] werd genoemd. Geen ander 'data-vakgebied' (bioinformatica, data science, et cetera) heeft zijn wortels zo goed verankerd in de wiskunde, waarmee we de mogelijkheden, maar juist ook de onmogelijkheden van veel hoog-dimensionale problemen kunnen aantonen. Dat laatste maakt ons gemiddeld pessimistischer dan andere data-broeders, en daarom ook minder populair bij (veel) biologen. Gelukkig is er hoop, in ieder geval in het genomisch onderzoek. Het concept van reproduceerbaarheid op onafhankelijke datasets is inmiddels behoorlijk goed verankerd in dit onderzoek: het is moeilijk om een mooie bevinding te publiceren zonder onafhankelijke validatie. Daarom moeten we ons wapenen tegen overoptimisme bij het bestuderen van de initiële dataset. En laat statistici daar nu juist goed in zijn, een kwaliteit die we moeten propageren bij andere wetenschappers, wellicht beter dan we tot nog toe doen.

Referenties

1 C.E. Bonferroni, Teoria statistica delle classi e calcolo delle probabilità, *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 8 (1936), 1–62.

2 R.A. Fisher, Has Mendel's work been rediscovered? *Ann. of Sci.* 1 (1936), 115–137.

3 R. Lister et al., Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells, Nature 471 (2011), 68–73.

4 G. Mendel, Experiments in Plant Hybridization (1865), beschikbaar via www.mendelweb.org/Mendel.html.

5 M.A. van de Wiel, Fictive replicates in epigenomic analysis (2014), doi:10.1038/nature09798.

6 M.A. van de Wiel, 'Het zit in de genen, toch?', inaugurele rede Vrije Universiteit, 23 mei 2014, www.few.vu.nl/~mavdwiel.

7 Citaat van Hal Varian, hoofdeconoom bij Google. Het exacte citaat varieert enigszins in de media.

8 Origineel citaat: "There are three kinds of lies: lies, damned lies, and statistics." Het citaat wordt vaak onterecht toegeschreven aan Mark Twain, maar de echte bron is onduidelijk. en.wikipedia.org/wiki/Lies,_damned_lies,_and_statistics.

Domenico Lahaye
Faculty of EEMCS
TU Delft
d.j.p.lahaye@tudelft.nl

Kees Vuik
Faculty of EEMCS
TU Delft
c.vuik@tudelft.nl

Research

Computational challenges in electrical power networks

Networks for the high-voltage distribution of electrical energy are currently undergoing far-reaching developments. National power grids are evolving from static entities, producing mainly a uni-directional flow from generation to loads, to more dynamic and decentralized structures. These emerging power systems should accommodate the local generation by renewable sources and peak demands of electrical vehicle charging. The cross-border interconnection of power grids further imposes new challenges in the design, planning and daily operation of these networks. In this paper Domenico Lahaye and Kees Vuik describe recent work by the chair of numerical analysis at the TU Delft on the numerical simulation of electrical power networks.

This work is intended to guide the transition towards nextgeneration of the power systems. We subdivided this paper into three parts.

We first describe the development of a dedicated load flow solver for the state in a network or in a collection of closely resembling networks. The networks we consider model the interconnection of a set of loads and generators through power lines. The term *state* refers to the set of voltage phasors associated with each node of the network. The ability to compute efficiently the nodal voltages is fundamental in any modeling and optimization of the power network. Our algorithm allows to solve large scale problems [3] faster than more commonly used techniques.

We subsequently outline a screening technique we developed to monitor the power system subject to critical events. We analyze how the system state alters after the removal of major interconnections and identify the need for measures to bring the network back to a save operation point. These measures include the redistribution of generation and the removal of loads in more severe cases. Unlike more commonly used techniques, our method accounts for both overloads of the interconnecting lines and the violation of the nodal voltage lower and upper bounds [4].

In the final part of the paper we describe a novel approach to compute transients caused by switching actions. Unlike other approaches in literature, our approach avoids the reassembly of the state-space representation at each switching event.

Noot van de redactie
Dit artikel is geschreven als bijdrage voor het themanummer over Netwerken, maar door een samenloop van toevalligheden heeft het de eindredactie helaas net te laat bereikt. Daarom publiceren we het alsnog in het voorliggende nummer.

Newton–Krylov load flow solver

Let us consider a network consisting of n nodes and m edges. Each of the nodes represents either a generator or a load. In a generator and load node electrical power is produced and consumed, respectively. An edge represents a transmission line that transports power from the generators to the loads. The electrical power can be expressed in terms of the voltage at the nodes and the current flowing through the edges. We consider a formulation in which the electromagnetic quantities are assumed to vary sinusoidally at a frequency of $50\,\mathrm{Hz}$. Ohm's Law then allows to eliminate the current in such a way that the power can be expressed in terms of solely the complex-valued voltage phasors.

Let $\mathbf{x}$ denote the n-dimensional vector that holds the voltage phasor at each node. The constraint that the total amount of power consumed should balance the total amount of power generated can be expressed as a system of n non-linear algebraic equations that we can express as

$$F(\mathbf{x}) = \mathbf{0} \in \mathbb{C}^n\,. \tag{1}$$

Let $J_i(\mathbf{x}_i)$ denote the Jacobian of F with respect to $\mathbf{x}_i$ at the i-th Newton iteration. Solving the system (1) using Newton's method re-

quires solving a sequence of linear systems of the form

$$J_i(\mathbf{x}_i)\,\mathbf{s}_i = -F(\mathbf{x}_i) \qquad (2)$$

for the step length $\mathbf{s}_i$. In literature this linear system is typically solved by a direct solution method.

Currently, however, there exists a pressing need to compute the power flow in networks that are large in size. Such networks originate for instance in the modeling of the interconnection of the network in Europe with that in Russia or in Northern Africa. It is well known that direct solution methods are not suited for such large scale problems [5]. We therefore developed a new generation of load flow solvers in which the LU-factorization of Jacobian $J_i(\mathbf{x_i})$ is replaced by a preconditioned Krylov subspace iteration. The term Newton–Krylov for these new solvers derives from the fact the accuracy of the inner linear solver is linked to the outer non-linear iteration residual.

Numerical results confirm that the Newton–Krylov solver is computationally more efficient than previously existing direct solution methods. The observed linear convergence of the outer GMRES iteration resulted in a convergence theory that provides insight in how to set the stopping criterion for the Krylov iteration. We also demonstrated that the flexibility of Newton–Krylov methods allows to efficiently solve a collection of closely resembling network models.

Monte Carlo security assessment

We extended previous work to the security assessment of power networks [4]. The network is said to be in a secure operating mode if the generator and load settings are such that the voltage at the nodes and the currents through the edges are within a priori defined ranges. In a security assessment one seeks to identify perturbations in the operating mode that bring the network away from being secure. Such perturbations might include changes in the generators and loads as well as changes in the topology of the network. In a subsequent stage one seeks actions that brings the network back to a secure state. Such actions include the redistribution of the total amount of power generated across the different generators, the installation of new interconnections and the removal of part of the load.

In our approach, the generated power and demanded load are stochastic variables. The load flow equations are solved within a Monte Carlo procedure. For each sample the safety of the operating point is evaluated after solving the load flow equations (1). Subsequently a tentative least invasive remedial action that brings the network operating back into save operation is proposed. The load flow equation are again solved for to validate action proposed. The procedure is repeated until the network is brought back to save operation or until the network is deemed to be collapsed. Data on the safety of the power network is gathered in the post-processing stage of the Monte Carlo simulation.

Switching induced transients

In a parallel project [8], we studied time integration methods to compute transients caused by switching actions in power systems [1, 7]. We investigate Runge–Kutta methods with adaptive time step to solve

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t) \quad \text{given } \mathbf{x}(t=0) = \mathbf{x}_0\,, \qquad (3)$$

for the current through the inductors and the voltage across the capacitors. A switching action induces a change of topology of the network. Existing time integration methods require the reconstruction of the state-space matrices after each switching event. We developed a new modeling method that avoids this need to reassemble the matrices. Computation time is therefore saved. We are currently looking into the possibility to replace the direct linear system solve at each time step by an approximate iterative solver.

Conclusions

In this paper we described recent work by the chair in numerical analysis of the TU Delft on the modeling of electrical power networks. We briefly outlined the development of a fast and versatile load flow solver, of a Monte Carlo based simulation tool for the security assessment of a power system and of a new modeling approach for the switching-induced transient.

References

1 W. Hundsdorfer and J.G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Springer, 2010.

2 R. Idema and D. Lahaye, *Computational Methods in Power System Analysis*, Atlantis Press, 2014.

3 R. Idema, G. Papaefthymiou, D. Lahaye, C. Vuik and L. van der Sluis, Towards Faster Solution of Large Power Flow Problems, *IEEE Transactions on Power Systems* 28(4) (2013), 4918–4925.

4 M. de Jong, G. Papaefthymiou, D. Lahaye, C. Vuik and L. van der Sluis, Impact of correlated infeeds on risk-based power system security assessment, *Power Systems Computation Conference (PSCC), Wroclaw, Poland, 2014*, doi 10.1109/PSCC.2014.7038439.

5 Y. Saad, *Iterative Methods for Sparse Linear Systems*, Second Edition, SIAM, 2003.

6 P. Schavemaker and L. van der Sluis, *Electrical Power System Essentials*, Wiley, 2008.

7 L. van der Sluis, *Transients in Power Systems*, Wiley, 2001.

8 R. Thomas, D. Lahaye, C. Vuik and L. Van der Sluis, *Transients in Power Systems: a Literature Survey*, TU Delft, 2013.

Margot Gerritsen
Computational and Mathematical Engineering
Stanford University, CA, USA
gerritsen@stanford.edu

Ying Wang
Computational and Mathematical Engineering
Stanford University, CA, USA
yw1984@stanford.edu

Farnaz Ronaghi
Management Science and Engineering
Stanford University, CA, USA
farnaz@stanford.edu

David F. Gleich
Informatics
Sandia National Labs, Livermore, CA, USA
dfgleic@sandia.gov

Xiangrui Meng
Computational and Mathematical Engineering
Stanford University, CA, USA
mengxr@stanford.edu

Amin Saberi
Management Science and Engineering
Stanford University, CA, USA
saberi@stanford.edu

Onderzoek

Licht in de digitale duisternis dankzij computertools voor digitaal beheer

Computertools zijn niet meer weg te denken bij tegenwoordige zoek- en aanbevelingstechnologieën. Moderne digitale archieven bestaan echter uit ongekend gevarieerde collecties van gedigitaliseerd materiaal en zogenaamde *born-digital content*. Het is nog altijd lastig om interessant materiaal in deze archieven op te zoeken. Vaak ontbreken hierin annotaties — of metagegevens — op basis waarvan mensen het interessantste materiaal kunnen vinden. David F. Gleich, Ying Wang, Xiangrui Meng, Farnaz Ronaghi, Margot Gerritsen en Amin Saberi van Computational Approaches to Digital Stewardship (CADS) werken aan een visie op een virtueel bibliotheeksysteem waarmee makkelijker de interessantste parels kunnen worden gevonden in gevarieerde collecties van digitale archieven. Zij beschrijven vier computertools die zij hebben ontwikkeld zodat digitale archieven beter kunnen worden verwerkt en onderhouden. De eerste tool is een verbeterd algoritme voor de indeling van grafen met honderdduizenden knooppunten. De tweede tool is een nieuw algoritme voor het afstemmen van databases met koppelingen tussen de objecten, ook bekend als netwerkalignmentprobleem. De derde tool is een heuristische optimalisatiemethode waarmee een reeks geografische verwijzingen in een boek worden gedesambigueerd. En de vierde tool is een techniek waarmee automatisch een titel wordt gegenereerd op basis van een beschrijving.

In de afgelopen 25 jaar is het karakter van documenten in onze samenleving veranderd. Voorheen werden documenten op papier of op een ander fysiek medium opgeslagen. Tegenwoordig worden onze documenten opgeslagen in digitale bestanden. Deze situatie stelt ons voor een subtiel probleem. Bedenk eens hoeveel van uw eigen — digitaal opgeslagen — werk niet langer toegankelijk is omdat:

– het programma waarin het bestand moet worden gelezen, niet meer beschikbaar is;
– het programma waarin het bestand moet worden gelezen, niet meer werkt met oude bestanden;
– er geen hardware meer bestaat om de fysieke media te lezen.

Kuny [30] zet de basis voor het probleem uiteen en bedacht de uitdrukking *een digitale duisternis* om de ernst van de situatie duidelijk te maken. Ook beschrijft hij enkele oplossingen die nodig zijn om dit aan te pakken. Deze ideeën zijn grotendeels gericht op het probleem om digitale bits, opslag en bestandsindelingen te behouden. Zo heeft Kuny als interessante uitdaging vastgesteld dat digitale opslag een openbaar goed moet worden. We zijn afhankelijk van historische documenten uit het verleden om het heden te informeren. Daarom moeten onze documenten voor dit doel behouden blijven. Het probleem met het bewaren van documenten is dat dit alleen nut heeft wanneer de informatie door iemand wordt gebruikt. Voor de meest succesvolle opslagactiviteiten moeten de gegevens dus beschikbaar en eenvoudig toegankelijk worden gemaakt.

Uitdagingen in digitale webarchieven

Alleen al het bieden van toegang tot de gegevens zorgt voor de nodige uitdagingen. Van oudsher was materiaal opgeslagen in een bibliotheek en gingen wetenschappers *naar* de bibliotheek om dit in te kijken. Eenmaal daar overlegden ze met archivarissen om te bepalen welk materiaal ze precies nodig hadden. Tegenwoordig verwachten gebruikers toegang vanaf elk apparaat met een internetverbinding. Eigenlijk — en misschien vooral als reactie op de efficiënte zoekmachine van Google — verwachten we een direct antwoord op onze slecht geformuleerde informatieverzoeken. Het probleem met een dergelijke werkwijze in deze digitale collecties is dat gebruikers vaak iets willen ontdekken in plaats van opzoeken. Met andere woorden, ze willen niet met systemen iets zoeken wat ze al weten, maar iets nieuws vinden wat ze interessant vinden. Zo zou het volgende gesprek in een bibliotheek kunnen hebben plaatsgevonden:

Bibliothecaris Kan ik u ergens mee helpen?
Bezoeker Ik ben onlangs vanuit Zweden hierheen verhuisd. Hebt u ook een goed boek over lokale geschiedenis?
Bibliothecaris Oh, veel van onze eerste immigranten kwamen uit Zweden. Ik heb precies het juiste boek voor u.

Onze hoop is dat we dergelijke hulp in een digitaal archief kunnen bieden. Laten we eens bekijken hoe dit scenario online zou kunnen verlopen om te begrijpen welke uitdagingen zich voordoen bij het bieden van toegang tot digitale archieven.

Gebruiker Voer een zoekopdracht voor 'lokale geschiedenis' in.
Systeem Geef een ranglijst met antwoorden weer om aan te geven wat de beste naslagwerken zijn voor informatie over lokale geschiedenis; samen met een lijst met belangrijke subonderwerpen, zoals Zweedse immigranten.
Gebruiker Klik op de lijst met subonderwerpen over Zweedse immigranten.
Systeem Geef een nieuwe ranglijst met antwoorden weer, waarvan er één is gemarkeerd als 'speciaal belichte selectie'.

Bedenk welke technologieën nodig zijn voor deze interactie. Ten eerste moet een dergelijk systeem weten dat de zoekopdracht 'lokale geschiedenis' verwijst naar de geschiedenis van de regio waar de zoeker zich bevindt, of dat deze een specifieke lokale geschiedenis impliceert. Ten tweede moet de zoekmachine in staat zijn om te zoeken naar het onderwerp of trefwoorden die verband houden met elk item in de collectie. Ten derde is een procedure nodig om de resultaten te classificeren zodat een nuttige ranglijst kan worden teruggestuurd naar de gebruiker. Ten vierde moet binnen de zoekopdracht een reeks subonderwerpen worden vastgesteld.

Voor boeken verloopt dit vrij goed via bestaande tools. Ook hebben veel bibliotheken hun *openbare webcatalogi* oftewel OPAC's (Online Public Access Catalogs) herzien om dergelijke zoekopdrachten mogelijk te maken. Raadpleeg de websites van de bibliotheek van de North Carolina State University, Queens en Stanford, bijvoorbeeld:

http://www.lib.ncsu.edu/summon
http://www.queenslibrary.org
http://searchworks.stanford.edu

Figuur 1 Een foto van de immigratiekaart van John van Neumann (Johann von Neumann), gemaakt in de Library of Congres in januari 2007. Wetenschapshistorici zouden dit artefact graag willen *ontdekken*, maar weten niet hoe ze hiernaar moeten zoeken.

De informatie over onderwerpen in een boek worden vaak verstrekt via de LCSH-descriptors (Library of Congress Subject Heading). Voor boeken die in de Verenigde Staten zijn gepubliceerd, zijn de LCSH-descriptors te vinden op de eerste paar pagina's van veel boeken met de catalogusgegevens van de Library of Congress. Zo heeft het boek *Handbook of Writing for the Mathematical Sciences* [18] van Nick Higham de volgende onderwerptitels: 'Mathematics–Authorship' en 'Technical writing' Hieruit blijkt dat het boek gaat over de problemen met het schrijven van wiskundige formules en met technisch schrijven. Deze descriptors waren een oud soort indexering die werd toegepast op boeken zodat onderwerpen konden worden opgezocht in kaartencatalogi. De ruimte van een kaartencatalogus was beperkt. Daarom moest het mogelijk zijn om met zo min mogelijk indexingangen allerlei onderwerpen op te nemen in de indexering. Recenter is het ook mogelijk om via *full text*-zoekopdrachten boeken op te zoeken omdat er steeds meer *born-digital*-inhoud beschikbaar is en boeken op grote schaal worden ingescand. Gezamenlijk ondersteunen deze technologieën dergelijke zoekopdrachten voor boeken, maar er is nog ruimte voor toekomstige verbeteringen. Zo is de bovenstaande zoekopdracht voor 'lokale geschiedenis' vooral problematisch omdat 'lokale geschiedenis' een specifiek soort geschiedenis is die wordt beschreven in de onderwerptitels van de Library of Congress. Met een dergelijke zoekopdracht op deze systemen worden meestal boeken over het concept 'lokale geschiedenis' opgehaald. Een zoekresultaat was een boek over hoe u meer informatie over de geschiedenis van uw regio te weten kunt komen, dus geen boeken over de geschiedenis van de regio zelf.

Digitale opslag gaat echter veel verder dan boeken of gedigitaliseerde boeken. Het omvat zowel monumentale als alledaagse digitale artefacten. Voor dergelijke objecten zijn waarschijnlijk geen gegevens over onderwerptitels beschikbaar. Bovendien bestaan de items zelf mogelijk niet uit tekst. De Library of Congress heeft meer dan 14 miljoen afbeeldingen (volgens de webpagina van de bibliotheek: http://www.loc.gov/rr/print, geraadpleegd op 13 augustus 2010). Andere mogelijkheden zijn: enquêteresultaten, kaarten, audio en video. In de volgende hoofdstukken neemt het ontbreken van tekstbeschrijving van deze soorten materiaal een belangrijke plaats in, want het is niet altijd duidelijk hoe we gebruikers het beste in staat kunnen stellen om interessante artefacten te ontdekken. Onze huidige technieken zijn erop gericht om gegevens te extraheren uit de weinige tekst die we mogelijk over het item hebben.

Digitale archieven voor historisch materiaal

Tot nu toe hebben we het probleem rond de toegang tot digitale archieven beredeneerd vanuit het oogpunt van digitale opslag. In bibliotheken worden echter ook vele zeldzame, cultureel belangrijke manuscripten, foto's en andere objecten bewaard. Deze items zijn vaak kwetsbaar en niet geschikt om door

allerlei handen te gaan; en toch is de missie van een bibliotheek om deze items te delen. Via digitalisering en beeldverwerking wordt een doeltreffende kopie verkregen die op brede schaal kan worden gedeeld. Dezelfde moeilijkheden doen zich echter voor wanneer mensen toegang krijgen tot deze items, zoals bij algemene digitale archieven. Laten we een voorbeeld geven. Tijdens een bezoek aan de manuscriptafdeling van de Library of Congress wees een van de inhoudsdeskundigen ons op een doos met artefacten van John von Neumann. Een van deze items was een kopie van zijn immigratiekaart (zie Figuur 1). Digitale opslag is bedoeld om interessant materiaal in een breed en gevarieerd archief te kunnen vinden. Op dezelfde manier zijn deze speciale digitale collecties bedoeld om parels te kunnen vinden, zoals — wat ons betreft — informatie over John von Neumann. We zouden niet weten hoe we anders zelf hiernaar hadden moeten zoeken.

Dit is inmiddels een acuut probleem in de Library of Congress. Rond 1994 startte deze bibliotheek een enorm project om enkele van de belangrijkste werken uit de Amerikaanse cultuur te digitaliseren. Het resultaat was de collectie *American Memory* met een webinterface. Tot de gedigitaliseerde collecties behoren het dagboek van George Washington, brieven van Abraham Lincoln, en de eerste films die door Thomas Edison zijn opgenomen. Het was echter moeilijk om mensen bij het materiaal in deze collectie te krijgen. Hoewel tijdens de eerste digitalisering enkele beperkte metagegevens werden verzameld, waren deze activiteiten vooral gericht op digitalisering in plaats van effectieve toegang tot het materiaal. Bijna twintig jaar later wilde de Library of Congress deze collecties aanpassen aan moderne standaarden voor digitale archieven. Hiermee bedoelen we toegangspatronen zoals hierboven. Hiervoor zijn in elk geval accurate metagegevens over onderwerp, plaats, tijd en mensen nodig.

Historisch gezien werden deze metagegevens door bibliothecarissen of inhoudsdeskundigen gemaakt. Omdat digitalisering tegenwoordig echter zo eenvoudig is, kunnen de deskundigen de hoeveelheid materiaal niet bijhouden om dit te annoteren. De UNESCO heeft onlangs de Digitale wereldbibliotheek opgezet om te proberen de belangrijkste artefacten ter wereld op te nemen in een digitaal webarchief. De grootte van de oorspronkelijke collectie werd beperkt omdat de UNESCO behoefte had aan goed georganiseerde metagegevens die handmatig werden vertaald in elk van de zeven talen van de VN. Moet onze toegang tot deze artefacten worden beperkt doordat deskundigen voor de lastige taak staan om alles te annoteren en vertalen?

OPAC	Online Public Access Catalog
MARC	MAchine Readable Cataloging
XML	eXtensible Markup Language
RDF	Resource Description Framework
LCSH	Library of Congress Subject Headings
HIT	Human Intelligence Task
born-digital	inhoud die altijd alleen maar in digitale vorm heeft bestaan
artefact	object in een digitaal archief
metagegevens	informatie *over* een digitaal object, met name *tijd*, *plaats* en *onderwerp*
crowd-sourced	een term waarmee gegevens worden beschreven die zijn verzameld uit officieuze bronnen
folksonomie	een specifiek type crowd-sourced gegevens met een reeks tags — korte beschrijvingen — die zijn toegepast op een reeks objecten in een database
tags	laagste niveau van een folksonomie

Tabel 1 Afkortingen en definities.

Overzicht

Laten we kort aangeven welke problemen zich voordoen bij het opbouwen van zoek- en bladertools in deze archieven. Ten eerste zijn de items zeer heterogeen: boeken zijn slechts een klein gedeelte van de collectie die kan worden doorzocht. Ten tweede zijn de metagegevens voor alles (behalve voor boeken) inconsistent en onvolledig, terwijl de nuttigste metagegevens mogelijk niet beschikbaar zijn. Ten derde bestaan er geen systeemeigen koppelingen tussen items. Ten vierde is de inhoud opgesteld in veel talen. Ten vijfde is het lastig om deze items te classificeren vanwege de zeer inconsistente metagegevens.

In dit artikel geven we geen uitgebreide oplossing voor deze problemen. In plaats hiervan halen we kleine, handelbare en interessante computerproblemen uit de visie op onze digitale bibliothecaris.

Hieronder beschrijven we enkele problemen uit ons onderzoek.

Als eerste probleem bespreken we de mengelmoes van beschikbare gegevens. Zie Tabel 1 voor een overzicht van de gegevens die we willen opzoeken en de gegevens die we kunnen gebruiken bij de zoekopdracht. We beschrijven elke gegevensset uitgebreider in de volgende paragraaf. Hoewel we als allesomvattend doel een uniforme zoek- en bladerinterface mogelijk willen maken, zijn de objecten waarnaar we willen zoeken en bladeren, divers. Een ander probleem is dat sommige gegevens die we mogelijk willen gebruiken, behoorlijk gecompliceerd zijn. Zo zijn de onderwerptitels van de Library of Congress een thesaurus waarmee een onderwerp uniek wordt geïdentificeerd. Deze wordt al meer dan honderd jaar gebruikt. Er worden volledige cursussen over deze database gegeven in curricula van informatiewetenschappen. Hoe kunnen we dan snel meer hierover te weten komen? Ons antwoord is visualisatie en we gaan in in de volgende paragraaf dieper op deze werkwijze in.

Toen we de structuur van de onderwerptitels van de Library of Congress eenmaal begrepen, viel het ons op dat deze verwant was aan de structuur van de categorieën die aan Wikipedia ten grondslag liggen. Naar aanleiding hiervan hebben we onderzocht hoe we de onderwerptitels van de Library of Congress konden *afstemmen* op de categorieën in Wikipedia. En bovendien hebben we hierdoor nagedacht over andere bronnen van *openbare* of *crowd-sourced* gegevens. In de paragraaf 'Openbare crowd-sourced gegevens' bespreken we ons idee om de onderwerptitels van de Library of Congress af te stemmen op Wikipedia-categorieën. Ook gaan we in op uitdagingen bij het gebruik van deze gegevenstypen.

Op dit punt doet zich een belangrijk probleem voor. Zoals we hebben opgemerkt, willen we vaak gegevens over de *plaats* en het *onderwerp* van elk object in onze collectie. Deze gegevens zijn echter niet altijd beschikbaar. In de volgende twee paragrafen stellen we ideeën voor om deze *ontbrekende metagegevens* te genereren. In de paragraaf 'Ambigue geografische verwijzingen' introduceren we een optimalisatieprobleem om geografische plaatsnamen te desambigueren. Er wordt dus geprobeerd de volgende vraag te beantwoorden: verwijst 'San Jose' naar San

(a) Een voorbeeld van een MARC-record in XML.

(b) De MARC-metagegevens voor een record in de speelfilmcollectie (*papr*) van American Memory, vertaald naar tekst.

(c) Een fragment van de SGML-annotaties in de essays van Abraham Lincoln (*mal*-collectie).

Figuur 2 Drie voorbeelden van onze gegevensbestanden. Deze blik op het binnenste van elk bestand toont hoe de bibliotheek records er in ongemaakte vorm uitzien.

Jose in Californië of naar San Jose in Costa Rica? In de paragraaf 'Metagegevens en titelremediatie' beschrijven we hoe automatisch een titel en een reeks trefwoorden kunnen worden gegenereerd op basis van een tekstbeschrijving.

We sluiten af met een samenvatting en richtingen voor toekomstig onderzoek.

Gegevens begrijpen

Zoals we hebben opgemerkt in de inleiding, is onze visie op een virtuele bibliothecaris vrijwel grenzeloos. Een van de gevolgen van deze visie is dat we allerlei bestaande gegevensbronnen moeten verwerken. Al deze gegevenssets hebben weer een andere indeling. Soms hebben ze zelfs niets met elkaar te maken. Desalniettemin is ons doel om de gegevens samen te voegen en onze virtuele bibliothecaris mogelijk te maken door bijvoorbeeld de onvolledige metagegevens van een bibliotheekrecord aan te vullen met gegevens uit openbare bronnen. Tabel 2 bevat een overzicht van de verschillende gegevenssets die we in dit document gebruiken. Er zijn drie algemene groepen:

1. eigen gegevens van Library of Congress,
2. openbare en crowd-sourced gegevens,
3. meertalige gegevens.

De eerste groep bevat informatie die de Library of Congress meestal niet deelt, zoals de onopgemaakte metagegevens achter de collectie American Memory, of informatie die de bibliotheek verkoopt om de kosten terug te verdienen. De tweede categorie bestaat uit gegevens die volledig vrij beschikbaar zijn. We vertellen meer over deze categorie in de paragraaf 'Openbare crowd-sourced gegevens'. De laatste categorie is ook eigendom van de Library of Congress, maar onderscheidt zich doordat de metagegevens beschikbaar zijn in meerdere talen. In dit document richten we ons op de eerste twee categorieën, maar we bespreken ideeën voor de meertalige gegevens in de paragraaf over toekomstig werk. We willen nadrukkelijk erop wijzen dat deze lijst met gegevensbronnen niet volledig is. Er zijn veel andere bronnen die we hadden kunnen gebruiken. Deze lijst bevat alleen maar bronnen die *wij* hebben gebruikt.

In elk van deze databases of collecties wordt informatie op een eigen manier opgeslagen, waarbij zelfs binnen een collectie verschillen bestaan. American Memory is in feite een collectie van collecties. Sommige metagegevens in verband met de items hebben de MARC-indeling; andere zijn in de XML-indeling. Figuur 2 bevat een voorbeeld van enkele onopgemaakte gegevens in deze databases. De details van de MARC- [44], RDF- en XML-indeling zijn niet relevant. Elke gegevensindeling biedt globaal een reeks records en velden over deze records. Tenslotte kunnen sommige items annotaties in nog een andere indeling bevatten. Bij de *mal*-collectie zijn bijvoorbeeld metagegevens in XML-bestanden en annotaties in SGML-bestanden (een voorganger van XML) opgeslagen. We noemen al deze details en gege-

Type	Collectie	Aantal objecten	Indeling	Opmerkingen
Eigendom van Library of Congress	Onderwerptitels	298.964	MARC	Autoriteitsbestanden van dec. 2006
	Naamautoriteiten	6.662.688	MARC	Autoriteitsbestanden van dec. 2006
	Catalogus	7.207.747	MARC	Boekencatalogus van Library of Congress
	American Memory	617.673	MARC of XML	101 heterogene collecties
	— *papr*	703	MARC	Speelfilms
	— *mal*	20.158	XML	Essays van Abraham Lincoln
	— *gmd*	6888	MARC	Kaartencollectie
	— *wpa*	2000	XML	American Life Histories
Openbaar en crowd-sourced	Wikipedia	3.799.337	XML	(Vanaf april 2007)
	Wikipedia-categorieën	226.221	(afgeleid)	(Vanaf april 2007)
	Geografische namen	6.914.549	Tekst	Een geografisch woordenboek
	Project-Gutenberg	24	Tekst	Tekstboeken
Meertalig	Global Gateways	21.274	MARC of tekst	
	Digitale wereldbibliotheek	196	XML	

Tabel 2 Een overzicht van de gebruikte gegevens tijdens ons onderzoek. Voor elke collectie vermelden we de grootte als het aantal 'dingen' in de collectie. American Memory is een groep collecties. *papr*, *mal*, *gmd* en *wpa* zijn dus subcollecties binnen American Memory.

vensindelingen om te benadrukken hoe heterogeen de onopgemaakte gegevens zijn, zelfs op het laagste niveau. We moeten doorlopend nieuwe interpreters voor elk van deze gegevenscollecties schrijven om eenvoudigweg de gegevens zelf te kunnen openen.

Nadat we de gegevens hebben geopend, stapelen de problemen zich op. In een ideale wereld zou elk item een volledige reeks consistent gespecificeerde metagegevens bevatten, inclusief datum, locatie, onderwerp en personen. De werkelijkheid laat echter veel te wensen over. In de paragraaf 'Metagegevens en titelremediatie' zullen we zien hoe inconsistent sommige metagegevens binnen deze bestanden zijn. Zodra we de gegevensbestanden kunnen lezen, doet zich echter een ander probleem voor: we moeten de inhoud begrijpen. Met begrijpen bedoelen we dat we vertrouwd moeten zijn met de bijzondere kenmerken van een gegevensset: idealiter zoals een deskundige die al jarenlang met de gegevens werkt. Zoals we eerder hebben opgemerkt, zijn sommige van deze gegevens in de afgelopen honderd jaar verzameld door de bibliotheek [4]. In de volgende subparagraaf duiken we in de onderwerptitels van de Library of Congress om te laten zien hoe we inzicht kunnen krijgen in de inhoud van deze databases.

Een grafische weergave van LCSH

De Library of Congress Subject Headings (LCSH) is een database van termen die worden bijgehouden door de Library of Congress en worden gebruikt voor de indexering van het onderwerp van bibliografische documenten en ook voor kruisverwijzingen tussen gerelateerde onderwerpen. Een onderwerptitel bevat bredere termen, smallere termen en 'zie ook'-termen.

De onderwerptitel 'Mathematics' houdt bijvoorbeeld verband met de bredere term 'Science' en de smallere termen 'Algebra', 'Economics', 'Mathematical' en 'Women in mathematics'. We kunnen de LCSH-database zien als een ongerichte graaf waarin elke onderwerptitel een knoop is en met elke relatie een ongerichte lijn wordt gedefinieerd.

Omdat we snel inzicht wilden krijgen in de informatie binnen deze verbindingen, wilden we de graaf visualiseren. Een grote graaf kan vaak op twee manieren worden gevisualiseerd: (i) in kleine delen [37]; of (ii) als volledige graaf. We hebben met beide technieken gewerkt en beschrijven alleen de tweede om ruimte te besparen. Door de volledige graaf te visualiseren krijgen we inzicht in de algehele verbindingsstructuur van het netwerk zodat we mogelijk gerichtere vragen kunnen stellen. Zie [21] voor inzichten in het Twitter-netwerk, als voorbeeld van dit type analyse. Als we een graaf met honderdduizenden knopen willen visualiseren, hebben we een middel nodig om een indeling (een toewijzing van punten aan coördinaten in het vlak) te berekenen. Dit is een uitdagende berekening waarnaar nog altijd onderzoek wordt verricht (zie [23] voor een recente bijdrage over de visualisatie van grote grafen).

We hebben het LGL-programma (Large Graph Layout) [1] gebruikt, dat ongeveer als volgt werkt:

1. zoek een minimale spanning-tree voor de graaf;
2. zoek de knoop met de minimale totale afstand van kortste paden naar alle andere knopen (we noemen deze knoop het midden);
3. tel voor $k = 1$ tot ... alle knopen op die op een afstand van k lijnen vanaf de middelste knoop liggen, en optimaliseer lokaal hun positie.

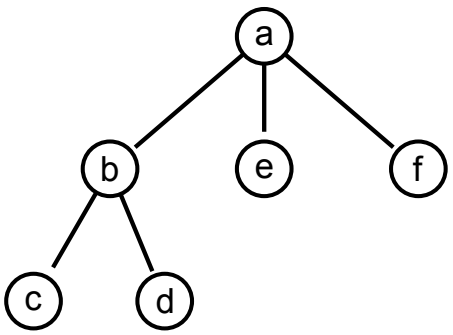
We kiezen LGL vanwege stap 1. Tijdens voorbereidend werk met de graaf achter LCSH ontdekten we dat grote delen hiervan een boomstructuur kunnen hebben. Daarom zou een indelingsalgoritme waarmee de boomstructuur in de graaf wordt onderzocht, een nuttige structuur moeten opleveren.

Tijdens het LGL-proces gaat het meeste werk in de laatste stap zitten: er moet een dynamische simulatie worden uitgevoerd om globaal een minimale energietoestand te berekenen. Zie het essay over LGL voor meer informatie over stap 3; wij richten ons op stap 2. Toen we met de code aan de slag gingen, duurde het ongeveer twee uur om het midden te vinden. In de oorspronkelijke implementatie van het LGL-algoritme werd de boomstructuur niet gebruikt bij het berekenen van de middelste knoop. Zoals we zo zullen zien, is het totaal van alle kortste paden voldoende voor een eenvoudige herhaling in een boomstructuur. We implementeerden een procedure om het midden efficiënt te berekenen. Nadat we deze wijziging hadden doorgevoerd, duurde het maar enkele seconden om stap 2 te berekenen.

We beschrijven nu onze optimalisatie om het midden efficiënt te berekenen. Stel dat $D_{u,v}$ het aantal lijnen langs het kortste pad tussen knoop u en v in de minimale spanning-tree is. We zijn op zoek naar de knoop c waarvoor $\sum_v D_{c,v}$ zo laag mogelijk is. Het belangrijkste idee achter onze optimalisatie is dat er altijd een uniek pad tussen twee knopen in een boomstructuur loopt. Formeel is de procedure als volgt. Neem $C_u = \sum_v D_{u,v}$ als de score voor het midden van knoop u. Wijzig in een boomstructuur eventueel C_u in C_w wanneer we een lijn hebben (u, w). We hoeven alleen maar te berekenen hoeveel paden die beginnen bij u, langer worden wanneer ze in plaats hiervan beginnen bij w, en hoeveel paden korter worden wanneer ze beginnen bij w. Zie Figuur 3 voor een voorbeeld van de manier waarop we van deze waarneming kunnen profiteren. Kies een willekeurige knoop a en laat de boomstructuur beginnen bij a ('root') om deze waarneming te implementeren voor een structuur T. Bereken vervolgens C_a. Voor elke knoop w die is verbonden met knoop a, geldt:

$$C_w = C_a - N_w + (n - N_w).$$

Hierin is N_w het aantal knopen in de substructuur dat begint bij w, en is n het totaal aantal knopen. Misschien wordt de formule duidelijker als u bedenkt dat alle paden vanaf w naar knopen in de subboomstructuur die beginnen bij w, één lijn korter zijn. Daarom verlagen we C_a ook met N_w. Bovendien bevatten alle overige knopen in de graaf ($(n - N_w)$ in totaal) paden die één lijn langer zijn wanneer ze beginnen bij w in plaats van a. Door deze procedure voor alle verdere niveaus te herhalen, kunnen we C_v voor elke knoop v berekenen in lineaire tijd. Voor het volledige proces moet de graaf driemaal worden doorlopen: eerst om C_a te berekenen voor de arbitraire root; dan om de grootte van elke subboomstructuur te berekenen (N_v voor elke v); en tenslotte om C_v te berekenen, waarbij C_a en N_v voor elke knoop zijn gegeven.



Figuur 3 Een klein voorbeeld van de manier waarop we de totale afstand van de kortste paden naar alle knopen efficiënt berekenen in een boomstructuur. We kunnen eenvoudig de totale afstand van de kortste paden berekenen wanneer we beginnen bij de valse root a: C_a=7. Als we nu C_b willen berekenen, zien we dat er drie knopen een lijn verder weg komen te liggen (a,e,f) en drie knopen één lijn dichterbij komen te liggen (b,c,d). Dus C_b=C_a−3+3=7. Evenzo vinden we dus C_c=C_b−1+5=11, en hetzelfde geldt voor C_d. Zowel C_e als C_f zijn ook 11. In dit voorbeeld kan a of b de rootknoop zijn. Ook N_a=7, N_b=3 en N_c=N_d=N_e=N_f=1.

Nadat we deze wijzigingen hadden aangebracht, voerden we het LGL-algoritme uit op de grootste verbonden component van de ongerichte graaf van LCSH. Figuur 8 aan het eind van dit artikel bevat een visualisatie waarin de door LGL berekende indeling wordt gebruikt. Lijnen zijn getekend met alfamenging om de lokale dichtheid te laten zien. Elk knooppunt is gekleurd op basis van een clustering die via het CLUTO-programma [24] is berekend. We zien grote vlakken met dezelfde kleur. Dit betekent dat met zowel CLUTO als LGL soortgelijke structuren in de graaf worden geïdentificeerd. Zie voor een uitgebreidere kijk op deze visualisatie

http://cads.stanford.edu/lcsh-galaxy

Op basis van deze visualisatie vinden we de volgende structuur in het LCSH-netwerk. Er is een dichte kern van onderwerptitels van algemeen belang, zoals 'Law', 'Science' en 'Art'. Rond deze kern zien we een aantal meer esoterische onderwerpen, inclusief een uitgebreide regio met geografische onderwerpen ten zuiden van de gele kern. Een ander inzicht is dat mogelijk niet alle regio's even goed zijn gecategoriseerd. Links op de afbeelding bevindt zich een grote, stervormige constructie rond de onderwerptitel Japan-Antiquities. Deze ster bevat meer dan duizend onderwerptitels, met slechts één verbinding terug naar het midden van de ster. Andere regio's van de graaf (zoals de onderwerptitels voor talen linksboven) blijken daarentegen beter te zijn georganiseerd.

Nadat we veel te lang deze visualisatie hadden bestudeerd en hiernaar hadden zitten staren, hadden we het gevoel dat we meer inzicht hadden in de onderwerptitels van de Library of Congress. In feite deden enkele kenmerken van de graaf ons denken aan een andere graaf: de categoriestructuur van Wikipedia. In het volgende hoofdstuk gaan we in op deze relatie nadat we kort crowd-sourced gegevens hebben uitgelegd.

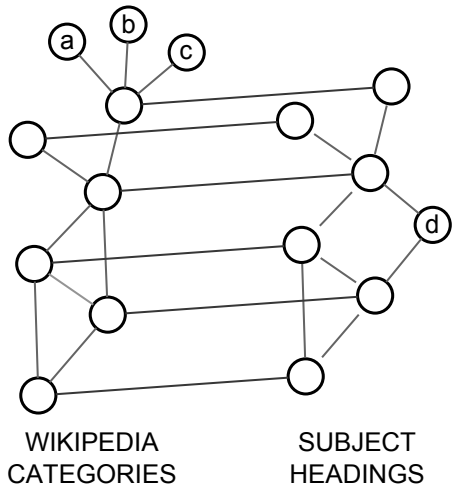
Openbare crowd-sourced gegeven

Ongeveer vijftien jaar geleden kostte een hoogwaardige encyclopedie een paar honderd euro. Tegenwoordig heeft iedereen met een internetverbinding gratis toegang tot Wikipedia. In feite mag de volledige inhoud van Wikipedia bulksgewijs worden gedownload voor niet-commercieel gebruik. De encyclopedie Wikipedia is misschien wel het beste voorbeeld van de tweede gegevenscategorie die we tijdens ons werk onderzoeken: openbare en crowd-sourced gegevens. Een openbare gegevensset is eenvoudig gezegd een gegevensset die gratis op internet beschikbaar is. Een voorbeeld van een openbare gegevensset is de website http://id.loc.gov/authorities, waar de onderwerptitels van de Library of Congress op een interactieve manier kunnen worden onderzocht en bulksgewijs kunnen worden gedownload. De records achter LCSH worden echter nog altijd beheerd door de Library of Congress.

