

## **Syllabus Vakantiecursus 2012**

Platform Wiskunde Nederland

Science Park 123, 1098 XG Amsterdam

Telefoon: 020-592 4006

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



# De exacte benadering

Vakantiecursus 2012

Eindhoven, 24 en 25 augustus 2012

Amsterdam, 31 augustus en 1 september 2012

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

Deze cursus wordt mede mogelijk gemaakt door een subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), en een bijdrage van 3TU.AMI, het toegepaste wiskunde instituut van de 3 Nederlandse technische universiteiten. Organisatie vindt plaats in nauwe samenwerking met het Centrum voor Wiskunde en Informatica (CWI) en de Technische Universiteit Eindhoven (TU/e).

## Docenten

Dr. Michiel Doorman, Freudenthal Instituut, Buys Ballotlaboratorium, Princetonplein 5, 3584 CC Utrecht

Dr. André Heck, Universiteit van Amsterdam, Faculty of Science, Korteweg-de Vries Institute for Mathematics, Science Park 904, 1098 XH Amsterdam

Prof. dr. Frans Keune, Institute for Mathematics, Astrophysics and Particle Physics, Afd. Algebra en Logica, Radboud Universiteit Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen

Dr. Cor Kraaikamp, Faculteit Electrotechniek, Wiskunde en Informatica, Technische Universiteit Delft, Mekelweg 4, 2628 CD Delft

Dr. Arno Kuijlaars, Departement wiskunde, Katholieke Universiteit Leuven, Celestijnenlaan 200B, bus 2400, B-3001 Leuven (Heverlee)

Dr. Boris Škorić, Technische Universiteit Eindhoven, Faculteit Wiskunde en Informatica, Postbus 513, 5600 MD Eindhoven

Dr. Jeroen Spandaw, Technische Universiteit Delft, Faculteit EWI, Mekelweg 4, 2628 CD Delft

*Helaas is de lezing door Dr. Joop Kolk door persoonlijke omstandigheden komen te vervallen.*

## Programmacommissie Vakantiecursus

Jan Wiegerinck (voorzitter)

Korteweg-de Vries Instituut voor Wiskunde, Universiteit van  
Amsterdam, Postbus 94248, 1090 GE Amsterdam

`J.J.O.O.Wiegerinck@uva.nl`

Benne de Weger

Technische Universiteit Eindhoven, Postbus 513, 5600 MB Eindhoven

`b.m.m.d.weger@tue.nl`

Frits Beukers

Mathematisch Instituut, Universiteit Utrecht, Postbus 80010,  
3508 TA Utrecht

`F.Beukers@uu.nl`

Marian Kollenveld (voorzitter NVvW)

Leeuwendaallaan 43, 2281 GK Rijswijk

`voorzitter@nvvw.nl`

Jan van Maanen

Faculteit Wiskunde en Natuurwetenschappen, Universiteit Groningen,  
Nijenborgh 4, 9747 AG Groningen

`j.a.van.maanen@rug.nl`

Kees Oosterlee

Centrum Wiskunde & Informatica, Postbus 94079, 1090 GB  
Amsterdam

`C.W.Oosterlee@cwi.nl`

Ionica Smeets

LIACS, Mathematisch Instituut, Universiteit Leiden, Postbus  
9512, 2300 RA Leiden

`ionica.smeets@gmail.com`

Jeroen Spandaw  
Technische Universiteit Delft, EWI, Mekelweg 4, 2628 CD  
Delft  
`J.G.Spandaw@tudelft.nl`

Marco Swaen  
Korteweg-de Vries Instituut voor Wiskunde, Universiteit van  
Amsterdam, Postbus 94248, 1090 GE Amsterdam  
`m.d.g.swaen@uva.nl`

## Historie van de vakantiecursus

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

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

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

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



*gestart en verheugen zich in blijvende belangstelling).”*

# Ten Geleide

Jan Wiegerinck  
Korteweg - de Vries Instituut voor Wiskunde  
Universiteit van Amsterdam  
e-mail: J.J.O.O.Wiegerinck@uva.nl