Wikipedia is daarentegen een voorbeeld van crowd-sourced gegevens. In de afgelopen tien jaar is de encyclopedie zonder toezicht geschreven en bewerkt door uiteenlopende personen. Ze ontwikkelden een zelfregulerend mechanisme waarmee vrijwel iedereen een bijdrage kon leveren aan de encyclopedie, terwijl personen minder mogelijkheden hadden om de inhoud voor eigen doeleinden te bewerken. Let eens op het verschil met oude modellen voor informatieverzameling. Op gegevensopslagplaatsen werd toezicht gehouden door een 'gezegende' groep deskundigen die wijzigingen beoordeelden en autoriseerden in een poging om fouten te voorkomen. In het geval van LCSH duurde het proces tientallen jaren, waarbij de regels voor het toevoegen van nieuwe items alleen bekend waren binnen een select gezelschap. Zoals we straks zullen zien, zette Wikipedia een soortgelijk categoriesysteem in slechts enkele jaren op.

Crowd-sourcing is een ongekend succes. Het is een pijler van zogenaamde Web 2.0-technologieën geworden en speelde een prominente rol op de websites van Flickr en Delicious. Volgens een theorie die wordt aangehangen door Surowiecki [41], kunnen veel betrouwbaardere voorspellingen worden gedaan op basis van de uiteenlopende perspectieven van veel mensen dan van enkele deskundigen. Deze theorie staat bekend als de 'wijsheid van de menigte'. Uit een recent onderzoek naar folksonomieën, een veelgebruikt type crowd-sourced gegevens waarmee items met enkele korte tags worden beschreven, zoals op Flickr en Delicious, blijkt dat de tags van 'breedsprakige beschrijvers' nuttiger zijn dan de tags van 'categoriseerders' [28]. Als we ervan uitgaan dat deskundigen eerder tot de laatste categorie behoren, kan dit worden gezien als een empirische validatie van de methodologie voor crowd-sourcing.

Ongeacht de theoretische ondersteuning is er tegenwoordig een enorme hoeveelheid gegevens beschikbaar uit deze wat meer onsystematische modellen voor informatieverzameling. We vroegen ons het volgende af: kan een bibliotheek deze gegevens gebruiken

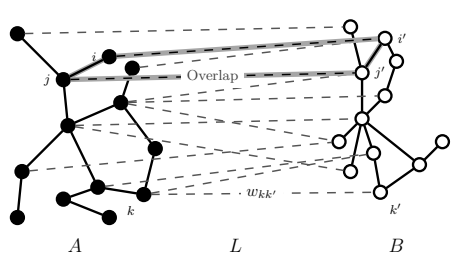


Figuur 4 We willen de Wikipediacategorieën afstemmen op de onderwerptitels van de Library of Congress. Via deze 1-op-1-afstemming (de horizontale verbindingen) kunnen we nieuwe onderwerptitels voorstellen (de knooppunten a, b, en c) en informatie toevoegen aan Wikipediacategorieën. (Het knooppunt d zou ons iets nuttigs moeten kunnen vertellen over de afgestemde buren.)

om onze virtuele bibliothecaris te implementeren? Het nut van deze openbare gegevensverzamelingen werd al vroeg erkend in [32]. We zullen in de rest van deze paragraaf laten zien hoe dit zou kunnen werken door de Wikipedia-categorieën af te stemmen op de onderwerptitels van de Library of Congress. In de volgende paragraaf bekijken we hoe openbare informatie rechtstreeks kan worden gebruikt om het probleem rond het zoeken naar de geografische verwijzingen in een document op te lossen.

Laten we nog eens teruggaan naar de structuur van de onderwerptitels van de Library of Congress: elk onderwerp is gerelateerd aan andere onderwerpen via de referenties 'Bredere term', 'Smallere term' en 'Zie ook'. We interpreteren deze relaties als een ongerichte graaf. De categoriepagina's in Wikipedia hebben een soortgelijke structuur.

Elke pagina in Wikipedia is lid van een of meer categorieën. Zo behoort de pagina over 'Singular Value Decomposition' tot de ca-



Figuur 5 Bij het algemene netwerkalignmentprobleem is het doel om de knopen van graaf A af te stemmen op de knopen van de graaf, terwijl B zoveel mogelijk lijnen probeert te *overlappen* en het gewicht van de lijnen in de overeenkomende $\sum w_{kk'}$ probeert te optimaliseren. Formeel overlapt een lijn bij een overeenkomst wanneer (i,j) een lijn in A is en het evenbeeld binnen de overeenkomst, $(m(i),m(j))$, ook een lijn in B is.

tegorieën 'Linear algebra', 'Matrix theory' en 'Functional analysis'. Categorieën hebben subcategorieën en gerelateerde categorieën die een hiërarchische structuur met enkele aanvullende lijnen vormen. Het lijkt misschien verrassend, maar de ongerichte graaf van Wikipedia-categorieën heeft ongeveer evenveel knopen als de LCSH-graaf: 205.948 tegenover 297.266. Andere kenmerken lijken ook op elkaar: de grootste verbonden component bestaat in beide grafen uit ongeveer 150.000 knopen, de gemiddelde afstand tussen knopenparen is ongeveer 7 in beide grafen, en ongeveer 6000 knooppunten hebben *identieke* tekstlabels.

Op basis van deze resultaten wilden we elke knoop in de LCSH-graaf *afstemmen* op of *toewijzen* aan een knoop in de Wikipediagraaf. We proberen een overeenkomst te vinden omdat de deskundigen die de LCSH hebben ontwikkeld, de overeenkomsten kunnen gebruiken om de dekking te verbeteren op nieuwe of snel ontwikkelende gebieden die mogelijk een betere dekking hebben in Wikipedia. Zie Figuur 4 voor een voorbeeld. We formaliseren het probleem als probleem bij het alignen van een licht ('sparse') netwerk [2]. Uit de oplossing blijkt hoe we de knopen van twee grafen moeten afstemmen wanneer we over een redelijke set *potentiële overeenkomsten* beschikken. Figuur 5 bevat de structuur van het netwerkalignmentprobleem. Voor deze kwestie stellen we ook in [2] een algoritme voor het doorgeven van berichten voor. Met ons algoritme worden binnen enkele minuten vrijwel optimale oplossingen voor het probleem bij het alignen van het LCSH- en Wikipedia-netwerk verkregen (zelfs indien geïmplementeerd in Matlab). Tabel 3 bevat enkele overeenkomsten die via deze werkwijze zijn vastgesteld. De overeenkomsten zijn ingedeeld in drie groepen: *goed*, *vreemd* en *onzinnig*. Alleen de goede overeenkomsten zijn juist. De vreemde reeks overeenkomsten klopt 'bijna' en verwijst naar gerelateerde, maar andere concepten. De onzinnige reeks is gewoon helemaal fout. Bij onze huidige formulering van het probleem worden geen 'strafpunten' toegekend voor het identificeren van een overeenkomst waar we niet zoveel aan hebben. Daardoor zitten onze resultaten vol valse overeenkomsten. In toekomstig werk hopen we een strafpuntensysteem op te nemen.

LCSH	↔	Wikipedia
	goed	
Dollar, American (Coin)	↔	United States dollar coins
Web sites	↔	Websites
Environmentalists	↔	Environmentalists by nationality
Peninsulas–Southeast Asia	↔	Peninsulas of Asia
	vreemd	
Cosby family	↔	Bill Cosby Songs
Peasants in literature	↔	Peasant foods
	onzinnig	
Hot tubs	↔	Hot dogs
Masques	↔	Vampire: The Masquerade

Tabel 3 Resultaten van onze afstemming van LCSH op Wikipedia. Zie de discussie in de tekst.

subject	werkwoord	object
Singular Value Decomposition	*is in categorie*	Linear algebra
Singular Value Decomposition	*is in categorie*	Matrix theory
Linear algebra	*is gerelateerd aan*	Affine geometry
Linear algebra	*is een subcategorie van*	Algebra

Tabel 4

Hoewel we ons algoritme zo hebben ontworpen dat dit werkt met honderdduizenden knooppunten, zijn er andere succesvolle technieken om knopen in een graaf af te stemmen. Dit probleem doet zich voor bij patroonherkenning, zie [6] voor een overzicht van dat werk. Er zijn ook allerlei fraaie matrixproblemen die zich voordoen. Zie [3, 12–13, 35–36] voor voorbeelden.

Uit recente onderzoeken komt één feit naar voren: *eenvoudige* algoritmen presteren net zo goed als of beter dan gecompliceerde algoritmen in het geval van aanvullende gegevens. Zie [38] voor een voorbeeld van dit verschijnsel in het probleem met Netflix-aanbevelingen en [15] voor een gedegen kijk op de rol van gegevens bij computergebruik. De krachtige mogelijkheid om twee gegevensverzamelingen af te stemmen lijkt handig om meer gegevens op te nemen. Hoewel de afstemming van LCSH op Wikipedia de aanzet gaf tot deze discussie, zijn we van mening dat netwerkalignment een middel is om openbare en crowd-sourced gegevens te gebruiken.

In algemenere zin staat het probleem rond het combineren van gekoppelde gegevens bekend als ontologieafstemming of -alignment. Een ontologie is een set beweringen waarmee relaties in een gestructureerde vorm worden uitgedrukt. Ze worden vaak beschreven als een set beweringen met een subject, werkwoord en object. Laten we nog eens kijken naar de Wikipedia-categorieën die eerder in dit hoofdstuk zijn genoemd. In Tabel 4 is te zien hoe ze als ontologie zouden worden uitgedrukt.

Algoritmen voor ontologie-alignment omvatten vaak *divide & conquer*-methoden om een soortgelijke doelstelling te optimaliseren als bij onze werkwijze voor netwerkalignment [10, 19].

Ambigue geografische verwijzingen

In de vorige paragraaf hebben we een techniek gezien waarmee we twee gerelateerde gegevensverzamelingen konden samenvoegen. We bekijken nu opnieuw een specifiek gegevenstype dat onze virtuele bibliothecaris nodig heeft: geografische metagegevens. De geografische context van een item is een essentieel stuk metagegevens om interessante informatie te ontdekken. Geografische correlaties zijn weliswaar meestal toevallig, maar fascinerend. Bovendien bieden geografische gegevens een eenvoudige manier om door een verzameling te bladeren of twee verschillende artefacten in verband te brengen. Niet alle items in *American Memory* hebben echter betrouwbare metagegevens. Een van de problemen waarmee we werden geconfronteerd, was hoe we de geografische entiteiten moesten extraheren uit een boek of document.

Hiervoor gaan we ervan uit dat we de tekst beschikbaar hebben of op een bepaalde manier aan beschrijvende tekst kunnen komen (misschien via spraakherkenning, Optical Character Recognition of crowd-sourced tags). Met deze tekst moet als eerste stap een lijst met locatienamen worden geëxtraheerd uit de tekst. Een locatienaam is een specifiek type benoemde entiteit. Bij een tekstverzameling kan een NER (*Named Entity Recognizer*) worden aangepast zodat alleen de namen worden uitgevoerd van tekst die waarschijnlijk staat voor naam van een locatie. We gebruikten de gratis beschikbare Stanford NER [11]. We hebben nog een stuk informatie nodig: de feitelijke geografische coördinaten van een locatienaam. Een database met toewijzingen tussen geografische coördinaten en locatienamen wordt een geografische index genoemd. We gebruikten GeoNames als onze geografische index. GeoNames is een gra-

tis beschikbare verzameling van ongeveer 7 miljoen plaatsnamen en de lengte- en breedtegraad van elke locatie. Gezamenlijk hebben we een verzameling plaatsnamen uit de Stanford NER-software en een verzameling cöordinaten van GeoNames. We hebben bijna ons doel bereikt: het vinden van alle plaatsen die in een boek of document worden genoemd. Plaatsnamen zijn echter niet gekoppeld aan unieke locaties. Verwijst de term 'San Jose' naar de hoofdstad van Costa Rica of naar het hart van Silicon Valley? Het antwoord op deze vraag is het probleem rond geografische desambiguering. Voor het antwoord moeten we context gebruiken.

Laten we het probleem in een wiskundige formule gieten. Stel dat $X = (x_1, \ldots, x_n)$ een reeks genoemde locaties is, gerangschikt op hun positie in de tekst. Deze reeks is de uitvoer van de NER-software. Formeel ging de locatienaam x_i vooraf aan x_j als $i < j$. Voor elke locatie x_i veronderstellen we dat er een set $Y_i = (y_{i,1}, \ldots, y_{i,k})$ is met bestaande locaties die overeenkomen met de tekstverwijzing x_i. Deze sets Y_i horen bij alle overeenkomsten in de GeoNames-database voor een locatienaam x_i. We noemen de set met alle mogelijke kandidaten $\mathcal{C}$, en daarmee elke $y_{i,r} \in \mathcal{C}$. Bovendien veronderstellen we dat we een afstandsfunctie tussen elementen in $\mathcal{C}$ hebben. Zie D nu als de geodetische afstand tussen de lengte- en breedtegraad van elke locatie. Deze functie wordt $D : \mathcal{C} \mapsto \mathbb{R}$. Ons doel is om een *bestaande* verwijzing voor elke kandidaat te kiezen. Deze verwijzingen kunnen op een natuurlijke manier worden gekozen door de afstand tussen de genoemde locaties tot een minimum te beperken. Dit idee wordt omgezet in het optimalisatieprobleem:

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^{n-1} D(z_i, z_{i+1}), \\ &\text{subject to} && z_i \in Y_i \quad \text{voor alle } i. \end{aligned}$$

Als we dit probleem willen oplossen, kunnen we een dynamisch programma gebruiken. Stel dat $f_{j,r}$ de optimale oplossing is van

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^{j-1} D(z_i, z_{i+1}), \\ &\text{subject to} && z_i \in Y_i \quad \text{voor alle } i, \\ & && z_j = y_{j,r}. \end{aligned}$$

Dan

$$f_{j+1,r} = \min_{s \in Y_j} \left(f_{j,s} + D\left(y_{j,s}, y_{j+1,r}\right) \right).$$

Met $\min_{r \in Y_n} f_{n,r}$ wordt het oorspronkelijke probleem tot een minimum beperkt. Voor dit Greedy Algoritme werkt $d = \max_j |Y_j|$ voor elke berekening van $f_{j+1,r}$. Er zijn hoogstens d van dergelijke berekeningen voor elke j, dus het totale werk van het algoritme is aan de bovenkant begrensd met nd^2. In praktijk zou d redelijk klein moeten zijn, want de meeste geografische entiteiten zullen een vrijwel unieke identifier hebben.

Onze zorg met dit algoritme is dat dit eenvoudig op het verkeerde been kan worden gezet door een verwijzing met één afstand. Bekijk het volgende fragment:

> Een Britse vakantieganger werd naar San Juan in Puerto Rico in plaats van San Jose in Costa Rica gestuurd door haar reisbureau. Andere toeristen die naar San Jose, Costa Rica wilden, kwamen in San Jose, Californië terecht en moesten toen de weg naar San Jose vragen.

(Geraadpleegd op http://www.skyscanner.net/news/articles/2010/09/007959-destination-doppelgangers-same-name-different-country.html op 8 september 2010.) Met het bovenstaande algoritme wordt bevestigd dat de eindverwijzing naar 'San Jose' betrekking heeft op 'San Jose, Californië' omdat de afstand 0 is, hetgeen onjuist is. Dit kan eenvoudig worden opgelost door aanvullende paarsgewijze afstanden op te nemen. Neem het gegeneraliseerde probleem:

$$\begin{aligned} &\text{minimize} && \sum_{\substack{0<j-i\le T \\ 0\le i,j\le n}} D(z_i, z_j), \\ &\text{subject to} && z_i \in Y_i \quad \text{voor alle } i. \end{aligned}$$

Nogmaals, we kunnen dit probleem oplossen met een variatie op het vorige dynamische programma. We tonen de generalisatie voor $T = 2$ en merken op dat grotere afstanden eenvoudiger af te leiden zijn. Stel dat $f_{k,(r,s)}$ de optimale oplossing is van

$$\begin{aligned} &\text{minimize} && \sum_{\substack{0<j-i\le 2 \\ j\le k}} D(z_i, z_j), \\ &\text{subject to} && z_i \in Y_i \text{voor alle } i, \\ & && z_{j-1} = y_{k-1,r}, \\ & && z_j = y_{k,s}. \end{aligned}$$

Dan

$$\begin{aligned} f_{j+1,(r,s)} = \min_{w \in Y_{j-1}} \Big(& f_{k,(w,r)} \\ & + D\left(y_{k-1,w}, y_{j+1,s}\right) \\ & + D\left(y_{k,r}, y_{j+1,s}\right) \Big). \end{aligned}$$

Indeling	Voorbeeld
####-##	1601-15
####-####	1862-1863
[Month] #, ####	Dec. 1, 1793
btw. #### and ####	btw. 1755 and 1762
#### [Season]	1939 Spring
anno ####	anno 1668
##/##/##	03/02/64
###-?	184-?
Bunka # ie ####	Bunka 1 ie 1804
Guangxu ## ####	Guangxu 30 1904
#########	185000930
United States	United States

Tabel 5 Deze tabel bevat een indeling van een type datumpatroon en een voorbeeld van dat patroon. De patronen worden weergegeven in vier groepen: duidelijk, ambigu, andere kalenders en verkeerd. Bunka 1 verwijst naar het eerste jaar van het Bunka-tijdperk in Japan, dus het jaar 1804. Evenzo is Guangxu 30 het dertigste jaar van het Guangxu-tijdperk in China, dus 1904. We hebben 'between' afgekort als 'btw'. om de tabel kort te houden.

Nu wordt met $\min_{(r,s)\in Y_{n-1}\times Y_n} f_{n,(r,s)}$ het 'lengte twee'-probleem tot een minimum beperkt. Laten we nog eens kijken naar het bovenstaande voorbeeld van San Jose. Er zijn vijf geografische verwijzingen: 'San Juan in Puerto Rico', 'San Jose in Costa Rica', 'San Jose, Costa Rica', 'San Jose, California' en 'San Jose'. Alleen de laatste verwijzing is ambigu. Stel dat we alleen San Jose, Californië en San Jose, Costa Rica als mogelijke alternatieven beschouwen. Als we T variëren, levert dit de volgende resultaten op:

$T = 1$	San Jose, Californië,
$T = 2$	San Jose, Californië *of*
	San Jose, Costa Rica,
$T = 3$	San Jose, Costa Rica,
$T = 4$	San Jose, Costa Rica.

Met een gematigde T wordt het algoritme minder gevoelig voor uitschieters.

Het algoritme levert vaak bevredigende resultaten op, maar heeft toch enkele zwakke punten. Ten eerste ligt aan het optimalisatieprobleem de veronderstelling ten grondslag dat de geografische verwijzingen in de tekst ertoe neigen om kleine clusters te vormen. Bovendien wordt ervan uitgegaan dat opeenvolgende locaties geografisch dicht bij elkaar zouden moeten liggen. Deze veronderstellingen houden mogelijk niet altijd stand. Ten tweede is geodetische afstand slechts een proxy voor de kans dat twee locaties vlakbij worden genoemd. Neem de volgende zin: "Ik ben net van New York naar Londen gevlogen."

Het is duidelijk dat de auteur van New York, New York naar Londen, Engeland is gevlogen en niet van New York, New York naar Londen, Ohio of van New York, Lincolnshire naar Londen, Engeland, terwijl beide bestemmingen geografisch dichterbij liggen. Voor de oplossing van dit probleem hebben we een betere afstandsfunctie tussen locaties nodig. Ook moeten we mogelijk aanvullende context opnemen in het algoritme. Het algemenere probleem bij het afleiden van gestructureerde gegevens uit ongestructureerde bronnen wordt gegevensextractie genoemd [7, 34]. Wat we in deze paragraaf doen, is een speciaal geval van extractie van geografische gegevens. In de conclusie stellen we een uitbreiding van ons algoritme voor disambiguering van benoemde entiteiten voor.

Metagegevens en titelremediatie

Zoals we in de inleiding hebben opgemerkt, beschikken we niet over tekst voor veel items waarmee we willen werken. Het is ook mogelijk om informatie uit de metagegevens zelf te proberen op te halen. De metagegevens bevatten vaak dubbelzinnige verwijzingen naar plaatsnamen of datums. Deze zouden we ter vervanging daarvan kunnen gebruiken. Het gebruik van metagegevens om de kwaliteit van deze gegevens te verbeteren wordt remediatie van metagegevens genoemd [8].

We bespreken eerst hoe het datumveld van een verzameling metagegevens kan worden geremedieerd. Het datumveld is met name belangrijk omdat mensen vaak naar items willen bladeren op basis van de tijdelijke relevantie. (Hoe vaak hebt u niet een e-mail opgezocht die u ongeveer twee maanden geleden hebt verstuurd?)

Het lijkt misschien een vreemd idee om metagegevens met zichzelf te remediëren. Per slot van rekening zijn metagegevens bedoeld om gestructureerde informatie over een artefact te verstrekken. Hoe kunnen we dit in hemelsnaam verbeteren? Dit is inderdaad mogelijk omdat de metagegevens mogelijk inconsistent zijn ingevoerd. Laten we een voorbeeld geven. Voor de collectie *gmd* in American Memory hebben we alle onderdelen van het MARC-veld onderzocht die de datumgegevens zouden moeten bevatten, bijvoorbeeld 260$c (publicatiedatum). In Tabel 5 wordt een overzicht weergegeven. Deze invoeritems zijn — als zodanig — enorm inconsistent en ongeschikt om een lijst met relevante items voor een bepaald jaar of een aantal jaren weer te geven. Om deze invoeritems te corrigeren, hebben we een ad-hocoplossing gekozen. In elk patroon dat we hebben aangetroffen, worden de jaargegevens vrijwel altijd aangegeven met de string ####. Omdat we dus de metagegevens wilden standaardiseren, converteerden we deze jaren naar een standaarddatumindeling en voerden we de gecorrigeerde metagegevens uit. We hoeven niet altijd ingewikkelde computertools te gebruiken.

Er is nog een uitdaging waarmee de Library of Congress wordt geconfronteerd bij veel van deze collecties: de metagegevens moeten in de loop der tijd worden verfijnd. Tijdens de eerste digitalisering van de *papr*-collectie van vroege speelfilms werd alleen een breedvoerige samenvatting van elke video verzameld. Zie Figuur 6 voor een voorbeeld. Op de meeste moderne websites, zoals YouTube, is vaak een korte titel voor elk item vereist. Deze titels moeten pakkend zijn en kunnen worden opgezocht om meer mensen te interesseren. Helaas waren de bestaande beschrijvingen te lang om als titel te fungeren. Omdat deze collectie minder dan duizend video's bevatte, kortte de Library of Congress handmatig elke beschrijving in tot een titel. We vroegen ons het volgende af: kunnen de beschrijvingen automatisch worden ingekort om een goede titel te verkrijgen? Nogmaals, zie de afbeelding voor een voorbeeld van onze titel van dezelfde video, vergeleken met de titel van de Library of Congress. In de gegenereerde titel wordt de essentie van de video beknopt vastgelegd. We bespreken in de volgende paragraaf hoe we onze gegenereerde titels hebben geëvalueerd, want hierbij deden zich enkele andere kwesties voor die we nader willen toelichten.

Titelsjablonen

Dan beschrijven we nu hoe we de titels genereren. Als eerste stap in het proces identificeren we gemeenschappelijke woordsoortpatronen in een bestaande database met titels. Deze patronen hebben de volgende vorm:

Samenvatting
Toont politieagenten en mannen met een hoge hoed en formele rijkleding terwijl ze grote bossen bloemen dragen tijdens een parade te paard. Wanneer de camerahoek enigszins verandert, verschijnt een marcherende band met een trommel waarop Bugle Corps, Lowell staat, gevolgd door een aantal gewapende militairen in uniform die in formatie marcheren. De camerahoek verandert zodat rijtuigen en de rest van de stoet in beeld worden gebracht. Het tafereel verplaatst naar een gebouw: de camera schuift langs de trap en toont een geestelijke in een lang gewaad die de kerk verlaat en teruggroet met zijn hoed in de hand. Geen titels.

Handmatige titel
St. Patrick's Day parade, Lowell, MA.

Onze titel
Parade van mannen te paard.

Figuur 6 Een voorbeeld van de manier waarop we automatisch de titel van een film genereren die in 1905 van een parade is gemaakt door Thomas Edison. De film is nu te zien op YouTube: http://www.youtube.com/watch?v=mKzcjKDgxHY.

Excavating	for	a	New	York	foundation
VBG	IN	DT	NNP	NNP	NN

Daarbij staan de codes voor respectievelijk: werkwoord, voorzetsel, lidwoord, eigennaam, eigennaam en zelfstandig naamwoord. We hebben deze berekend met de woordsoorttagger van Stanford [43]. Het idee is dat een grote titelverzameling gemeenschappelijke patronen in woordsoortreeksen zal bevatten. We kunnen de meest voorkomende patronen identificeren en als titelsjabloon gebruiken. Vervolgens kunnen we tekst uit de beschrijving afstemmen op de titelsjablonen en hopen dat het resultaat uit nuttige titels bestaat. Als eerste stap bij het genereren van titels moeten we dus een reeks titelsjablonen berekenen. Wer gebruikten de Newswire-collectie voor deze taak. Deze collectie bevat $1{,}3$ miljoen artikelen. Voor de titel van elk artikel berekenden we de woordsoortreeks en analyseerden we de patronen. Het resultaat is een database van 225.000 titelsjablonen.

Scores toewijzen aan woordgroepen

Voor het opbouwen van betekenisvolle titels moeten we betekenisvolle woordgroepen uit de beschrijving halen. We gebruiken een idee van Tomokiyo en Hurst [42]. Daarbij worden titels gezocht door een score toe te kennen aan een woordreeks op basis van twee maateenheden: de informatiewaarde en de woordgroepcohesie. Een reeks heeft een hoge informatiewaarde als de kans erg klein is dat deze reeks voorkomt in normale tekst. Een voorbeeld is de beschrijving 'singular value decomposition'. Het is zeer onwaarschijnlijk dat deze woordreeks voorkomt in een dagelijkse tekst, waardoor deze woordgroep zeer informatief is. De kans is echter aanzienlijk dat 'singular value decomposition' voorkomt in artikelen in het *SIAM Journal of Matrix Analysis*. De informatiewaarde van deze woordgroep staat dus in verhouding tot een achtergrondverzameling van 'standaardtekst'. Een woordgroep heeft een grote woordcohesie als de statistische eigenschappen van de woordgroep drastisch veranderen wanneer we de woordgroep opsplitsen. De woordgroep 'New York' heeft een hoge cohesie omdat in een document over 'New York' de woorden 'New' en 'York' vrijwel altijd samen zullen voorkomen. De statistische gegevens van 'New' en 'York' worden gekoppeld in dit document.

Deze concepten zijn geformaliseerd met een op n-grammen gebaseerd taalmodel tegen een achtergrondverzameling van tekst. Stel dat $\mathcal{C}$ een verzameling documenten is die als standaard worden beschouwd. De keuze van $\mathcal{C}$ is bepalend voor welke woorden als belangrijk worden gekozen in het bovenstaande voorbeeld met 'singular value decomposition', maar niet voor welke woorden als woordgroep worden beschouwd. Elke $d \in \mathcal{C}$ is in feite een reeks van woordtokens $d = (w_1, \ldots, w_m)$. Een op unigrammen gebaseerd taalmodel is de kans dat elk afzonderlijk woord voorkomt in de verzameling documenten. Een op bigrammen gebaseerd taalmodel is de kans dat elke woordenreeks voorkomt in de verzameling documenten.

Neem nu de woordenreeks in de beschrijving van een item: $d = (w_1, \ldots, w_m)$. Voor een woordenreeks (w_i, w_{i+1}, w_{i+2}) bedraagt de score voor de informatiewaarde:

$$P(w_i, w_{i+1}) = \text{Prob}\left[(w_i, w_{i+1}, w_{i+2}) \text{ in } d\right] \cdot \log\left(\frac{\text{Prob}\left[(w_i, w_{i+1}, w_{i+2}) \text{ in } d\right]}{\text{Prob}\left[(w_i, w_{i+1}) \text{ in } d\right] \cdot \text{Prob}\left[(w_{i+1}, w_{i+2}) \text{ in } d\right]}\right).$$

De score voor de informatiewaarde is:

$$I(w_i, w_{i+1}) = \text{Prob}\left[(w_i, w_{i+1}) \text{ in } \mathcal{C}\right] \cdot \log\left(\frac{\text{Prob}\left[(w_i, w_{i+1}) \text{ in } \mathcal{C}\right]}{\text{Prob}\left[w_i \text{ in } \mathcal{C}\right] \cdot \text{Prob}\left[w_{i+1} \text{ in } \mathcal{C}\right]}\right).$$

Deze scores zijn slechts de waarden voor de Kullback–Leibler-divergentie tussen het trigram- en bigrammodel in de beschrijving voor informatiewaarde en tussen het bigram- en unigrammodel in de achtergrondverzameling voor informatiewaarde. Met extreem korte beschrijvingen gebruiken we de achtergrondverzameling om de scores voor informatiewaarde te berekenen in plaats van de tekst van de beschrijving.

Een probleem met deze modellen is dat we gebeurtenissen met een kans van nul kunnen tegenkomen. Afgevlakte modellen zijn de standaardcorrectie voor deze kans van nul. Het idee achter een afgevlakt model is dat de kans op gebeurtenis niet nul is, zelfs niet als deze nog nooit is waargenomen. Een eenvoudig type afvlakking dat in statistiek wordt gebruikt, staat bekend als pseudo-count, waarvan Laplaciaanse afvlakking op basis van Laplace's regel van opeenvolging het klassieke voorbeeld is. Voor de kans dat een n-gram voorkomt in taal worden twee technieken veel gebruikt: Katz-afvlakking [25] en Kneser–Ney-afvlakking [27]. Bij Katz-afvlakking worden de gemeten aantallen verminderd met een vermenigvuldigingsfactor kleiner dan 1. De verwijderde aantallen worden gedistribueerd over de niet-waargenomen n-grammen op basis van het aantal n-grammen van een lagere orde, bijvoorbeeld het aantal unigrammen in plaats van het aantal bigrammen. Bij Kneser–Ney-afvlakking wordt additieve reductie in plaats van een vermenigvuldigingsfactor gebruikt. Ook kunnen hiermee beter n-grammodellen van een lagere orde worden opgebouwd waarin combinaties van meerdere woorden beter worden verwerkt. Stel dat 'San Francisco' veel voorkomt, maar dat 'Francisco' alleen voorkomt na 'San'. Kneser-Ney wijst aan 'Francisco' een lagere kans op een unigram toe omdat het woord alleen voorkomt in bepaalde bigram-combinaties, hetgeen tot uiting komt in hoge kansen op een bigram.

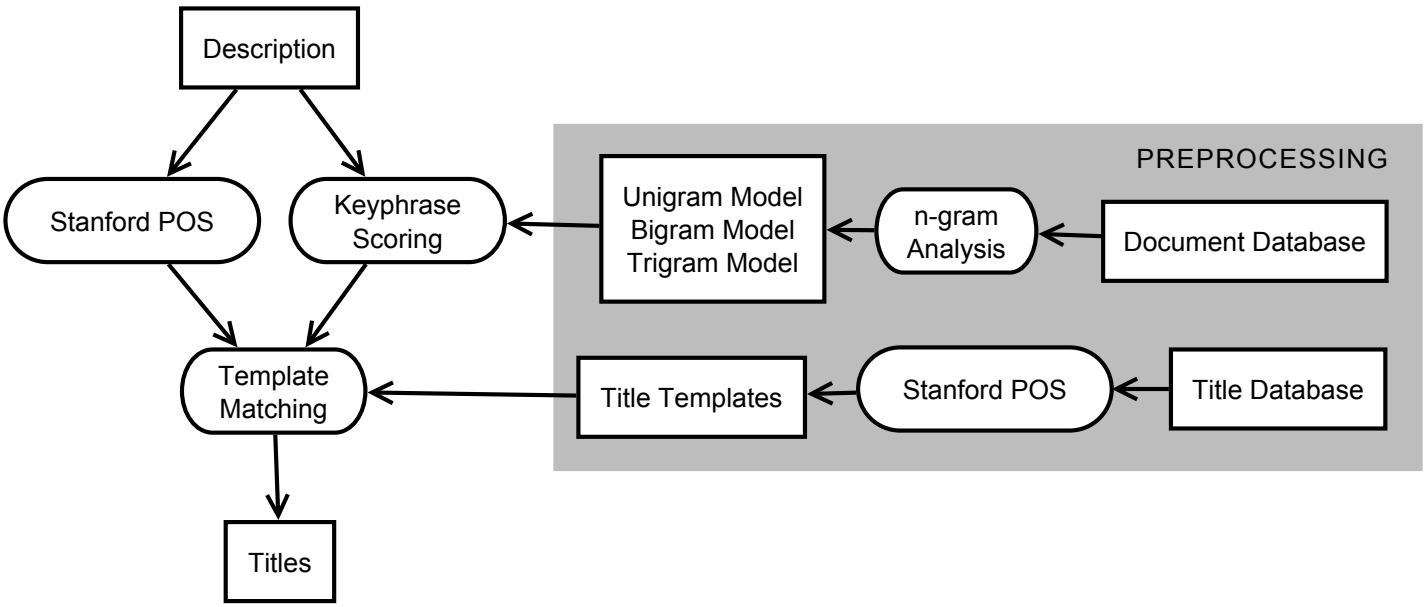
Samenvatting

Nadat we de scores van het nut van een bepaalde woordgroep hebben verkregen, hoeven we alleen nog maar de woordgroepen af te stemmen op de titelsjablonen om een titel te genereren. De titel met het hoogste gewicht (som van scores) is waarschijnlijk de beste titel.

Laten we het proces nog even samenvatten. Stel dat $\mathcal{C}$ een verzameling is van tekst met algemene informatie. Dit is de achtergrondverzameling. Bereken de kans op unigrammen, bigrammen en trigrammen in deze achtergrondverzameling. Stel vervolgens dat $\mathcal{T}$ een titelverzameling is. Bereken voor elke titel $t \in \mathcal{T}$ de woordsoortreeks voor de titel met behulp van de woordsoort-tool van Stanford (of een andere tool waarmee woordsoorten worden geïdentificeerd). Stel een reeks titelsjablonen samen op basis van deze woordsoortreeksen. Bereken nu op basis van een beschrijving de woordsoortreeks voor deze beschrijving. Bereken voor elke bigram in de beschrijving de score voor woordgroepcohesie en informatiewaarde. Neem de som van deze scores als totaalscore voor de sleutelwoordgroep van dit bigram. Stel vervolgens voor elke titelsjabloon een reeks bigrammen samen die overeenkomen met de woordsoortreeks. We vatten het proces samen in Figuur 7.

Conclusies en ideeën

Nog even onze motivering. Voor moderne verzamelingen van digitale gegevens zijn nieuwe zoektechnologieën nodig om deze gegevens relevant te maken zodat ze het waard zijn om te worden opgeslagen. Voor historische verzamelingen van gedigitaliseerde gegevens zijn geavanceerde zoekmethoden vereist zo-



Figuur 7 Ons proces voor het opbouwen van titels. Cirkels geven verwerking aan en rechthoeken geven data aan. Het grijze gebied geeft eenmalige voorverwerking aan. De rest van het proces moet voor elke beschrijving worden uitgevoerd.

dat mensen de artefacten kunnen vinden die ze interessant vinden. In beide scenario's zijn interessante metagegevens over de objecten van de digitale verzamelingen nodig. In dit artikel hebben we een algemene visualisatie van de onderwerptitels van de Library of Congress gepresenteerd. Dankzij deze visualisatie kregen we snel inzicht in een nieuwe verzameling gekoppelde gegevens. Op basis van onze ervaring met deze gegevensset onderzochten we een algoritme waarmee we de onderwerptitels in de collectie van de Library of Congress konden *afstemmen* op de categorieën van de Wikipedia-encyclopedie. Ons algoritme, dat is beschreven in [2], leverde bijna-optimale theoretische resultaten op; en een potentieel nuttige reeks overeenkomsten tussen de twee.

Vervolgens hebben we gekeken naar de desambiguering van geografische verwijzingen in een tekst. Dit leidde tot een eenvoudig, dynamisch programma waarbij we gebruikmaakten van de afstanden tussen mogelijke locaties die we eenvoudig kunnen oplossen. Voor de oplossing van geografische verwijzingen gaat het met name om het remediëren van metagegevens. We gingen verder met een ander onderzoek naar hoe we betere titels konden genereren op basis van korte beschrijvingen.

Met deze ideeën worden slechts enkele mogelijkheden onderzocht en onderzoek in digitaal beheer is in volle zwang, niet alleen bij ons, maar ook bij vele andere onderzoeksgroepen. Wij zijn bijvoorbeeld nu bezig met het verbeteren van geografische desambiguering, waar we niet alleen van locatieverwijzingen gebruik maken, maar ook gerelateerde informatie die we vinden in Wikipedia, of andere informatiebestanden. In de toekomst leggen we ons voornamelijk toe op meertalig zoeken en ontdekken [5, 9, 16]. In dit artikel hebben we nog niet gesproken over evaluatie van onze werkwijzen. In ons onderzoek maken we zelf vaak gebruik van Mechanical Turk van Amazon. Expliciete feedback van gebruikers kan ook worden benut en dit is een tweede richting die we hopen in te slaan in de nabije toekomst [33].

Gegevensbronnen

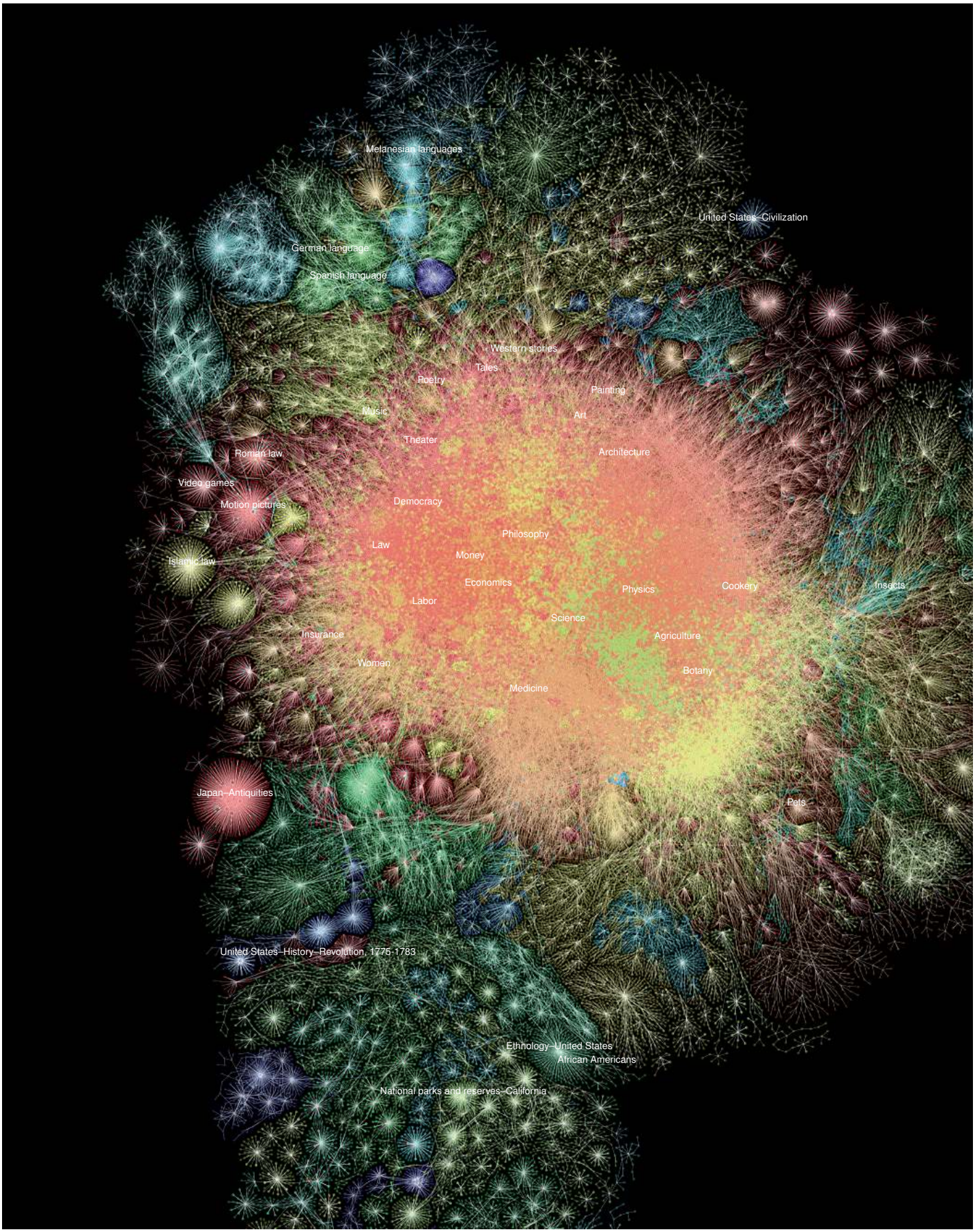
Een groot deel van dit artikel was gericht op hoe we met openbare gegevens de zoekervaring in de eigen gegevens van de Library of Congress kunnen verbeteren. Daardoor vragen lezers zich mogelijk af hoe ze een bijdrage kunnen leveren. Zoals we eerder hebben opgemerkt, heeft het succes van openbare gegevens geleid tot een overvloed aan vrij beschikbare gegevenssets. Dit zijn enkele van onze favorieten:

- Onderwerptitels van de Library of Congress: nu vrij beschikbaar, http://id.loc.gov/authorities.
- Rameau: de onderwerptitels van de Franse nationale bibliotheek, http://www.cs.vu.nl/STITCH/rameau/dump.
- Freebase: een grote verzameling gestructureerde en semigestructureerde informatie, http://freebase.com.
- Open Library: metagegevens over boeken, http://openlibrary.org.

Elke gegevensset biedt de gegevens bulksgewijs. Daardoor wordt het eenvoudig om de gegevens te interpreteren. Freebase bestaat uit allerlei kleine verzamelingen. Het ontwerpen van zoek- en bladertechnieken voor deze individuele verzamelingen lijkt enigszins op de eigen metagegevens van de Library of Congress.

Daarnaast willen we nog een mogelijkheid aanbevelen: neem contact op met verantwoordelijken op uw universiteit of in de nationale bibliotheek. We hebben de ervaring dat deze instellingen openstaan voor nieuwe ideeën en benaderingen. Als u deze weg volgt, moet u echter het geduld opbrengen om meer te leren over bibliotheek- en informatiewetenschappen (de historische thuisbasis voor het bestuderen van gegevensorganisatie en -toegang).

Dankwoord
We zijn zeer veel dank verschuldigd aan de behulpzame mensen van de Library of Congress, de Digitale wereldbibliotheek en het National Digital Information Infrastructure Preservation Program omdat ze ons hebben verteld over hun problemen bij het beheer van digitale gegevens en ons geduldig hebben geholpen tijdens onze voorlichting over hun problemen. Speciale dank gaat uit naar Laura Campbell, George Coulbourne, Beth Dulabahn, Jane Mandelbaum, Barbara Tillett en de bibliothecaris van het Congres, James Billington. Ook gaat onze dank uit naar Mohsen Bayati, die ons heeft geholpen bij het opstellen van een schaalbaar algoritme voor het netwerkalignmentprobleem, en Les Fletcher, die ons vaak heeft voorzien van nuttige adviezen.



Figuur 8 Deze tekening toont de grootste verbonden component van de ongerichte graaf met koppelingen in de onderwerptitels van de Library of Congress, waarbij knooppunten zijn gekleurd op basis van een cluster uit het CLUTO-programma. Enkele knooppuntlabels worden weergegeven om de onderwerpen in een specifieke regio te laten zien.

Referenties

1 A.T. Adai, S.V. Date, S. Wieland en E.M. Marcotte, LGL: creating a map of protein function with an algorithm for visualizing very large biological networks, *J. Mol. Biol.* 340(1) (2004), 179–190.

2 M. Bayati, M. Gerritsen, D.F. Gleich, A. Saberi en Y. Wang, Algorithms for large, sparse network alignment problems, in *Proceedings of the 9th IEEE International Conference on Data Mining*, 2009, pp. 705–710.

3 V.D. Blondel, A. Gajardo, M. Heymans, P. Senellart en P.V. Dooren, A measure of similarity between graph vertices: Applications to synonym extraction and web searching, *SIAM Review* 46(4) (2004), 647–666.

4 L.M. Chan, Still robust at 100: A centry of LC Subject Headings, in *Library of Congress Information Bulletin*, 1998.

5 P.A. Chew, B.W. Bader, T.G. Kolda en A. Abdelali, Cross-language information retrieval using PARAFAC2, in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007, pp. 143–152.

6 D. Conte, P.P. Foggia, C. Sansone en M. Vento, Thirty years of graph matching in pattern recognition, *International Journal of Pattern Recognition and Artificial Intelligence* 18(3) (2004), 265–298.

7 H. Cunningham, D. Maynard, K. Bontcheva en V. Tablan, GATE: A framework and graphical development environment for robust nlp tools and applications, in *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, 2002.

8 G. de Groat, Future directions in metadata remediation for metadata aggregators, Technical Report, Digital Library Federation, 2009.

9 S.C. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas en R.A. Harshman, Indexing by latent semantic analysis, *Journal of the American Society of Information Science* 41(6) (1990), 391–407.

10 M. Ehrig en S. Staab, QOM – quick ontology mapping, in *Third International Semantic Web Conference*, 2004, pp. 683–697.

11 J.R. Finkel, T. Grenager en C. Manning, Incorporating non-local information into information extraction systems by gibbs sampling, in *ACL'05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 2005, pp. 363–370.

12 C. Fraikin, Y. Nesterov en P.V. Dooren, A gradient-type algorithm optimizing the coupling between matrices, *Linear Algebra and its Applications* 429(5–6) (2008), 1229–1242. Special Issue devoted to selected papers presented at the *13th Conference of the International Linear Algebra Society*.

13 C. Fraikin, Y. Nesterov en P. Van Dooren, Optimizing the coupling between two isometric projections of matrices, *SIAM J. Matrix Anal. Appl.* 30(1) (2008), 324–345.

14 F. Göbel en A.A. Jagers, Random walks on graphs, *Stochastic Processes and their Applications* 2(4) (1974), 311–336.

15 A. Halevy, P. Norvig en F. Pereira, The unreasonable effectiveness of data, *IEEE Intelligent Systems* 24(2) (2009), 8–12.

16 R.A. Harshman, PARAFAC2: Mathematical and technical notes, *UCLA Working Papers in Phonetics* 22 (1972), 30–44.

17 J. Heer en M. Bostock, Crowdsourcing graphical perception: using mechanical turk to assess visualization design, in *CHI'10: Proceedings of the 28th international conference on Human factors in computing systems*, 2010, pp. 203–212.

18 N. J. Higham, *Handbook of Writing for the Mathematical Sciences*, SIAM, 1998.

19 W. Hu, Y. Qu en G. Cheng, Matching large ontologies: A divide-and-conquer approach, *Data Knowl. Eng.* 67(1) (2008), 140–160.

20 B.A. Huberman, D.M. Romero en F. Wu, Social networks that matter: Twitter under the microscope, *First Monday* 14(1) (2008).

21 A. Java, Twitter social network analysis, *UMBC ebquity blog*, 2007, http://ebiquity.umbc.edu/blogger / 2007 / 04 / 19 / twitter-social-network-analysis.

22 A. Java, X. Song, T. Finin en B. Tseng, Why we Twitter: understanding microblogging usage and communities, in *WebKDD/SNA-KDD'07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 2007, pp. 56–65.

23 Y. Jia, J. Hoberock, M. Garland en J. Hart, On the visualization of social and other scale-free networks, *IEEE Transactions on Visualization and Computer Graphics* 41(6) (2008), 1285–1292.

24 G. Karypis, CLUTO – a clustering toolkit, Technical Report 02-017, University of Minnesota, 2002.

25 S.M. Katz, Estimation of probabilities from sparse data for the language model component of a speech recognizer, *IEEE Transactions on Acoustics, Speech and Signal Processing* 35(3) (1987), 400–401.

26 A. Kittur, E.H. Chi en B. Suh, Crowdsourcing user studies with mechanical turk, in *CHI'08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 2008, pp. 453–456.

27 R. Kneser en H. Ney, Improved backing-off for m-gram language modeling, in *International Conference on Acoustics, Speech, and Signal Processing, 1995. ICASSP-95*, 1995, pp. 181–184.

28 C. Körner, D. Benz, A. Hotho, M. Strohmaier en G. Stumme, Stop thinking, start tagging: tag semantics emerge from collaborative verbosity, in *WWW'10: Proceedings of the 19th international conference on World wide web*, 2010, pp. 521–530.

29 B. Krishnamurthy, P. Gill en M. Arlitt, A few chirps about Twitter, in *WOSP'08: Proceedings of the first workshop on Online social networks*, 2008, pp. 19–24.

30 T. Kuny, A digital dark ages? challenges in the preservation of electronic information, in *63rd International Federation of Library Associations and Institutions Council and General Conference (IFLA1997)*, 1997.

31 H. Kwak, C. Lee, H. Park en S. Moon, What is Twitter, a social network or a news media?, in *WWW'10: Proceedings of the 19th international conference on World wide web*, 2010, pp. 591–600.

32 A.Y. Levy, A. Rajaraman en J.J. Ordille, Querying heterogeneous information sources using source descriptions, in *VLDB'96: Proceedings of the 22th International Conference on Very Large Data Bases*, 1996, pp. 251–262.

33 S. Levy, How google's algorithm rules the web, *Wired Magazine* 18(3), 2010.

34 A. McCallum, Information extraction: Distilling structured data from unstructured text, *Queue* 3(9) (2005), 48–57.

35 S. Melnik, H. Garcia-Molina en E. Rahm, Similarity flooding: A versatile graph matching algorithm and its application to schema matching, in *Proceedings of the 18th International Conference on Data Engineering*, 2002, p. 117.

36 L. Ninove, *Dominant Vectors of Nonnegative Matrices: Application to Information Extraction in Large Graphs*, Ph.D. thesis, Université Catholique de Louvain, 2008.

37 D. Rafiei en S. Curial, Effectively visualizing large networks through sampling, *Visualization Conference, IEEE*, 2005, p. 48.

38 A. Rajaraman, More data usually beats better algorithms, *Datawocky Blog*, 2008, http://anand.typepad.com/datawocky/2008/03/more-data-usual.html.

39 J. Ross, L. Irani, M.S. Silberman, A. Zaldivar en B. Tomlinson, Who are the crowdworkers?: shifting demographics in mechanical turk, in *CHI EA'10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, 2010, pp. 2863–2872.

40 M.S. Silberman, J. Ross, L. Irani en B. Tomlinson, Sellers' problems in human computation markets, in *HCOMP'10: Proceedings of the ACM SIGKDD Workshop on Human Computation*, 2010, pp. 18–21.

41 J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, Little, Brown, 2004.

42 T. Tomokiyo en M. Hurst, A language model approach to keyphrase extraction, in *Proceedings of the ACL 2003 workshop on Multiword expressions*, 2003, pp. 33–40.

43 K. Toutanova, D. Klein, C.D. Manning en Y. Singer, Feature-rich part-of-speech tagging with a cyclic dependency network, in *NAACL'03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 2003, pp. 173–180.

44 Various, The MARC Standard, 2007, http://www.loc.gov/marc, accessed on 17 September 2007.

45 J. Weng, E.-P. Lim, J. Jiang en Q. He, TwitterRank: finding topic-sensitive influential twitterers, in *WSDM'10: Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 261–270.

46 K.-P. Yee, K. Swearingen, K. Li en M. Hearst, Faceted metadata for image search and browsing, in *CHI'03: Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 401–408.

47 E. Yeh, D. Ramage, C.D. Manning, E. Agirre en A. Soroa, Wikiwalk: random walks on wikipedia for semantic relatedness, in *TextGraphs-4: Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, 2009, pp. 41–49.

Neda Sepasian
Department of Biomedical Engineering
Eindhoven University of Technolgy
n.sepasian@tue.nl

Jan ten Thije Boonkkamp
Department of Mathematics and Computer Science
Eindhoven University of Technolgy
j.h.m.tenthijeboonkkamp@tue.nl

Anna Vilanova
Faculty EEMCS
Delft University of Technolgy
a.vilanova@tudelft.nl

Onderzoek

Diffusion tensor imaging: brain pathway reconstruction

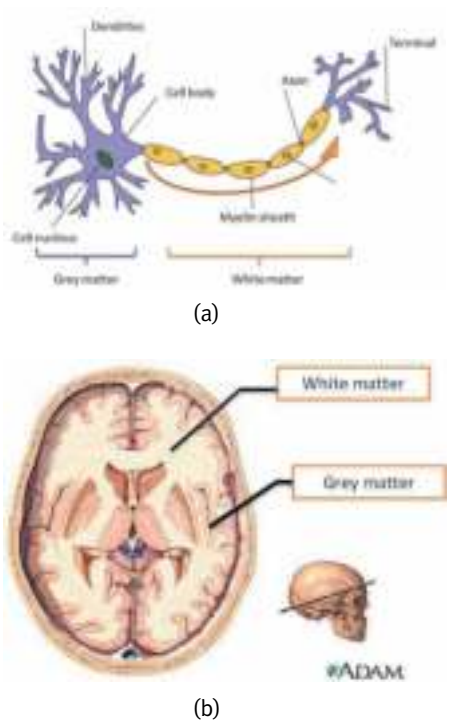
Diffusion tensor imaging (DTI) is a recently developed modality of magnetic resonance imaging (MRI), which allows producing in-vivo images of biological fibrous tissues such as the neural axons of white matter in the brain. The techniques for reconstructing connections between different brain areas using DTI are collectively known as fibre tracking or tractography. The brain connectivity map, derived from the tractography visualization and analysis, is an important tool to diagnose and analyse various brain diseases, and is of essential value in providing exquisite details on tissue microstructure and neural networks. In this article Neda Sepasian, Jan ten Thije Boonkkamp and Anna Vilanova describe a technique for the reconstruction of fibre pathways.

Assuming that fibre pathways follow the most efficient diffusion propagation trajectories, we specifically develop a geodesic-based tractography technique for the reconstruction of fibre pathways. Results we obtain using our technique, based on finding multiple geodesics connecting two given points or regions are encouraging and give confidence that this method can be used for practical purposes in the near future.

Brain structure

The nervous system functions as the body's communication and decision centre. The brain and spinal cord are collectively known as central nervous system. Brain and spinal cord are made of grey matter and white matter. White matter consist mostly of myelinated axons and non-neural cells. Grey matter is a type of neural tissue mainly consisting of dendrites and both unmyelinated and myelinated axons. The grey matter takes care of the processing functions whereas the white matter provides the communication between different grey matter areas and the rest of the body. Sensory nerves gather the information from the environment and send them to the spinal cord. The spinal cord sends this information to the brain.

The white matter axons are surrounded by myelin; see Figure 1. The myelin gives the whitish appearance to the white matter. Myelin increases the speed of transmission of all nerve signals and is distributed diffusely or is concentrated in bundles. These bundles are often referred to as tracts or fibre pathways. Our goal is to develop accurate mathematical models for in-vivo reconstruction of brain fibre bundles to study a host of various disorders and neurodegenerative diseases including, among others, Parkinson, Alzheimer and Huntington.



(a)

(b)

Illustration: www.adameducation.com

Figure 1 (a) Structure of a typical core component (neuron) of the central nervous system. (b) Axial view of a brain illustrating white and grey matter.

Diffusion process

It is known that a significant amount of the human body consists of water. At a microscopic scale water molecules move freely and collide with each other. This movement of water molecules is known as Brownian motion, which implies that molecules in a uniform volume of water will diffuse randomly in all directions. At a macroscopic scale, this phenomenon is known as diffusion. Diffusion is the thermal motion of all (liquid and gas) molecules at temperatures above absolute zero. Depending on the medium, diffusion can be either *isotropic* or *anisotropic*. Figure 2 illustrates the difference between diffusion processes in different media. For free or isotropic diffusion, the probability distribution of a single molecule located at position $\mathbf{x}_0$ to reach another position $\mathbf{x}_1$ after a given time t is spherically symmetric, i.e., every direction is equally probable. This is illustrated in Figure 2(a).

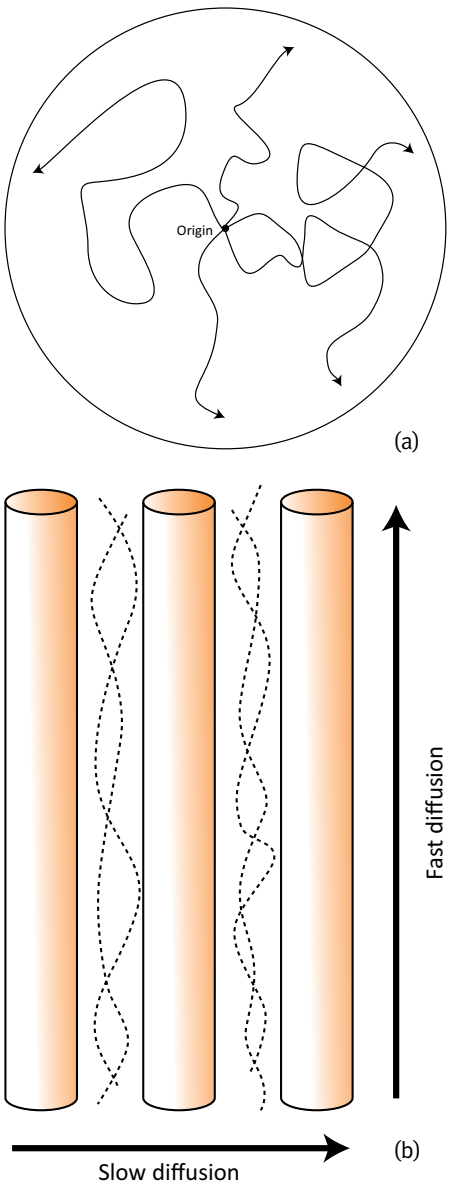


Figure 2 (a) Isotropic diffusion. (b) Anisotropic diffusion. The arrows indicate possible trajectories which molecules may follow. In the presence of barriers, e.g. axons, diffusion is restricted in certain directions.

Einstein [9] describes diffusion by relating the diffusion coefficient D, which characterizes the mobility of the molecules, to the root mean square of the diffusion displacement, i.e.,

$$D = \frac{1}{6t}\langle \mathbf{R}^T\mathbf{R}\rangle. \tag{1}$$

In this expression $\mathbf{R}$ is the net displacement vector $\mathbf{R} = \mathbf{x}_1 - \mathbf{x}_0$. The bracket $\langle\rangle$ denotes the ensemble average. In the isotropic case, the scalar D depends on the molecule type and the medium properties but not on the direction.

Using Fick's law of diffusion, the diffusion process can be approximated as follows [8]:

$$\frac{\partial P(\mathbf{R},t)}{\partial t} = D\nabla^2 P(\mathbf{R},t). \tag{2}$$

Here, ∇^2 is the Laplacian in $\mathbf{R}$ and $P(\mathbf{R},t)$ represents the probability of a water molecule displacement $\mathbf{R}$ in time t, and is known as the diffusion displacement probability density function (PDF). Under the condition

$$\int_{\mathbf{R}^3} P(\mathbf{R},t)\mathrm{d}\mathbf{R} = 1, \tag{3}$$

the solution to equation (2) is a Gaussian distribution and can be written as

$$P(\mathbf{R},t) = \frac{1}{\sqrt{(4\pi Dt)^3}}\exp\left(\frac{-1}{4Dt}\mathbf{R}^T\mathbf{R}\right). \tag{4}$$

In anisotropic biological tissues, the mobility of water molecules is restricted by obstacles formed by surrounding structures, such as the axons in the brain; see Figure 2(b). It is known that myelin sheaths have a property to modulate the anisotropy of diffusion [6].

Several models have been proposed for the PDF of anisotropic diffusion. Amongst these models, the most popular one is known as the diffusion tensor (DT) model [2]. In this model of water diffusion, Einstein's law of diffusion is generalized to anisotropic diffusion by replacing the scalar diffusion coefficient D in (1) by a symmetric positive definite matrix $\mathbf{D}$, as follows

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & d_{13} \\ d_{12} & d_{22} & d_{23} \\ d_{13} & d_{23} & d_{33} \end{pmatrix} = \frac{1}{6t}\langle \mathbf{R}\mathbf{R}^T\rangle. \tag{5}$$

Analogously, equation (2) generalizes to

$$\frac{\partial P(\mathbf{R},t)}{\partial t} = \nabla\cdot\left(\mathbf{D}\nabla P(\mathbf{R},t)\right). \tag{6}$$

The solution of (6) is a Gaussian distribution and gives the diffusion PDF of water molecules. Given condition (3), it can be written as

$$P(\mathbf{R},t) = \frac{1}{\sqrt{(4\pi t)^3|\mathbf{D}|}}\exp\left(\frac{-1}{4t}\mathbf{R}^T\mathbf{D}^{-1}\mathbf{R}\right), \tag{7}$$

where $|\mathbf{D}| > 0$ is the determinant of $\mathbf{D}$.

Acquisition and reconstruction of diffusion

Diffusion-weighted magnetic resonance imaging (DWMRI) is an acquisition technique to measure the random Brownian motion of water molecules within a voxel of tissue. This technique provides a unique non-invasive tool for measuring the local characteristics of tissues. The first diffusion weighted imaging (DWI) acquisition was done by Taylor et al. [19] using a hen's egg as a phantom in a small bore magnet. Later, Le Bihan et al. [12] applied the first DWI acquisition for the human brain on a whole body scan.

To obtain diffusion weighted images, a pair of strong gradient pulses of a magnetic field, which defines the direction in which the diffusion is measured, is applied. The diffusion weighting sequence is commonly known as the so-called Stesjkal–Tanner sequence. Since the diffusion probability distribution function has assumed to be Gaussian, the attenuated signal of the Stesjkal–Tanner sequence in relation to $\mathbf{D}$ is specified as follows:

$$S(\mathbf{y}) = S_0 e^{-b\mathbf{y}^T\mathbf{D}\mathbf{y}}, \tag{8}$$

where $\mathbf{y}$ is a unit vector in the diffusion gradient direction and $S(\mathbf{y})$ is the associated signal. Here b represents the so-called b-value and is the diffusion weighting factor depending on scanner parameters and S_0 is the reference nuclear magnetic resonance signal.

Given multiple diffusion weighted images, we can measure quantitative scalars such as the apparent diffusion coefficient (ADC), which describe the diffusion process. The ADC is given by the relation

$$D(\mathbf{y}) = -\frac{1}{b}\ln\left(\frac{S(\mathbf{y})}{S_0}\right). \tag{9}$$

The ADC in anisotropic tissues varies depending on the direction $\mathbf{y}$ in which it is measured; see Figure 3. To model the intrinsic diffusion properties of biological tissues, Basser et al. proposed to fit the DWI data to a second order symmetric and positive-definite tensor $\mathbf{D}$ [4]. To this end, one can write the relation between the ADC and the diffusion tensor $\mathbf{D}$ as follows:

$$D(\mathbf{y}_i) = \mathbf{y}_i^T\mathbf{D}\mathbf{y}_i = \sum_{\beta=1}^{3}\sum_{\alpha=1}^{3} d_{\alpha\beta}y_i^\alpha y_i^\beta, \tag{10}$$

for $i = 1, 2, \ldots, n$ with n the number of sampled gradient directions. Using relation (10), the six unknown coefficients of the diffusion tensor $\mathbf{D}$ can be computed by choosing at least six gradient directions, typically we take $20 \le n \le 60$. Relation (10) gives rise to an over-determined system and can be solved

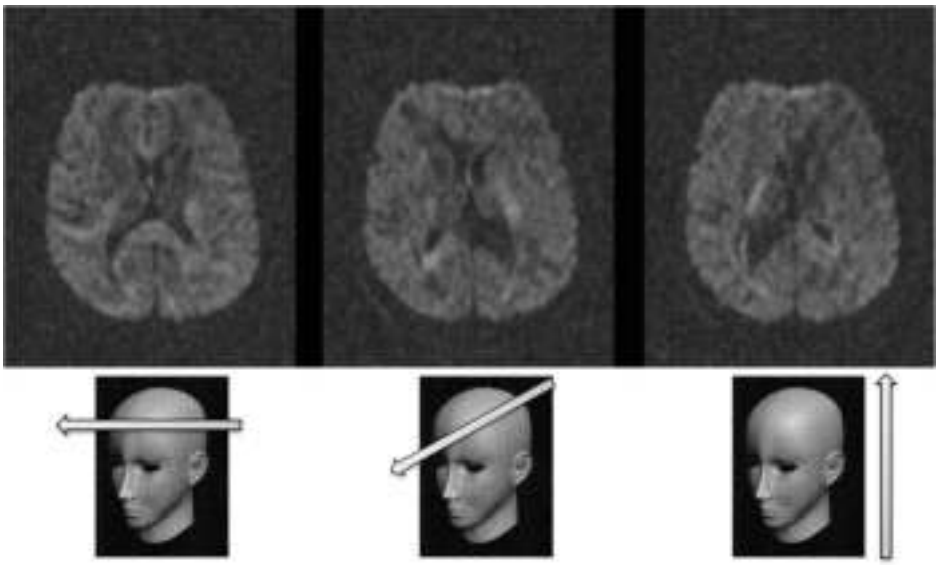


Figure 3 Diffusion-weighted images with three different acquisition directions. Note the differences in contrast as the gradient direction is changing. Arrows indicate the gradient directions. Adapted from Campbell [8].

using least squares fitting [1]. This is in fact the same diffusion tensor (DT) as introduced earlier in Einstein's equation (5) for anisotropic diffusion.

The DT is determined by its three eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$ and its three corresponding orthogonal eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. The largest eigenvalue λ_1 gives the principal direction $\mathbf{e}_1$ of the diffusion tensor. Note that the other two eigenvectors span the plane orthogonal to the main eigenvector. Using the three eigenvalues and their corresponding eigenvectors a DT can be visualized as an ellipsoid which corresponds to the implicit surface $\{\mathbf{R} : \mathbf{R}^T\mathbf{D}^{-2}\mathbf{R} = \text{const}\}$ [3, 7]. Figure 4(a) and 4(b) are illustrations of these procedures. Figure 4(c) shows the diffusion tensor field for a slice of a brain image.

Diffusion tensor images are useful when the tissue of interest is dominated by isotropic water movement such as grey matter in the cerebral cortex, where the diffusion time appears to be the same along any axis. Therefore, they are for example applicable to diagnose vascular strokes in the brain. However, in the cases where the direction and shape of the diffusion propagation is important, the resulting image using this technique is difficult to interpret directly and does not provide much information about the underlying fibrous structure. This is particularly crucial for analysing the white matter structure. Therefore, further reconstruction techniques have been developed in order to extract more useful information from these images.

Reconstruction of brain fibre tracts

Due to the fibrous structure of white matter, diffusion of water molecules is dominant in the direction of the fibres. As we described before, diffusion and its directional variation can be measured by DWI. The process of reconstructing fibres using DWI is commonly known as tractography or fibre tracking.

In clinics, the most commonly used DTI tractography algorithms are principal diffusion direction (PDD) methods [8] where the fibres are integrated along the main eigenvector field $\mathbf{e}_1(\mathbf{x})$ of the diffusion tensor. This is numerically equivalent to solving the initial value problem

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{e}_1(\mathbf{x}(t)), \quad t > 0, \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases} \tag{11}$$

Here $\mathbf{x}_0$ denotes the initial position or seed point and t is the time. The initial value problem (11) can be solved using a fourth-order Runge–Kutta method. Figure 5 illustrates an example of PDD tractography.

The PDD methods just employ local information and are therefore sensitive to noise. Small changes can produce completely different results or undesired fibre pathways; see Figure 6. A relatively small amount of noise in the diffusion tensor field causes accumulative errors in the trajectory of the fibres. Tackling this problem in tractography algorithms has been a main inspiration for introducing many variations of PDD tractography. Moreover, it has been recently shown that the expected properties of actual fibres, such as fanning, cannot be reconstructed using PDD based tractography [5]. To overcome this problem, advanced models, such as global geometric tractography methods, were developed to deduce connectivity in the white matter by globally optimizing a certain cost function on the basis of the diffusion tensor information. The goal of global geometric tractography is to find optimal paths that connect two given regions/points. This can potentially overcome accumulative errors in-

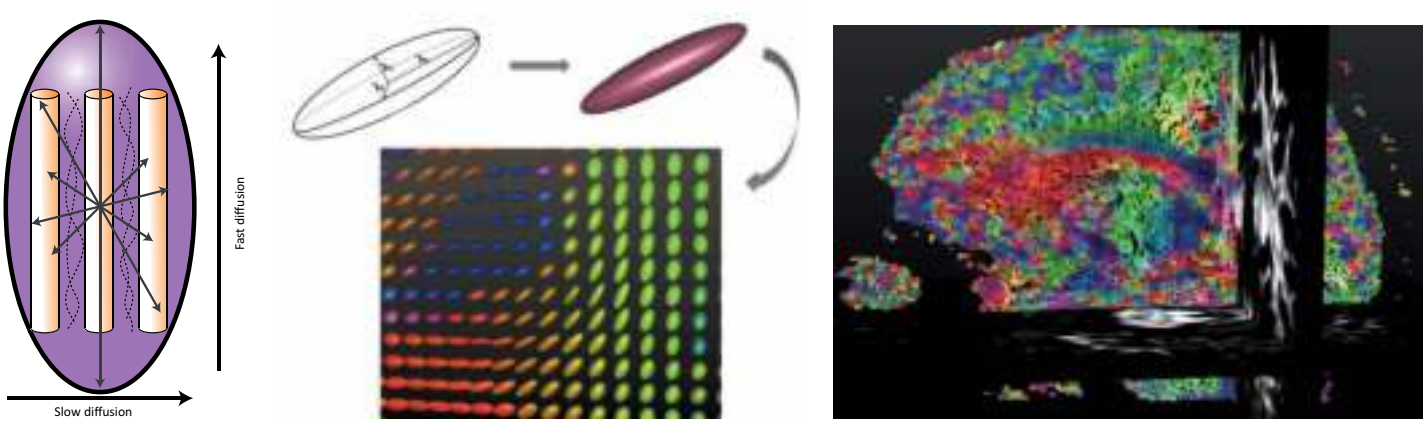
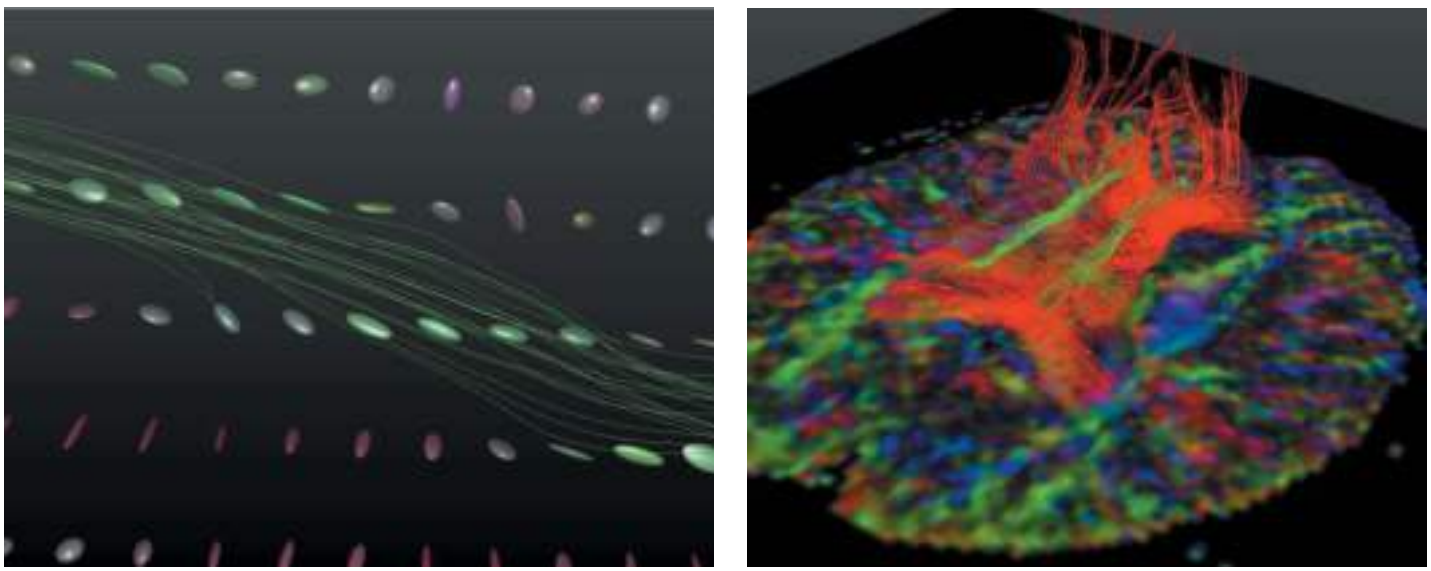


Figure 4 (a) Restricted diffusion process. (b) DT ellipsoids. (c) Saggital view of DT ellipsoids generated for a healthy Human Brain.



(a) (b)

Figure 5 (a) Small group of fibres generated using PDD tractography. The main eigenvectors of the diffusion tensors determine the local orientation of the fibres. (b) fibres generated by PDD tractography showing part of the cingulum and the corpus callosum.

troduced in tractography due to local noise. Besides, these models use the whole diffusion tensor profile instead of reducing this information into a single direction of the main eigenvector.

Geodesics for tractography

In order to reconstruct the globally optimal pathways we assume that fibre tracts coincide with geodesics in the Riemannian manifold defined using the inverse of the diffusion tensor as metric. The rationale behind this assumption is that water molecules move freely along fibre tracts, and their movement is restricted in the perpendicular directions. Therefore, it is assumed that the fibre connecting two points follows the most efficient diffusion path for water molecules. We are searching for a path that maximizes diffusion. This can be achieved by inverting the metric that converts the largest eigenvalue into the smallest one. Therefore, we choose $\mathbf{G} = (g_{\alpha\beta}) = \mathbf{D}^{-1}$ with $\mathbf{D}$ defined in (5). Consequently, the geodesics for this metric represent the fibres [20].

Thus, consider a bounded curve $\mathcal{C}$, with parametrization $\mathbf{x} = \boldsymbol{\chi}(\tau)$, $a \leqslant \tau \leqslant b$. A geodesic between two points $\boldsymbol{\chi}(a)$ and $\boldsymbol{\chi}(b)$ is the smooth curve whose length is the minimum of all possible lengths. In the following we use the Einstein notation, i.e., we sum over repeated indices, one in the upper (superscript) and one in the lower (subscript) position. For a general metric $\mathrm{d}s^2 = g_{\alpha\beta}\mathrm{d}x^\alpha \mathrm{d}x^\beta$, the length of $\mathcal{C}$ is given by

$$J[\boldsymbol{\chi}] = \int_{\mathcal{C}} \mathrm{d}s = \int_a^b \left(g_{\alpha\beta}(\boldsymbol{\chi}(\tau)) \cdot \dot{\chi}^\alpha(\tau)\dot{\chi}^\beta(\tau) \right)^{1/2} \mathrm{d}\tau. \tag{12}$$

The metric tensor $(g_{\alpha\beta})$ only depends on $\mathbf{x}$, and is symmetric positive definite. The solution to the so-called geodesic equations minimizes $J[\boldsymbol{\chi}]$. These are given by

$$\ddot{x}^\alpha + \Gamma^\alpha_{\beta\gamma}\dot{x}^\beta\dot{x}^\gamma = 0, \tag{13}$$

where $\Gamma^\alpha_{\beta\gamma}$ is the Christoffel symbol of the second kind, defined by

$$\Gamma^\alpha_{\beta\gamma} = g^{\alpha\delta}[\beta\gamma, \delta], \tag{14}$$

where $[\beta\gamma, \alpha]$ is the Christoffel symbol of the first kind, and given by

$$[\beta\gamma, \alpha] = \frac{1}{2}\left(\frac{\partial g_{\alpha\beta}}{\partial x^\gamma} + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta} - \frac{\partial g_{\beta\gamma}}{\partial x^\alpha} \right). \tag{15}$$

Alternatively, we consider the functional that minimizes the length of all curves joining the fixed point $\boldsymbol{\chi}(a)$ and the time variable end point $\boldsymbol{\chi}(t)$, i.e.,

$$T(\mathbf{x}, t) = \min_{\boldsymbol{\chi}} \int_a^t \left(g_{\alpha\beta}(\boldsymbol{\chi}(\tau)) \cdot \dot{\chi}^\alpha(\tau)\dot{\chi}^\beta(\tau) \right)^{1/2} \mathrm{d}\tau, \tag{16}$$

with $\mathbf{x} = \boldsymbol{\chi}(t)$. The geodesic connecting $\boldsymbol{\chi}(a)$ with $\boldsymbol{\chi}(t)$ can be determined from the Hamilton–Jacobi (HJ) equation, given by

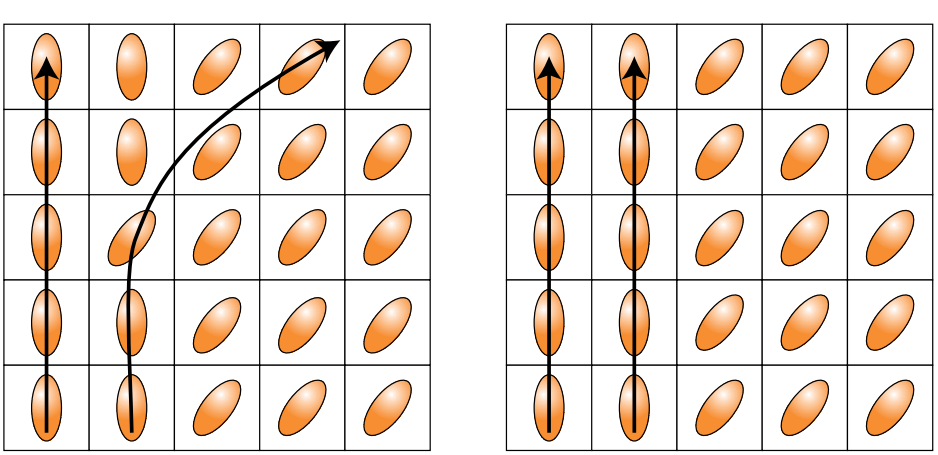
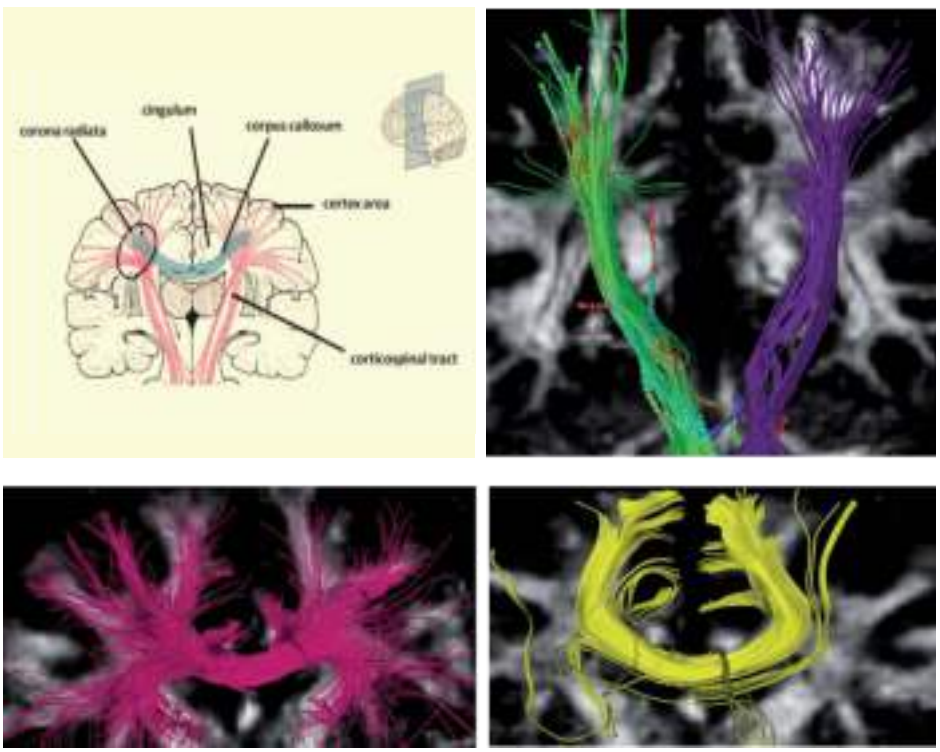


Figure 6 Illustration of the influence of noise in PDD tractography. Small changes in the direction of the tensor can cause deviation of the fibre from the actual pathway.



(a)

(b)

Figure 7 Illustration of a healthy human brain fibre bundle reconstructions. (a) The motor tracts in red and corpus callosum tracts in blue modified from [11] (top-left), fibres reconstructed for corticospinal tracts using multi-valued geodesics (top-right) and postcentral gyri areas of the corpus callosum using PDD (bottom-right) and multi-valued geodesics (bottom-left). (b) Results for the complete area of corpus callosum multi-valued geodesics.

$$H\left(\mathbf{x}, \frac{\partial T}{\partial \mathbf{x}}\right) = 1, \tag{17}$$

where the Hamiltonian H is given by [15]

$$\begin{aligned} H^2(\mathbf{x}, \mathbf{p}) &= g^{\alpha\beta}(\mathbf{x}) p_\alpha p_\beta, \\ p_\alpha &:= g_{\alpha\beta}(\mathbf{x}) \dot{x}^\beta. \end{aligned} \tag{18}$$

The HJ-equation may generate multi-valued solutions when, for example, there are discontinuities in the gradient field. Therefore, the viscosity solution is needed to ensure the existence and uniqueness of the solution to the HJ-equation [13]. This implies the viscosity solution is the minimum time; i.e. the first arrival time, for any curve from a given initial point to reach any other points inside the domain. Using the viscosity solution will not ensure that the solution we obtain is the real physically meaningful one; e.g., shortcuts when they are not desired. In order to tackle this issue, multi-valued solutions of the arrival time can be approximated by computing the geodesics directly from the geodesic equations [16–17].

The geodesic equation (13) can be rewritten as the system of ordinary differential equations

$$\begin{aligned} \dot{x}^\alpha &= u^\alpha, \\ \dot{u}^\alpha &= -\Gamma^\alpha_{\beta\gamma} u^\beta u^\gamma. \end{aligned} \tag{19}$$

Consider $(x^1(0), x^2(0), x^3(0))$ as given initial point and $(u^1(0), u^2(0), u^3(0))$ as initial direction. We compute the solution of (19) for the given initial position and multiple initial directions using the standard fourth order explicit Runge–Kutta method. This gives us a set of geodesics for the given initial point and integrate till they hit the boundary of a given domain. Here, the domain is the outer surface of the brain.

Human brain fibre reconstruction

We applied our proposed multi-valued geodesic tractography to reconstruct the fibrous tissue structure of the underlying neural axons of the white matter of a healthy human brain. Using available atlases of the human brain map [14], we select the region of interests. Geodesics are then computed until they meet one of the boundaries. To determine the fibre connecting two given regions we apply the line-plane intersection [17]. This allows us to cut off the geodesics once they enter one of the selected end regions.