De vakantiecursus is na een jaar onderbreking weer terug op het oude schema. Twee dagen in Eindhoven, twee dagen in Amsterdam. Voor wiskundeleraren uit Zuid-Nederland en Vlaanderen maakt dit deelname weer eenvoudig. Daarmee lijkt alles weer bij het oude, maar toch verandert er veel. De taakopvatting van het CWI is gewijzigd, en het Platform Wiskunde Nederland (PWN) is opgericht door het Koninklijk Wiskundig Genootschap en de Nederlandse Vereniging voor Wiskundeleraren. Als gevolg hiervan zal de vakantiecursus vanaf dit jaar door het PWN georganiseerd worden, terwijl CWI zich uit de algemene organisatie zal terugtrekken. Als deelnemer zult u hier niet veel van merken. Coby van Vonderen zal minder bij de cursus betrokken zijn, verder zal er weinig veranderen. Graag wil ik Coby bedanken voor haar inzet voor de vakantie-cursus in de afgelopen jaren.

Het thema dit jaar is “De exacte benadering”. Wat dat kan betekenen wordt het best duidelijk door iets over het programma te vertellen. Benaderingen in de zin van “de wijze waarop men een (wiskundig) onderwerp benadert”, bijvoorbeeld in een onderwijssituatie, komen in de voordrachten van Michiel Doorman en André Heck aan de orde; Doorman bespreekt hoe je voor een niet-beta gehoor de abstracte benadering van de

wiskunde meer concreet kunt maken, Heck behandelt een alternatieve benadering voor het invoeren van goniometrische functies op het VWO met behulp van het pakket GeoGebra.

Frans Keune heeft een zeer klassiek, maar niet zo eenvoudige onderwerp. Hij bespreekt de derde-graadsvergelijking, en het nut van de gesloten oplossing van Cardano. Jeroen Spandaw's voordracht gaat over reeksen. Hij laat ons zien hoe divergente reeksen verrassende eindige uitkomsten kunnen hebben, die nog zin hebben ook! Arno Kuijlaars spreekt over het Riemannvermoeden en Random Matrices. Random Matrices zijn de laatste 15 jaar een belangrijk onderwerp van onderzoek geworden, en het lijkt er op dat deze iets met het vermoeden van Riemann over de ligging van de nulpunten van de zeta-functie te maken hebben. Ook Boris Škorić heeft een modern onderwerp. Hij bespreekt methoden uit de coderingstheorie om films of muziek te beschermen tegen illegale verspreiding, zelfs als dat door samenwerkende piraten wordt geprobeerd. Cor Kraaikamp spreekt over het benaderen van irrationale getallen door rationale getallen.

# Inhoudsopgave

Docenten	vi
Programmacommissie Vakantiecursus	vii
Historie van de vakantiecursus	ix
Ten Geleide	xi
Hoeveel is oneindig minus oneindig?	1
Het benaderen van irrationale getallen door ‘rationale’	29
Eigenwaarden van random matrices	57
Traitor Tracing	81
Wiskunde voor Dichters	93

Een GeoGebra-ondersteunde benadering van sinus en cosinus	114
Derdegraads vergelijkingen oplossen	145

# Hoeveel is oneindig minus oneindig?

Jeroen Spandaw, TU Delft

e-mail: [.G.Spandaw@tudelft.nl](mailto:.G.Spandaw@tudelft.nl)

*Ich glaube, dass jede Reihe einen bestimmten Wert haben müsse.*  
(Leonhard Euler)

*Divergent series are the invention of the devil, and it is a shame to base on them any demonstration whatsoever.* (Niels Abel)

*Just because something is infinite, doesn't mean it's zero!* (Wolfgang Pauli)

## Voorwoord

In dit hoofdstuk kijken we naar divergente reeksen. De grootmeester van de divergente reeksen was Leonhard Euler (1707 – 1783). We zullen zien hoe optimistisch Euler omgaat met rijen, reeksen en integralen. In de 19e eeuw verdween die onbekommerdheid, zoals het bovenstaande citaat van Abel illustreert. Er verschenen preciese definities van het begrip convergentie en limiet. Wiskundigen bewezen stellingen onder welke voorwaarden de bekende regels die bij eindige sommen gelden, zoals het veranderen van sommatievolgorde, ook voor reeksen en integralen geldt.