Figure 7(a) shows the geodesics reconstructed for corticospinal tracts (top-right) and postcentral gyri areas of the corpus callosum using PDD (bottom-right) and multi-valued geodesics (bottom-left). Figure 7(b) illustrates the results for the complete area of corpus callosum multi-valued geodesics.

Since there is no available ground truth for fibre bundles, simulated diffusion tensor data sets or white matter brain atlases are used for validating the tractography methods. We validate the results for our method with simulated phantoms and the report can be found in [16–17]. Moreover, multi-valued geodesics tractography has been applied for various human brain data sets. According to clinical experts, multi-valued geodesics are more coherent with expected fibre tracts associated with the underlying bundles. Our model reconstructs the fanning tracts, particularly the ones connecting the cortex area, while those were completely missing using the PDD approach. Our proposed model has been integrated as a part of Vist/e biomedical visualization software and is publicly available, see https://sourceforge.net/projects/viste.

Future work

Despite the simplicity of the DTI model, tractography techniques using the DT are shown to be very promising to reveal the structure of brain white matter. However, DTI assumes that each voxel contains fibres with only one single main orientation and it is known that brain white matter has multiple fibre orientations, which can be in many directions. High angular resolution diffusion imaging (HARDI) acquisition and its modelling techniques have been developed to overcome this limitation. The models applied to HARDI data result in a function on the sphere that gives information about the diffusion profile within the voxel [10]. An ongoing extension of geodesic based tractography models is to apply the previously discussed geodesic based models for the HARDI data [18]. Nevertheless, DTI is still widely used in clinical research due to either unavailability of the scanning protocols for HARDI or computationally expensive data processing. Therefore, improving existing methods and algorithms for DTI processing is beneficial for clinical purposes.

Referenties

1 P.J. Basser et al., Estimation of the effective self-diffusion tensor from the NMR spin echo, *J. Magn Reson B.* 103(3) (1994), 247–254.

2 P. Basser et al., MR diffusion tensor spectroscopy and imaging, *Biophysical Journal* 66(1) (1994), 259â‰-267.

3 P.J. Basser et al. Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review, *NMR Biomed.* 15(7-8) (2002), 456–467.

4 P.J. Basser et al., Spectral decomposition of a 4th-order covariance tensor: Applications to diffusion tensor MRI, *Signal Processing* 87 (2007), 220–236.

5 M. Bastiani et al., Human cortical connectome reconstruction from diffusion weighted MRI: the effect of tractography algorithm, *Neuroimage* 62(3) (2012), 1732–1749.

6 C. Beaulieu, The basis of anisotropic water diffusion in the nervous system – a technical review, *NMR Biomed.* 15 (2002), 435–455.

7 T. Brox et al., Nonlinear structure tensors, *Image and Vision Comput.* 24(1) (2006), 41–55.

8 J.S.W. Campbell, *Diffusion Imaging of White Matter Fibre Tracts*, PhD thesis, McGill University, 2004.

9 A. Einstein, *Investigations on the Theory of the Brownian Movement*, Dover Publications, 1956.

10 A. Fuster et al., Riemann–Finsler geometry for diffusion weighted magnetic resonance imaging, in *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization, Springer, 2014, pp. 189–208.

11 H. Gray, *Anatomy of the Human Body*, The Bartleby edition, 1918.

12 D. LeBihan et al., Imagerie de diffusion in-vivo par résonance magnétique nucléaire, *Physics in Med. and Bio.* 30(4) (1985).

13 C. Mantegazza et al., Hamilton–Jacobi equations and distance functions on Riemannian manifolds, *App. Math. and Optim.* 47(1) (2002), 1–25.

14 S. Mori et al., *MRI Atlas of Human White Matter*, Elsvier, 2005.

15 H. Rund, *The Hamilton–Jacobi Theory in the Calculus of Variations: Its Role in Mathematics and Physics*, Van Nostrand, 1966.

16 N. Sepasian, *Multi-valued Geodesic Tractography for Diffusion Weighted Imaging*, PhD thesis, Eindhoven University of Technology, 2011.

17 N. Sepasian et al., Multivalued geodesic raytracing for computing brain connections using diffusion tensor imaging, *SIAM Journal on Imaging Sciences* 5(2) (2012), 483–504.

18 N. Sepasian et al. Riemann–Finsler multivalued geodesic tractography for HARDI, in *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*, Mathematics and Visualization, Springer, 2014, pp. 209–225.

19 D.G. Taylor et al., The spatial mapping of translational diffusion coefficients by the nmr imaging technique, *Physics in Med. and Bio.* 30(4) (1985).

20 V. Wedeen et al., Mapping fibre orientation spectra in cerebral white matter with Fourier transform diffusion MRI, *Proceedings of the Intern. Soc. of Mag. Res. in Med.*, 2000.

Jacob Fokkema
Technische Universiteit Delft
j.t.fokkema@tudelft.nl

Eugène Bernard
Ons Middelbaar Onderwijs, Tilburg
e.bernard@omo.nl

Petra de Bont
NWO, Den Haag
p.debont@nwo.nl

Frank den Hollander
Universiteit Leiden
denholla@math.leidenuniv.nl

Christiane Klöditz
NWO, Den Haag
c.kloditz@nwo.nl

Barry Koren
Technische Universiteit Eindhoven
b.koren@tue.nl

John Koster
ASML, Veldhoven
john.koster@asml.com

Jan Karel Lenstra
Centrum Wiskunde & Informatica, Amsterdam
jan.karel.lenstra@cwi.nl

Ieke Moerdijk
Radboud Universiteit Nijmegen
i.moerdijk@math.ru.nl

Gerrit Timmer
ORTEC, Zoetermeer
gtimmer@ortec.nl

Nellie Verhoef
Universiteit Twente
n.c.verhoef@utwente.nl

Nieuws

Een Deltaplan voor de Nederlandse wiskunde

Het Platform Wiskunde Nederland (PWN) publiceerde in 2014 zijn Visiedocument, dat de visie en de ambitie van de Nederlandse wiskunde voor de middellange termijn presenteert. Het Ministerie van Onderwijs, Cultuur en Wetenschap verzocht NWO en PWN de aanbevelingen uit het Visiedocument uit te werken, concrete acties te formuleren en zich te verzekeren van draagvlak voor de realisering daarvan. NWO en PWN hebben een commissie onder leiding van Jacob Fokkema, voormalig rector magnificus van de Technische Universiteit Delft, gevraagd dit werk op zich te nemen. De Commissie Fokkema geeft in dit artikel een samenvatting van het Deltaplan voor de Nederlandse wiskunde.

Het Deltaplan [2] is gebaseerd op consultaties met rectoren, decanen, wiskundigen, industriële partners en gerelateerde partijen. Inspiratiebronnen zijn onder andere de Europese Onderzoeks- en Innovatieagenda Horizon 2020, de Wetenschapsvisie van de overheid [3], de Strategische Agenda Hoger Onderwijs en Onderzoek 2015–2025 van het Ministerie van OCW [4] en de Nationale Wetenschapsagenda [5].

Het Deltaplan stelt een geïntegreerde aanpak voor van onderzoek, onderwijs, lerarenopleiding, en maatschappij en innovatie. Een profilering van het wiskundeonderzoek via de clusters, verdere samenwerking in het universitaire wiskundeonderwijs, vernieuwing van de lerarenopleiding, aandacht voor het innovatieve vermogen van de wiskunde en een sterke organisatie van de discipline weerspiegelen de ambities van de Wetenschapsvisie. De Nederlandse wiskunde is bij uitstek in staat prominente vragen uit de Nationale Wetenschapsagenda te adopteren en een stevige rol te spelen in de publiek-private samenwerking.

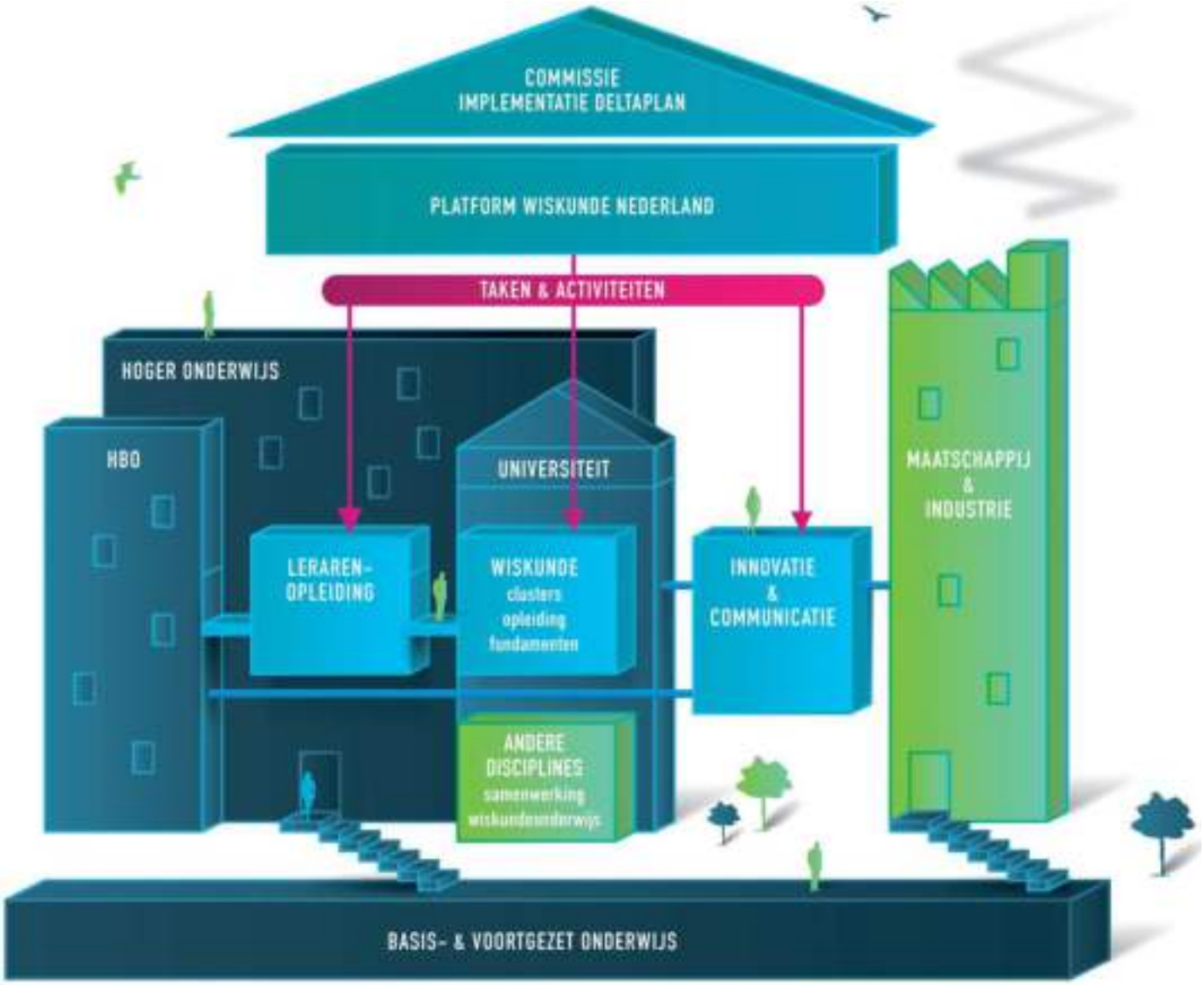
Het Deltaplan geeft een beschrijving en analyse van de Nederlandse wiskunde in relatie tot zijn omgeving en doet voorstellen tot acties. Het beeld dat daarbij ontstaat is samengevat in Figuur 1. De voorgestelde acties hebben betrekking op de drie centrale lichtblauwe velden. Voor een nadere onderbouwing en uitwerking van de acties verwijzen wij de lezer naar het volledige Deltaplan.

Onderzoek in het wiskundehuis

Wij bouwen een wiskundehuis rond drie recente landelijke verworvenheden: de wiskundeclusters, Mastermath en PWN; zie Figuur 2. Het huis rust op de wiskundige fundamenten en heeft bruggen naar andere disciplines en naar maatschappij en innovatie.

In internationaal opzicht is het wiskundehuis klein. Het groeiende aantal wiskundestudenten, de stagnerende omvang van de wiskundestaf, de teruglopende deelname van Nederland aan het mondiale wiskundeonderzoek en de geringe financiering van individuele promotieplaatsen tasten de fundamenten van het Nederlandse wiskundehuis aan. Om dit proces ten goede te keren en het personeelstekort op te heffen, vragen wij de Colleges van bestuur, de decanen en NWO om een *geleidelijke uitbreiding van de universitaire wiskundestaf en van het aantal promotieplaatsen*.

De vier wiskundeclusters hebben de laatste tien jaar gezorgd voor meer focus van het onderzoek en meer samenwerking tussen wiskundigen. Om de bundeling van krachten van de Nederlandse wiskunde en haar inter-



Figuur 1 De wiskunde in zijn omgeving.

nationale positie een extra impuls te geven, vragen wij het Ministerie van OCW en NWO om *structurele financiering van de clusters*. Deze zal de clusters in staat stellen bij te dragen aan de profilering van het wiskundeonderzoek en aan de aansluiting van de wiskunde bij de economische topsectoren.

Een belangrijk voorstel is de oprichting van een *nationale onderzoekschool Mastermath*, op basis van een evaluatie van de bestaande landelijke samenwerking in het masteronderwijs. Het nieuwe Mastermath verzorgt een breed onderwijsaanbod voor masterstudenten en promovendi, alsmede *cursussen op grensvlakken* met andere disciplines en *bijscholingscursussen* voor leraren en voor het bedrijfsleven. Mastermath stimuleert *alternatieve promotiemodellen*, zoals duopromoties en industrial doctorates. Het werkt nauw samen met de clusters en de instituten en zorgt in overleg met de decanen voor een goede aansluiting op de lokale graduate schools. Alleen een hechte landelijke samenwerking kan de vereiste inhoud, breedte en kwaliteit bieden.

Wij bepleiten tevens het *promotierecht voor ud's en uhd's*, nog meer aandacht voor de *deelname van vrouwen* aan de wiskunde, en het verder uitbouwen van het *Nederlands Mathematisch Congres* (NMC) tot een grote bijeenkomst, die prominente aandacht geeft aan de activiteiten van de clusters, aan de beoefening en het gebruik van de wiskunde buiten de eigen kring en aan ontwikkelingen in het onderwijs. Ook doen wij voorstellen tot het beter benutten van de kansen die *Europese programma's* aan de Nederlandse wiskundigen bieden.

De landelijke coördinatie van het wiskundehuis ligt in handen van een door OCW in te stellen *Commissie Implementatie Deltaplan Wiskunde* (CIΔ) en PWN. De CIΔ bepleit het belang van de wiskunde bij overheid en bedrijfsleven en verzorgt de monitoring van de in het Deltaplan voorgestelde acties. Zij voert regelmatig overleg met de decanen en helpt hen en de wiskundigen keuzen te maken. *PWN* ontwikkelt zich tot een platform voor alle Nederlandse wiskundigen werkzaam in onderzoek, onderwijs en industrie. Het stelt een businessplan op voor zijn bestuurlijke en financiële opschaling. Er komt een bureau van passende omvang dat professionele ondersteuning verzorgt van landelijke activiteiten, waaronder Mastermath.

Onderwijsvernieuwing

De universitaire wiskundeopleidingen kennen veel dynamiek qua inhoud en inrichting en een toenemende samenwerking. Wij stel-

len voor dat zij in het eerste jaar een college opnemen over de *wiskunde in zijn omgeving* en ook daarnaast aandacht besteden aan presentatietechnieken en communicatieve vaardigheden. Tevens vragen wij om aandacht voor het *wiskundeonderwijs in andere opleidingen*. De betrokkenheid van wiskundigen daarbij is op veel plaatsen gering, terwijl de rol van ons vak in andere disciplines groeit. Het is aan de wiskunde-instituten dit onderwerp te agenderen bij hun decanen en Colleges van bestuur, met integrale financiering van het service-onderwijs als uitgangspunt.

Wat betreft het voortgezet onderwijs pleiten wij ervoor *docenten* te betrekken bij universitair onderzoek en *studenten en promovendi* in te zetten in de bovenbouw van het vwo. Wij vragen OCW een *permanente curriculumcommissie* voor de wiskunde in te stellen.

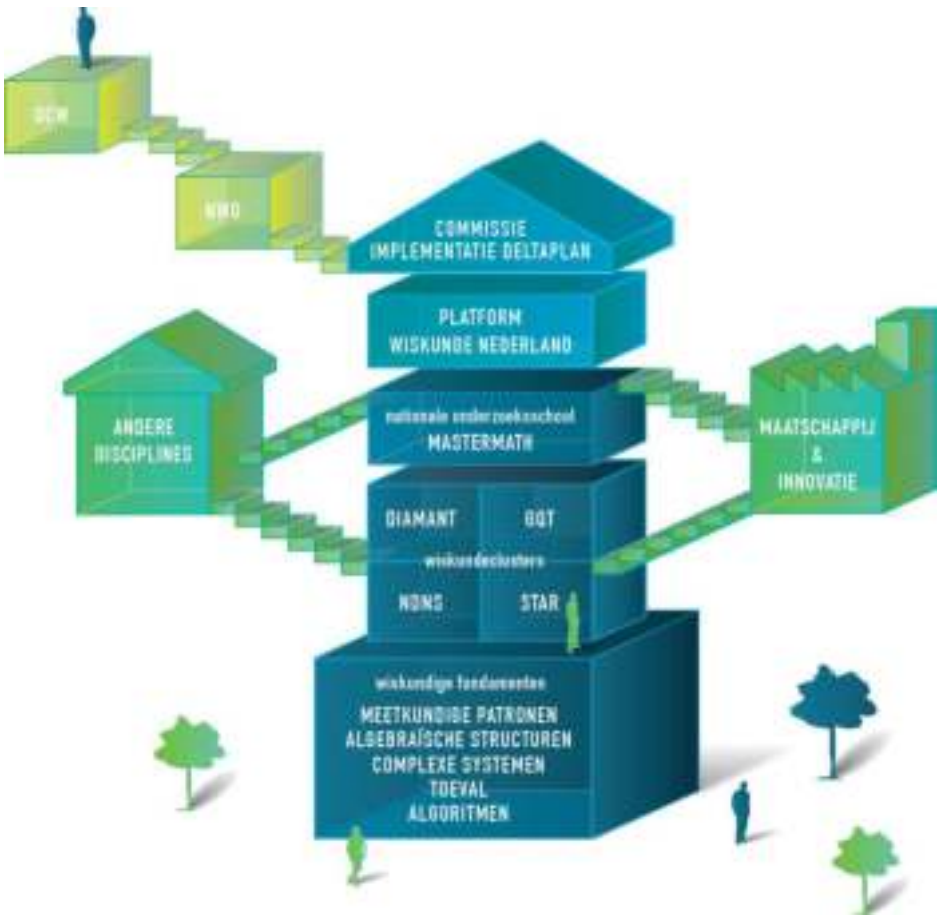
Lerarenopleiding

Om de *universitaire vorming van leraren* te bevorderen, dienen de universiteiten de lerarenopleiding een prominente plaats in hun beleid en communicatie te geven en een premie te stellen op het behalen van de eerstegraadsbevoegdheid als deel van de master of de promotie. De wiskunde-instituten moeten samen met de instituten voor de lerarenopleiding aantrekkelijke en flexibele routes naar de eerstegraadsbevoegdheid creëren en de vakdidactiek dichter bij de vakinhoud plaatsen.

Het aanbod van *nascholing* is divers maar onoverzichtelijk, de vraag ernaar is groot. Wij stellen voor dat PWN, samen met de Nederlandse Vereniging van Wiskundeleraren en de vaksteunpunten wiskunde, een digitale nascholingscatalogus opzet en een financieel stabiel model ontwikkelt voor de organisatie van nascholingscursussen en voor de certificering van eerste- en tweedegraadsbevoegdheden wiskunde.

Maatschappij en innovatie

Een enquête onder twintig bedrijven leert dat er behoefte is aan meer structurele contacten tussen bedrijfsleven en academia. Naast de succesvolle Studiegroep Wiskunde met de Industrie denken wij aan regionale bedrijvendagen, een symposium over wiskunde en innovatie als onderdeel van het NMC, en lifelong learning activiteiten voor het bedrijfsleven, verzorgd door Mastermath.



Figuur 2 Het wiskundehuis.

De positie van de wiskunde in de publiek-private samenwerking en in het bijzonder de *aansluiting bij de economische topsectoren* moet worden versterkt. De decanen kunnen gezamenlijk en in overleg met de CIΔ strategische thema's voor de wiskunde identificeren en het opstellen van wiskundige roadmaps initiëren, die door de wiskundeclusters worden geadopteerd. PWN kan helpen de publieke en private partners tot elkaar te brengen.

Ten slotte merken we op dat er veel goeds gebeurt op het gebied van *communicatie en outreach*. De activiteiten zijn echter versnipperd. Wij stellen voor dat PWN in samenwerking met de communicatiemedewerkers van de instellingen een professionele communicatiegroep opzet, die een strategie ontwikkelt om met één gezicht en één boodschap tot de verschillende doelgroepen te spreken, eigen activiteiten ontplooit en lokale initiatieven stimuleert.

Dankwoord
Wij zijn veel dank verschuldigd aan 177 gesprekspartners: rectoren, bètadecanen en 3TU.AMI-decanen, instituuts- en clusterdirecteuren, de leiding van Mastermath, WONDER en PWN, diverse groepen wiskundigen, twintig industriële partners, medewerkers van de Ministeries van OCW en EZ en van NWO, de leiding van het Platform Bèta Techniek, het ICT-onderzoek Platform Nederland en het Nederlands Instituut voor Biologie, en vele anderen.

Referenties

1 *Formulas for Insight and Innovation; Mathematical Sciences in the Netherlands: Vision document 2025*, Platform Wiskunde Nederland, Amsterdam, 2014.

2 *Een Deltaplan voor de Nederlandse wiskunde*, NWO, Den Haag, en Platform Wiskunde Nederland, Amsterdam, 2015.

3 *Wetenschapsvisie 2025: keuzes voor de toekomst*, Ministerie van OCW, Den Haag, 2014.

4 *De waarde(n) van weten: Strategische Agenda Hoger Onderwijs en Onderzoek 2015–2025*, Ministerie van OCW, Den Haag, 2015.

5 *Nationale Wetenschapsagenda*, Ministerie van OCW, Den Haag, wordt verwacht in herfst 2015.

Arjeh Cohen
Department of Mathematics and Computer Science
Technische Universiteit Eindhoven
a.m.cohen@tue.nl

Afscheidsrede

Wiskunde in het web

Na bijna 22 jaar hoogleraarschap nam Arjeh Cohen op 28 maart 2014 afscheid van de Technische Universiteit Eindhoven. In 1992 werd hij er benoemd tot hoogleraar discrete wiskunde aan de Faculteit Wiskunde en Informatica. Dit artikel is een licht bewerkte versie van zijn afscheidsrede.

Tussen mijn afstuderen in 1971 en deze afscheidsrede zijn veel beroemde wiskundige vermoedens opgelost. Twee daarvan zijn zó bekend dat u er vast wel eens van hebt gehoord.

Het eerste is de Laatste Stelling van Fermat. In 1637 schreef Fermat, in de marge van een bladzij uit een boek, dat hij een bewijs had. Maar dat bewijs is nooit door een ander gevonden of gereconstrueerd. Het eerste algemeen geaccepteerde bewijs van de stelling komt van Wiles en is pas verschenen in 1995. Het vermoeden stond dus 358 jaar open.

Het tweede is de vierkleurenstelling. Die gaat over het zodanig kleuren van landkaarten dat geen enkel tweetal aangrenzende landen dezelfde kleur heeft. De uitspraak is dat je bij elke landkaart daarvoor met vier kleuren toekan. In 1840 vroeg Möbius al of elke landkaart vierkleurbaar is. Het is voor het eerst bewezen door Appel en Haken in 1976. De vraag stond dus 136 jaar open. Het bewijs van Appel en Haken berust op computerwerk. Het is een bewijs uit het ongerijmde. Dat betekent dat het bestaan van een kleinste landkaart die niet vierkleurbaar is, wordt aangenomen en dat daaruit een tegenspraak wordt afgeleid. Met bekende redeneringen kwamen Appel en Haken uit op een lijst van 1936 landkaarten. Ten minste één daarvan zou in een kleinste landkaart voor moeten komen die niet vierkleurbaar is. Met behulp van computerberekeningen hebben ze vastgesteld dat dit voor geen van de 1936 kaarten het geval is. Zo verkregen ze de gewenste tegenspraak.

Het contrast met de laatste stelling van Fermat kan haast niet groter zijn. Ten eerste is de rol van de computer in het bewijs van Wiles minimaal. Ten tweede strekken de gevolgen voor de wiskunde van zijn resultaten veel verder dan alleen Fermat. Ik vraag met deze observatie dus eigenlijk: wat maakt het ene resultaat beter dan het andere? Zijn de resultaten van Wiles belangrijker, interessanter, of dieper dan de vierkleurenstelling van Appel en Haken? Bestaat er zoiets als goede wiskunde?

Mijn promotor Springer zei lang geleden dat je als wiskundige wel aanvoelt wat goede wiskunde is. Misschien vraagt u zich af waarom ik deze vraag dan toch stel. Dat is omdat bestuurders en geldschieters ons al jaren analyseren met methoden als citatie-analyses, waarbij de beoordeling wel met maar niet door wiskundigen plaatsvindt. Daar voelen velen zich ongemakkelijk bij.

Ter aansporing tot een beoordeling geheel door wiskundigen zelf, stel ik vijf criteria voor. Aan de hand daarvan ga ik op een derde grote stelling in. Daarna gebruik ik de vijf criteria om de positieve bijdrage van het web aan de wiskunde te bespreken.

Vijf criteria

Volwassenheid
De nadruk die ik legde op de lange tijd dat de twee genoemde problemen open stonden, suggereert dat de lengte van die periode een criterium voor een goed wiskundig resultaat is. Maar een vraag kan best lang open staan omdat niemand zich er voor interesseert. Daarom is het verstandig dit criterium ruimer te nemen. Een goede omschrijving lijkt me volwassenheid van het probleem. Van de achterliggende vermoedens zijn al veel gevolgen geïdentificeerd en al veel pogingen tot falsificatie gestrand.

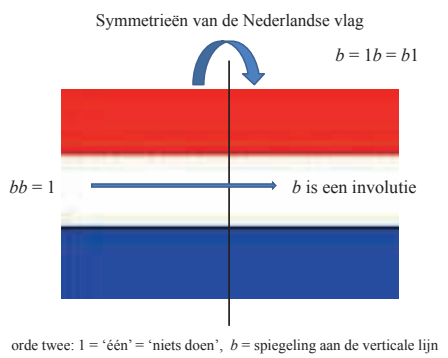
Het gaat bij het oplossen van een probleem duidelijk niet alleen om de vastgestelde uitspraak. Ook telt hoe het resultaat afgeleid wordt. Dit verwijst direct naar de uitdaging die wiskundigen in hun werk zien: slim zijn.

Resistentie
Zodra een bewijs gevonden is, kan het beeld van het achterliggende probleem drastisch veranderen. Fermat zelf verwees naar een bewijs dat hij had, en dat nooit door anderen gevonden is. Wat zou de waardering van de stelling geweest zijn als dat wel het geval was geweest? Hoeveel mensen zouden dan niet beweren dat het probleem eigenlijk triviaal is? Dat is wiskundig jargon voor ‘te flauw om los te lopen.’ Het is een favoriete bezigheid van wiskundigen om dat te roepen. Het bevestigt je status als slimmerd.

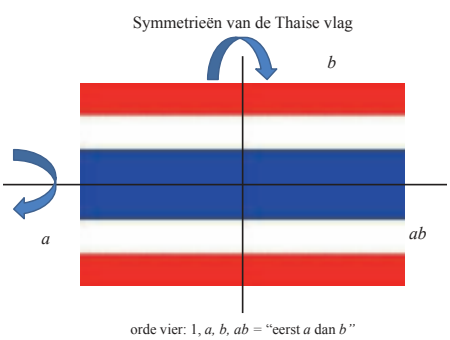
Het bewijs van een stelling is soms alleen maar lang omdat het onhandig is opgezet. De waardering van het resultaat kun je dus niet alleen maar aflezen aan de lengte van het bewijs. Daarom wil ik dit criterium graag als resistentie omschrijven: de weerstand tegen belagers die het bewijs proberen te vereenvoudigen.

Originaliteit

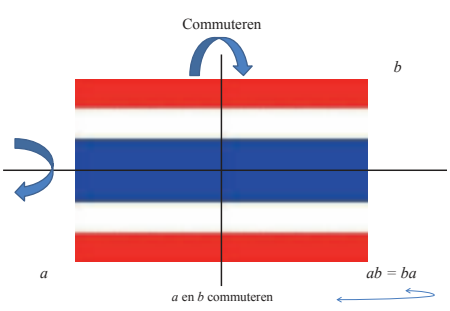
Het komt ook voor dat een bewijs drastisch ingekort wordt door verrassend gebruik van een sterke stelling uit een heel ander gebied. De waardering hiervoor deel ik in bij het criterium originaliteit. Het is natuurlijk ook van toepassing op het resultaat zelf. Goede wetenschap brengt een schok te weeg, een plots verkregen inzicht dat uitstijgt boven het normale. Nu zou je kunnen denken dat dat bij Fermat toch nauwelijks het geval is, omdat de uitspraak al eeuwen bekend is.



Figuur 1



Figuur 2



Figuur 3

Toch is de bewering al wonderbaarlijk voor iedereen die er voor het eerst naar kijkt, vooral omdat er voor exponent n gelijk aan twee heel veel gehele oplossingen zijn, wat sterk contrasteert met het ontbreken van dergelijke oplossingen voor n groter dan twee. De vierkleurenstelling biedt echter geen grote verrassing. De professionele kaartenmaker ziet bij vrijwel elke concrete landkaart snel hoe je vaak al met drie maar in ieder geval met vier kleuren toekunt.

Unificatie

Dicht bij de verrassing zit een ander aspect: goede wetenschap unificeert. Unificatie berust op abstractie en kracht. Abstractie helpt bij het terugbrengen van een model naar de kale essentie. Door alle afleidende ruis weg te snijden, krijg je beter boven tafel wat de kern van het probleem is. Bovendien kan een oplossing in meerdere situaties gebruikt worden.

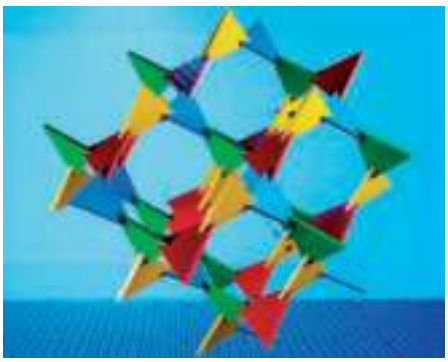
Een goed voorbeeld van abstractie is een netwerk. Als wiskundig concept is een netwerk een verzameling knooppunten waarvan bepaalde tweetallen met elkaar verbonden zijn, al of niet voorzien van informatie over de lengte of sterkte van die verbinding.

Of het netwerk nu een web voorstelt, een sociaal netwerk of een gasdistributie, maakt daarbij niet uit. De abstractie zorgt ervoor dat het model in verschillende situaties kan worden toegepast.

Een tweede voorbeeld is het wiskundige begrip groep. In mijn onderzoek heb ik daar veel mee te maken gehad. Het is een abstractie van het concept symmetrie. Een symmetrie van een meetkundige figuur is een transformatie die de figuur als geheel niet verandert.

De Nederlandse vlag, bijvoorbeeld, neergelegd in het platte vlak, heeft twee symmetrieën. Een daarvan is de spiegeling aan de verticale lijn door het midden van de vlag. Deze spiegeling geven we aan met b. De ander is 'niets doen', ofwel 'alles op zijn plaats laten'. Deze luie actie noemen we de één en geven we aan met het symbool 1, net als het cijfer. Zie Figuur 1.

De spiegeling b en de één samen vormen, zo zeggen we dan, een symmetriegroep die orde twee heeft. We kunnen de groep algebraïsch beschrijven met de symbolen 1 en b. Het uitvoeren van twee of meer symmetrieën achter elkaar heet een samenstelling. Samenstellingen noteren we als woorden in de symbolen, hier 1 en b. Dus bb staat voor het twee keer uitvoeren van de spiegeling b. De formule $bb = 1$ drukt uit dat twee keer spiegelen aan dezelfde lijn hetzelfde resultaat oplevert als niets doen. Elementen met deze eigenschap noemen we involuties. De formule $1b = b$ drukt uit dat samenstelling van eerst b en dan de één, weer de symmetrie b geeft. Immers, de 1 staat voor niets doen. Het geheel van symmetrieën en hun samenstellingen is een groep.



Figuur 4

De symmetriegroep van de Thaise vlag heeft orde vier. Zie Figuur 2. Naast de elementen 1 en b als in de Nederlandse vlag, hebben we nog een spiegeling a en de samenstelling van a en b.

Of we nu eerst a en dan b uitvoeren, of andersom, het resultaat hier is hetzelfde: $ab = ba$. Als deze gelijkheid geldt, zeggen we dat a en b commuteren. Zie Figuur 3. We zullen later zien dat dit lang niet altijd het geval is. De groep van symmetrieën bestaat hier dus uit de vier elementen 1, a, b en ab. Met andere woorden, de groep heeft orde vier.

Groepen worden gebruikt om verschillende patronen zoals kristalstructuren, naar hun symmetrieën te onderscheiden. Een mooi stukje van een dergelijk patroon hangt op onze campus. Zie Figuur 4.

Het begrip groep staat centraal in het Erlanger Programm, dat in 1872 door Felix Klein is gepubliceerd. Het gaat ervan uit dat alle relevante natuurkundige wereldbeelden afgeleid kunnen worden van een speciaal soort oneindige groepen, die Lie-groepen genoemd worden, naar de Noorse wiskundige Sophus Lie. In hun zoektocht naar de vereniging van de vier fundamentele krachten in één unificerend model, hebben natuurkundigen al verschillende Lie-groepen gebruikt.

Naast abstractie koppel ik aan unificatie het begrip kracht. Dat drukt uit in hoeveel gevallen de stelling een uitspraak van betekenis doet. Bij toegepaste wiskunde kun je de kracht van een resultaat meten aan situaties waar het een voorspellende uitspraak over doet. In meer abstracte wiskunde kun je

tellen hoeveel andere resultaten direct uit de gegeven stelling volgen.

Fermat staat hierin sterk ten opzichte van de vierkleurenstelling. Taylor, die meewerkte aan de laatste loodjes van het bewijs van Fermat, heeft later Wiles' methoden gebruikt om een erg krachtige uitspraak te bewijzen. In tegenstelling tot zijn kleurige karakter, zijn van de vierkleurenstelling weinig consequenties bekend, noch in de wiskunde noch daarbuiten.

Waardering

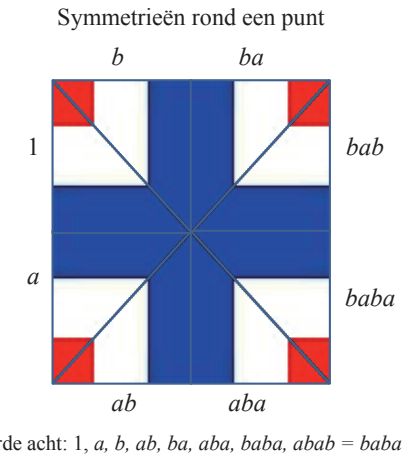
Het laatste criterium ter beoordeling van wiskundige resultaten dat ik wil voorstellen, is de waardering van de vakgenoten.

Wiskundigen zien zichzelf graag als onafhankelijke denkers. Toch hebben ze vrijwel allemaal een bepaalde gemeenschap van vakgenoten als referentiekader. Een stel collega's van Wiles keek gezamenlijk het werk van hem na vóórdat het bewijs van Fermat gepubliceerd werd. De waardering van deze onderzoekers voor zijn werk en die speciale behandeling getuigen van een gemeenschap van vakgenoten die zeer stimulerend kan werken, zelfs voor een individualist als Wiles.

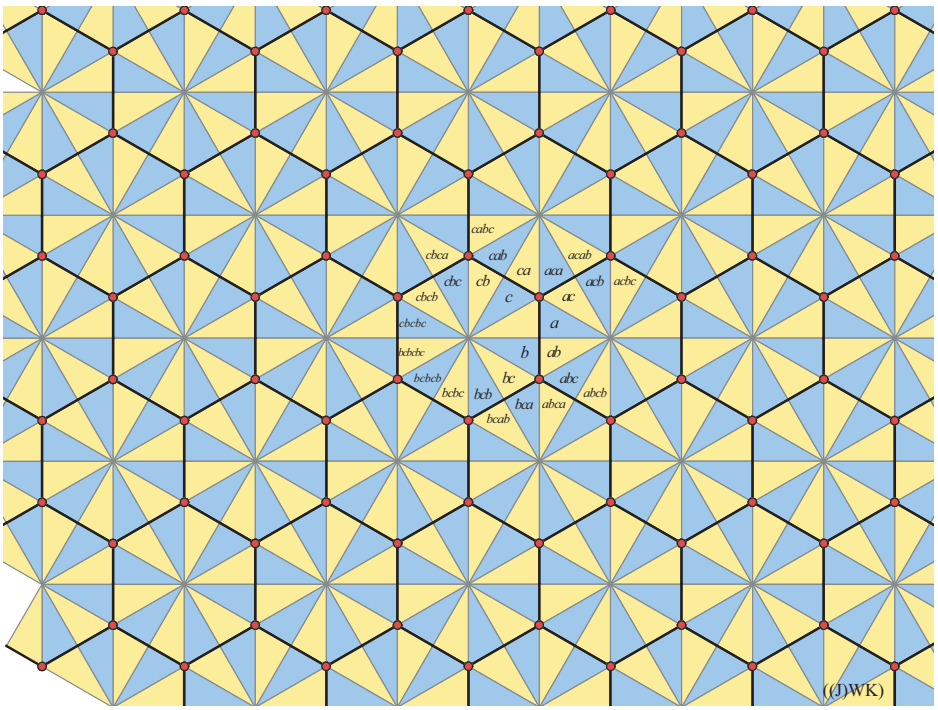
Hoewel ook wiskundigen aan hypes en modes blootgesteld zijn, bepalen de bevindingen van de eigen gemeenschap mede het duurzame belang van een stelling. Die waardering spreekt ook uit prijzen, waarvan er steeds meer in omloop zijn.

Conclusie

Zo zijn we tot vijf criteria gekomen om in te schatten hoe goed een wiskundig resultaat is. Ik heb de benamingen zo gekozen dat de eerste letters een aansprekend acroniem vormen: VROUW.



Figuur 5



Figuur 6

Eindige groepen

We gaan deze vijf criteria uitproberen op een derde groot resultaat: de classificatie van de eindige enkelvoudige groepen. Ik heb namelijk een groot deel van de geschiedenis van dichtbij meegemaakt en mijn favoriete onderzoeksterrein, een bepaald soort meetkunde, heeft hier veel mee te maken.

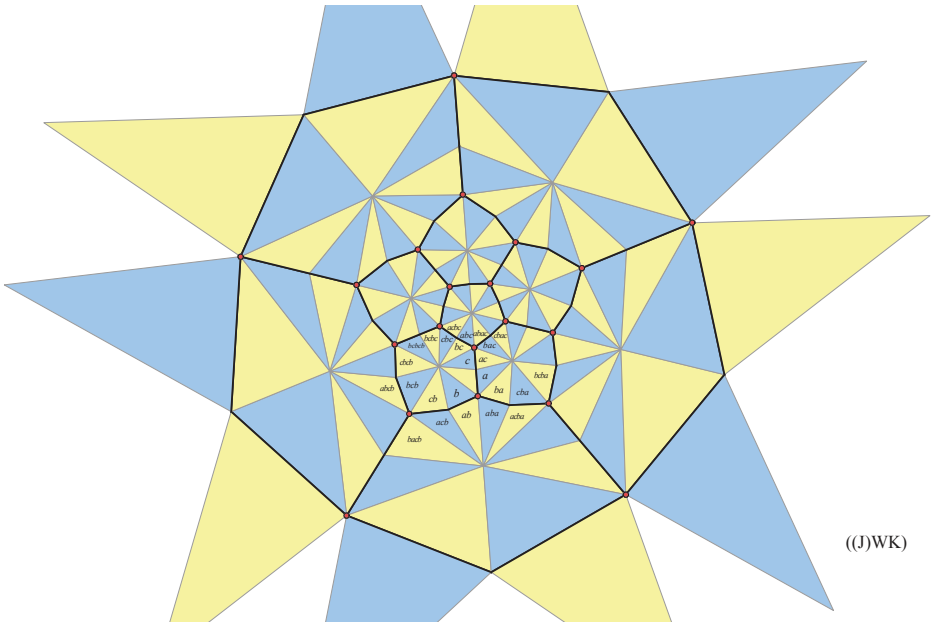
Om een idee te geven waar het vak over gaat, visualiseren we een abstracte groep met behulp van meetkunde. Denk weer even terug aan de twee vlaggen. De Nederlandse vlag heeft een symmetriegroep van orde twee en de Thaise van orde vier. Maar met behulp van de groep van symmetrieën van de Thaise vlag, kunnen we de Nederlandse vlag Thais maken. Daartoe passen we eerst twee spiegelingen toe op de Nederlandse vlag. De ene, die we b noemen, is de spiegeling aan de oostzijde van de vlag en de andere, die we a noemen, aan de zuidzijde van de vlag. Passen we daarna ook hun samenstelling toe, dan ontstaat de Thaise vlag. We kunnen dus aan de hand van een abstracte groep een concrete meetkundige figuur maken waarin alle symmetrieën van de groep zichtbaar zijn. Van de abstracte groep hebben we hier gebruikt dat de twee involuties a en b commuteren, zodat de twee samenstellingen — eerst de één en dan de ander of andersom — het zelfde resultaat leveren, namelijk de vierde vlag.

We gaan nu over op driehoeken in plaats van vlaggen. We nemen een driehoek uit de vlag waarin een hoek zit van 45 graden. Als we die driehoek gaan spiegelen om de zijden door een hoekpunt van 45 graden en dat blijven doen, ontstaan acht driehoeken op dat hoekpunt. Zie Figuur 5. Dit heeft te maken met het feit dat 45 gelijk is aan 360 gedeeld door 8. Niet alle elementen uit de groep commuteren met elkaar, maar de vier draaiingen wel. Deze groep heet de cyclische groep van orde vier. Net zo krijgen we, uitgaande van een driehoek met een hoek van $\frac{360}{2m}$ graden, voor elk natuurlijk getal m, de cyclische groep van orde m.

Dergelijke constructies van meetkunde en groep met spiegelingen voeren we nog drie keer uit, elke keer met een driehoek als uitgangspunt en met spiegelingen aan alle drie zijden van de driehoek in plaats van twee, zoals daarnet.

We kijken eerst naar een driehoek in het platte vlak met hoeken van 30, 60 en 90 graden. Zie Figuur 6, waarin we beginnen met de witte driehoek binnen de zijden waar a, b en c bij staan.

Steeds spiegelen we aan de zijden van bekende driehoeken om nieuwe driehoeken te maken. In sommige driehoeken staat een samenstelling van de spiegelingen a, b en c beschreven die de begindriehoek naar die driehoek transformeert. Rond elk hoekpunt ontstaan steeds vier driehoeken, zes driehoeken of twaalf driehoeken. Het eindresultaat is een betegeling van het platte vlak met driehoeken. Elke spiegeling wordt een symmetrie van de hele betegeling. De groep van symme-



Figuur 7

trieën en de meetkundige figuur ontstaan zo hand in hand. Er zijn oneindig veel driehoeken, en de groep heeft oneindig veel symmetrieën. Met andere woorden, de groep heeft een oneindige orde. In Figuur 6 is de rand van elk twaalftal driehoeken op één hoekpunt aangedikt om het honingraat-motief naar voren te laten komen.

Dit voorbeeld laat een van de vele mooie interacties zien tussen de meetkunde, de plaatjes dus, en de algebra, de groep waarvan symmetrieën geregistreerd worden in woorden, met de letters a, b en c.

De tweede constructie starten we met een driehoek die hoeken van 36, 60 en 90 graden heeft. Zie Figuur 7. Dit betekent dat het spiegelen aan lijnen door een hoekpunt respectievelijk tien, zes en vier driehoeken op dat hoekpunt oplevert. We maken weer de drie buren van de oorspronkelijke driehoek en gaan door met spiegelen aan de zijden om nieuwe driehoeken te creëren.

Na 119 stappen is de opbouw compleet: er zijn 120 driehoeken ontstaan. Het eindresultaat is wat moeilijk te tekenen in het platte vlak. Daarom zijn de toppen van de tien uitstekende driehoeken apart getekend, terwijl ze eigenlijk samenvallen. Ook de lijnstukken die bij dat buitenste punt eindigen en in hetzelfde punt starten, vallen eigenlijk samen. We dikken de rand van elk tiental driehoeken op één hoekpunt aan om te accentueren dat we de dodecaëder, het beroemde regelmatige twaalfvlak, hebben gevonden.

Het plaatje is letterlijk en figuurlijk rond. Zie Figuur 8. Dit is te zien dankzij de transformaties van de figuur, die het driehoekenpatroon intact houden. Het aantal gevonden driehoeken is gelijk aan het aantal symmetrieën van de groep. De bijbehorende groep heeft dus orde 120.

In de eerste constructie begonnen we met een driehoek waarvan de som van de hoeken gelijk is aan 180 graden, zoals we gewend zijn. In de tweede constructie was de som 186 graden, dus meer dan de gebruikelijke 180; de ontstane figuur in het platte vlak bleek tot een boloppervlak om te krommen.

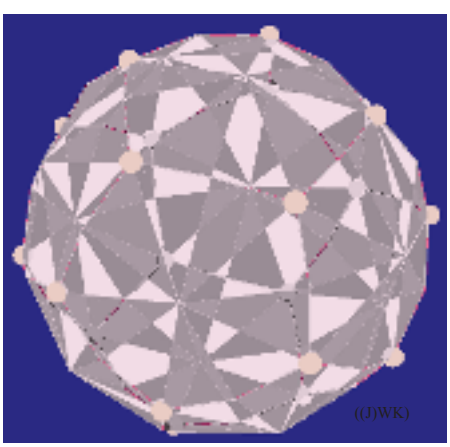
Voor het derde voorbeeld starten we met een driehoek waarvan de hoeken $\frac{360}{14}$, 60 en 90 graden zijn. De som van de hoeken is dan minder dan 180 graden. Zie Figuur 9.

Ook hier blijven we spiegelen om nieuwe driehoeken te maken, waarbij er steeds vier driehoeken, zes driehoeken of veertien driehoeken op een hoekpunt komen. Ten slotte dikken we de rand van elk veertiental driehoeken op één punt weer aan. De figuur die hier ontstaat, ligt in een zogenaamd hyperbolisch vlak, dat getekend is in een cirkelschijf. De cirkelschijf geeft een klassiek model voor wie zich afvraagt waar de rand van de wereld ligt. In het midden van de schijf kun je je bewegen zoals je gewend bent. Maar naarmate je dichter bij de rand van de schijf komt, worden je bewegingen trager. Zó traag dat je nooit echt bij de rand kunt komen. Dit model heeft Poincaré gebruikt om het gevoel over te brengen dat je in een wereld leeft waarvan je de rand nooit zult bereiken. De groep van symmetrieën van deze Poincaré-schijf is een Liegroep en het wereldmodel past in het eerder genoemde Erlanger Programm van Felix Klein. De groep van symmetrieën van de verenigde driehoeken is een deel van deze Lie-groep.

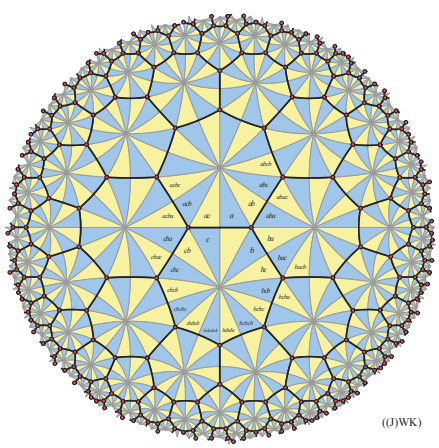
Er ontstaan oneindig veel driehoeken. Maar je kunt eindige groepen vinden, door quotiënten te nemen. Een eindig quotiënt krijg je door de rand om een stuk met eindig veel driehoeken van de figuur uit te knippen en de zijden van telkens twee driehoeken uit die opengeknipte rand zó aan elkaar te plakken dat er geen rand overblijft. Niet elke aanpak leidt onmiddellijk tot een interessante groep, maar we weten waar we het zoeken moeten. In Figuur 10 staat een voorbeeld, waarin het plakvoorschrift voor twee zijden uit de rand door een roze pad is aangegeven.

Het resultaat is te zien in Figuur 11. De groep in dit voorbeeld heeft orde 336.

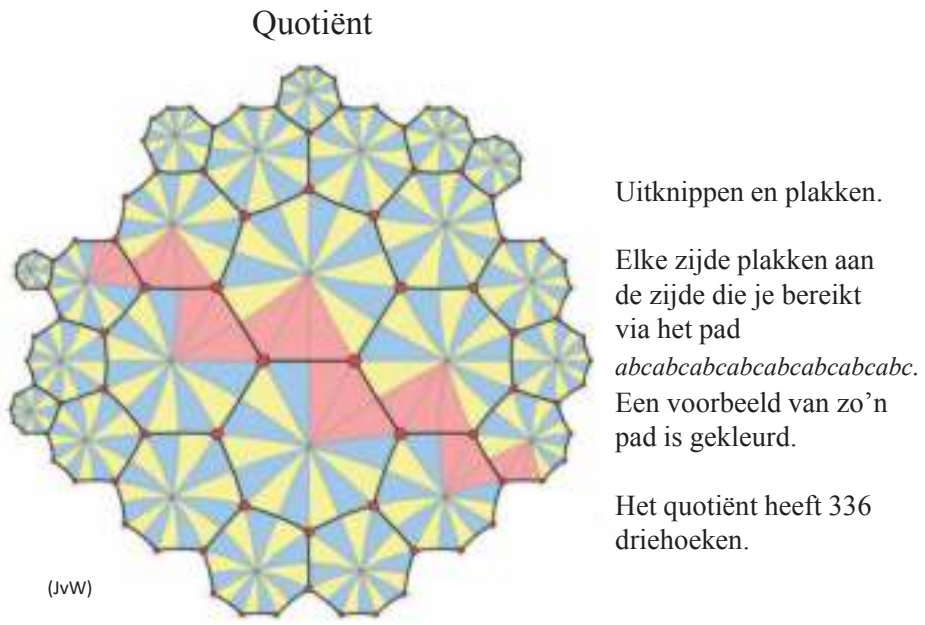
Een eindige groep bestaat uit eindig veel symmetrieën. Dat aantal heet, zoals gezegd, de orde van de groep. Veruit de meeste groepen kunnen worden opgebouwd uit kleinere groepen, waarbij het product van de ordes van de kleinere groepen gelijk is aan de orde



Figuur 8



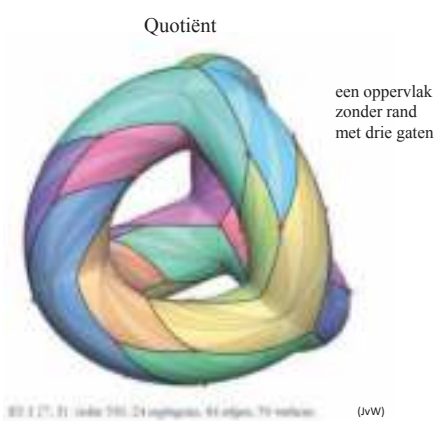
Figuur 9



Figuur 10

van de oorspronkelijke groep. Bijvoorbeeld, de cyclische groep van orde 4 is opgebouwd uit twee cyclische groepen van orde 2, en de cyclische groep van orde 6 is opgebouwd uit een cyclische groep van orde 2 en een cyclische groep van orde 3. Een groep die niet opgebouwd kan worden uit twee kleinere groepen, heet enkelvoudig.

Tussen 1888 en 1894 zijn de enkelvoudige Lie-groepen geclassificeerd. Deze groepen kennen varianten die eindige enkelvoudige groepen zijn. Er zijn precies 18 oneindige reeksen van die eindige varianten. Veruit de eenvoudigste van de 18 reeksen bestaat uit de cyclische groepen waarvan de orde een priemgetal is. De twee kleinste voorbeelden die niet cyclisch zijn maar wel in de oneindige reeksen voorkomen, zijn de groepen van draaiingen in de eerder besproken groep van orde 120 van de dodecaëder en de



Figuur 11

quotiëntgroep van orde 336 van het oppervlak met de drie gaten. Er zijn precies 26 eindige enkelvoudige groepen bekend die niet in een van de 18 oneindige reeksen voorkomen. Deze groepen worden sporadisch genoemd. De grootste van deze sporadische groepen heet het monster. Zijn orde, het aantal symmetrieën van deze groep dus, is ongeveer gelijk aan het aantal kilogrammen dat het universum weegt.

Na al deze voorbereidingen zijn we klaar voor het derde grote resultaat: de classificatie van de eindige enkelvoudige groepen. Die zegt dat elke eindige enkelvoudige groep bekend is, dat wil zeggen, tot een van de 18 oneindige reeksen behoort of een van de 26 sporadische groepen is.

In 1892 vroeg Hölder of een dergelijke stelling mogelijk was. Het bewijs was rond in 2005, dus 113 jaar later.

We gaan nu de VROUW-criteria voor de classificatie nalopen.

Volwassenheid van CFSG

De enige eindige enkelvoudige groepen waarvan elk tweetal symmetrieën commuteert, zijn de cyclische die als orde een priemgetal hebben. Dit is een betrekkelijk eenvoudig deelresultaat van de classificatie. In 1899 publiceerde Burnside een moeilijker geval: hij classificeerde alle eindige enkelvoudige groepen die een involutie hebben waar alleen maar de één en enkele andere involuties mee commuteren. Deze classificatie, zo zag hij het, zou meeromvattend kunnen worden als het waar is dat elke eindige enkelvoudige groep die niet cyclisch is, een involutie heeft. Dit vermoeden leek lange tijd te moeilijk om op te lossen, maar het is in 1963 bewezen door Feit en Thompson. Het was het startsein voor een grote gezamenlijke inspanning om de volledige classificatie van eindige enkelvoudige groepen rond te krijgen. Sindsdien zijn stelling en bewijs langzaam maar zeker volwassen geworden.

Resistentie van CFSG

Omdat het begrijpen van de stelling al zo veel werk is, zal het niemand verbazen dat het bewijs erg lang is. Het bewijs beslaat ongeveer 4000 pagina's. Ter vergelijk: de publicatie van Wiles en Taylor telt niet meer dan 140 pagina's.

De meest aansprekende blunder van de classificatie zou een eindige enkelvoudige groep zijn die niet bekend is. Deze gedachte kwam vaak op en is al vroeg uitgesproken door Conway. Maar het bewijs lijkt redelijk robuust tegen de komst van een nieuwe groep; het zal dan hoogstwaarschijnlijk maar op een paar puntjes aanpassing behoeven.

Originaliteit van CFSG

Een verrassend aspect aan het bewijs is het 'van lokaal naar globaal'-principe, waarbij je verregaande conclusies kunt trekken over een groep als geheel uit informatie die alleen maar iets zegt over kleine onderdeeltjes ervan. Een illustratie van dit principe zit in een vlucht vogels in V-formatie. Een prachtige globale structuur als deze komt tot stand door de lokale sturing van elke deelnemende vogel. Het globale beeld ontstaat zonder centrale organisatie!

Bij de figuren die uit spiegelingen van driehoeken ontstonden, zagen we ook een voorbeeld van dit principe: de som van de hoeken in de startdriehoek bepaalt op voorhand of het betegelde oppervlak dat de driehoeken vormen, recht blijft als in het platte vlak, krom trekt tot een bol, of uitdijt tot het hyperbolisch vlak als in de Poincaré-schijf.

Dan nu het 'lokaal naar globaal'-principe in de classificatie. In de opzet van een bewijs uit het ongerijmde, nemen we aan dat er een kleinste eindige enkelvoudige groep G bestaat die niet bekend is. De stelling van Feit en Thompson uit 1963 zegt dat die groep G een involutie heeft. De symmetrieën van G die met deze involutie commuteren, vormen een groep die kleiner is dan G. Uit de aannames volgt dat deze kleinere groep opgebouwd is uit bekende groepen. De lokale informatie van de involutie is dus min of meer bekend. Volgens het 'van lokaal naar

globaal'-principe worden alle eindige enkelvoudige groepen met deze bekende lokale structuur bepaald, om te constateren dat die zelf ook bekend zijn. In het bijzonder is de groep G bekend, in tegenstelling tot de aanname. Een tegenspraak is afgeleid en de stelling is bewezen uit het ongerijmde. Dit is natuurlijk een hele grove indicatie van het bewijs, maar het geeft het verrassende 'van lokaal naar globaal'-principe goed weer.

Voor de oplossing waren heel veel nieuwe concepten en methoden nodig, meer in de orde van Fermat dan de vierkleurenstelling. Er moest ook veel parallel gewerkt worden aan allerlei verschillende gevallen, wat weer meer overeenkomst vertoont met de vierkleurenstelling.

Unificatie van CFSG

Veel stellingen die in de groepentheorie niet zonder meer bewezen konden worden, zijn nu bewezen dankzij de classificatie. Want de lijst van voorbeelden is uitputtend, zodat elke uitspraak over eindige enkelvoudige groepen op juistheid gecontroleerd kan worden door haar te verifiëren voor elke groep uit de lijst. De kracht van de stelling is dus dat het een abstract begrip als eindige enkelvoudige groep zeer concreet maakt.

Daar staat tegenover dat op dit moment de stelling weinig toepassingen buiten de groepentheorie heeft. Ik verwacht dat dat over vijftig jaar wel anders zal zijn en hoop het mee te maken.

Waardering van CFSG

Het onderzoek werd uitgevoerd door een relatief grote gemeenschap, geleid door een soort van generaal: Gorenstein. Hij was een groot wetenschapper, maar ook een groot leider.

Op de Santa Cruz-conferentie in 1979, waar ik bij was, werd door Gorenstein de op handen zijnde classificatie aangekondigd. Het nieuws had een enorm effect op de gemeenschap. Er was veel waardering voor die grote stelling die er aan zat te komen, en de gemeenschap was trots op wat er bereikt was. Maar er was ook ontreddering omdat voor velen hun dagelijkse werk, het bewijzen van deelresultaten voor die classificatie, uit handen leek geslagen. Sommigen besloten acuut naar een ander gebied van onderzoek te verhuizen. Anderen gingen gewoon door in de eindigegroepentheorie. Deze personen hadden gelijk in de zin dat de aankondiging van Gorenstein in 1979 achteraf wat voorbarig bleek. Tegenwoordig wordt het in 2005 verschenen werk van Aschbacher en Smith als de afronding van het bewijs gezien.

Er waren ook onderzoekers die hun heil zochten in meetkundig onderzoek dat de classificatie vanuit een ander gezichtspunt zou kunnen benaderen. Het was de ambitie van Jacques Tits en Francis Buekenhout de classificatie van de eindige enkelvoudige groepen zoveel mogelijk in meetkunde te vatten. Ook die activiteiten hebben bijgedragen aan de grote waardering voor het resultaat.

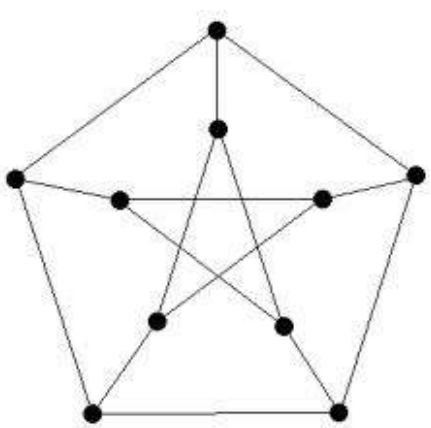
Meetkunde

Ik gaf al aan dat ik de recente geschiedenis van de classificatie van dichtbij heb meegemaakt.

Tijdens mijn mastersstudie in Utrecht volgde ik colleges over groepen bij Freudenthal en Strooker, en combinatoriek bij Klarner, die destijds in Eindhoven te gast was. Zij wezen me de weg naar de discrete wiskunde. Ik werd promovendus bij Springer en schreef een proefschrift over eindige complexe spiegelingsgroepen. Dit gaat over een uitbreiding van groepen van de gewone wereld van de reële getallen naar die van de complexe getallen.

Een paar jaar later, aan de Universiteit Twente, stortte ik me op nog complexere groepen en classificeerde ik de eindige spiegelingsgroepen in de wereld van de quaternion-getallen. Hierbij kwam een sporadische groep uit de classificatie naar voren. Seidel, de oprichter van de Faculteit Wiskunde en Informatica in Eindhoven, vertelde het nieuws aan de eerder genoemde Buekenhout. Na een seminar bij hem in Brussel om het verhaal te vertellen, leerde ik snel de gemeenschap kennen die aan de classificatie werkte. De Santa Cruz-conferentie van 1979 was een van de eerste gelegenheden daarvoor.

Van de classificatie fascineerde me het werk van Tits en Buekenhout het meest. Het 'van lokaal naar globaal'-principe uit de classificatie werd op verschillende manieren in meetkunde vertaald. In een daarvan wordt uit een groep een netwerk gemaakt waarin de knooppunten de involuties van de groep zijn en twee involuties verbonden worden als ze commuteren. In Figuur 12 ziet u een voorbeeld van een netwerk dat zo tot stand gekomen is, de zogenaamde Petersen-graaf.



Figuur 12

Het 'van lokaal naar globaal'-principe van de classificatie is ook naar deze netwerken te vertalen. Zo zijn bepaalde stukjes van de classificatie pure meetkunde geworden. Aan dit soort resultaten heb ik een bijdrage geleverd.

Computers

Naast deze fraaie meetkunde heb ik al vroeg gewerkt met computers. Van 1979 tot 1992 werkte ik in het Centrum voor Wiskunde en Informatica. Daar vroegen natuurkundigen me regelmatig bepaalde getallen uit te rekenen die met de eerder genoemde Lie-groepen te maken hebben. Om nieuwe vragen voor te zijn, schreef ik, met hulp van enkele collega's, het softwarepakket LiE, dat nog steeds in gebruik is.

Later op het CWI begon ik me te interesseren voor interactieve wiskundige documenten. In 1991 ging ik met collega's als Meertens en Pemberton aan de slag om een prototype daarvan te ontwerpen. Het bleek moeilijker dan we dachten. Een belangrijke oorzaak was de opkomst van het web, waardoor er steeds nieuwe technische mogelijkheden ontstonden, die ons tot herijking van de aanpak dwongen.

De inspanning heeft uiteindelijk geleid tot een *blended learning*-aanpak bij meer dan tien colleges aan deze universiteit, en tot het platform dat de firma SOWISO gebouwd heeft voor een zeer interactieve en soepele wijze van exact onderwijs.

Het web

Door deze activiteiten werd ik vanaf het begin met mijn neus gedrukt op het zich steeds verder ontwikkelende web. Zo heb ik de positieve inwerking gezien van het web op de wiskunde. Ik wil daarvan een aantal ervaringen met u delen aan de hand van de VROUW-criteria.

Volwassenheid dankzij het web

Volwassenheid bereik je door je werk zoveel mogelijk te toetsen aan dat van anderen. Maar in de grote hooiberg aan informatie is het niet altijd gemakkelijk de goede spelden te vinden. Op MathSciNet staan reviews van bijna alle wiskundige wetenschappelijke publicaties, compleet met allerlei online zoekmiddelen. Die mogelijkheid om relevante pu-

blicaties te vergaren voor eigen onderzoek, maakt nieuwe resultaten volwassener.

Ook wordt de voortgang van wiskunde nu beter geregistreerd dan ooit. Veel overwegingen en leerzame misstappen achter bewijzen staan voor iedereen zichtbaar op het web. Een belangrijke bijdrage in die registratie levert de website arXiv. Bijna iedereen kan er een artikel op plaatsen, waarbij de indieningsdatum geregistreerd wordt. Zo kunnen auteurs hun rechten (zoals prioriteit) zeker stellen.

Resistentie dankzij het web

Hoeveel overeenstemming er ook is over de correctheid van een bewijs, als er een stap is die je niet kan volgen, komt er weer twijfel naar boven. Had je dit gewoon in moeten zien, of heeft de auteur een misstap begaan? Deze twijfel is weg te nemen dankzij werk van onze eigen De Bruijn uit het eind van de jaren zestig. Zijn programma *AutoMath* is de eerste van een serie kleine overzichtelijke programma's, waarmee de verificatie van de correctheid van een bewijs volledig automatisch kan worden uitgevoerd. Daar is dan wel voor nodig dat het bewijs in groot detail en met veel precisie formeel is opgeschreven.

In 2005 heeft Gonthier een bewijs van de Vierkleurenstelling zodanig formeel opgeschreven dat het volledig automatisch door het computerprogramma COQ geverifieerd kon worden. In 2012 deed hij hetzelfde voor de grote stelling van Feit en Thompson over het bestaan van een involutie.

Hoe lang zal het nog duren voor alle wiskundigen zullen werken met een bewijsontwikkelomgeving, waarin formele bewijzen aan menselijke intuïtie gekoppeld worden? Kort geleden betoogde de Fieldsmedaillewinnaar Vladimir Voevodksy dat het gebruik ervan niet alleen gewenst maar zelfs noodzakelijk is. Veel collega's willen er niet aan denken. Toch verwacht ik dat over vijftig jaar iedere wiskundige op die manier wetenschappelijke publicaties zal voorbereiden.

Originaliteit dankzij het web

Hoe kun je een aansprekend wiskundig probleem vinden, je oriënteren op een oplossing, en laten inspireren door anderen? Daarvoor zijn sites als MathOverflow. Dit is een vraag-en-antwoord-kennisbank voor wiskundigen. Er is een ludiek systeem van bonuspunten aan verbonden voor goede antwoorden, en de begeleiding door experts is indrukwekkend.

Verder zijn er websites waarop zeer verrassende reacties op door jezelf gegeven informatie terug kunnen verschijnen. Een legendarisch voorbeeld is een website van Sloane, waar je een rij getallen kunt intypen om terug te krijgen welke bekende oneindige rijen zo beginnen.

Unificatie dankzij het web

Ik heb al gesproken over automatische bewijsverificatie, maar nog weinig over computeralgebra. Vrijwel alle denkbare wiskunde waaraan exact te rekenen valt, is in een computeralgebrapakket uit te voeren. Deze software, die in de jaren tachtig commercieel werd, heeft het mogelijk gemaakt op grote schaal wiskundig te experimenteren. Aan de pakketten GAP en Magma heb ik ook enkele bescheiden bijdragen geleverd.

Computeralgebrapakketten hebben allemaal hun eigenaardigheden, zoals afwijkende opvattingen over de meest eenvoudige vorm van een eindresultaat. Om de gebruiker de mogelijkheid te geven zich uit te drukken onafhankelijk van het te gebruiken pakket, heb ik meegewerkt aan de ontwikkeling van een wiskundig Esperanto, de taal *OpenMath*. Vanuit deze semantisch rijke taal zijn formules en opdrachten eenvoudig over te zetten naar willekeurig welk softwarepakket.

Waardering dankzij het web

De waardering voor een stelling kan groter worden als je zelf aan de totstandkoming van het bewijs meewerkt. In 2009 formuleerde Gowers op zijn blog een wiskundig probleem. Hij nodigde zijn lezers daarbij uit om zelfs de kleinst denkbare bijdragen aan een oplossing op zijn website aan te leveren. Dit experiment is het begin geworden van het Polymath project, dat een webgebaseerde samenwerkingsvorm is waarin iedereen aan het oplossen van een bepaald probleem kan bijdragen. Het lijkt te werken. Resultaten worden onder pseudoniem in gerespecteerde tijdschriften gepubliceerd, met verwijzing naar de Polymathwebsite.

Bijna alles wat ik over het web gezegd heb, geldt niet alleen voor onderzoek maar ook voor onderwijs. Ook daar is veel van terug te vinden op het web. De MOOCs zijn daarbij wel heel spraakmakend. In dat format bestaan online colleges over vele onderdelen van de wiskunde, waarin erg goede docenten optreden en prachtige animaties worden vertoond. Ze vormen een verrijking van het studiemateriaal voor iedereen die weet wat zelfstandig werken is.

Maar de online cursussen zijn geen volwaardige vervangers van het bestaande onderwijs. Het contact in twee richtingen tussen studenten en goede leermeesters ontbreekt nog. Een nieuwe trend, waarin MOOCs gekoppeld worden aan bereikbare docenten, gaat hier wat aan doen.

De gegeven voorbeelden laten zien hoe sterk het web helpt in het bedrijven van onderzoek en het volgen van onderwijs in de wiskunde. De wiskunde heeft veel bijgedragen aan de totstandkoming van het web, dat is bekend. Maar aan de hand van de VROUW-criteria heb ik betoogd dat het er ook veel voor terugkrijgt.

Dankwoord

De afgelopen 22 jaar heb ik aan de Technische Universiteit Eindhoven mogen werken. Dat heb ik als een voorrecht ervaren. Heel vaak als ik van het station naar de campus liep, deed ik een klein privé-onderzoekje. Was dit een campus waar ik met plezier naartoe liep? Een enkele keer spande het erom, en koesterde ik de gedachte aan een nieuwe onderneming. Maar het overgrote deel van al die loopjes werd ik vervuld van plezier vanwege de omgang met studenten en collega's, de vrijheid om eigen ideeën te verwezenlijken, en de vooruitgang die de universiteit heeft geboekt. Ik ben de TU/e daar zeer dankbaar voor.

Slot

Veel redes van dit soort beginnen met de klassieken. Deze eindigt er mee.

Euclides werd wel eens gevraagd waar zijn wiskunde nou eigenlijk goed voor was. Hij reageerde ooit met de woorden "Geef de vragensteller wat geld, want hij heeft er kennelijk niet genoeg aan zoiets moois te leren." Gelukkig kunnen we op de vraag naar het nut voor de maatschappij, tegenwoordig antwoorden dat achter dertig procent van het nationale inkomen, wiskunde zit.

Pythagoras introduceerde het woord mathematicus voor ingewijden in zijn elitaire school. Zij waren de enigen die volledig kennis mochten nemen van de wiskundige resultaten van die lokale school. Het web stelt je in staat als mathematicus te participeren in vele globale kringen.

Zelf ben ik natuurlijk van plan het web grotelijks te benutten bij mijn activititeiten vanachter de geraniums. Ik hoop dat u, met welke achtergrond dan ook, zult meegenieten van de verworvenheden van de Wiskunde in het web.

Figuur 6–9 zijn gemaakt door Wil Kortsmit met medewerking van Jan Willem Knopper. Figuur 10 en 11 zijn gemaakt door Jack van Wijk.

Christine van Vredendaal
Faculteit Wiskunde en Informatica
Technische Universiteit Eindhoven
c.v.vredendaal@tue.nl

Column Masterscriptie

Zekere veiligheid door kangoeroes

Elk jaar reikt de Technische Universiteit Eindhoven prijzen uit voor de beste afstudeerwerken van het afgelopen kalenderjaar. Dit jaar mocht Christine van Vredendaal de prijs voor de beste masterscriptie van het jaar 2014 in ontvangst nemen. In dit artikel zet zij uiteen wat haar afstudeerwerk inhield.

Tegenwoordig is een wereld zonder elektromagnetische cryptografie niet meer voor te stellen. Cryptografie wordt gebruikt voor het versleutelen van je telefoonverkeer, bankinformatie, e-mails en nog veel meer. Versleuteling werkt als volgt (zie Figuur 1). Als Alice een bericht naar haar vriend Bob wil versturen, versleutelt ze dit eerst met een *sleutel*. Als ze het versleutelde bericht dan verzendt, kan niemand het lezen, behalve Bob die het met een sleutel weer kan ontcijferen. In dit artikel zullen we het hebben over asymmetrische cryptografie. Hier heeft iedereen een privésleutel $k \in \mathbb{N}$ en een publieke sleutel $p_k = f(k) \in \mathbb{N}$ voor een bepaalde functie f.

Alice versleutelt haar bericht met Bobs publieke sleutel, waarna alleen Bob met zijn privésleutel het bericht kan lezen. De veiligheid zit in het feit dat f^{-1} zeer moeilijk te evalueren is en dus uit de publieke sleutel de privésleutel niet gevonden kan worden én dat er in principe zoveel mogelijkheden voor de sleutel zijn dat hij niet zomaar te raden valt. Voor elliptische-kromme-cryptosystemen is een veilige sleutel minstens een 256-bits getal. Voor RSA moet men eerder aan een 2048-bits getal denken.

Aanvallen van het nevenkanaal

Een aanvaller Eve probeert toch de sleutel van een cryptosysteem te vinden in de hoop deze te gebruiken om geheimen te achterhalen. Een van de manieren waarop ze dit kan doen is door naar de wiskunde van de versleuteling te kijken, maar we zullen ervan uitgaan dat deze veilig is. Een andere manier om meer informatie over de sleutel te vinden is door het doen van *nevenkanaal-aanvallen* (Engels: side-channel attacks). Van machines die het versleutelen uitvoeren, kan het stroomverbruik, het geluidsniveau of soms zelfs straling gemeten worden. Deze gemeten waarden zijn gecorreleerd aan de operaties die de machine uitvoert en dus ook aan de bits van de waarde van de sleutel. Eén klasse van dergelijke metingen resulteert in informatie die er zo uitziet:

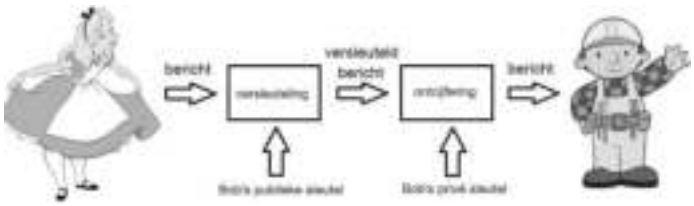
$$k = \underbrace{k_1|\cdots|k_{o-1}}_{\text{groen}}\,|\,\underbrace{k_o|\cdots|k_{r-1}}_{\text{oranje}}\,|\,\underbrace{k_r|\cdots|k_n}_{\text{rood}}.$$

Van de privésleutel k kunnen de meest significante bits (groen) precies achterhaald worden (dit kunnen er ook 0 zijn). Dit betekent dat het interval in $\mathbb{N}$ van 2^n mogelijke sleutels tot een interval van grootte 2^{n-o+1} gereduceerd wordt. Over de volgende bits (oranje) kan men *partiële* informatie vinden. Dit betekent dat er een kansverdeling is die aangeeft wat de kans is dat de sleutel in elk van 2^{r-o} intervallen van grootte $2^{n-r+1} := \ell$ zit. Over de minst significante bits (rood) kan geen informatie gevonden worden. De hoeveelheid groene, oranje en rode bits is afhankelijk van hoe goed het systeem tegen dergelijke aanvallen beschermd is.

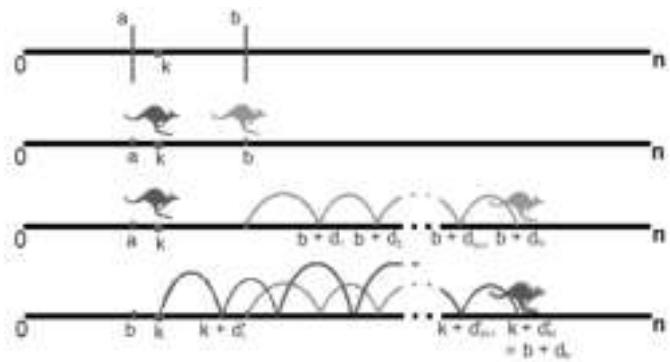
Het ordenen van sleutels

Nadat Eve een nevenkanaal-aanval heeft gebruikt, kan ze de resultaten gebruiken om de sleutel slimmer te zoeken. Wat ze zou doen is begin-

Christine van Vredendaal



Figuur 1



Figuur 2

nen met het interval dat de grootste kans krijgt van de aanval, deze ℓ sleutels doorzoeken en als ze de goede sleutel niet vindt, doorgaan naar het volgende interval met de grootste kans. De zoektijd die Eve op deze manier nodig heeft om de sleutel te vinden, heet de *rang* van een sleutel [3]. Voor Eve zelf is dit concept niet nuttig, omdat ze de sleutel en dus de rang niet weet. Bedrijven die hun creditkaarten, laptops of telefoons met encryptie willen verkopen, weten de sleutel echter wel en kunnen via metingen bepalen welke rang hij heeft. Als de rang van een sleutel laag is dan is de cryptografie niet veilig en kunnen de spullen niet verkocht worden. Als de rang van een sleutel hoger is dan de computerkracht die we veronderstellen dat een aanvaller heeft, dan is de cryptografie veilig genoeg om te gebruiken. Zo kan een gemiddelde laptop in een week 2^{40} sleutels proberen, een computercluster van een universiteit 2^{50} sleutels en misschien dat bepaalde overheidsinstanties wel 2^{80} sleutels kunnen testen. Een rang geeft dus aan hoe veilig een sleutel is, maar dit is alleen relevant als die berekend wordt op basis van de best mogelijke aanvallen. Als men de rang zou berekenen door met brute force alle mogelijkheden te proberen en zou concluderen dat een bepaald systeem veilig is, kan men bedrogen uitkomen als er een slimmere manier bestaat om de sleutels te doorzoeken.

Kruskals kaarttruc

Voor de beschreven nevenkanaal-aanvallen bestaat er een slimmere manier. Voor we bij de methode komen om sneller een interval te doorzoeken, beschouwen we eerst de volgende kaarttruc [1]. Neem een standaard stok kaarten en leg ze open gedraaid in een lange rij op de tafel. Ga denkbeeldig met je vinger op de eerste kaart staan en loop als volgt naar rechts:

- Als je op een getal staat, ga evenveel stappen naar rechts.
- Als je op een Aas staat, ga een stap naar rechts.
- Als je op een plaatje (Boer, Vrouw, Heer) staat, ga vijf stappen naar rechts.

Herhaal dit totdat je van de rij af zou vallen met de volgende stap en onthoud de kaart waar je eindigt. Vertel je slachtoffer dat hij een van de eerste tien kaarten van de rij mag kiezen en op dezelfde manier naar rechts moet lopen als hierboven beschreven. Als jij kan gokken waar hij terechtkomt (voordat hij zijn keuze bekend maakt), win jij de weddenschap. Wijs vervolgens de kaart aan die jij vanaf de eerste kaart had gevonden. In $5/6$ van de keren zal je slachtoffer op dezelfde plaats terechtkomen. Bij dezelfde truc met twee stokken kaarten is dit zelfs in meer dan 95 procent van de gevallen.

Pollards kangoeroe-algoritme

Het concept van Kruskals kaarttruc legt goed uit hoe je een sleutel in een interval vindt met Pollards kangoeroe-algoritme [2]. We plaatsen twee kangoeroes in het interval van mogelijke sleutels, zie Figuur 2. Van de donkergrijze kangoeroe weten we niet wat de sleutel k is (het slachtoffer in de kaarttruc). Van de lichtgrijze kangoeroe weten we de sleutel b wel (de eerste kaart van Kruskal). Vervolgens laten we beide volgens vaste regels naar rechts springen.

Hoe langer ze springen, hoe groter de kans dat ze botsen en op dezelfde waarden terechtkomen. Zo'n botsing kunnen we detecteren met behulp van de functie f en er is dan een methode om te bepalen wat de oorsprong van de donkergrijze kangoeroe was zonder dat we f^{-1} nodig hebben.

Het schatten van de rang

Pollards kangoeroe-algoritme is het best bekende algoritme om een sleutel in een interval te vinden. Als de lengte van het interval ℓ was en allebei de kangoeroes in het interval begonnen, verwachten we dat ze in $O(\sqrt{\ell})$ stappen botsen. Sterker, we kunnen gegeven X stappen van de kangoeroes de kans uitrekenen dat k in het interval zat, maar nog niet gevonden is. Deze kans is voor $X = c \cdot \sqrt{\ell}$ (constante c) zo dicht bij 0 dat we beter het volgende interval kunnen gaan doorzoeken dan alle ℓ sleutels controleren. Daarom weten we dat een aanvaller in elk interval maar $O(\sqrt{\ell})$ berekeningen hoeft te doen om de sleutel te vinden (of niet) en kunnen we nu ook de rang van een sleutel bepalen. We tellen stappen in elk van de intervallen waar de sleutel niet inzit, maar die volgens de nevenkanaal-aanval wel een grotere kans hebben, op en dit is de rang van de sleutel. Deze rang is vele malen realistischer dan simpelweg alle sleutels te tellen in de intervallen. Stel immers dat $\ell = 2^{40}$ en de gebruikte sleutel zit in het 1025ste interval. Als in elk interval alle sleutels getest moeten worden, dan is de rang $1024 * 2^{40} = 2^{50}$. Echter met Pollards kangoeroe-algoritme is een rang van $1024 \cdot c \cdot 2^{20} \approx 2^{30}$ veel realistischer.

Ten slotte

Om de veiligheid van onze cryptografische systemen te garanderen is onderzoek naar de rang erg belangrijk. Mijn bijdrage hierin was het modelleren van nevenkanaal-aanvallen waarin Pollards kangoeroe-algoritme gebruikt kan worden en vervolgens de theorieën ontwikkelen die nodig zijn om de rang van een sleutel te schatten in dit model. Dit zorgt ervoor dat we zekerder kunnen zijn over de veiligheid van onze laptops, creditkaarten en telefoons.

Referenties

1 M. Gardner, Mathematical games, *Scientific American*, februari 1978.

2 J.M. Pollard, Kangaroos, monopoly and discrete logarithms, *Journal of Cryptology*, Volume 13, 2000.

3 C. van Vredendaal, *Rank Estimation Methods in Side-Channel Attacks*, masterscriptie, 2014.

François Genoud
Delft Institute of Applied Mathematics
Delft University of Technology
s.f.genoud@tudelft.nl

Column Tenure-tracker

Bifurcations in nonlinear PDEs

In this column holders of a tenure track position introduce themselves. The tenure track positions in mathematics became available in 2013. Excellent researchers could apply in several expertise areas of mathematics. François Genoud has a tenure track position at Delft University of Technology.

I am a Swiss mathematician from Lausanne, where I did my undergraduate studies in physics and my PhD in mathematics with Charles A. Stuart. Since January 2015 I am an Assistant Professor Tenure Track in the Analysis Group of Delft University of Technology. I arrived here after a six-year postdoctoral journey through Oxford, Edinburgh and Vienna, which allowed me to diversify my research interests in important areas of mathematical physics.

My research lies in the rigorous mathematical analysis of differential equations. An important part of my work revolves around bifurcation theory, which is a powerful tool to understand qualitative properties of nonlinear partial differential equations. Partial differential equations (PDEs) are a natural language to describe many physical phenomena. The solutions of the equations represent physical quantities characterising the state of a given system. Bifurcation theory explains how the possible states of the system change when some physical parameters are varied.

The core of my work is in the analysis of nonlinear PDEs. I develop and apply abstract functional analytic methods (e.g. topological degree theory, min-max methods from the calculus of variations, implicit function theorems) to study PDEs in a rigorous mathematical framework. Thanks to my early education in physics, I am always keen on understanding the underlying physical models as well. Many important phenomena in nature involve some sort of oscillatory motion, modelled by 'wave equations' that are typically nonlinear. Even though it is in general not possible to solve the equations explicitly, the mathematical analyst can prove theorems about existence and properties (regularity, stability, blow-up, et cetera) of the solutions. This is essential for a deep understanding of the physical theories formulated through the equations.