Natuurkundigen laten zich vaak minder weerhouden door dat soort vervelende voorwaarden en gaan opportunistischer te werk (zie het citaat van Pauli). Men gaat het erom om getallen te berekenen die experimenteel bevestigd of weerlegd kunnen worden. Wiskundige strengheid is van minder belang. De rechtvaardiging ligt niet in een wiskundig bewijs, maar in het experiment. Hiervan zullen we in §5 een spectaculair voorbeeld zien.

Maar ook wiskundigen hebben zich door divergente reeksen niet uit het veld laten slaan. Er zijn allerlei methoden ontworpen om divergente reeksen in het gareel te krijgen. Ik behandel er drie in §3. In §4 leg ik de link met de oudere ideeën van Euler.

Het natuurkundig intermezzo van §5 brengt ons bij het verschijnsel “asymptotische benadering”. In §6 keren we terug naar de wiskunde om dit uit te leggen in een eenvoudig voorbeeld, dat overigens weer teruggaat op Euler. In §7 maken we

kennis met een curieuze eigenschap van een vertrouwde benadering van  $\pi$ . Hier blijkt asymptotische benadering weer een rol te spelen, net als in §5 en §6.

De laatste paragraaf gaat ook over divergente machtreeksen, maar nu kijken we er door een totaal andere (20e-eeuwse, getaltheoretische) bril tegenaan.

Hoewel de thema's divergentie, sommatiemethode en asymptotische benadering dus een verbindend element vormen tussen de verschillende paragrafen, zouden deze toch redelijk onafhankelijk van elkaar te lezen moeten zijn.

Tot slot een waarschuwing aan wiskundigen: Het is niet mijn doel om alles in de grootste algemeenheid en met de waterdichtste argumenten te presenteren. Integendeel, ik wil u laten kennismaken met een aantal interessante wiskundige en natuurkundige zaken rond het thema divergentie. Sommige beweringen zullen u wellicht de gordijnen in jagen. Toch verkoop ik u geen onzin. Als ik bijvoorbeeld beweer dat  $1 + 2 + 3 + 4 + \dots = -\frac{1}{12}$ , evidente onzin, dan leg ik u uit wat die divergente reeks precies met  $-\frac{1}{12}$  te maken heeft.

## 1 Warming up

Als warming up kijken we naar een beroemde divergente reeks en zijn even beroemde convergente broer.



## 1.1 Oresme en $\sum_{n=1}^{\infty} \frac{1}{n}$

Nicolas Oresme (1320 - 1384) bewees dat de harmonische reeks divergeert:

$$\begin{aligned} 1 + \frac{1}{2} + \frac{1}{3} + \dots &= 1 + \left(\frac{1}{2} + \frac{1}{3}\right) + \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right) + \left(\frac{1}{8} + \dots + \frac{1}{15}\right) + \dots \\ &\geq 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots \\ &= \infty. \end{aligned}$$

Dus zelfs als de termen naar 0 convergeren, kan de reeks toch divergeren.

## 1.2 Euler en $\sum_{n=1}^{\infty} \frac{1}{n^2}$

De functie  $f(x) = \sin(\pi x)/\pi x$  heeft enkelvoudige nulpunten in  $\pm 1, \pm 2, \dots$  en  $f(x) \rightarrow 1$  als  $x \rightarrow 0$ . De grote wiskundige Euler, die nog vaak ter sprake zal komen in dit hoofdstuk, concludeerde daaruit:

$$\begin{aligned} \frac{\sin(\pi x)}{\pi x} &= (1-x)(1+x) \left(1 - \frac{x}{2}\right) \left(1 + \frac{x}{2}\right) \left(1 - \frac{x}{3}\right) \left(1 + \frac{x}{3}\right) \dots \\ &= (1-x^2) \left(1 - \frac{x^2}{4}\right) \left(1 - \frac{x^2}{9}\right) \left(1 - \frac{x^2}{16}\right) \dots \\ &= 1 - \left(1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots\right) \cdot x^2 + \dots \end{aligned}$$

De Taylor-ontwikkeling van  $\sin(x)$  is  $x - \frac{1}{3!}x^3 + \dots$ , dus we vinden

$$\frac{\sin(\pi x)}{\pi x} = \frac{\pi x - \frac{1}{6}\pi^3 x^3 + \dots}{\pi x} = 1 - \frac{\pi^2}{6} \cdot x^2 + \dots$$

Vergelijk de coëfficiënten van  $x^2$ :

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots = \frac{\pi^2}{6}.$$

Naar moderne maatstaven is deze afleiding onvolledig. Het bovenstaande argument is ook toepasbaar op  $e^x f(x)$ , dat dezelfde nulpunten heeft als  $f$  en dezelfde limiet voor  $x \rightarrow 0$ . Verder moet een theorie van oneindige producten worden ontwikkeld. Voor een upgrade van Eulers argument, zie bijvoorbeeld [4], hoofdstuk 9.