I have applied nonlinear analysis to various important PDEs coming from mathematical physics, for instance in models of large-scale oceanic waves based on the Euler equation [7], or the study of phase transitions in nematic liquid crystals [1]. A large part of my research has been concerned with nonlinear Schrödinger (NLS) equations, which arise in the modelling of a variety of wave motions, including the propagation of light in optical fibres, Langmuir waves in plasma, Bose–Einstein condensates, or water waves on the sea. For NLS equations, I have proved the existence of stable nonlinear waves, known as 'solitons'. These

are idealisation of waves encountered in real-world systems, characterised by strong localisation in space (and/or time) and strong stability properties. Such waves can for instance represent narrow laser/light beams in nonlinear optical media, solitary waves on a water surface, rogue waves, et cetera.

Nonlinear wave guides

Bifurcation theory has proved especially useful to study equations having a nontrivial spatial dependence, sometimes referred to as *inhomogeneous NLS*. In the context of a planar nonlinear waveguide, they take the general form

$$\begin{aligned} &i\partial_z\psi + \partial_{xx}^2\psi + f(x,|\psi|^2)\psi = 0,\\ &\psi = \psi(x,z) : \mathbb{R}\times\mathbb{R}\to\mathbb{C}, \end{aligned} \tag{1}$$

where z is the direction of propagation of the wave and $\partial_{xx}^2\psi$ is the Laplacian of the (complex envelope of the) electric field ψ in the transverse direction x. In this model, the nonlinear response $f(x,|\psi|^2)$ represents the electric permittivity of the material. In *self-focusing* media, this is a positive function, increasing in the field intensity $|\psi|^2$. A laser beam travelling in the material locally modifies its permittivity, thereby focusing itself along the propagation axis $x = 0$. The dependence on x accounts for inhomogeneities in the medium. The most commonly used materials are the *Kerr media*, for which $f(x,|\psi|^2) = V(x)|\psi|^2$, for some $V : \mathbb{R}\to\mathbb{R}_+$.

We call *soliton* a standing wave solution of the form $\psi(x,z) = u(x)e^{ikz}$, where $k\in\mathbb{R}$ and $u:\mathbb{R}\to\mathbb{R}$ is localized — typically $u\in H^1(\mathbb{R})$ and $u(x)\to 0$ exponentially as $|x|\to\infty$. Such a solution of (1) exists if and only if u satisfies the nonlinear ordinary differential equation

$$u''(x) + f(x,u^2(x))u(x) = ku(x), \quad u\in H^1(\mathbb{R}). \tag{2}$$

Soliton curves $k\mapsto\psi_k(x,z) = u_k(x)e^{ikz}$ can be obtained by bifurcation techniques applied to (2). Heuristically, the existence of solitons is allowed by a balance in (1) between the diffraction modelled by the Laplacian and the self-focusing effects due to the nonlinear term $f(x,|\psi|^2)\psi$. Their stability then depends on the monotonicity of the function $k\mapsto\|u_k\|_{L^2}$ and on the spectral properties of linearised operators associated with (1)–(2).

The combination of space-dependent coefficients and nonlinearities more general than the pure-power law $f(x,|\psi|^2) = |\psi|^{p-1}$ $(p>1)$ is of major interest for applications, but has only been scarcely investigated in the mathematical literature. I have obtained curves $k\mapsto u_k\in H^1(\mathbb{R})$ of stable solitons for a nonlinear response of the form

$$f(x,|\psi|^2) = V(x)|\psi|^{p-1} \quad \text{or}$$
$$f(x,|\psi|^2) = V(x)\frac{|\psi|^{p-1}}{1+|\psi|^{p-1}} \quad (1<p<5)$$

under appropriate regularity and decay assumptions on the coefficient $V:\mathbb{R}\to\mathbb{R}$, see [2–4]. Another model of interest in nonlinear optics is given by

$$f(x,|\psi|^2) = \epsilon\delta(x) + 2|\psi|^2 - |\psi|^4,$$

where $\epsilon>0$ is a coupling constant and the Dirac mass $\delta(x)$ models a narrow attractive potential centred at $x=0$. A remarkable feature of this model is that explicit solutions are available, that can be expressed in terms of elementary functions. Their stability can be proved by bifurcation and spectral analysis [5].

Wave collapse

What is meant here by stability is that, given an 'initial condition' at $z=0$, $\psi(\cdot,0)\in H^1(\mathbb{R})$, close to the initial condition u_k of the standing wave $\psi_k(x,z) = u_k(x)e^{ikz}$, the corresponding solution $\psi(x,z)$ of (1) remains close (in an appropriate sense) to $\psi_k(x,z)$, for all $z>0$. In particular $\psi(x,z)$ exists for all $z>0$. However, in some situations, the focusing effects will beat the diffraction in the dynamics of (1), giving rise to solutions which *blow up* at a finite propagation distance $Z>0$ in the waveguide, in the sense that

$$\lim_{z\uparrow Z}\|\partial_x\psi(x,z)\|_{L^2} = \infty.$$

This phenomenon of 'wave collapse' — where, typically, all the beam power concentrates on the axis of propagation at the blow-up point — has been well-known since the early days of nonlinear optics. Of course, the collapse indicates that the physical relevance of the model breaks down at the blow-up point. However, the dynamics leading to the blow-up give valuable information on the behaviour of the beam undergoing intense self-focusing. The formation of singularities in NLS equations has attracted substantial interest in the past twenty years, but mostly in the pure-power case, $f(x,|\psi|^2) = |\psi|^{p-1}$. I have contributed to extend the theory to inhomogeneous NLS equations [6], and further work in this direction is in progress with Elek Csobo, my PhD student.

References

1 S. Bachmann and F. Genoud, Effective theories for liquid crystals and the Maier–Saupe phase transition, under review, preprint arxiv.org/abs/1508.05025.

2 F. Genoud, Existence and orbital stability of standing waves for some nonlinear Schrödinger equations, perturbation of a model case, *J. Differential Equations* 246 (2009), 1921–1943.

3 F. Genoud, Bifurcation and stability of travelling waves in self-focusing planar waveguides, *Adv. Nonlinear Stud.* 10 (2010), 357–400.

4 F. Genoud, Orbitally stable standing waves for the asymptotically linear one-dimensional NLS, *Evolution Equations and Control Theory* 2 (2013), 81–100.

5 F. Genoud, B.A. Malomed and R.M. Weishäupl, Stable NLS solitons in a cubic-quintic medium with a delta-function potential, to appear in *Nonlinear Anal.*, preprint arxiv.org/abs/1409.6511.

6 F. Genoud and V. Combet, Classification of minimal mass blow-up solutions for an L^2 critical inhomogeneous NLS, to appear in *J. Evol. Equ.*, preprint arxiv.org/abs/1503.08915.

7 F. Genoud and D. Henry, Instability of equatorial water waves with an underlying current, *J. Math. Fluid Mech.* 16 (2014), 661–667.

Piet Groeneboom
Instituut voor Toegepaste Wiskunde
TU Delft
p.groeneboom@tudelft.nl

Geurt Jongbloed
Instituut voor Toegepaste Wiskunde
TU Delft
g.jongbloed@tudelft.nl

Onderzoek

Statistiek met vormrestricties

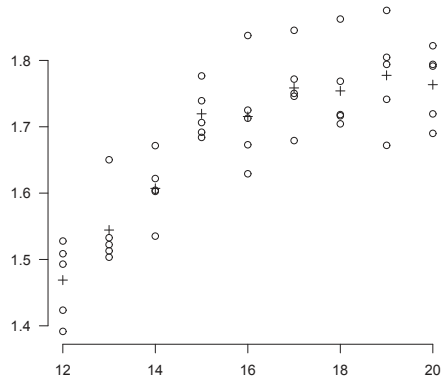
In december 2014 verscheen het boek *Nonparametric Estimation under Shape Constraints* **van de hand van Piet Groeneboom en Geurt Jongbloed. Dit boek gaat over statistiek onder vormrestricties, een onderwerp dat in de jaren vijftig van de vorige eeuw ontstond. Vanaf de jaren tachtig zijn de methoden doorontwikkeld en heel nuttig gebleken in niet-parametrische problemen, met name in censureringsproblemen en andere statistische inverse problemen. In dit artikel geven Groeneboom en Jongbloed in vogelvlucht ontwikkelingen in het vakgebied weer.**

We willen de gemiddelde lengte van alle vijftienjarige jongens in Nederland weten. We kunnen niet van al die jongens de lengte bepalen, maar doen dit alleen voor een willekeurige groep van omvang n uit die populatie. Dit voorbeeld is een klassiek probleem uit de statistiek. Een veelgebruikt model in deze situatie is dat de n metingen $x_1, \ldots, x_n$ worden opgevat als steekproef uit een normale verdeling met onbekende verwachtingswaarde μ en (onbekende) variantie σ^2. De beste manier om de onbekende parameter μ te schatten is dan door het gemiddelde van de metingen te nemen. Als de steekproefomvang n groot is, zal dit steekproefgemiddelde in de buurt van μ liggen.

In lijn met dit eenvoudige voorbeeld, beschouwen we nu de situatie waarin er lengtemetingen zijn van twaalf- t/m twintigjarige scholieren. Ook dan kan per leeftijd het steekproefgemiddelde worden genomen als schatting voor de gemiddelde lengte over de populatie. Wanneer de steekproeven groot zijn, zullen ook die schattingen dicht in de buurt van de echte populatiegemiddelden liggen. De veronderstelling dat er een ordening in de populatieparameters zit, in de zin dat de gemiddelde lengte zal toenemen met leeftijd, ligt voor de hand. Bij grote steekproeven zal dezelfde ordening te zien zijn in de steekproefgemiddelden, maar bij kleinere steekproeven hoeft dit niet zo te zijn. In die gevallen kan de ordening die er in de populatieparameters is, gebruikt worden om de schattingen te verbeteren. Zie Figuur 1 voor een synthetische dataset met lengtemetingen. De steekproefgemiddelden van de vijf lengten per leeftijd worden gegeven door

$$1{,}469,\ 1{,}544,\ 1{,}607,\ 1{,}720,$$
$$1{,}715,\ 1{,}759,\ 1{,}754,\ 1{,}778,\ 1{,}763.$$

Het is duidelijk dat deze gemiddelden geen stijgende rij vormen.



Figuur 1 Per leeftijd vijf lengtemetingen alsmede per leeftijd de gemiddelde lengte (+).

Isotone regressie heeft te maken met het benutten van de ordeningsinformatie. In de volgende paragraaf wordt uitgelegd hoe zo'n isotone regressie kan worden berekend. De paragraaf daarna gaat over een specifiek censureringsprobleem. Binnen dat probleem is het van belang om een (per definitie monotone) verdelingsfunctie te schatten, op basis van gecensureerde gegevens. In plaats van concrete metingen zijn alleen intervallen bekend waarin die metingen liggen. Een schatter die geen gebruik maakt van die monotonie heeft zeer slechte eigenschappen. Echter, als isotone regressie wordt toegepast, verandert die schatter in een schatter met heel goede eigenschappen. Recent zijn ook combinaties van vormrestricties (als monotonie) met andere regulariseringstechnieken (bijvoorbeeld gladmaken) in de niet-parametrische statistiek ontwikkeld. Die ontwikkelingen leiden weer tot nieuwe mogelijkheden als het gaat om de constructie van betrouwbaarheidsintervallen en toetsen onder vormrestricties. Deze recente ontwikkelingen en nieuwe uitdagingen komen aan de orde in de laatste twee paragrafen.

Isotone Regressie

Beschouw de volgende eenvoudige deelverzameling van $\mathbb{R}^n$:

$$C_n = \{x = (x_1, , \ldots, x_n) \in \mathbb{R}^n : x_1 \leq \cdots \leq x_n\}.$$

Dit is een *convexe kegel* in $\mathbb{R}^n$, dus als twee punten in C_n liggen, liggen alle positieve li-

neaire combinaties van die punten ook in C_n. Isotone regressie heeft te maken met projecteren van een punt $y \in \mathbb{R}^n$ op C_n. Zij $w = (w_1, \ldots, w_n) \in (0, \infty)^n$ een gewichtsvector en definieer de gewogen euclidische norm op $\mathbb{R}^n$ door

$$\|x\|_{w,2} = \sqrt{\sum_{i=1}^{n} x_i^2 w_i}.$$

Definieer vervolgens de (gekwadrateerde) afstand van een punt x tot y door

$$Q_w(x) = \|x - y\|_{w,2}^2.$$

De projectie $\hat{y}$ van y op C_n kan dan worden gedefinieerd door

$$\hat{y} = \text{argmin}_{x \in C_n} Q_w(x),$$

het argument dat in Q_w moet worden ingevuld om de minimale waarde over C_n aan te nemen.

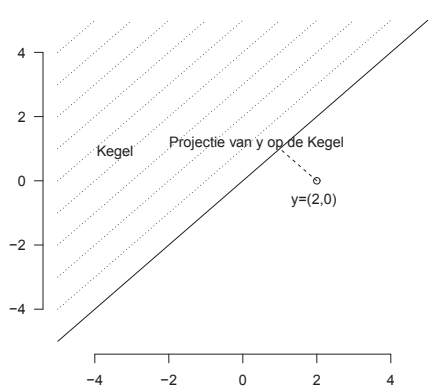
Beschouw ter illustratie een punt $y \in \mathbb{R}^2$ (dus $n = 2$) en $w = (1, 1)$. We willen het punt $\hat{y} \in C_2$ bepalen waarvoor

$$\|x - y\|_2^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2$$

minimaal is. Als $y \in C_2$, dan is duidelijk dat $\hat{y} = y$. De afstand van $\hat{y}$ tot y is dan nul, en kleiner kan die niet worden. Als $y \notin C_2$, dan laat een schets van de situatie (zie Figuur 2) zien dat de projectie van y op de rand van C_2, de lijn $x_1 = x_2$ ligt. Hieruit volgt eenvoudig dat

$$\hat{y} = \begin{pmatrix} (y_1 + y_2)/2 \\ (y_1 + y_2)/2 \end{pmatrix}. \qquad (1)$$

In het bijzonder wordt de projectie van het punt $(2, 0)$ op C_2 gegeven door $\hat{y} = (1, 1)$, een



Figuur 2 De kegel C_2, met projectie $(1,1)$ van $(2,0)$.

observatie die ook puur op basis van Figuur 2 kan worden gedaan.

Voor algemene $n \geq 2$ is er ook een verrassend eenvoudige oplossing van het projectieprobleem. Noodzakelijke voorwaarden voor optimaliteit voor de oplossing kunnen worden verkregen uit variationele overwegingen: stel $\hat{y}$ minimaliseert de afstand tot y over C_n, dan is iedere richtingsafgeleide van Q_w in toegestane richtingen binnen C_n niet-negatief. Concreet, voor alle $x \in C_n$ geldt dat

$$\lim_{\epsilon \downarrow 0} \frac{Q_w(\hat{y} + \epsilon(x - \hat{y})) - Q_w(\hat{y})}{\epsilon} \geq 0.$$

De ongelijkheden die hieruit voortvloeien blijken ook voldoende te zijn en zijn equivalent met de volgende (on-)gelijkheden:

$$\sum_{j=1}^{i} \hat{y}_j w_j \begin{cases} = \sum_{j=1}^{i} y_j w_j, & i \in I, \\ \leq \sum_{j=1}^{i} y_j w_j, & i \notin I, \end{cases} \qquad (2)$$

waar I de verzameling van 'spronglocaties' van $\hat{y}$ is:

$$I = \{1 \leq i \leq n : \hat{y}_{i+1} > \hat{y}_i\} \cup \{n\}.$$

Zie voor de details Lemma 2.1 in [7]. Deze (on-)gelijkheden kunnen eenvoudig worden gebruikt om na te gaan of een voorgestelde x de functie Q_w over C_n minimaliseert. De (on-)gelijkheden kunnen echter ook worden gebruikt om de oplossing van het projectieprobleem te construeren. Startpunt is om op basis van het rechterlid in (2) een cumulatief somdiagram op te stellen. Dit is een verzameling van $n + 1$ punten in het platte vlak: $P_0 = (0, 0)$ en

$$P_i = \left(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} y_j w_j \right), \quad 1 \leq i \leq n.$$

Denk aan deze $n+1$ punten als spijkers op een bord en verbind vervolgens deze punten door een strak gespannen koord dat P_0 met P_n verbindt en 'onderlangs' gaat. Deze constructie geeft de zogenaamde (grootste) convexe minorant van het puntendiagram. De oplossing $\hat{y}_i$ van het projectieprobleem wordt nu gegeven door de linkerafgeleide van deze convexe minorant in punt P_i $(1 \leq i \leq n)$. Beschouw, om dit in te zien, de tweede coördinaat van het cumulatieve somdiagram in punt P_i. Deze komt overeen met het rechterlid in (2), met index i. Het cusum-diagram wordt van links naar rechts opgebouwd door vanuit punt P_{i-1} over een horizontale afstand w_i een lijn te volgen met richtingscoëfficiënt y_i, en zo aan te komen in P_i $(1 \leq i \leq n)$. Als de $y \notin C_n$, zal de grafiek die door het verbinden van de punten ontstaat niet convex zijn. De convexe minorant blijft altijd onder het cusum-diagram en raakt eraan als de linkerafgeleide van deze minorant (omhoog) springt. Definieer nu $\hat{y}_i$ als de linkerafgeleide van het de convexe minorant van het cumulatieve somdiagram in P_i. Op dezelfde manier als het cumulatieve somdiagram met behulp van de waarden y_i en w_i, wordt de tweede coördinaat van de convexe minorant verkregen door $\hat{y}_i$ en w_i te gebruiken. De waarde van de convexe minorant horend bij P_i is dan gelijk aan het linkerlid van (2). Hieruit volgt dat de (on-)gelijkheden die de convexe minorant van het cusum-diagram karakteriseren precies de (on-)gelijkheden zijn in (2).

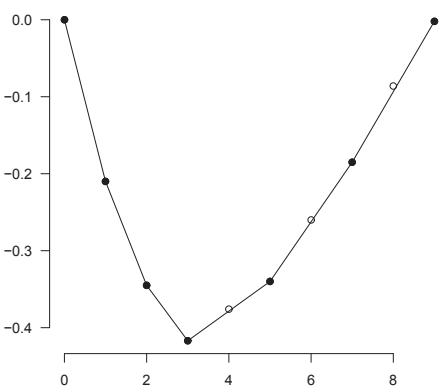
Voor het eerder gegeven voorbeeld is eenvoudig te controleren dat naast P_0, het cumulatieve somdiagram bestaat uit

$$P_1 = (1, y_1) \quad \text{en} \quad P_2 = (2, y_1 + y_2).$$

Als $y_2 \geq y_1$, ofwel $y \in C_2$, dan loopt de convexe minorant van het diagram via een rechte lijn van P_0 naar P_1 en vervolgens via een rechte lijn van P_1 naar P_2. De linkerafgeleiden in de punten P_1 en P_2 zijn dan respectievelijk y_1 en y_2. Dus (als eerder gezien), $\hat{y} = y$. Ingeval $y_2 < y_1$, dan wordt de convexe minorant van het resulterende diagram gegeven door de lijn die P_0 via een rechte lijn verbindt met P_2. De linkerafgeleiden in P_1 en P_2 zijn dan gelijk, wat leidt tot de oplossing gegeven in (1).

Deze constructie kan dus ook in hogere dimensies worden gebruikt. Bijvoorbeeld om het punt

$$y = (-0{,}210,\ -0{,}135,\ -0{,}072,\ 0{,}041,\ 0{,}036,\ 0{,}080,\ 0{,}075,\ 0{,}099,\ 0{,}084)$$



Figuur 3 Cumulatieve som diagram en convexe minorant.

in $\mathbb{R}^9$ te projecteren op de kegel C_9, met gewichtsvector $w = (w_1, \ldots, w_9)$ en $w_i \equiv 1$. Figuur 3 laat de constructie in dit geval zien. De knikpunten van de convexe minorant in de figuur zijn als zwarte punten weergegeven. De projectie wordt gegeven door de linkerafgeleide van de convexe minorant te bepalen in de punten P_1 t/m P_9:

$$\begin{aligned}\hat{y} = (&-0{,}2100,\ -0{,}1350,\ -0{,}0720,\\ &0{,}0385,\ 0{,}0385,\ 0{,}0775,\\ &0{,}0775,\ 0{,}0915,\ 0{,}0915).\end{aligned}$$

Wat heeft deze projectie nu met statistiek te maken? Denk aan het voorbeeld uit de introductie, met gemiddelde lengten van scholieren in leeftijd variërend van 12 tot en met 20 jaar. We modelleren de individuele metingen als onafhankelijke normaal verdeelde stochastische variabelen met gelijke varianties σ^2 en verwachtingswaarden $\mu = (\mu_{12}, \ldots, \mu_{20})$ die afhangen van de leeftijd. De zogenaamde maximum likelihood schatting voor μ minimaliseert dan de functie

$$\mu \mapsto \sum_{i=12}^{20} (\bar{y}_i - \mu_i)^2.$$

De oplossing is dus de projectie van de vector van geobserveerde gemiddelden $y = (\bar{y}_{12}, \ldots, \bar{y}_{20})$ op de kegel C_9 met gewichten 1. Via de genoemde constructie kan de oplossing worden bepaald en deze blijkt

$$\begin{aligned}&(1{,}4690,\ 1{,}5440,\ 1{,}6070,\ 1{,}7175,\\ &1{,}7175,\ 1{,}7565,\ 1{,}7565,\ 1{,}7705,\ 1{,}7705)\end{aligned}$$

te zijn. Deze oplossing hangt samen met het voorbeeld van Figuur 3 omdat de y die daar geprojecteerd wordt precies gelijk is aan de vector van gemeten gemiddelden minus het overall gemiddelde van de metingen.

Drie vroege artikelen waarin isotone regressie aandacht kreeg zijn [1], [3] en [4]. De boeken [2] en [12] geven een goed overzicht van de stand van zaken ten tijde van de publicatie. Recente boeken over het onderwerp zijn [15] en [7].

Een censureringsprobleem

Een belangrijk deelgebied van de statistiek gaat over de analyse van 'time-to-event data' of 'survivalanalyse'. Deze analyses komen veel voor in medische studies en kwaliteitsanalyses in de industrie. Men is geïnteresseerd in het tijdstip waarop een bepaalde gebeurtenis plaatsvindt. Denk hierbij aan het begin van een bepaalde ziekte of het moment waarop een onderdeel van een product kapot gaat. Vaak zijn deze tijdstippen niet precies te observeren, en heeft men met censurering te maken. Dit houdt in dat alleen bekend is dat het moment plaatsgreep in een bepaald interval van positieve lengte. Een speciaal censureringsprobleem staat bekend onder de naam '*current status*'-censurering. Bij dit probleem zijn de intervallen ofwel van de vorm $[0, t]$ ofwel van de vorm (t, ∞). De gedachte bij de naam van dit model is dat op inspectietijdstip t wordt bepaald of het onderdeel al kapot was (dan is bekend dat het moment van stukgaan in het interval $[0, t]$ lag) of nog niet (in dat geval ligt het moment van stukgaan in (t, ∞)). Op tijdstip t wordt dus de 'huidige toestand' of 'current status' van het onderdeel bepaald.

Wanneer er voor n onderdelen (een sample) *current status*-data zijn, worden deze vaak weergegeven als

$$\{(t_i, \delta_i) : 1 \le i \le n\},$$

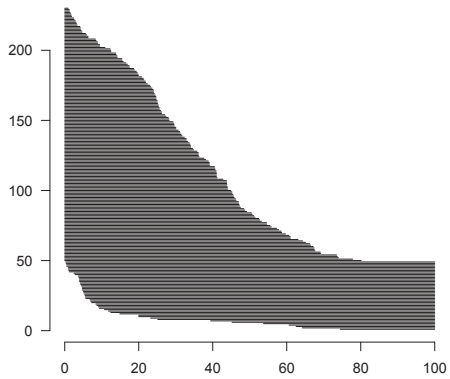
waar $t_1 \le t_2 \le \cdots \le t_n$ en $\delta_i = 0$ als het i-de observatie-interval gelijk is aan (t_i, ∞) en $\delta_i = 1$ als het gegeven wordt door $[0, t_i]$. Op basis van deze data moet de *verdelingsfunctie* van de event-time worden geschat. Deze verdelingsfunctie F is een monotone functie, en de maximum likelihood schatting voor deze functie maximaliseert de functie

$$\ell(F) = \sum_{i=1}^{n} \delta_i \log F(t_i) + (1 - \delta_i) \log(1 - F(t_i))$$

over alle rechtscontinue stijgende functies met waardenbereik in $[0, 1]$. Zonder beperking van algemeenheid kan de maximalisering worden beperkt tot stapfuncties met sprongen beperkt tot de verzameling $\{t_1, t_2, \ldots, t_n\}$. De functie ℓ hangt namelijk alleen van de waarden van F in deze punten af. Nu blijkt het zo te zijn dat de oplossing van het maximaliseringsprobleem dezelfde is als de kleinste kwadratenschatter die de volgende functie minimaliseert over de vector $(F(t_1), \ldots, F(t_n)) \in C_n$:

$$Q(F) = \sum_{i=1}^{n} (F(t_i) - \delta_i)^2 .$$

Dit is precies het projectieprobleem uit de vorige paragraaf. Specifiek hieraan is dat de gewichten allemaal gelijk zijn en dat de te projecteren vector uit nullen en enen bestaat.

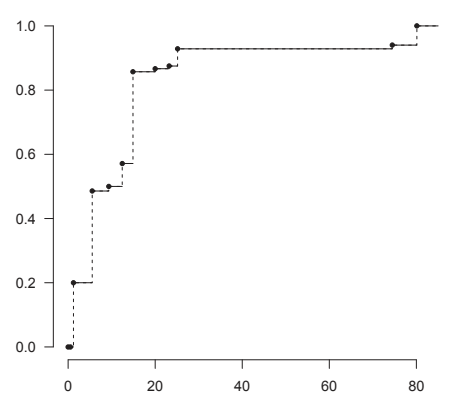


Figuur 4 Representatie van de rubella infectiedata.

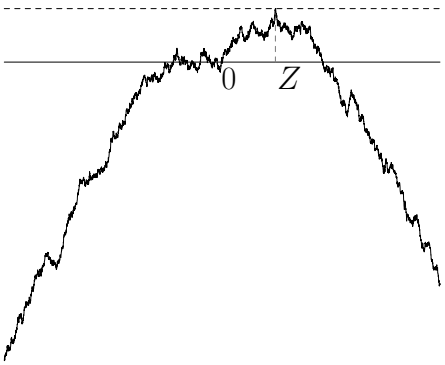
Als alle δ_i'tjes al in stijgende volgorde zouden staan (startstuk bestaand uit nullen gevolgd door een staartstuk bestaande uit enkel enen), is de oplossing eenvoudig. Anders moet de convexe minorant van het eerder beschreven cumulatieve somdiagram worden bepaald.

In Oostenrijk is een studie gedaan naar de leeftijd waarop mannen immuun worden voor de rodehond. In maart 1988 zijn daartoe 230 mannen gecontroleerd op antistoffen. Dit levert *current status*-gegevens op. Ofwel er werden antilichamen aangetroffen, of niet. De data, die ook geanalyseerd zijn in [10], zijn weergegeven in Figuur 4. Op de verticale as zijn de mensen weergegeven (genummerd 1 tot en met 230). Het horizontale lijntje op hoogte i geeft het tijdsinterval aan waarin de infectie optrad. Sommige intervallen beginnen bij nul en eindigen bij t_i. Zo'n interval correspondeert met een observatie $(t_i, 1)$. Andere intervallen beginnen op t_i en lopen naar rechts tot het einde door. Dit betekent dat de betreffende persoon op inspectietijdstip t_i nog niet geïnfecteerd was.

Op basis van deze gegevens kan de maximum likelihood schatter voor de (monotone) verdelingsfunctie F worden berekend. Deze is gegeven in Figuur 5.



Figuur 5 Maximum likelihood schatter voor de verdelingsfunctie van de infectietijdstippen.



Figuur 6 Brownse beweging met parabolische drift en locatie Z van het maximum.

In zowel het klassieke isotone regressieprobleem als het *current status*-model, wordt de limietverdeling van de maximum-likelihood-schatter gegeven door de locatie Z van het maximum van tweezijdige Brownse beweging met een negatieve parabolische trend. Deze verdeling is zeer verschillend van een normale verdeling. Een idee van deze stochastische variabele Z wordt gegeven in Figuur 6.

De verdeling werd analytisch gekarakteriseerd, gebruikmakend van Airy-functies en gerelateerd aan een stationair proces van locaties van maxima van de Brownse beweging ten opzichte van verschuivende parabolen in [5]. Deze preprint kreeg in 1985 de Rollo Davidson-prijs. De Rollo Davidson-prijzen worden jaarlijks in Cambridge uitgereikt ter nagedachtenis van de jonge wiskundige Rollo Davidson, die omkwam bij een bergbeklimming.

Recent is een nieuwe afleiding van deze karakterisering in termen van Airy-functies gegeven in [9], waarbij ook nieuwe relaties tussen speciale functies aan de dag kwamen via analyse van bepaalde partiële (parabolische) differentiaalvergelijkingen. Karakteristiek voor onderzoek in de isotoneregressie is dat technieken uit andere gebieden van de wiskunde nodig zijn om verder te komen. Om asymptotische eigenschappen voor maximum-likelihood-schatters in deconvolutieproblemen af te leiden, zijn bijvoorbeeld resultaten nodig voor bepaalde vrij gecompliceerde integraalvergelijkingen.

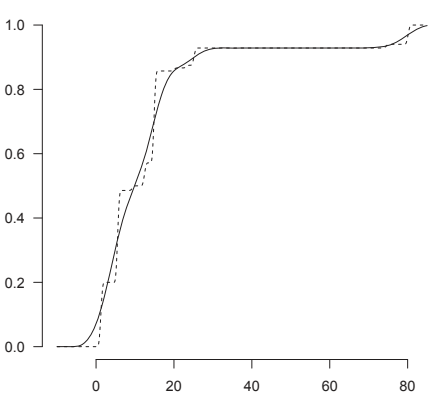
Recente ontwikkelingen

Binnen de niet-parametrische statistiek is er sinds de artikelen [13] en [11] veel gewerkt met kernschatters voor kansdichtheden en regressiefuncties. Deze schatters zijn per definitie glad en hoe meer gladheid (mits terecht, dat is wel belangrijk) wordt aangenomen van de onderliggende functie, hoe nauwkeuriger de schatter is. In modellen waar functies aan vormrestricties moeten voldoen, hebben kernschatters het nadeel dat deze in het algemeen niet aan die restricties voldoen. De combinatie van gladheidsaannamen en vormrestricties levert een interessant type schatters op. In het *current status*-model van de vorige paragraaf zijn dit type schatters bestudeerd in [6]. Zie Figuur 7.

De resulterende schatters kunnen dikwijls gebruikt worden in (bootstrap-) procedures om de verdeling van een toetsingsgrootheid te benaderen. Ook voor het construeren van betrouwbaarheidsintervallen voor functies die aan vormrestricties moeten voldoen, kunnen de gladde schatters worden gebruikt (zie bijvoorbeeld [8] en [14]). Het laten zien dat de bootstrapmethode (asymptotisch) het correcte resultaat geeft, is slechts in enkele modellen gedaan. De C/C++-programma's die we geschreven hebben voor de berekening van de schatters en (bootstrap- en niet-bootstrap-) betrouwbaarheidsintervallen en hieraan verbonden R-procedures worden beschikbaar gesteld op de site http://statistics.tudelft.nl/CUPbook en hieraan gelinkte sites.

Een andere ontwikkeling vindt plaats in meerdimensionale censureringsproblemen. Er zijn modellen waarbinnen de grootheid waarin men is geïnteresseerd meerdimensionaal is. Dan zijn er veel manieren waarop deze grootheid gecensureerd kan zijn. In twee dimensies kan bijvoorbeeld de eerste coördinaat gecensureerd zijn en de tweede niet. Voor enkele typen censurering zijn karakteriseringen afgeleid voor schatters en asymptotische verdelingen afgeleid. Onder die asymptotische verdelingen komen verdelingen voor die niet eerder zijn bestudeerd.

Waar de asymptotische analyse van schatters met vormrestricties voornamelijk puntsgewijs gebeurt, zijn er ook ontwikkelingen bij het onderzoek naar asymptotisch gedrag van globale afstandsmaten. Denk aan een L_2-afstand tussen de schatter en de onderliggende functie die geschat wordt. Hierbij komen methoden aan de orde die gebruik maken van slim gekozen opsplitsingen van de globale afstandsmaat (denk aan partities van het interval van integratie waarover de L_2-afstand wordt berekend).



Figuur 7 Twee gladde monotone schattingen op basis van de rodehond data.

Uitdagingen

Veel uitdagingen sluiten aan bij de recente ontwikkelingen. De constructie van betrouwbaarheidsbanden voor functies die aan vormrestricties voldoen is pas bij enkele modellen bestudeerd. Bij meerdimensionale problemen ontbreekt die studie in het geheel. Nieuwe schatters moeten worden gedefinieerd, uitgerekend en asymptotisch bestudeerd. Hiervoor moeten nieuwe methoden worden ontwikkeld, die de vaak impliciete karakteriseringen van schatters gebruiken om het asymptotisch gedrag te bepalen.

De hier genoemde censureringsproblemen zijn net als deconvolutieproblemen en problemen uit de stereologie zogenaamde statistische inverse problemen. In dat type probleem komen vormrestricties heel natuurlijk voor. De mogelijkheden van het combineren van gladheid en vormrestricties is in die modellen nog maar weinig onderzocht. Juist als datasets niet heel groot zijn, is hier winst te halen. Welke schatters moeten dan gebruikt worden, en wat kan over de nauwkeurigheid van die schatters worden gezegd? Allemaal vragen die op antwoord wachten.

Tot slot

In deze bijdrage hebben we aan de hand van eenvoudige voorbeelden laten zien waar statistiek onder vormrestricties over gaat. De vormrestrictie is dan met name die van monotonie. Restricties als convexiteit, k-monotonie, volledige monotoniciteit en log-concaafheid zijn ook belangrijk en zorgen ten opzichte van de 'monotone problemen' vaak voor extra complicaties. Zo zijn de karakteriserende ongelijkheden voor de maximum-likelihood- (of kleinste kwadraten-) schatters vaak niet direct geschikt om deze schatters te bepalen. Het berekenen van die schatters vergt iteratieve algoritmen en ook de asymptotische eigenschappen van de schatters kunnen alleen op basis van de impliciete karakterisering worden afgeleid. De resulterende asymptotische verdelingen leiden vervolgens ook weer tot uitdagende vragen.

Al met al zal het duidelijk zijn dat het gebied van de statistiek met vormrestricties volop in beweging is!

Referenties

1 M. Ayer, H.D. Brunk, G.M. Ewing, W.T. Reid en E. Silverman, An empirical distribution function for sampling with incomplete information, *Annals of Mathematical Statistics* 26 (1955), 641–647.

2 R.E. Barlow, D.J. Bartholomew, J.M. Bremner en H.D. Brunk, *Statistical inference under order restrictions*, Wiley, New York, 1972.

3 C. van Eeden, Maximum likelihood estimation of ordered probabilities, *Indagationes Mathematicae* 18 (1956), 444–455.

4 U. Grenander, On the theory of mortality measurement, II, *Skandinavisk Aktuarietidskrift* 39 (1956), 125–153.

5 P. Groeneboom, Brownian motion with a parabolic drift and Airy functions, CWI Technical Report, oai.cwi.nl/oai/asset/6435/6435A.pdf, 1984.

6 P. Groeneboom, G. Jongbloed en B.I. Witte, Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model, *The Annals of Statistics* 38 (2010), 352–387.

7 P. Groeneboom en G. Jongbloed, *Nonparametric estimation under shape constraints*, Cambridge University Press, New York, 2014.

8 P. Groeneboom, G. Jongbloed, Nonparametric confidence intervals for monotone functions, te verschijnen in *Annals of Statistics* 43 (2015).

9 P. Groeneboom, S.P. Lalley en N.M. Temme, Chernoff's distribution and differential equations of parabolic and Airy type, *Journal of Mathematical Analysis and Applications* 423 (2015), 1804–1824.

10 N. Keiding, K. Begtrup, T.H. Scheike en G. Hasibeder, Estimation from current statust data in continuous time, *Lifetime Data Analysis* 2 (1996), 119–129.

11 E. Parzen, On Estimation of a probability density function and mode, *Annals of Mathematical Statistics* 32 (1962), 1065–1076.

12 T. Robertson, F.T. Wright en R.L. Dykstra, *Order restricted statistical inference*, Wiley, Chichester, 1988.

13 M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics* 27 (1956), 832–837.

14 B. Sen en G. Xu, Model based bootstrap methods for interval censored data, *Computational Statistics & Data Analysis* 81 (2015), 121– 129.

15 M.J. Silvapulle en P.K. Sen, *Constrained statistical inference*, Wiley, Hoboken, 2005.

Ieke Moerdijk
Afdeling Wiskunde
Radboud Universiteit Nijmegen
i.moerdijk@math.ru.nl

In Memoriam Daniël Marinus Kan (1927–2013)

Simpliciale verzamelingen en geadjungeerde functoren

Daniël Marinus Kan werd geboren in Amsterdam op 4 augustus 1927. Na een lang en vruchtbaar leven overleed hij in 2013 vredig in zijn eigen huis in Waban (Newton, MA, Verenigde Staten), onder de rook van Boston, op zijn zesentachtigste verjaardag. In 1982 werd Daan Kan benoemd tot correspondent van de Koninklijke Nederlandse Akademie van Wetenschappen. Dit levensbericht, geschreven door Ieke Moerdijk, is oorspronkelijk verschenen in de KNAW-publicatie *Levensberichten en herdenkingen 2014*.

Daan Kan groeide op als enig kind in een liberaal joods advocatengezin in Amsterdam-Zuid. Na de Montessori-kleuterschool en de toentertijd bekende lagere 'Openluchtschool voor het Gezonde Kind' in de Cliostraat te hebben doorlopen, ging hij in 1939 naar het Barlaeus Gymnasium. Daar kon hij in eerste instantie maar twee jaren blijven, omdat joodse kinderen na het schooljaar 1940–1941 van het Barlaeus af moesten. Hij heeft toen een jaar op het Joods Lyceum doorgebracht, waarna een akelige tijd voor Kan en zijn familie aanbrak.

In de zomer van 1943 werd hij samen met zijn ouders opgepakt en naar Westerbork gebracht. Daar bleven zij een half jaar, om vervolgens naar Bergen-Belsen te worden weggevoerd. Daan Kan heeft vijftien maanden in dat kamp doorgebracht. Zijn beide ouders zijn daar door uitputting en ondervoeding om het leven gekomen. Daan zelf heeft het ternauwernood overleefd en is na de ontzetting van het kamp nog drie maanden in Duitsland gebleven om enigszins te herstellen. In de zomer van 1945 keerde hij terug naar Amsterdam. Daar werd hij samen met drie andere joodse jongens in eerste instantie slechts voorwaardelijk toegelaten tot de zesde klas van het Barlaeus. Hierover sprak Kan tot op hoge leeftijd als iets wat hij als uitermate beledigend heeft ervaren, te meer daar deze vier jongens als besten eindexamen deden.

Intussen beraadde Kan zich op de studiekeuze. Hij was geïnteresseerd in wiskunde, maar wilde liever (in zijn eigen woorden) geen leraar worden of bij een verzekeringsmaatschappij gaan werken. Het was L.E.J. Brouwer die hem op andere beroepsmogelijkheden wees, zodat hij in 1946 met de studie wiskunde en natuurkunde met scheikunde aan (wat nu heet) de Universiteit van Amsterdam begon. In 1948 legde Kan met succes het kandidaatsexamen af, en eind 1950 het doctoraal.

Een tweetal zaken uit zijn studietijd zijn wellicht vermeldenswaard, daar deze van grote invloed zijn geweest op Kans latere ontwikkeling. Ten eerste sprak Kan altijd lovend over de colleges integratie en differentiatie van professor De Groot. In het voorjaar van 1949 gaf dezelfde De Groot een lezing over een zojuist verschenen boek over topologie van de hand van de wiskundige Lefschetz uit Princeton, en organiseerde vervolgens bij hem thuis een werkgroep met studenten om dit boek door te werken. Kan woonde genoemde lezing bij en sloot zich aan bij de werkgroep, net als bijvoorbeeld het in 2011 overleden Akademielid T.A. Springer. Ten tweede kreeg Kan na zijn kandidaatsexamen het aanbod om assistent bij Brouwer te worden, een assistentschap dat hij mede vanwege de grote geboden vrijheid als zeer stimulerend heeft ervaren. Zo hebben Brouwer en De Groot de verdere keuzes van Daan Kan tot in hoge mate bepaald.

Israël

In februari 1951 is Kan met aanbevelingsbrieven van Brouwer en De Groot op zak naar Israël vertrokken, alwaar hij na enkele weken een aanstelling kreeg aan het Weizmann Instituut. Er werd toen met seismische methoden naar olie in de woestijn gezocht, en Kan moest de wiskundige berekeningen daarvoor doen. (De zoektocht naar olie bleef overigens zonder succes.) Kan beschreef het werk als niet erg intellectueel uitdagend en enigszins eentonig. Na een jaar moest hij in militaire dienst, maar hij mocht zijn dienstplicht aan het Weizmann Instituut vervullen en kon zo nog tweeënhalf jaar daar blijven. Zijn werk bood hem veel vrije tijd, die hij gebruikte om de topologie met succes weer op te pakken.

In het voorjaar van 1954 kwam Samuel Eilenberg vanuit Columbia University voor een bezoek van twee maanden naar de Hebrew University in Jeruzalem. (Eilenberg was toen al een van de leidende figuren op het gebied van de algebraïsche topologie, door zijn werk met Saunders Mac Lane en het kort daarvoor verschenen zeer invloedrijke boek *Foundations of Algebraic Topology* dat hij samen met Norman Steenrod had geschreven.) Eilenberg sprak met Kan, en moedigde hem aan zijn resultaten op te schrijven. Er werd ad hoc een status van graduate student aan de Hebrew University voor Kan geregeld, alwaar hij in de zomer van 1954 een proefschrift indiende (de formele PhD-graad werd hem in 1955 toegekend).

Intussen was Kan in 1953 in het huwelijk getreden met Nora Poliakof, dochter van een Amsterdamse huisarts en net als Kan overlevende van Bergen-Belsen. Ook Nora had beide ouders in de oorlog verloren, en was al direct na de oorlog naar Israël verhuisd. De twee kenden elkaar uit hun kindertijd en hadden altijd wat contact gehouden. Zo had Kan Nora in 1949 in Israël opgezocht. Het echtpaar kreeg vier kinderen, Ittay (1956), Michael (1957), Tamara (1962) en Jonathan (1965). Jonathan overleed in 1973 aan leukemie, een zware klap voor het gezin Kan. Kans vrouw Nora overleed in augustus 2007.

Verenigde Staten

Na zijn promotie kreeg Kan een postdocaanstelling voor een jaar (1955–1956) aan Columbia University, bij Eilenberg. In dat jaar schreef hij drie baanbrekende artikelen over simpliciale verzamelingen en twee over geadjungeerde functoren (cf. [4–8]), die hem later grote faam en naamsbekendheid in het vakgebied bezorgden. Hierover later meer. Vervolgens verbleef Kan een jaar in Princeton, om daarna in 1957 naar Israël terug te keren op wat wij nu een *tenure track position* zouden noemen, aan de Hebrew University. In 1959 keerde hij echter terug naar de Verenigde Staten, waar hij een positie aan MIT aannam. Daar werd hij al na vier jaar tot Full Professor bevorderd, vanwege een concurrerend aanbod uit Amsterdam.

Kan is heel zijn verdere werkzame leven aan MIT verbonden gebleven, tot aan zijn pensionering in 1993. Onder zijn invloed ontstond de MIT-school in de algebraïsche topologie, met nadruk op simpliciale en categorische methoden, methoden die ook tegenwoordig dominant zijn in veel ontwikkelingen in het vakgebied. Kan heeft een aantal promovendi opgeleid die zich later tot belangrijke onderzoekers in de topologie ontwikkelden, waaronder Pete Bousfield, Ed Curtis, Emmanuel Dror Farjoun, Bill Dwyer, Phil Hirschhorn, Dave Rector en Jeff Smith. De grote verschillen in aard van begeleiding van deze promovendi zijn tekenend voor Kans originele en onafhankelijke manier van werken. Zo schreef Rector een proefschrift van zeven pagina's ('een creatieve jongen', zei Kan), terwijl hij Bousfield beschreef als 'a collaborator from day one'. Samen met enkelen van zijn promovendi schreef Kan twee belangrijke boeken, waarnaar hij zelf altijd verwees als *The Yellow Monster* (1972) en *The Blue Beast* (2004), zie [1] en [2].

Ook op een aantal van zijn collega's op MIT heeft het werk van Kan veel invloed gehad. Het bekendste voorbeeld hiervan is het werk van Daniel Quillen over onder meer rationale homotopietheorie en K-theorie, dat doordrenkt is van categorietheoretische en simpliciale technieken.

Na zijn pensionering is Kan niet veel meer op MIT geweest. Hij hield op andere manier contact met jongere collega's, die veel bij hem thuis op bezoek kwamen.

Nederland

Kan heeft na zijn vertrek weinig contact met Nederland gehouden. Hij had zijn familie verloren en enkele van zijn beste jeugdvrienden waren ook geëmigreerd (zijn vrouw Nora ging nog wel regelmatig terug om vriendinnen uit haar jeugd op te zoeken). Maar Kan koesterde zijn band met de Akademie, en publiceerde regelmatig in *Indagationes Mathematicae*. En hij was een fervent fietser: tot op hoge leeftijd klom hij in weer en wind, en gekleed in knalrood tenue, op zijn fraaie mountainbike, om op de koffie te gaan bij een van zijn medewiskundigen en daar de laatste nieuwtjes op te doen, of om samen aan een artikel te werken.

Wiskundig werk

Topologie is de bestudering van ruimtelijke objecten en hun mogelijke vervormingen, in al hun algemeenheid. In het bijzonder kunnen deze objecten een willekeurig grote dimensie hebben. De resultaten en methoden van de topologie worden overal in de wiskunde gebruikt. Een belangrijk onderwerp van studie is de manier waarop een mogelijkerwijs wat vervormde kopie van één zo'n object in een ander gemaakt kan worden. Wiskundigen noemen de objecten topologische ruimten, en spreken van afbeeldingen tussen die objecten. De algebraïsche topologie probeert de objecten en de mogelijke afbeeldingen te classificeren met behulp van algebraïsche kenmerken van deze objecten. In het algemeen kunnen deze kenmerken twee objecten niet onderscheiden als de een in de ander vervormd kan worden door 'duwen en trekken' (in tegenstelling tot 'plakken en knippen'). Na funderend werk van onder anderen Brouwer en Freudenthal beleefde de algebraïsche topologie na de oorlog een heel snelle en zeer succesvolle ontwikkeling, onder meer door het werk van Steenrod, Lefschetz, Eilenberg en Mac Lane in de Verenigde Staten, Henri Cartan in Frankrijk en vele anderen.

Simpliciale verzamelingen

Het contrast tussen de veelheid van topologische ruimten en hun continue vervormingen enerzijds, en de starheid van de algebraïsche invarianten anderzijds, vroeg om een meer

discrete of combinatorische aanpak van topologische ruimten, en hier lag het baanbrekende werk van Kan. Een simpliciale verzameling ('complete semi-simplicial complex' was de Engelstalige terminologie in de vijftiger jaren) kan men zich voorstellen als een ruimte, opgebouwd uit punten, lijnstukjes, driehoekjes, tetraëders, enzovoort, die slechts op een zeer beperkte manier aan elkaar geplakt mogen worden, bijvoorbeeld door voor te schrijven welke lijnstukjes als rand van welke driehoekjes figureren, welke driehoekjes als zijvlak van welk tetraëders, enzovoort. Een simpliciale verzameling kan dus beschreven worden door voor elke dimensie $0, 1, 2, 3, \ldots$ een aantal 'simplices' (dat wil zeggen punten, lijntjes, driehoekjes, enzovoort) te specificeren, samen met de 'plakinstructies'. Dit geeft een structuur die in allerlei delen van de wiskunde op blijkt te duiken. Een fundamenteel resultaat uit de topologie is dat voor elke topologische ruimte er een simpliciale verzameling gevonden kan worden die bijna niet te onderscheiden is van de oorspronkelijke ruimte, in de zin dat de gebruikelijke algebraïsche kenmerken volledig overeenkomen. Kans fundamentele bijdragen uit de vijftiger jaren [4–6] gaven onder andere een methode om de belangrijkste en meest onderscheidende van die algebraïsche kenmerken, de zogenaamde homotopiegroepen, van een topologische ruimte volledig te beschrijven in termen van de daarbij passende simpliciale verzameling. Kan ontdekte dat men zich hierbij kan beperken tot simpliciale verzamelingen die een bepaalde volledigheidseigenschap hebben. Deze staan nu bekend als 'Kan-complexen'. De ontdekking van dit begrip baande de weg tot een volledige beschrijving van de algebraïsche topologie (preciezer, van de 'homotopie-categorie') in termen van simpliciale verzamelingen en hun onderlinge relaties, waarbij de Kan-vezelingen ('Kan fibrations'), een generalisatie van de Kan-complexen, tot op de dag van vandaag een centrale rol spelen. Een meer gedetailleerde beschrijving van simpliciale verzamelingen en Kan-complexen wordt gegeven in een kader aan het eind van dit artikel.

Categorieën en functoren

Rond de oorlog ontstond in de topologie steeds meer het besef dat er een formalisme nodig was dat beter in staat was om wiskundige objecten niet zozeer in hun isolement te beschrijven, maar de nadruk legt op de mogelijke afbeeldingen tussen verschillende objecten, en bovendien het gedrag van de algebraïsche kenmerken van objecten onder die afbeeldingen beter kan beschrijven. Dit leidde tot de geboorte van de theorie van categorieën en functoren (zie [9]). Wiskundige objecten van een bepaalde soort en afbeeldingen daartussen vormen een 'categorie' (bijvoorbeeld topologische ruimten, of verzamelingen, of groepen, respectievelijk ringen uit de algebra) en constructies als de algebraïsche kenmerken van topologische ruimten kunnen efficiënt beschreven worden als 'functoren' van de ene categorie naar de andere. Gegeven zo'n functor tussen categorieën is er vaak een 'zuinigste' functor in de omgekeerde richting. Kan destilleerde dit begrip van 'adjoint functor', en liet zien dat veel belangrijke constructies in de wiskunde voorbeelden van adjoint functors zijn. Ook beschreef hij, gegeven een functor, precieze criteria voor het bestaan van zo'n geadjungeerde functor in de andere richting, met behulp van een later veel gebruikte methode bekend als 'Kan-extensie' (zie [7–8]). Een meer gedetailleerde beschrijving van geadjungeerde functoren wordt gegeven in een kader aan het eind van dit artikel.

Later werk

Met deze twee ontdekkingen, van Kan-complexen en Kan-extensies, had Kan reeds aan het begin van zijn wetenschappelijke carrière een naam verworven. Maar ook in latere jaren bleef Kan productief en leverde belangrijke bijdragen aan het vakgebied. Ik wil twee voorbeelden noemen. Ten eerste het werk met Bousfield over 'lokalisatie en completering' [1]. Enigszins vereenvoudigd gesteld is het idee als volgt: een topologische ruimte (of simpliciale verzameling!) heeft zoals gezegd algebraïsche kenmerken, die zich laten verenigen in bepaalde algebraïsche structuren zoals bijvoorbeeld een ring. Als je nu aan de algebraïsche kant die structuur enigszins verandert (als je de ring bijvoorbeeld lokaliseert of completeert), is er dan een topologische ruimte die precies de nieuwe kenmerken heeft? En hoe kun je die maken uit de eerste gegeven ruimte? Bousfield en Kan hebben in hun boek op deze vragen hele precieze antwoorden gegeven, op een manier die niet mogelijk was geweest zonder voorgaande ontdekkingen over categorieën en simpliciale verzamelingen. Een tweede voorbeeld is een serie van drie artikelen die Kan samen met Bill Dwyer schreef, gepubliceerd rond 1980. Quillen had eerder een axiomatische manier gevonden om aan twee objecten in een categorie een topologische ruimte (of beter, een simpliciale verzameling) toe te kennen die de structuur van alle afbeeldingen van het ene object in het andere beschreef. Deze constructie (voor vakgenoten: van de 'derived mapping space') is van groot belang in allerlei contexten. Samen met Dwyer vond Kan een alternatieve constructie, bekend als de simpliciale of 'hammock'-lokalisatie van een categorie, die hetzelfde resultaat oplevert als de constructie van Quillen, maar veel minder informatie gebruikt en dus veel algemener toepasbaar is gebleken.

Referenties

1 A.K. Bousfield en D.M. Kan, *Homotopy limits, completions and localizations*, Springer, Berlijn , 1972.

2 W.G. Dwyer, P.S. Hirschhorn, D.M. Kan en J.H. Smith, *Homotopy Limit Functors on Model Categories and Homotopical Categories*, American Mathematical Society, Providence, RI, 2004.

3 W.G. Dwyer en D.M. Kan, Simplicial localizations of categories, *Journal of Pure and Applied Algebra* 17 (1980), 267-284.

4 D.M. Kan, On c.s.s. complexes, *American Journal of Mathematics* 79 (1957), 449–476.

5 D.M. Kan, A combinatorial definition of homotopy groups, *Annals of Mathematics* 67 (1958), 282–312.

6 D.M. Kan, On homotopy theory and c.s.s. groups, *Annals of Mathematics* 68 (1958), 38–53.

7 D.M. Kan, Adjoint functors, *Transactions of the American Mathematical Society* 87 (1958), 294–329.

8 D.M. Kan, Functors involving c.s.s. complexes, *Transactions of the American Mathematical Society* 87 (1958), 330–346.

9 S. Mac Lane, *Categories for the Working Mathematician*, Springer, Berlijn, 1971, 2de herziene druk, 1998.

Simpliciale verzamelingen en Kan-complexen

Een simpliciale verzameling is een systeem X van verzamelingen en afbeeldingen daartussen van de volgende vorm: voor elk natuurlijk getal $n \geq 0$ is er een verzameling X_n, waarvan de elementen de n-simplices van X genoemd worden. Verder is er voor elke niet-dalende functie

$$\alpha : \{0, \ldots, n\} \longrightarrow \{0, \ldots, m\}$$

een afbeelding α^* of $X(\alpha) : X_m \to X_n$. Deze afbeeldingen moeten bovendien aan twee voorwaarden voldoen: ten eerste moet gelden dat als α een identiteitsfunctie is, α^* dat ook is; verder moet voor een α als hierboven en een tweede functie $\beta : \{0, \ldots, m\} \to \{0, \ldots, k\}$ gelden dat

$$(\beta \circ \alpha)^* = \alpha^* \circ \beta^* : X_k \longrightarrow X_n.$$

Voorbeelden van zulke niet-dalende functies α zijn de injectieve functies

$$\delta_i : \{0, \ldots, n-1\} \to \{0, \ldots, n\}$$

die de waarde i overslaan (voor $i = 0, \ldots, n$) en de surjectieve functies

$$\sigma_j : \{0, \ldots, n\} \to \{0, \ldots, n-1\}$$

die de waarde j twee keer aannemen (voor $j = 0, \ldots, n-1$). Elke α kan geschreven worden als samenstelling van enkele δ_i's en σ_j's en daarom kan een simpliciale verzameling X ook beschreven worden door alleen de verzamelingen X_n en de operaties δ_i^* en σ_j^* te geven, zolang de laatste maar voldoen aan de vergelijkingen voor de samenstelling, zoals $\delta_i^* \delta_j^* = \delta_{j-1}^* \delta_i^*$ voor $i < j$. Men schrijft gewoonlijk $d_i : X_n \to X_{n-1}$ voor δ_i^* en $s_j : X_{n-1} \to X_n$ voor σ_j^*. Deze afbeeldingen d_i en s_j heten de kant-afbeeldingen en degeneratie-afbeeldingen en passen samen in een diagram:

$$X_0 \underset{d_0}{\overset{d_1}{\rightleftarrows}} X_1 \underset{d_0}{\overset{d_2}{\rightleftarrows}} X_2 \underset{d_0}{\overset{d_3}{\rightleftarrows}} \cdots$$

terwijl de genoemde vergelijkingen voor de samenstellingen van deze afbeeldingen de simpliciale identiteiten heten.

Een simpliciale verzameling X moet geïnterpreteerd worden als een combinatorische beschrijving van een topologische ruimte $|X|$, de zogenaamde meetkundige realisatie van X. Elk element x in X representeert een n-simplex in $|X|$, en de kantafbeeldingen d_i geven aan hoe die simplices aan elkaar geplakt moeten worden. Bovendien geven de degeneratie-afbeeldingen s_j aan hoe je een $(n-1)$-simplex op kunt vatten als een dichtgeklapt (gedegenereerd) n-simplex. Voor een precieze beschrijving definiëren we eerst het standaard n-simplex

$$\Delta^n = \{(t_0, \ldots, t_n) \in \mathbb{R}^n \mid 0 \leq t_i \leq 1, \Sigma_i t_i = 1\}.$$

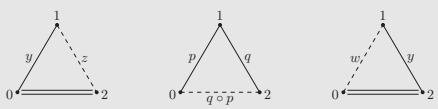
Dit n-simplex heeft $n+1$ hoekpunten van de vorm $(0, \ldots, 0, 1, 0, \ldots, 0)$. Elke functie $\alpha : \{0, \ldots, n\} \to \{0, \ldots, m\}$ zoals hierboven induceert een affiene lineaire afbeelding $\alpha_* : \Delta^n \to \Delta^m$ die het i-de hoekpunt van Δ^n naar het $\alpha(i)$-de hoekpunt van Δ^m stuurt:

$$\alpha_*(t_0, \ldots, t_n) = (u_0, \ldots, u_m)$$
$$\text{met} \quad u_j = \Sigma_{\alpha(i)=j}\, t_i.$$

Vervolgens definieert men $|X|$ als het quotiënt van de disjuncte vereniging van kopieën van de verschillende Δ^n, precies een kopie voor elke $x \in X_n$ en elke $n \geq 0$. Dit quotiënt wordt verkregen door, voor elke $\alpha : \{0, \ldots, n\} \to \{0, \ldots, m\}$ en elke $x \in X_m$, een punt $(t_0, \ldots, t_n)$ in de kopie van Δ^n behorende bij $(n, \alpha^* x)$ te identificeren met het punt $\alpha_*(t_0, \ldots, t_n)$ in de kopie van Δ^m behorende bij (m, x).

Zo geeft elke simpliciale verzameling X een topologische ruimte $|X|$. Omgekeerd krijgen we voor elke topologische ruimte T een simpliciale verzameling $\mathrm{Sing}(T)$, het zogenaamde singuliere complex van T. De verzameling $\mathrm{Sing}(T)_n$ van n-simplices in $\mathrm{Sing}(T)$ is hierbij de verzameling van alle continue afbeeldingen $\Delta^n \to T$, en de operaties $\alpha^* : \mathrm{Sing}(T)_m \to \mathrm{Sing}(T)_n$ worden eenvoudig gedefinieerd door samenstelling met α_*: voor een continue afbeelding $f : \Delta^n \to T$ geldt $\alpha^*(f) = f \circ \alpha_*$.

Er bestaat een evaluatie-afbeelding $|\mathrm{Sing}(T)| \to T$ die $(t_0, \ldots, t_n)$ in Δ^n in de kopie behorende bij $(g : \Delta^n \to T)$ in $\mathrm{Sing}(T)_n$ stuurt naar $g(t_0, \ldots, t_n) \in T$. Een fundamenteel resultaat in de topologie is dat via deze afbeelding nauwelijks onderscheid kan worden gemaakt tussen T en $|\mathrm{Sing}(T)|$, in de zin dat deze isomorfismen in homotopie en cohomologie induceert. In feite is deze evaluatie-afbeelding een homotopie-equivalentie als T een voldoende nette ruimte is (een manifold, bijvoorbeeld). Mede om deze reden ligt het voor de hand te trachten om de homotopie-theorie van topologische ruimten te ontwikkelen op een zuiver combinatorische manier, alleen met simpliciale verzamelingen. Het werk van Daan Kan heeft aangetoond dat dit inderdaad mogelijk is, mits men zich maar beperkt tot simpliciale verzamelingen die aan een zekere extensieconditie voldoen. Deze conditie staat tegenwoordig bekend als de Kan-conditie, en de simpliciale verzamelingen die eraan voldoen heten Kan-complexen. Voordat we de conditie formuleren, merken we op dat een n-simplex $x \in X_n$ een $(n+1)$-tal kanten $d_i(x) \in X_{n-1}$ heeft (voor $i = 0, \ldots, n$), die aan elkaar passen op grond van de simpliciale identiteiten, zoals $d_i d_j x = d_{j-1} d_i x$ voor $i < j$. De Kan-conditie is nu de voorwaarde dat indien, omgekeerd, n zulke potentiële kanten x_i in X_{n-1} zijn gegeven voor elke $i = 0, \ldots, n$ op één na, zeg k, die aan elkaar passen alsof ze de kanten van een n-simplex x zijn, er dan ook inderdaad een n-simplex x in X bestaat met $d_i(x) = x_i$ voor elke i behalve k. Deze Kan-conditie stelt ons in staat om 'paden' samen te stellen of van richting te veranderen zoals voor paden in een topologische ruimte. Ter illustratie hiervan geven we voorbeelden van diagrammetjes die horen bij de Kan-condities voor $n = 2$ en $k = 0, 1, 2$. Zie Figuur 1.



Figuur 1

Het gelijkheidsteken hier geeft een gedegenereerd 1-simplex aan. Dus de Kan-conditie voor het eerste plaatje zegt: gegeven een $y \in X_1$ bestaat er een $x \in X_2$ met $d_2 x = y$ en $d_1 x$ gedegenereerd. Dan is $z = d_0 x$ dus een 1-simplex met de eigenschap dat $d_1 z = d_0 y$ en $d_0 z = d_1 y$; met andere woorden, z kan worden opgevat als het pad y doorlopen in omgekeerde richting, een soort links-inverse voor y. Het tweede plaatje beschrijft de samenstelling van paden p en q; het derde plaatje beschrijft, analoog aan het eerste plaatje, een rechts-inverse voor y.

Aanbevolen literatuur: Er zijn vele goede leerboeken over simpliciale verzamelingen geschreven aan het eind van de zestiger jaren nadat de basistheorie was ontwikkeld, bijvoorbeeld:

- P. Gabriel en M. Zisman, *Calculus of Fractions and Homotopy Theory*, Springer.
- K. Lamotke, *Semisimpliziale algebraische Topologie*, Springer.
- J.P. May, *Simplicial Objects in Algebraic Topology*, University of Chicago Press.

Een moderner boek is:

- P.G. Goerss en J.F. Jardine, *Simplicial Homotopy Theory*, Springer.

Geadjungeerde functoren

Diverse constructies in de algebraïsche topologie, zoals homologie- en homotopiegroepen, hebben goede eigenschappen met betrekking to continue afbeeldingen tussen topologische ruimten. Om deze eigenschappen te beschrijven introduceerden Eilenberg en Mac Lane de taal van categorieën en functoren. Een categorie is een structuur bestaande uit 'objecten' en 'morfismen' daartussen. In het bijzonder is er voor elk object X in een categorie een identiteitsmorfisme gegeven en is er een associatieve compositie van morfismen, die voor twee morfismen $f : X \to Y$ en $g : Y \to Z$ een morfisme $gf : X \to Z$ geeft. De identiteitsmorfismen werken als neutrale elementen voor deze compositieoperatie. De meeste typen wiskundige objecten vormen categorieën: zo is er de categorie van topologische ruimten en continue afbeeldingen daartussen, de categorie van groepen en homomorfismen, de categorie van (bijvoorbeeld) reële vectorruimten en lineaire afbeeldingen daartussen, de categorie van verzamelingen en functies daartussen, enzovoort, enzovoort.

Een 'functor' is een soort afbeelding tussen categorieën: zo'n functor stuurt objecten van de ene categorie naar objecten van de andere en net zo voor morfismen, op zo'n manier dat identiteiten en compositie gerespecteerd worden. Enkele voorbeelden van functoren vanuit de categorie van topologische ruimten en continue afbeeldingen zijn de 'onderliggende verzameling', naar de categorie van verzamelingen en functies, en de n-de homologie, naar de categorie van abelse groepen en homomorfismen (voor elke $n \geq 0$ een functor). De fundamentaalgroep geeft een voorbeeld van een functor van de categorie van topologische ruimten met basispunt naar de categorie van groepen.

Veel van deze zogenaamde functoriële constructies hebben speciale eigenschappen als het gaat om morfismen vanuit of naar waarden van de betreffende functor. Zo is voor een vectorruimte V een morfisme van commutatieve $\mathbb{R}$-algebra's van de symmetrische algebra $S(V)$ naar een andere $\mathbb{R}$-algebra A 'hetzelfde als' een morfisme van V naar A opgevat als vectorruimte. Deze correspondentie kan worden beschreven door een tweetal functoren

$$S : (\text{reële vectorruimten}) \rightleftarrows (\mathbb{R}\text{-algebra's}) : U$$

waarbij $U(A)$ gedefinieerd is als de onderliggende vectorruimte van een algebra A. Er is een een-op-eencorrespondentie tussen morfismen $S(V) \to A$ in de ene categorie en $V \to U(A)$ in de ander. (Bovendien gedraagt deze correspondentie zich netjes ten opzichte van de compositie aan beide kanten.) In zo'n soort situatie zegt men dat de functor S links geadjungeerd is aan U, en U rechts geadjungeerd aan S. Het was Daan Kan die in het begin van de vijftiger jaren dit begrip van geadjungeerde functor ontdekte, geïnspireerd door de analogie tussen twee bekende een-op-eencorrespondenties: de ene tussen afbeeldingen $\Sigma(X) \to Y$ en $X \to \Omega(Y)$, dat wil zeggen van de suspensie van een topologische ruimte X naar een andere ruimte Y en van X naar de lussenruimte van Y; de andere tussen morfismen van modulen $L \otimes M \to N$ en morfismen $M \to \mathrm{Hom}(L, N)$. Het werd al snel daarna duidelijk dat dit soort paren van geadjungeerde functoren in heel veel verschillende situaties in de wiskunde opduiken en hoe belangrijk ze daar zijn. Omgekeerd is het voor een gegeven constructie vaak nuttig om te kijken of er soms een links of rechts geadjungeerde voor bestaat. De constructies van het singuliere complex van een topologische ruimte en van de meetkundige realisatie, besproken in het andere kadertje, zijn ook voorbeelden hiervan: realisatie is links geadjungeerd aan het singuliere complex. Dit betekent dat er een bijectieve correspondentie is tussen morfismen van simpliciale verzamelingen $X \to \mathrm{Sing}(T)$ en continue afbeeldingen $|X| \to T$. In het bijzonder krijgen we als we deze correspondentie toepassen op een identiteitsmorfisme $\mathrm{Sing}(T) \to \mathrm{Sing}(T)$ een continue afbeelding $|\mathrm{Sing}(T)| \to T$. Zoals vermeld in het andere stukje induceert deze afbeelding isomorfismen op het niveau van allerlei invarianten uit de algebraïsche topologie en drukt op een meer formele manier uit dat simpliciale verzamelingen en topologische ruimten vanuit het perspectief van de algebraïsche topologie equivalente categorieën zijn.

Cor Baayen
emeritus hoogleraar VU, Almelo
pcbaayen@planet.nl

In Memoriam Piet Mullender (1917–2014)

Flitsend wiskundige en bekwaam bestuurder

Op 31 juli 2014 overleed Pieter Mullender, hoogleraar wiskunde aan de Vrije Universiteit van 1952 tot zijn emeritaat in 1981, op de leeftijd van 97 jaar. Piet Mullender werd op 18 juli 1917 in Amsterdam geboren. Hij ging in Amsterdam naar de middelbare school en hij studeerde wiskunde aan de Vrije Universiteit. Tijdens de Tweede Wereldoorlog schreef hij het proefschrift *Toepassing van de meetkunde der getallen op ongelijkheden in* $K(1)$ *en* $K(i\sqrt{m})$**, waarop hij in juli 1945 bij J.F. Koksma promoveerde. Na zijn promotie studeerde hij op aanraden van Koksma enkele jaren in Cambridge. In 1949 werd hij benoemd tot lector aan de VU. De meetkunde der getallen bleef zijn onderzoeksterrein. Zijn onderwijs bestreek een breed spectrum van onderwerpen, waaronder getallentheorie, algebra, verschillende analyse-vakken en, althans in de beginfase, ook mechanica. Voor twee studenten was hij de promotor: Willem Kuijk (1960) en Jan van de Lune (1984). Hij heeft een belangrijke rol gespeeld bij de ontwikkeling van wiskunde en informatica aan de VU. Na zijn emeritaat verhuisden Piet Mullender en zijn vrouw Fini naar Doorwerth. In 2013 verhuisden ze naar een verzorgingshuis in Driebergen. Fini overleed in Driebergen in 2013 en Piet in 2014. Bij de begrafenis sprak Cor Baayen, die Mullender als student aan de VU en later als collega ook aan de VU heeft meegemaakt, de volgende rede uit.**

Wij nemen afscheid van een geliefd en gewaardeerd mens, die heel veel heeft betekend voor zijn vrouw, zijn kinderen en hun gezinnen, maar ook voor generaties studenten, voor collega's en andere vrienden. En daarbij: hij was de architect van de huidige afdeling Wiskunde en Informatica van de Vrije Universiteit. Piet Mullender was een heel veelzijdig mens: een geliefde pater familias niet alleen, maar ook een flitsende wiskundige, een boeiende docent, een goede en hartelijke collega en vriend, en een bekwaam bestuurder die heel veel tot stand heeft gebracht voor de wiskunde en informatica aan de VU.

In 1951, toen ik ging studeren in Amsterdam aan de Vrije Universiteit, ontmoette ik dr. Mullender voor het eerst. Doctor, niet professor Mullender. De wiskundegroep aan de VU was toen nog heel klein. Er waren twee hoogleraren (Koksma en Grosheide) één bijzonder hoogleraar (van Rooijen) één student-assistent (Dirk Kruyswijk), en een lector: dr. Mullender. Klein was wiskunde niet alleen in personele omvang: de accommodatie was er ook naar. In het Natuurkundig Laboratorium aan de De Lairessestraat waren twee collegezalen beschikbaar: een kleine en een grote. Die grote zaal was eigenlijk van natuurkunde, maar omdat natuurkundestudenten destijds ook wiskundige vaardigheden moesten aanleren werd in die zaal college gegeven aan wis- en natuurkundestudenten samen; er waren ook nog scheikundigen bij. Mullender gaf analyse in die grote zaal, maar ter wille van die fysici en chemici werden in dat 'massacollege' de grondslagen en de moeilijkere bewijzen weggelaten: de zeven wiskundestudenten kregen dat één middag per week aangevuld in de kleine collegezaal. Docent Mullender bracht ons verder de mechanica bij, en gaf ook het college beschrijvende meetkunde (toen nog!). Achter die kleine zaal was een nog veel kleiner kamertje: daar konden de wiskunde-docenten zich terugtrekken in de kwartiertjes tussen de collegeuren en eventueel hun meegebrachte boterham eten of hun aantekeningen nog eens doornemen.

Mullender was een goed docent, hij kon heel flitsend en speels met de stof omgaan. Maar hij was wel eens wat nonchalant en het kon gebeuren dat hij vastliep in een ingewikkelde bewijsvoering. Een paar jaar later heb ik meegemaakt, bij een doctoraal caputcollege (een keuzevak, zogezegd) dat Mullender na een maand zei: het lukt zo niet, doet u uw aantekeningen maar weg, we beginnen helemaal opnieuw. Het onderwerp van dat caputcollege was de meetkunde der getallen, het specialisme waarin Piet Mullender gepromoveerd was. Het gaat daarbij om wiskundige problemen die leven op verzamelingen punten in een regelmatig patroon (een rooster), zoals bijvoorbeeld de hoekpunten van de tegels in een regelmatige betegeling. Toen Piet Mullender, na al snel benoemd te zijn tot hoogleraar, zijn inaugurele oratie hield in de Woestduinkerk (de VU had nog geen aula), sprak hij daar ook over. Speels en heel verhelderend maakte Mullender gebruik van de aanwezigheid van glas-in-loodbeglazing in de kerkramen om aan zijn toehoorders uit te leggen wat een puntenrooster was en welke problemen daarmee verbonden konden worden. De zon scheen die middag, en ik herinner mij hoe het zonlicht een kleurig rooster wierp op de vloer en over de toehoorders.

Wiskunde en informatica aan de VU
Meer over de ontwikkeling van wiskunde en informatica aan de VU is te vinden in Hendrik Blauwendraat, *Worsteling naar Waarheid, De opkomst van Wiskunde en Informatica aan de VU*, Uitgeverij Meinema, Zoetermeer, 2004.

Piet Mullender in 2013

Piet Mullender in de vijftiger jaren

Ja, en als je destijds tentamen ging doen dan moest je natuurlijk bij de docent thuis komen. Waar anders: op het lab was immers geen plaats! Voor mijn allereerste tentamen, analyse, werd ik vriendelijk binnengelaten door mevrouw Mullender, die ons ook koffie bracht en mij bemoedigend toeknikte.

Waarom ik hier zo uitvoerig bij stil sta? Niet zozeer omdat ik met dankbaarheid en plezier terug denk aan die vroege tijd en aan Mullenders colleges (ook daarom), maar vooral om u een indruk te geven van de beginsituatie van waaruit Piet Mullender, als hoogleraar-directeur van het door hem opgezette Wiskundig Seminarium, in de jaren 1962–1973 een instituut heeft opgebouwd dat in 1973, bij zijn aftreden als directeur, een wetenschappelijke en administratieve staf telde van bijna vijftig personen. Een seminarium, een zaaistation, dat wilde hij tot stand brengen. Hij kreeg voor elkaar dat een pand werd gehuurd, niet ver van het laboratorium in de De Lairessestraat, en voor het eerst kregen de wiskundigen eigen werkruimte. Paul Noordzij was aangetrokken als docent, en er werd iemand benoemd voor eigen administratie en beheer: Corry van Rossum. In 1962 ook werd mij een leeropdracht (om te beginnen functionaalanalyse) verleend.

In december 1964 overleed Koksma, de eerste en meest eminente wiskundehoogleraar van de Vrije Universiteit, en de leermeester en promotor van Piet Mullender. En in de loop van 1965 konden wij, in fasen, de nieuwbouw in Buitenveldert betrekken. Piet Mullender heeft zich in die tijd geweldig ingespannen voor de ontwikkeling en inrichting van de wiskunde. Hij vroeg mij een ontwikkelingsplan voor het onderwijs te schrijven, en wist dat te gebruiken om zelfs twee lectoraten toegewezen te krijgen (later omgezet in professoraten). Die werden bezet door Rien Kaashoek en Maarten Maurice. Later verwierf hij verdere formatieruimte voor toegepaste wiskunde, voor stochastiek, er kwam later een lector algebra bij, er werden meer studentassistenten en promotiemedewerkers aangesteld. Bij zijn aftreden als hoogleraar-directeur in 1973 was er zelfs al een kleine maar veelbelovende groep Informatica. Onder Mullenders vriendelijke maar krachtige leiding ontwikkelde Corry van Rossum zich tot een heel bekwame en deskundige administrateur: zij was jarenlang Piets rechterhand.

Bovenal, Piet Mullender was een buitengewoon vriendelijk bestuurder. Hij ontving zijn nieuwe collega's met open armen, vaak ook 's avonds bij hem thuis, met zijn vrouw Fini (ook een wiskundige en voormalige VU-studente) als fantastische gastvrouw. Als regel hadden we het dan niet zozeer over ons werk: cultuur kwam breed aan de orde: muziek, literatuur, beeldende kunsten. Wat vonden wij van *Le Petit Prince* van Saint-Exupéry, of hoe waardeerden wij de uitvoering van Rossini's 'Sonata's voor Strijkers' in de uitvoering van de Academy of St Martin-in-the-Fields onder Neville Mariner? Hoe hield Piet Mullender zich staande in al die drukte? (In die tijd gaf hij ook nog leiding aan het bestuur van het Christelijk Lyceum Buitenveldert: ook dat moet hem veel tijd en energie hebben gekost.) Het toverwoord was *delegeren*. Piet bood zijn nieuwe collega's veel ruimte, maar er was een taakverdeling: de nieuwe collega's schreven de ontwikkelingsplannen, hij verdedigde ze — met succes — bij het universitaire bestuur, en dan moesten de nieuwe collega's natuurlijk de plannen ook uitvoeren.

Veel van de oplossingen voor al die praktische en organisatorische problemen van die jaren van ontwikkeling en groei zijn, tot ver na Piets emeritaat in 1981, de VU-wiskunde en informatica tot groot voordeel geweest. In de jaren na zijn pensionering — meer dan dertig jaar! — is (en dat spreekt vanzelf) distantie, afstand ontstaan tussen Piet Mullender en de nu werkzame wiskundigen en informatici aan de VU. Maar hij is en wordt niet vergeten, zijn naam wordt met ere en respect genoemd. En de weinigen van zijn collega's van destijds die er nog zijn, gedenken hem met grote waardering en dankbaarheid.

Boekbesprekingen

| Book Reviews

Johannes Ueberberg
Foundations of Incidence Geometry Projective and Polar Spaces
Springer Monographs in Mathematics
Springer, 2011
248 p., prijs € 85,55
ISBN 978-3-642-20971-0

This book provides an introduction into projective and affine spaces as well as polar spaces in the modern language of diagram geometry.

The classical theory of projective spaces is discussed in the first part of the book. These spaces are introduced as point-line incidence geometries satisfying the axioms that any two points are incident with a unique line, that lines are incident with at least three points, and the axiom of Veblen–Young (if l and m are two lines incident with a point p, then any two lines meeting both l and m not in p, will meet at a point).

Both the first Fundamental Theorem for Projective spaces, stating that a Desarguesian projective space is isomorphic to a projective space of a vector space, as well as the second Fundamental Theorem, stating that all automorphisms of Desarguesian projective spaces are induced by semi-linear transformations of the underlying vector space, are discussed in full detail. The corresponding results for affine spaces are also included in this first part. The second part of the book treats polar spaces. A polar space is a point-line incidence geometry in which any two points are on at most one line, lines contain at least three points, and a point is collinear with one or all points of any line. The author discusses polar spaces defined by polarities of projective spaces, the corresponding sesquilinear forms and (pseudo-) quadratic forms. Several classical (classification) results on polar spaces are provided with detailed proofs. But, unfortunately, the main result on polar spaces, the classification of nondegenerate polar spaces containing projective planes, based on work by Tits, Buekenhout, Shult and Johnson, is stated only for spaces of finite rank and without a proof.

The book is well written and contains many enlightning pictures. It is mainly directed to students. However, for them it might be disappointing that the book contains no exercises. Although researchers can also find various interesting results in the book, they probably find the book *Diagram Geometry* by Cohen and Buekenhout (Springer, 2013) more appealing.

Hans Cuypers

Rectificatie

In de bespreking van het boek *België + wiskunde* door Robert van der Waall op p. 146 van het juninummer is een door de recensent doorgegeven wijziging per abuis niet verwerkt. Op de plaats van de laatste zin van de eerste alinea, “Nergens ben ik zo’n boeiende beschrijving van hun leven en werk tegengekomen; ook niet dus in internationale periodieken of boekwerken”, dient gelezen te worden “In 2013 is de Abelprijs aan Deligne toegekend. Naar aanleiding daarvan heeft Frans Oort in het NAW van september 2014 verhaald over de ‘flow of mathematics’, die Deligne tot het bewijs bracht van de Weil-vermoedens in de jaren zeventig van de vorige eeuw. Ik durf te stellen dat combinering van de bijdragen van Oort en van Huylebrouck aangaande Deligne een buitengewoon volledig beeld oplevert van Delignes leven en werk.”

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

Machiel van Frankenhuijsen

The Riemann Hypothesis for Function Fields
Frobenius Flow and Shift Operators
London Mathematical Society Student Texts 80

Cambridge University Press, 2014
xii + 152 p., prijs £22.99
ISBN 9781107685314

"This book grew out of an attempt to understand a paper by Alain Connes in which he constructs a beautiful noncommutative space with a view to proving the Riemann hypothesis."

In 1859 Riemann wrote an astonishing paper, 'Über die Anzahl der Primzahlen unter einer gegebenen Grösse', *Monatsberichte der Berliner Akademie*, November 1859 (see http://www.claymath.org/sites/default/files/zeta.pdf). These merely nine pages are the only testimony of Riemann's interest in number theory. His question, 'the Riemann hypothesis', is still unsolved, although it received an impressive amount of attention (Hilbert 1900, the Millennium problems 2000, et cetera). This problem I will indicate by RH.

This charming book is an attempt to understand some modern approaches to the RH. In the period starting with the PhD thesis (1921) by Emil Artin, an analogous problem was formulated: the Riemann hypothesis for function fields in characteristic p. In order to avoid misunderstanding, I will indicate this problem (in various variants) by pRH (often referred to by the terminology 'the RH in the function field case'). Several generalizations of the RH can be given, and the pRH is a special case of one of these generalizations. Did we make any progress, once we know how to prove the pRH, to an understanding, or perhaps even an approach to a proof of the RH?

The pRH has a rich history, with formulations and proofs and new conjectures by Emil Artin, Hasse, Weil, Serre, Grothendieck, many others, and eventually Deligne, in the period 1920–1974, with several Fields medals, astonishing developments and deep results. This work was the origin of modern algebraic geometry. And still there seems to be no end yet to questions, conjectures and new developments in this direction.

Does a solution to pRH give any clue for a possible proof of the RH? In a direct sense it does not: the pRH concentrates on one characteristic, whereas the RH involves a (convergent) sum taken over all primes. However, it might be that a *method* in proving pRH could give a clue where to look for a proof of the RH. This is the already classical analogy between number fields and function fields in one variable over a finite field, an analogy that often suggests results, but usually has no direct implications either way.

Modern attempts to prove RH by Deninger, Haran and Connes, are courageous quests. In this book the author tries to convince the reader that methods for proving the pRH are underlying these modern approaches. However, this disguise is not so easy to decipher. Yet that is what the book tries to do. We see an explanation of results in Tate's PhD thesis, approaches to pRH (such as proofs by W.M. Schmidt, Stepanov, Bombieri for a curve over a finite field) and possible translations to the number field case.

Exercises and problems challenge the reader to further research in this area. At various stages the author indicates obstacles for a translation from the function field case to the number field case. Such as, pages 3–4: $S = Spec(\mathbb{Z})$ should be a curve, and then "what is the dimension of $S \times S$?" Or (page 101): "How can $\mathbb{Z}/p\mathbb{Z}$ be viewed as an extension of degree log p of $\mathbb{F}_1$?" Working through this material can be rewarding. We will see whether the contagious optimism of the author for this point of view will materialize in the future. This intriguing material is recommended, e.g., for an advanced student seminar.

"The author believes that Connes' approach provides the first truly convincing heuristic argument for the Riemann hypothesis. He also believes that working out this argument for the function field case is the key to getting it to work for the integers." *Frans Oort*

Ivan Arzhantsev, Ulrich Derenthal, Jürgen Hausen, Antonio Laface

Cox Rings

Cambridge Studies in Adv. Mathematics 144
Cambridge University Press, 2014
472 p., prijs £50.00
ISBN 9781107024625

Historisch gezien begint de theorie van Cox-ringen met het artikel 'The homogeneous coordinate ring of a toric variety' van David Cox in de *Journal of Algebraic Geometry* van 1995. Het gaat dus om een relatief nieuw concept. Natuurlijk dateren de wortels van het begrip veel verder terug: de auteurs van dit boek noemen een tweetal artikelen van de hand van Colliot-Thélène en Sansuc die al uit 1976/77 stammen.