## 2 Twee curiositeiten

We bekijken twee opmerkelijke fenomenen die illustreren dat oneindige integralen en reeksen zich vreemd kunnen gedragen.

### 2.1 De trompet van Toricelli

De trompet van Toricelli (1608 – 1647) is het omwentelingslichaam van  $y = 1/x$  voor  $x \geq 1$  om de  $x$ -as. De inhoud van die trompet is

$$V = \pi \int_1^\infty y^2 dx = \pi \int_1^\infty \frac{1}{x^2} dx = \pi.$$

De oppervlakte van de trompet is

$$A = 2\pi \int_1^\infty y \sqrt{1 + (y')^2} dx \geq 2\pi \int_1^\infty y dx = \infty.$$

De eindige hoeveelheid verf die in de trompet van Toricelli past, is dus niet voldoende om de oneindige buitenkant te verven!

## 2.2 De herschikkingsstelling van Riemann

Een leuke reeks is de alternerende harmonische reeks:

$$A := 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots .$$

Het is gemakkelijk te bewijzen dat alternerende reeksen convergeren als de termen naar nul convergeren. De bovenstaande reeks is gemakkelijk als volgt uit te rekenen

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \dots ,$$

dus door te integreren vinden we

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

voor  $-1 < x \leq 1$ . We vinden  $A = \ln(2) = 0.693\dots$

Laten we nu eens een ander getal nemen, zeg  $\pi = 3.1415926535\dots$ . Tel net zolang positieve termen bij elkaar op totdat je boven  $\pi$  zit:

$$1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{1111} = 3.142\dots .$$

Tel hierbij net zolang negatieve termen op totdat je weer onder  $\pi$  zit:

$$1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{1111} - \frac{1}{2} = 2.642\dots .$$

Nu gaan we weer verder met de positieve termen totdat we weer boven  $\pi$  uitkomen:

$$1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{1111} - \frac{1}{2} + \frac{1}{1113} + \frac{1}{1115} + \dots + \frac{1}{3019} = 3.1416\dots .$$

Tel nu weer negatieve termen op totdat je weer onder  $\pi$  zit, enzovoorts. Omdat

$$1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \cdots = \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \cdots = \infty$$

(dat volgt eenvoudig uit Oresme's resultaat), weten we zeker dat het altijd weer zal lukken om afwisselend onder en boven  $\pi$  uit te komen. Door dit proces eindeloos te herhalen, krijgen we een reeks die naar  $\pi$  convergeert. Als we de termen van de alternerende reeks voor  $\ln(2)$  in een andere volgorde optellen, komt er dus een ander antwoord uit!

Natuurlijk kunnen we in het bovenstaande  $\pi$  door ieder ander getal vervangen. Niet alleen hangt de som van de reeks dus af van de volgorde van optelling, je kunt zelfs ieder reëel getal als som van de reeks krijgen door een geschikte volgorde van optelling te kiezen. Dit heet de *herschikkingsstelling van Riemann*. Deze stelling geldt niet alleen voor de alternerende harmonische reeks, die we hierboven hebben gebruikt, maar voor iedere *voorwaardelijk convergente reeks*, dat wil zeggen convergente reeksen  $\sum a_n$  waarvoor  $\sum |a_n|$  divergeert. Als  $\sum |a_n|$  daarentegen convergeert, dan hangt  $\sum a_n$  niet van de sommatievolgorde af.

### 3 De sommeermethoden van Cesàro, Abel en Euler

We bekijken drie methoden om divergente reeksen te temmen.

### 3.1 De noodzaak van definities

Volgens Hardy (zie [3], de bron voor deze en de volgende paragraaf) was onze held Euler van mening dat iedere reeks, ook divergente reeksen, een waarde kan worden toegekend (zie het citaat van Euler). Om bijvoorbeeld

$$S_1 := 1 - 1 + 1 - 1 + \dots$$

“uit te rekenen”, kunnen we als volgt argumenteren. Er geldt:

$$\begin{aligned} S_1 &= 1 - 1 + 1 - 1 + \dots \\ &= 1 - (1 - 1 + 1 - 1 + \dots) \\ &= 1 - S_1, \end{aligned}$$

dus  $S_1 = \frac{1}{2}$ . Volgens moderne maatstaven is dit geen correct argument, want we gaan er bij voorbaat al van uit dat de som  $S_1$  van de reeks bestaat. In de tijd van Euler ontbrak de *definitie* van de som van een reeks. Wiskundigen gingen ervan uit dat de som bestond en dat je hem alleen nog maar moest uitrekenen. De hedendaagse wiskundige werkt niet zo: die start met een definitie van de som.

Per definitie is de som van een reeks  $\sum_{n=0}^{\infty} a_n$  gelijk aan de limiet van de rij van deelsommen

$$s_0 = a_0, \quad s_1 = a_0 + a_1, \quad s_2 = a_0 + a_1 + a_2, \quad \dots$$

als die limiet bestaat. De reeks convergeert dus per definitie precies dan wanneer de rij van deelsommen convergeert. De reeks kun je formeel zelfs definiëren als de rij van deelsommen. In het bovenstaande geval is de rij van deelsommen gelijk aan  $1, 0, 1, 0, 1, 0, \dots$ . Die rij convergeert niet en de reeks  $S_1$  convergeert volgens de definitie dus ook niet.

## 3.2 De methode van Cesàro

Toch lijkt het antwoord  $S_1 = \frac{1}{2}$  nog niet zo gek. Allereerst hebben we het bovenstaande argument voor de identiteit  $S_1 = 1 - S_1$ . Ten tweede alterneert de rij van deelsommen tussen 0 en 1, dus  $\frac{1}{2}$  lijkt de eerlijke keuze. Ten derde kunnen we het recept van Ernesto Cesàro (1859 – 1906) volgen. Dat bestaat eruit dat we de gemiddeldes gaan nemen van de rij van deelsommen. In het onderhavige geval vinden we

$$\begin{aligned}g_0 &:= s_0 = 1 \\g_1 &:= \frac{1}{2}(s_0 + s_1) = \frac{1}{2}(1 + 0) = \frac{1}{2} \\g_2 &:= \frac{1}{3}(s_0 + s_1 + s_2) = \frac{1}{3}(1 + 0 + 1) = \frac{2}{3} \\g_3 &:= \frac{1}{4}(s_0 + s_1 + s_2 + s_3) = \frac{1}{4}(1 + 0 + 1 + 0) = \frac{1}{2} \\&\vdots\end{aligned}$$

De rij van deelsommen  $1, 0, 1, 0, \dots$  vervangen we door de minder springerige rij  $1, \frac{1}{2}, \frac{2}{3}, \frac{1}{2}, \frac{3}{5}, \frac{1}{2}, \frac{4}{7}, \dots$ . Deze rij convergeert naar  $\frac{1}{2}$  en daarom noemen we  $\frac{1}{2}$  de *Cesàro-som* van de reeks  $1 - 1 + 1 - 1 + \dots$ . We hebben Cesàro's methode nu met succes op een divergente reeks toegepast en deze zodoende alsnog een waarde toegekend, zoals dat Euler voor ogen stond. Als we Cesàro's methode op een convergente reeks toepassen, dan is het duidelijk dat Cesàro's methode hetzelfde antwoord oplevert. De sommatiemethode van Cesàro wordt daarom *regulier* genoemd.

### 3.3 De methode van Abel

Hier is nog een vierde argument om aan de divergente reeks  $1 - 1 + 1 - 1 + \dots$  de waarde  $\frac{1}{2}$  toe te kennen. We volgen nu het recept van Niels Abel (1802 – 1829), de grote Noorse wiskundige die bewees dat er geen *abcde*-formule voor de algemene vijfdegraads vergelijking  $ax^5 + bx^4 + cx^3 + dx + e = 0$  bestaat (en die, tussen twee haakjes, waarschuwde tegen het gebruik van divergente reeksen). Abel keek naar de machtreeks