Om de term *Cox-ring* uit te leggen moet ik iets zeggen over torische variëteiten. Hier slaat 'torisch' op een algebraïsche torus en moet men dus aan de niet-nulelementen van een lichaam denken en niet aan een topologische torus. Voor het gemak beperk ik me nu tot de complexe getallen, $\mathbb{C}$. Een torische variëteit is een algebraïsche variëteit X van dimensie d waar de torus $T = (\mathbb{C}^\times)^d$ op werkt. Men kan deze combinatorisch beschrijven via een d-dimensionaal polytoop. Meetkundige eigenschappen vind je daarin terug. Het polytoop vertelt hoe je via knippen en plakken X terug kan vinden door affiene stukken langs tori te plakken. De Cox-aanpak is daarentegen globaal en start met een homogene ring $R(X)$ die X beschrijft als een quotiënt, zeg $C(X)/T(X)$, waarbij $C(X)$ gelijk is aan $\mathbb{C}^{d+r}$ minus 'exceptionele' variëteit en $T(X)$ een torus is van dimensie r. Denk hierbij aan $X = \mathbb{P}^d = \mathbb{C}^{d+1} \setminus \{0\}/\mathbb{C}^*$. Hier is $r = 1$, de exceptionele variëteit is de oorsprong en $R(X) = \mathbb{C}[X_0, \ldots, X_d]$, waarbij alle variabelen dezelfde graad, namelijk 1 hebben.

In het boek worden Cox-ringen in hoofdstuk 1 abstract ingevoerd en worden de basale meetkundige en algebraïsche eigenschappen afgeleid. In het tweede hoofdstuk wordt het boven aangestipte verband met torische variëteiten uit de doeken gedaan. Boven werd al vermeld dat torische meetkunde volledig via combinatorische methodes bedreven kan worden. Hoofdstuk 3 doet daarvan verslag. In hoofdstuk 4 worden bepaalde klassen van meetkundige objecten besproken die zich goed lenen als testgrond voor de besproken technieken. Met name dien ik hier de Mori-droomruimtes, de sferische en prachtige ('wonderful') variëteiten te noemen. Deze klassen van variëteiten waar ik hier de definities niet van zal geven, spelen een belangrijke rol in andere takken van de meetkunde: Mori-droomruimtes komen uit de classificatietheorie van hogerdimensionale variëteiten, prachtige en sferische ruimtes komen uit de theorie van algebraïsche groepen en zijn ooit door Dominique Luna ingevoerd.

Hoofdstuk 5 gaat in op speciale oppervlakken die Mori-droomruimtes kunnen zijn, zoals bepaalde K3-oppervlakken, Enriques-oppervlakken en Del Pezzo-oppervlakken. Ook hiervan laat ik definities achterwege. Wat de lezer wel moet beseffen, is dat niet al dit soort oppervlakken Mori-droomruimtes kunnen zijn en dat het een uitdaging is om te bepalen welke dat wel zijn. Daar zijn de auteurs wonderwel in geslaagd.

Het laatste hoofdstuk is van aritmetische aard en gaat terug naar de eerder genoemde artikelen van Jean-Luc Colliot-Thélène en Jean-Jaques Sansuc. Het gaat hier om het tellen van rationale punten op speciale variëteiten. Een beroemd vermoeden van Yuri Manin zoals verfijnd door Emmanuel Peyre is hier de leidraad. Dit vermoeden zegt heel precies hoe het aantal over $\mathbb{Q}$ gedefinieerde punten op een Fano-oppervlak met logaritmische hoogte $\leq B$ groeit met B.

Het besproken boek verdient alle lof. Het is helder geschreven, de onderwerpen zijn alle actueel en sluiten goed aan bij heel uiteenlopende takken van de meetkunde. De auteurs proberen ook zo veel mogelijk zelfvoorzienend te zijn zodat je niet eerst vele artikelen en boeken hoeft te raadplegen voordat je de tekst begrijpt. Handig voor jonge onderzoekers waar het boek voor geschreven lijkt. Om die ter wille te zijn, staan er bijvoorbeeld ook heel veel opgaven in van uiteenlopende moeilijkheidsgraad.

Uit het bovenstaande blijkt het wel: dit boek is een echte aanrader. In het bijzonder voor algebraïsch meetkundigen die willen zien hoe de abstracte theorie van Cox-ringen met vrucht toegepast kan worden op allerlei klassiek bekende expliciete variëteiten. Maar ook de combinatoricus en de wiskundige met een meer algebraïsche inslag kan in dit boek ideeën opdoen en zien wat de meetkunde te brengen heeft.

Chris Peters

Vincenzo De Risi (ed.)
Mathematizing Space
The Objects of Geometry from Antiquity to the Early Modern Age
Trends in the History of Science
Birkhäuser/Springer, 2015
ix + 318 p., prijs € 128,39
ISBN 9783319121017

In de vroegmoderne tijd heeft zich een radicale omwenteling voltrokken in het meetkundig begrip van ruimte. De klassieke meetkunde was een wetenschap van voorwerpen en hun onderlinge verhoudingen die de ruimte opspannen. In de moderne meetkunde is de ruimte een eigenstandige, abstracte structuur die plaats en identiteit aan objecten verleent. De moderne meetkunde onderzoekt de ruimte als zodanig en dat is een cruciaal verschil met de meetkunde tot in de achttiende eeuw. Vincenzo De Risi zegt het fraai (p. 5): "het schoolbord waarop de figuren werden getekend ... werd zelf niet gethematiseerd als onderwerp van meetkundig onderzoek." De Risi is groepsleider op het Max-Planck Institut für Wissenschaftsgeschichte in Berlijn en instigator van het project waaruit het boek voorkomt. Hij combineert geschiedenis en wijsbegeerte van de wiskunde, wat ook tot uitdrukking komt in de opzet van het boek waarin de wisselwerking tussen wiskundige praktijk en wijsgerige reflectie centraal staat. Concepten van de meetkundige ruimte kunnen niet los gezien worden van de manieren waarop wiskundigen objecten en structuren gebruiken en uitwerken.

Het boek opent met een prikkelende en verhelderende inleiding in de vorm van een programmatische schets van een geschiedenis van het ruimtebegrip. De Risi onderscheidt vier fases in de transformatie van het meetkundige object: een meetkunde zonder ruimte bij de oude Grieken; een Neo-Platoonse meetkunde in materiële uitgebreidheid die opkwam in de late Oudheid en doorwerkte tot in de zeventiende eeuw; een Renaissance meetkunde in de ruimte die herkenbaar is in Newtons opvattingen over absolute plaats, tijd en ruimte; en de meetkunde van de ruimte waarvan Leibniz de eerste conceptuele schreden zette en die culmineerde in de transcendentale theorie van Kant. Daarmee was de grondslag voor het moderne ruimtebegrip gelegd. De verdere uitwerking van nieuwe meetkundes en ruimtelijke structuren vanaf Lambert, Monge, enzovoort, valt buiten het bestek van het boek.

Dit boek bevat een aantal uitstekende bijdragen op het snijvlak van geschiedenis van de meetkunde en wijsbegeerte van de ruimte. De nadruk ligt daarbij op de ideeën van filosofen; vernieuwende denkers zoals Desargues en Lambert worden hooguit genoemd. De bijdragen zijn tamelijk losstaand: er is weinig onderlinge discussie en ook op plekken waar dat voor de hand ligt wordt niet naar elkaar gerefereerd. Daarbij is de focus op de vraagstelling zoals De Risi die formuleert niet altijd even scherp; een aantal artikelen gaat over klassieke thema's als oneindigheid en continuïteit en maar zijdelings over plaats en ruimte. Afgezien daarvan is er voldoende lezenswaardigs. Henry Mendell opent met een verfrissend nuchter stuk waarin hij laat zien dat het zinloos is een filosofie uit wiskundige teksten te reconstrueren. Door systematisch te kijken naar het gebruik van begrippen als positie en lengte door Griekse wiskundigen komt hij tot de conclusie dat er geen specifiek idee van plaats of ruimte te vinden is. "The constructional nature of Greek mathematics ... tells us no more about Greek mathematicians' conceptions of space than the activity of a pâtissière producing an elaborately layered cake would tell us about her conception of space." (p. 17)

In een kort essay legt Jeremy Gray uit hoe Euclides lijn en vlak definieert in termen van rechtheid en vlakheid, waardoor een onhelder begrip van ruimte en ruimtelijkheid ontstaat. Van een heel andere orde is het artikel van Alexander Jones over Ptolemaios' theorie van hemelse sferen, dat eerder over dimensies en modelleren gaat dan over ruimte en meetkunde, maar daarmee niet minder leerzaam is. Op een vergelijkbare manier werpt Gary Hatfield nieuw licht op de waarnemingsopvatting van Descartes, waarin hij uitlegt dat de waargenomen positie van voorwerpen niet voortkomt uit een rationele beoordeling maar uit de psychofysiologische ervaring door het waarnemingsorgaan: de meetkunde is zodoende geworteld in de zintuiglijke ervaring. De worstelingen met de mechanische filosofie en Descartes' identificatie van materie en uitgebreidheid leveren fascinerende resultaten op: het radicale materialisme van Hobbes waarin beweging het primaat krijgt, of het radicale empirisme van Hume waarin continuïteit op het waarneembare wordt teruggebracht. Dit mondt uit in een fraai stuk van Daniel Garber over Leibniz' innovatieve ideeën over kracht als grondslag voor materialiteit en vervolgens een begrip van abstracte ruimte in wat hij analysis situs noemde.

Over dit laatste heeft De Risi een doorwrochte studie geschreven, *Geometry and Monadology. Leibniz's Analysis Situs and Philosophy of Space* (2007), waarbij Garber de aantekening maakt dat gewaakt moet worden voor het terugprojecteren van Kantiaanse opvattingen op Leibniz' leer. Dat laat onverlet dat Leibniz een breuk in het meetkundig denken bewerkstelligde die de deur naar een nieuwe ruimte opende.

Fokko Jan Dijksterhuis

Ian Hacking
Why is there Philosophy of Mathematics at all?
Cambridge University Press, 2014
290 p., prijs £17.99
ISBN 9781107658158 (Paperback)

Ian Hacking, emeritus hoogleraar aan het Collège de France en emeritus universiteitshoogleraar aan de universiteit van Toronto, heeft met het voorliggende boek een opmerkelijk werk geschreven. Wijsbegeerte van de wiskunde, met direct daaraan gekoppeld de vraag waartoe die wijsbegeerte dan wel dient, zeer creatief verwoord in één titel waarin de beide onderwerpen nauw aan elkaar verbonden worden. Wijsbegeerte van de wiskunde is op zich een mooi wetenschapsgebied, maar in Nederland hoor je toch telkens de vraag naar het nut daarvan. Word je door te filosoferen over/in de wiskunde een beter wiskundige? Of doet het er allemaal niet toe? Maar waar hebben we het dan over? Voor veel collega-wiskundigen is de wijsbegeerte van de wiskunde een terra incognita. Dat was in de eerste helft van de twintigste eeuw wel anders. Denk bijvoorbeeld aan Mannoury, Beth, Vollenhoven (een theoloog (!) die gepromoveerd was in de wijsbegeerte van de wiskunde) en onze Luitzen Brouwer. Wanneer we tegenwoordig in Nederland over wijsbegeerte van de wiskunde spreken, dan gaat de discussie al heel snel in de richting van de logica. Het boek van Hacking daarentegen is een echt filosofisch werk over de wiskunde. Het neemt de lezer mee naar een moderne behandeling van de oude fundamentele vragen, zoals die met betrekking tot bewijzen, en de vragen over de raadselachtige samenhang van de wiskunde met de realiteit. In een aanstekelijke stijl stelt de auteur je voor mooie vragen, zoals "Wat maakt wiskunde tot wiskunde, oftewel wat is wiskunde eigenlijk?" Doe maar eens eenzelfde experiment met je eigen studenten/leerlingen, als waarvan Hacking verslag doet: vraag eens aan je studenten wat zij vinden dat wiskunde nu precies is. Een Socratisch gesprek leidt bij hem tot de niet mis te verstane serieuze conclusie dat wiskunde datgene is wat wiskundigen bedrijven. En dan dat onderscheid waarvan we de mond vol hadden/hebben, het verschil tussen zuivere en toegepaste wiskunde. Bestaat een dergelijk onderscheid of is dat grote flauwekul?

De grote delen waarin het boek is ingedeeld zijn getiteld 'A cartesian introduction (Application; Proof)', 'What makes mathematics mathematics?', 'Why is there philosophy of mathematics? (An answer from the ancients: proof and exploration; An answer from the Enlightenment: application)', 'Proofs (Little contingencies; Proof)', 'Applications (The emergence of a distinction; A very wobbly distinction)', 'In Plato's name (Alain Connes, Platonist; Timothy Gowers, anti-Platonist)', 'Counter-platonisms (Totalizing platonism as opposed to intuitionism; Today's platonism/nominalism)'.

Dit overzicht doet op zich al vermoeden dat de genoemde onderwerpen in extenso en met een behoorlijke diepgang behandeld zullen worden. De schrijver maakt dit vermoeden meer dan waar. Daarbij legt Hacking zich niet vast op een bij voorbaat gekozen wijsgerige visie, maar hij benadert de problematiek met een open mind, hij beschouwt de diverse onderwerpen vanuit verschillende wijsgerige standpunten. Hij maakt dit expliciet bij de behandeling van de gekozen onderwerpen, waarbij hij zijn persoonlijke mening en standpunt wel vermeldt, maar niet opdringt. Uiteraard vergeet de auteur niet om de onderwerpen ook vanuit historisch perspectief te bestuderen. De filosofische behandeling is zowel doordrenkt van het verleden, als op een minder gebruikelijke wijze ook in overeenstemming met de elkaar bestrijdende filosofische ideeën van de huidige wiskunde. Hij laat zien dat bewijzen en andere vormen van wiskundige onderzoekingen nog steeds levend zijn en tevens passen binnen onze nieuwe technologieën. Hij onderscheidt verschillende soorten van toepassingen van de wiskunde en hij toont aan dat elk daarvan kan leiden tot een verschillend filosofisch onderzoeksgebied. Het boek biedt door dit alles een opmerkelijk geheel van wijsgerig denken over bewijzen, toepassingen en andere wiskundige activiteiten. De diversiteit en de openheid van behandeling van de onderwerpen maken het boek extra aantrekkelijk om te bestuderen.

Een uitvoerig referentie-overzicht en een nauwkeurige index besluiten dit boek. Terugkomend op de vraag uit het begin: word je door de bestudering van de wijsbegeerte van de wiskunde en van dit boek een beter wiskundige? Dat hangt uiteraard samen met wat onder een 'betere wiskundige' wordt verstaan. Van en met de wijsbegeerte leer je geen nieuwe technieken, methodieken en ook geen nieuwe inzichten in de wiskunde zelf. Maar het beoefenen van de wijsbegeerte van de wiskunde verdiept wél het inzicht in wat wiskunde is, hoe wiskunde werkt en wat de plaats is van de wiskunde in het geheel van het wetenschappelijk denken. Uw recensent is daarom van mening dat het antwoord op de gestelde vraag voluit met een 'ja' beantwoord moet worden. Wijsbegeerte van de wiskunde is beslist geen zweverig vak, het is net zo precies en abstract als de wiskunde zelf. Het is te hopen dat de typisch Nederlandse koudwatervrees voor (vermeende) vage metafysische zaken, de bestudering van de wijsbegeerte van de wiskunde niet in de weg zal staan. Het boek van Hacking is een schoolvoorbeeld van een boeiende betoogtrant, van exact en abstract denken over zaken die iedere wiskundige direct aangaan.

Het boek is geschreven in een heldere en toegankelijke stijl. De niet wijsgerig geschoolde lezer zal zich enige moeite moeten getroosten om zich in de stof in te werken, maar haar/zijn inspanningen zullen dan ruimschoots beloond worden. De zorgvuldigheid waarmee het werk is geschreven straalt van iedere bladzijde af en datzelfde geldt ook voor het plezier dat de auteur zelf tijdens het schrijven aan het werk kennelijk heeft beleefd. Ik wens het boek een goede toekomst: dat het ook binnen de wiskundefaculteiten van onze universiteiten en hogescholen naarstig bestudeerd zal worden. *Wim Kleijne*

David Reimer
Count Like an Egyptian
A Hands-on Introduction to Ancient Mathematics
Princeton University Press, 2014
xiii + 237 p., prijs $ 29.95
ISBN 9780691160122

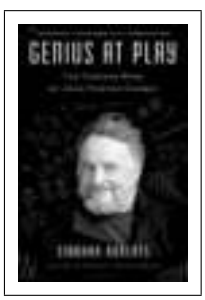
Een blik in dit boek leert meteen dat het onderwerp van studie de auteur zeer na aan het hart ligt. De manier van beschrijven die hij hanteert is zonder meer briljant te noemen. Het is een absoluut meeslepende vertelling geworden over geschiedenis en cultuur van Egypte uit een lang vervlogen tijd, dit alles uiteraard der zaak gedompeld in het onderwerp van onderzoek, het rekenen der Egyptenaren in theorie en in toepassingen op de praktijk. Alles wordt aan de hand van honderden gemakkelijk te volgen voorbeelden en (reken)opgaven verhaald en uitgelegd. Formules, figuren en berekeningen staan zodanig afgedrukt

dat dit alles een lust voor het oog is, meestal in kleur en, zo schijnt het toe, als in- of uitspringende reliëfs op het papier. Prachtig suggestief, net zoals de Egyptenaren dat destijds met hun hiëroglyfen deden. De afgedrukte tekst op elke bladzijde in het boek meet in de lengte doorgaans iets meer dan 20 cm (verdeeld over twee kolommen), terwijl de breedte der afgedrukte tekst iets meer dan 18 cm bedraagt. Ik vermeld dit detail omdat deze layout bewust zo gekozen is en buitengewoon prettig overkomt. Als ik de inhoud van het boek vergelijk of zou willen vergelijken met die van het reeds enige tijd geleden verschenen boek van Richard J. Gillings, *Mathematics in the time of the Pharaohs* (MIT Press, 1972), dan kan ik de inhoud van het laatste boek, ofschoon correct, slechts als gortdroog betitelen.

Voorbeelden en uitleg van het Egyptisch rekenen in Reimers boek zijn deels ontleend, toegelicht en uitgewerkt aan de hand van de volgende bronnen die, zij het wat verscholen, in het boek worden genoemd, namelijk: 1) De Onomasticon van Amenepet, 2) Ahmose's precieze rekenmethodes genaamd de invoering in de kennis van alle be staande dingen en verscholen geheimen (het is trouwens een verhandeling over breukenleer), 3) de Egyptische Mathematische Lederen Rol, 4) Archimedes' Methode, 5) de Objecten in de Duat, 6) het Boek der Pylonen, 7) het Gilgamesj Epos, 8) het Dodenboek, 9) de Chinese I'Ching. Maar dit zijn zomaar wat kapstokken waaraan sommige delen van de tekst zijn opgehangen. Feitelijk staat er in het boek veel meer om van te genieten.

Samenvattend: een goed en bijzonder boek; voor nieuwelingen in het vakgebied absoluut een openbaring, voor bekenden met het onderwerp een frisse en onverwachte manier om met het onderwerp van studie om te gaan. De schrijver en de uitgever hebben de wiskundige gemeenschap, en niet alleen deze, een dienst bewezen met het realiseren van dit boek. *Robert van der Waall*

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.



Siobhan Roberts
Genius at Play
Bloomsbury; 2015
ISBN 9781620405932
www.bloomsbury.com/9781620405932

Frans Keune
Galoistheorie
Beginselen van de theorie der veeltermvergelijkingen in één onbekende
Epsilon Uitgaven, deel 79, 2015
ISBN 9789050411509
www.epsilon-uitgaven.nl/E79.php

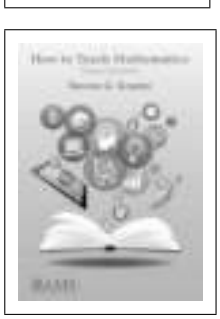
Birgit van Dalen, Quintijn Puite
Wanneer is Cheryl jarig?
+ 99 andere wiskunderaadsels
Bertram + de Leeuw uitgevers, 2015
ISBN 9789461561961
www.bertramendeleeuw.nl/boek/wanneer-cheryl-jarig

Heinz Hanßmann
Functionaalanalyse
Epsilon Uitgaven, deel 81, 2015
ISBN 9789050411523
www.epsilon-uitgaven.nl/E81.php

Ton Langendorff
Denken wiskundigen wel zo exact?
Observaties en gesprekken
Athenaeum, 2015
ISBN: 9789025307677
www.singeluitgeverijen.nl/denken-wiskundigen-wel-zo-exact/

Terence Tao
Expansion in Finite Simple Groups of Lie Type
American Mathematical Society, 2015
ISBN 9781470421960
www.ams.org/bookstore-getitem/item=gsm-164

Lewis Carroll
Illustrated by Salvador Dalí, edited by Mark Burnstein
Alice's Adventures in Wonderland
150th anniversary edition
Princeton University Press, 2015
ISBN 9780691170022
press.princeton.edu/titles/10538.html



Steven G. Krantz
How to Teach Mathematics (3rd edition)
American Mathematical Society, 1999
ISBN 9781470425524
www.ams.org/bookstore-getitem/item=HTM-2

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. At times there will be a Star Problem, to which the proposer does not know any solution. For the first correct solution sent in within one year there is a prize of €100.
When proposing a problem, please either include a complete solution or indicate that it is intended as a Star Problem.
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 2016.

Problem A (folklore)

Let n be a positive integer. Given a $1 \times n$-chessboard made out of paper, one is allowed to fold it along grid lines, and in such a way that the end result is a flat rectangle, say $1 \times m$. For example, the following figure shows side views of valid ways of folding a 1×7-chessboard (gray lines depict white squares).

Let a_i for $i = 1, 2, \ldots, m$ be the number of black squares under the i-th square of the resulting rectangle, and consider the tuple $(a_1, a_2, \ldots, a_m)$. So in our examples, the respective corresponding tuples are $(1, 1, 1)$ and $(2, 1, 1)$.
Show that for any positive integer m the m-tuple $(a_1, a_2, \ldots, a_m)$ of non-negative integers can be obtained via the above process if and only if for all $i, j \in \{1, 2, \ldots, m\}$ such that $i + j$ is odd, we have $(a_i, a_j) \neq (0, 0)$.

Problem B (proposed by Jinbi Jin)

Let A be a commutative ring with unit, and let I be an ideal of A with $I \neq 0$ and $I^2 = 0$. Let B be the ring of which the elements are triples (a_1, a_2, a_3) where $a_1, a_2, a_3 \in A$ are such that $a_1 + I = a_2 + I = a_3 + I$, with coordinate-wise addition and multiplication. Show that there exist at least four distinct ring homomorphisms $B \to A$.

Problem C (proposed by Hendrik Lenstra)

Does there exist a non-trivial abelian group A that is isomorphic to its automorphism group?

Edition 2015-2 We received solutions from Raymond van Bommel, Alex Heinis, Jos van Kan, Thijmen Krebs, Julian Lyczak, Tejaswi Navilarekallu, Traian Viteam and Robert van der Waall.

Problem 2015-2/A (proposed by Gabriele Dalla Torre)
Show that there are infinitely many primes that divide at least one integer of the form

$$2^{n^3+1} - 3^{n^2+1} + 5^{n+1}.$$

Solution We received solutions from Raymond van Bommel, Alex Heinis, Jos van Kan, Thijmen Krebs, Tejaswi Navilarekallu, Traian Viteam and Robert van der Waall. The following is based on that of Tejaswi Navilarekallu, who also receives the book token.
Suppose for a contradiction that there are only finitely many primes that divide at least one

Redactie:
Gabriele Dalla Torre
Christophe Debry
Jinbi Jin
Marco Streng
Wouter Zomervrucht

Problemenrubriek NAW
Mathematisch Instituut
Universiteit Leiden
Postbus 9512
2300 RA Leiden
problems@nieuwarchief.nl
www.nieuwarchief.nl/problems

Oplossingen

Solutions

integer of the form $f(n) = 2^{n^3+1} - 3^{n^2+1} + 5^{n+1}$. Let $P = \{p_1, p_2, \ldots, p_s\}$ be the odd primes amongst them, and let $n = 4(p_1 - 1)(p_2 - 1)\cdots(p_s - 1)$. Then $f(n) \equiv 2 - 3 + 5 \equiv 4 \not\equiv 0 \mod p$ for all $p \in P$, and $f(n) \equiv 0 - 3 + 5 \equiv 2 \mod 8$ (using that n is even and that odd squares are 1 modulo 8). Moreover, as $n > 1$, we have $f(n) = 2^{n^3+1} - 3^{n^2+1} + 5^{n+1} > 2^{(n^2+1)(n-1)} - 3^{n^2+1} + 5^{n+1} > 5^{n+1} > 8$. So by assumption, 2 is the only prime dividing $f(n)$; i.e. $f(n)$ is a power of 2 that is greater than 8 and that is 2 modulo 8, this is a contradiction.

Problem 2015-2/B (proposed by Jinbi Jin)

Let n be a positive integer. Two players, Ann and Bill, play the following game. First, Ann distributes a number of balls over boxes numbered from 1 up to n. Then Bill chooses one of the boxes, and adds a ball to it. Finally, Ann attempts to empty all boxes, using only the following moves.

- Taking one ball from three consecutive boxes.
- Taking three balls from one box.

Ann wins if she succeeds in doing so, otherwise Bill wins.

1. Determine (as a function in n) the maximum number of losing moves Bill can have. What is the minimum number of balls Ann needs to attain this number?
2. Do the same as in point 1, if Ann in addition is allowed *only once* to remove two balls from one box.

Solution We received solutions from Raymond van Bommel and Julian Lyczak, Alex Heinis and Thijmen Krebs. The following solution is loosely based on that of Raymond van Bommel and Julian Lyczak, who also receive the book token.

We will view a distribution of balls over n boxes as a map from $\{1, 2, \ldots, n\}$ to $\mathbb{Z}_{\geq 0}$; and will often write them as words of length n with alphabet $\mathbb{Z}_{\geq 0}$. We denote the distribution with a single ball in box m by $[m]$ (so that Bill adding a ball in box m to some distribution corresponds to adding $[m]$ to the distribution).

Given a distribution D of balls over n boxes, we denote by

- $\#D = \sum_{i=1}^{n} D(i)$, the number of balls in D;
- $\#_{<d}D = \sum_{i=1}^{d-1} D(i)$, the number of balls in D "to the left of" box d;
- $\#_{>d}D = \sum_{i=d+1}^{n} D(i)$, the number of balls in D "to the right of" box d;
- $a_s(D) = \sum_{i \equiv s \mod 3} D(i)$, the number of balls in D in boxes that are congruent to s modulo 3.

Moreover, let

- A_d for $1 \leq d \leq n-2$ denote the distribution $0^{d-1}1^3 0^{n-d-2}$ (a *move of type A*);
- B_d for $1 \leq d \leq n$ denote the distribution $0^{d-1}30^{n-d}$ (a *move of type B*);
- C_d for $1 \leq d \leq n$ denote the distribution $0^{d-1}20^{n-d}$ (a *move of type C*);

these correspond to the three moves Ann can perform.

Ad 1. Let D be a distribution of balls over n boxes. Here a losing move m for Bill is one such that $D + [m]$ can be written as the sum of moves of types A and B.

We show that the maximum number of losing moves Bill can have is $\lceil \frac{n}{3} \rceil$, and that the minimum number of balls Ann needs to achieve this is 2 balls if $n \leq 3$, and $3\lceil \frac{n}{3} \rceil - 4$ balls otherwise.

We first give examples that attain these bounds. If $n \leq 3$, then $D = 20^{n-1}$ works, as Bill has one losing move $[1]$ ($D + [1] = B_1$), and Ann uses 2 balls. If $n > 3$, then writing $\alpha = 3\lceil \frac{n}{3} \rceil - 4$, we see that $D = 01^{\alpha}0^{n-\alpha-1}$ works, as for all $i \equiv 1 \mod 3$, we have

$$D + [i] = 01^{i-1}0^{n-i} + 0^{i-1}1^{\alpha-i+2}0^{n-\alpha-1}.$$

As $\alpha \equiv -1 \mod 3$, both terms on the right can be written as a sum of moves of type A, so this shows that Bill has $\lceil \frac{n}{3} \rceil$ losing moves.

We show that $\lceil \frac{n}{3} \rceil$ is the maximum number of losing moves Bill can have. Let D be any distribution of balls over n boxes. Then note that $a_s(A_d) - a_t(A_d)$ and $a_s(B_d) - a_t(B_d)$ are both divisible by 3 for all s, t. Hence if putting a ball in box m is a losing move for Bill, then $a_s(D) - a_m(D) - 1 = a_s(D + [m]) - a_m(D + [m]) \equiv 0 \mod 3$ for all $s \not\equiv d \mod 3$. So putting a ball in any box i with $i \not\equiv m \mod 3$ is a winning move for Bill, as $a_i(D + [i]) - a_m(D + [i]) = a_i(D) + 1 - a_m(D) \equiv 2 \mod 3$, so D cannot be written as the sum of moves of types A and

Oplossingen

Solutions

B. Therefore Bill can have at most $\lceil \frac{n}{3} \rceil$ losing moves, all of which must have the same value modulo 3.
Next, we show that 2 is the minimum number of balls Ann needs if $n \leq 3$, and that $3\lceil \frac{n}{3} \rceil - 4$ is the minimum number of balls she needs otherwise. Let D be any distribution of balls over n boxes. Note that $\#A_d = 3$ and $\#B_d = 3$, so for D to have losing moves, one needs $\#(D+[m]) \equiv 0 \mod 3$ for some box m. Hence $\#D = 2 \mod 3$. This proves the lower bound for $n \leq 3$.
Now suppose that $n \geq 4$, and let D be a distribution with $\lceil \frac{n}{3} \rceil$ losing moves. By the proof that $\lceil \frac{n}{3} \rceil$ is the maximum number of losing moves, these must all be the same modulo 3, so there must be at least $3\lceil \frac{n}{3} \rceil - 4$ boxes (strictly) between them. We show that all of these must contain at least one ball.
Suppose for a contradiction that box d is an empty box lying between the leftmost losing move and the rightmost losing move. Then Ann has no move that removes balls from both sides of box d at the same time. Hence for any losing move m of Bill, we must have both $\#_{<d}(D+[m])$ and $\#_{>d}(D+[m])$ divisible by 3. But since there is a losing move to the left of box d (which contributes 1 to the former), as well as one to its right (which contributes 1 to the latter), this is a contradiction.
This shows that at least $3\lceil \frac{n}{3} \rceil - 4$ balls are needed (if $n \geq 3$) in order to give Bill $\lceil \frac{n}{3} \rceil$ losing moves.

Ad 2. Let D be a distribution of balls over n boxes. Here a losing move m for Bill is one such that $D+[m]$ can be written as the sum of moves of types A and B, plus at most one move of type C.
We show that the maximum number of losing moves Bill can have is n, and that the minimum number of balls Ann needs to achieve this is 1 ball if $n = 1$, and $3\lfloor \frac{n}{3} \rfloor + 4$ balls otherwise.
We first give examples that attain these bounds.
For $n = 1$, we see that $D = 1$ works, since $D+[1] = C_1$. If $n = 2$, then $D = 22$ works, since $D+[1] = B_1 + C_2$ and $D+[2] = C_1 + B_2$. If $n \geq 3$ and $n \equiv 2 \mod 3$, then $D = 031^{n-4}30$ works;
- if $i \equiv 1 \mod 3$, then $D+[i] = C_2 + B_{n-1} + 01^{i-1}0^{n-i} + 0^{i-1}1^{n-i-1}0^2$, and since $n \equiv 2$ the last two terms are sums of A_d's, by mirror symmetry, all $i \equiv 2 \mod 3$ are losing for Bill as well;
- if $i \equiv 0 \mod 3$, then $D+[i] = B_2 + B_{n-1} + C_i + 0^2 1^{i-3} 0^{n-i+2} + 0^i 1^{n-i-2} 0^2$, and since $n \equiv 2$ the last two terms are sums of A_d's.

Moreover, since the first and last boxes of D are empty, by removing one or both of those we get examples for all $n \geq 3$.
We show that any distribution of balls over n boxes that has n losing moves must have at least 1 ball if $n = 1$, and $3\lfloor \frac{n}{3} \rfloor + 4$ balls otherwise. We first prove the following lemma.

Lemma 1. *Let D be any distribution of balls over $n \geq 2$ boxes that has n losing moves. Then $\#D \equiv 1 \mod 3$, and some move of type C occurs in $D+[m]$ for all $m \in \{1, 2, \ldots, n\}$.*

Proof. If no move of type C occurs, then by point 1, Bill can have at most $\lceil \frac{n}{3} \rceil < n$ losing moves. Moreover, every move of types A and B contribute 3 to $\#(D+[m])$ for all $m \in \{1, 2, \ldots, n\}$, and a move of type C contributes 2 to $\#(D+[m])$ for all $m \in \{1, 2, \ldots, n\}$. Hence $\#D \equiv 1 \mod 3$. □

The case $n = 1$ is trivial. For $n = 2$, we note that by Lemma 1, it suffices to show that $\#D > 1$. This follows since $[1]+[2]$ cannot be written as sum of moves of types A, B, or C.
Let D be a distribution of balls over $n \geq 3$ boxes that has n losing moves. Note that by Lemma 1, it suffices to show that $\#D \geq n+2$, as $n+2 \geq 3\lfloor \frac{n}{3} \rfloor + 2$.
We first show that all boxes between 2 and $n-1$ (inclusive) contain at least one ball. Suppose for a contradiction that for some $2 \leq d \leq n-1$, box d is empty. Note that Ann still has no moves that simultaneously takes away balls from both sides of box d. Therefore $\#_{<d}(D+[1]), \#_{<d}(D+[n]), \#_{>d}(D+[1]), \#_{>d}(D+[n])$ must all be either 0 or 2 modulo 3, depending on where the move of type C comes from. Hence both $\#_{<d}D$ and $\#_{>d}D$ must both be 2 modulo 3.
For $n = 3$, note that A_1 must occur in $D+[2]$, since $B_2(2), C_2(2) > 1$ and no other moves contribute to box 2. But this implies that $\#_{<2}(D - A_1), \#_{>2}(D - A_1) \equiv 1 \mod 3$, so by the argument above, we have a contradiction.
For $n \geq 4$, the above implies that if a move $m < d$ is losing, then the pair must come from the boxes to the right of d, and vice versa. So both the part of D to the left of d and that to the right of d are instances of the situation in point 1. At least one of these instances involve at least

Oplossingen

Solutions

two boxes, but this implies that Bill has winning moves, which is a contradiction. Therefore all boxes between 2 and $n-1$ (inclusive) contain a ball, so $\#D \geq n-2$.
We now take a closer look at $a_1(D), a_2(D), a_3(D)$.

Lemma 2. *Let $m \in \{1,2,\ldots,n\}$. If $i \neq j \in \{1,2,3\}$ such that $a_i(D+[m]) \equiv a_j(D+[m]) \mod 3$, then a move of the form C_z, with z congruent to the unique element of $\{1,2,3\}-\{i,j\}$ modulo 3, occurs in $D+[m]$.*

Proof. Moves of type A and B do not contribute to the difference $a_i(D+[m]) - a_j(D+[m])$ modulo 3. The move C_z contributes to the difference $a_i(D+[m]) - a_j(D+[m])$ modulo 3 if and only if $z \not\equiv i, j \mod 3$. □

Lemma 3. *Let $m \in \{1,2,\ldots,n\}$, and let C_z be a move of type C occurring in $D+[m]$. If $i \neq j \in \{1,2,3\}$ such that $a_i(D+[m]-C_z) \geq a_j(D+[m])+3$, then some move of the from B_y with $y \equiv i$ occurs in $D+[m]-C_z$.*

Proof. Moves of type A do not contribute to the difference $a_i(D+[m]-C_z) - a_j(D+[m]-C_z)$. The move B_y contributes positively to $a_i(D+[m]-C_z) - a_j(D+[m]-C_z)$ if and only if $y \equiv i \mod 3$. □

Note that two of $a_1(D)$, $a_2(D)$, $a_3(D)$ must be congruent modulo 3, since otherwise their sum must be divisible by 3, which by the above is not the case. Moreover, as their sum is 1 modulo 3, it follows that the third one must be one more than the other two, modulo 3. Now let $s,t,u \in \{1,2,3\}$ be three distinct elements such that $a_t(D) \equiv a_u(D)$ modulo 3, and $a_t(D) \geq a_u(D)$.
If $a_s(D) \geq a_u(D)+1$, then let $m \equiv t \mod 3$. Consider $D+[m]$. Then $a_s(D+[m]) \equiv a_t(D+[m]) \mod 3$, so by Lemma 2, a move of the form C_z with $z \equiv u \mod 3$ occurs in $D+[m]$. As $a_s(D+[m]-C_z) \geq a_u(D+[m]-C_z)+3$ and $a_t(D+[m]-C_z) \geq a_u(D+[m]-C_z)+3$, we see that by Lemma 3, moves of the form B_y and $B_{y'}$ with $y \equiv s \mod 3$ and $y' \equiv t \mod 3$ occur in $D+[m]-C_z$. Since $y \not\equiv m \mod 3$, it follows that $D(y) \geq 3$, and therefore that $\#D \geq n$. We also have $y' \equiv m \mod 3$, so $D(y') \geq 2$. If $n = 3,4,5$, then this implies that $\#D \geq 5 \geq 3\lfloor \frac{n}{3} \rfloor + 2$. As $\#D \equiv 1 \mod 3$, it follows that $\#D \geq 3\lfloor \frac{n}{3} \rfloor + 4$. If $n \geq 6$, then either there exists some $i \equiv t \mod 3$ with $D(i) \geq 3$, or for all $i \equiv t \mod 3$ we have $D(i) = 2$. So either way, because $n \geq 6$, we find that now $\#D \geq n+2$, as desired.
If $a_s(D) < a_u(D)+1$, then let $m \equiv s \mod 3$. Consider $D+[m]$. As $a_t(D+[m]) \equiv a_u(D+[m]) \mod 3$, we see that by Lemma 2, a move of the form C_z with $z \equiv s \mod 3$ occurs in $D+[m]$. As $a_s(D+[m]-C_z)+3 \leq a_u(D+[m]-C_z) \leq a_t(D+[m]-C_z)$, it follows by Lemma 3 that moves of the form B_y and $B_{y'}$ with $y \equiv t \mod 3$ and $y' \equiv u \mod 3$ occur in $D+[m]-C_z$. Since $y, y' \not\equiv m \mod 3$, it follows that $D(y), D(y') \geq 3$, and therefore that $\#D \geq n+2$, as desired.

Problem 2015-2/C (proposed by Hendrik Lenstra)
Let p be a prime number and let k be a positive integer. Prove that for every integer n there exist integers w, x, y, z such that

$$n \equiv w^p + x^p + y^p + z^p \mod p^k.$$

Solution We received solutions from Raymond van Bommel, Thijmen Krebs and Tejaswi Navilarekallu. All received solutions started similarly, and the first part of the following solution is based on those. The last part of the following solution is based on that of Raymond van Bommel. The book token goes to Thijmen Krebs.
For $p = 2$ this is well known; any positive integer can be written as the sum of four squares. Therefore assume p is odd. We prove by induction on k the following stronger statement.

Lemma 4. *For every integer n there exist integers w, x, y, z, with w coprime to p, such that*

$$n \equiv w^p + x^p + y^p + z^p \mod p^k.$$

Oplossingen

Solutions

Our base case is the case $k = 2$; this case automatically implies the case $k = 1$.
First note that if $n^p \equiv n \mod p^2$, then we're done; we can take $(1, -1, 0, n)$ as solution in that case. Therefore assume that $n^p \not\equiv n \mod p^2$; by Fermat's little theorem, $n - n^p$ is divisible by p, and then our assumption implies that $m = \frac{n-n^p}{p}$ is coprime to p.
Note that $(\mathbb{Z}/p^2\mathbb{Z})^\times$ is a cyclic group of order $(p-1)p$. Therefore there exists $a \in \mathbb{Z}_{>0}$ such that $a^p \not\equiv a \mod p^2$. Let a be the smallest such positive integer and note that $a > 1$. Then $1^p + (a-1)^p + (-a)^p \equiv rp \mod p^2$ for some integer r coprime to p. So there exists an integer s coprime to p such that $m \equiv rs \mod p$, and since $s^p \equiv s \mod p$, it follows that $s^p + \big(s(a-1)\big)^p + (-sa)^p \equiv rsp \equiv mp \equiv n - n^p \mod p^2$. Hence $n \equiv s^p + \big(s(a-1)\big)^p + (-sa)^p \mod p^2 + n^p$, so $\big(s, s(a-1), -sa, n\big)$ is a solution of the desired form. This completes the base case.
As our induction hypothesis, suppose that our claim holds if $k = i \geq 2$. Then consider the case $k = i + 1$.
Suppose that n is an integer. By our induction hypothesis, there exist integers w, x, y, z, with w coprime to p, such that

$$n \equiv w^p + x^p + y^p + z^p \mod p^i.$$

Note that $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is cyclic of order $(p-1)p^{k-1}$ for all integers $k \geq 2$. Therefore the number of p-th powers in $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is equal to $(p-1)p^{k-2}$. Now note that the reduction map $(\mathbb{Z}/p^{i+1}\mathbb{Z})^\times \to (\mathbb{Z}/p^i\mathbb{Z})^\times$ that is p-to-1, and maps p-th powers to p-th powers. So comparing the number of p-th powers on each side, we see that all pre-images of p-th powers s in $(\mathbb{Z}/p^i\mathbb{Z})^\times$ are again p-th powers.
In particular, we see that $n - x^p - y^p - z^p$ defines a p-th power in $(\mathbb{Z}/p^i\mathbb{Z})^\times$, therefore also defines a p-th power in $(\mathbb{Z}/p^{i+1}\mathbb{Z})^\times$ as well, which completes the induction.