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

bij de reeks  $\sum a_n$  en liet vervolgens  $x$  van onderen naar 1 convergeren. Als  $f(x)$  voor  $x \uparrow 1$  convergeert naar een limiet  $A$ , dan noemen we dat getal  $A$  de Abel-som van de reeks  $\sum a_n$ . Abel bewees dat als de reeks  $\sum a_n$  convergeert in de gebruikelijke zin, dat zijn methode dan hetzelfde resultaat oplevert. (De methode van Abel is dus regulier, net als de methode van Cesàro.) Maar de lol is natuurlijk dat de methode van Abel ook werkt bij sommige divergente reeksen, net als Cesàro's methode. Voor de reeks  $1 - 1 + 1 - 1 + \dots$  vinden we  $f(x) = \sum_{n=0}^{\infty} (-1)^n x^n = 1/(1+x)$ , dus Abels methode geeft opnieuw  $1 - 1 + 1 - 1 + \dots = \frac{1}{2}$ . Het is geen toeval dat Cesàro en Abel hetzelfde antwoord vinden. Als een reeks Cesàro-sommeerbaar is met som  $L$ , dan is de reeks ook Abel-sommeerbaar met dezelfde som  $L$  (zie [3], theorem 55).

We keren terug naar het eerste argument  $S_1 = 1 - 1 + 1 - 1 + \dots = 1 - (1 - 1 + 1 - 1 + \dots) = 1 - S_1$ , dus  $S_1 = \frac{1}{2}$ . Dit argument is correct als we de sommatie van divergente reeksen oppvatten in de zin van Cesàro of Abel! Deze twee methoden

voldoen immers aan

$$a_0 + a_1 + a_2 + \dots = L \quad \implies \quad a_1 + a_2 + \dots = L - a_0 \quad (1)$$

en

$$a_0 + a_1 + a_2 + \dots = L \quad \implies \quad (ka_0) + (ka_1) + (ka_2) + \dots = kL. \quad (2)$$

(De tweede eigenschap hebben we gebruikt voor  $k = -1$ .) Iedere sommatiemethode die aan deze twee axioma's voldoet en die aan  $1 - 1 + 1 - 1 + \dots$  een eindige waarde toekent, moet dus  $1 - 1 + 1 - 1 + \dots = \frac{1}{2}$  opleveren.

We kijken nu naar

$$S_2 := 1 - 2 + 3 - 4 + \dots .$$

De deelsommen zijn  $1, -1, 2, -2, 3, -3, \dots$ , dus de reeks divergeert. De Cesàro-gemiddelden zijn  $1, 0, \frac{2}{3}, 0, \frac{3}{5}, 0, \frac{4}{7}, 0, \dots$ , dus de reeks is niet eens Cesàro-sommeerbaar. (Als we de Cesàro-truc om gemiddelden te nemen nogmaals zouden toepassen, dan zouden we  $\frac{1}{4}$  vinden als som.)

Wat doet Abel? De reeks  $f(x) = 1 - x + 2x^2 - 3x^3 + \dots$  is de afgeleide van  $-1 + x - x^2 + \dots = -(1+x)^{-1}$ , dus  $f(x) = (1+x)^{-2}$ . De methode van Abel werkt dus wel en levert  $S_2 = f(1) = \frac{1}{4}$ . Dit voorbeeld laat zien dat de methode van Abel echt sterker is dan de methode van Cesàro.



Hier is nog een argument voor  $S_2 = \frac{1}{4}$ :

$$\begin{aligned} S_2 &= 1 - 2 + 3 - 4 + \dots \\ &= 1 - (2 - 3 + 4 - 5 + \dots) \\ &= 1 - (1 - 1 + 1 - 1 + \dots) - (1 - 2 + 3 - 4 + \dots) \\ &= 1 - S_1 - S_2 \\ &= 1 - \frac{1}{2} - S_2. \end{aligned}$$

Dit argument werkt voor iedere sommatiemethode waarvoor  $S_1$  en  $S_2$  convergeren en die behalve aan de bovenstaande axioma's ook voldoen aan

$$\sum a_n = s \quad \text{en} \quad \sum b_n = t \quad \implies \quad \sum (a_n + b_n) = s + t. \quad (3)$$

De methoden van Cesàro en Abel voldoen beide aan de drie axioma's. Toch geldt niet  $S_2 = \frac{1}{4}$  volgens Cesàro. De reden:  $S_2$  is niet Cesàro-sommeerbaar, zoals we hebben gezien.

### 3.4 De methode van Euler

We hebben gezien dat Cesàro en Abel sommige divergente reeksen kunnen temmen. De methode van Abel is sterker dan die van Cesàro, maar soms faalt zelfs Abel. Hier is een voorbeeld:

$$S_3 := 1 - 2 + 4 - 8 + \dots$$

De bijbehorende reeks  $f(x) = 1 - 2x + 4x^2 - 8x^3 + \dots = 1 - (2x) + (2x)^2 - (2x)^3 + \dots = (1 + 2x)^{-1}$  divergeert voor  $x > \frac{1}{2}$ , dus we kunnen niet de limiet  $x \uparrow 1$  nemen. De reeks is dus niet sommeerbaar in de zin van Abel (en dus zeker niet in

de zin van Cesàro). Toch is  $f(x) = 1/(1 + 2x)$  keurig netjes gedefinieerd in  $x = 1$ . Het ligt dus voor de hand om

$$S_3 = f(1) = \frac{1}{3}$$

te definiëren. We noemen dit de *Euler-som* van de divergente reeks en we zeggen dat de reeks Euler-sommeerbaar is. (Voor de kenners: het wezenlijke punt is dat  $f(z) = 1/(1 + 2z)$  analytisch is buiten  $z = -\frac{1}{2}$ .)

Net als de methoden van Cesàro en Abel, voldoet de methode van Euler aan de drie axioma's. Voor iedere sommatiemethode die voldoet aan deze axioma's en waarvoor  $S_3$  bestaat, geldt

$$S_3 = 1 - 2 + 4 - 8 + \dots = 1 - (2 - 4 + 8 - \dots) = 1 - 2 \cdot (1 - 2 + 4 - 8 + \dots) = 1 - 2S_3,$$

$$\text{dus } S_3 = \frac{1}{3}.$$

De lezer kan nagaan dat voor

$$S_4 := 1 + 2 + 4 + 8 + \dots$$

vergelijkbaars geldt: niet Cesàro-sommeerbaar, zelfs niet Abel-sommeerbaar, maar wel Euler-sommeerbaar met som  $S_4 = -1$ . Dit is toch wel een vrij verrassend resultaat: de som van positieve termen is negatief!

## 4 De droom van Euler

Zoals eerder vermeld, meende Euler dat aan iedere divergente reeks op een zinnige manier een (eindige?) waarde kan worden toegekend. Voor sommige reeksen, zoals

$$S_5 := 1 + 1 + 1 + 1 + \dots$$

falen echter zowel de methode van Cesàro, Abel en Euler. De bijbehorende machtreeks geeft namelijk  $f(x) = 1/(1-x)$  met een singulariteit bij  $x = 1$ . We kunnen ook algebraïsch redeneren: als een sommeermethode voldoet aan het eerste axioma, dan kan  $1+1+1+1+\dots$  volgens die methode niet convergeren. Dat zou immers

$$S_5 = 1 + S_5$$

impliceren! Toch wist Euler zelfs die reeks klein te krijgen met een ander argument (geen algemene methode!):

$$1 + 1 + 1 + 1 + \dots = \zeta(0) = -\frac{1}{2}.$$

Hierbij is  $\zeta(s) = \sum n^{-s}$  de beroemde Riemann-zèta-functie, waarover later meer. Met dezelfde methode zullen we ontdekken dat

$$S_6 := 1 + 2 + 3 + 4 + \dots = -\frac{1}{12}.$$

Om de virtuositeit van Euler te illustreren, is hier nog een kunststuk van de meester:

$$0! - 1! + 2! - 3! + \dots = \int_0^\infty \frac{e^{-t}}{1+t} dt = 0.596347 \dots$$

Dit is de bijbehorende redenering:

$$\int_0^\infty \frac{e^{-t}}{1+t} dt = \int_0^\infty \sum_{n=0}^\infty (-t)^n e^{-t} dt = \sum_{n=0}^\infty \int_0^\infty (-t)^n e^{-t} dt = \sum_{n=0}^\infty (-1)^n n!.$$

Zeer dubieus: mogen we bijvoorbeeld reeks en integraal wel verwisselen? Erger nog, de reeks  $\sum (-t)^n$  is divergent op een groot gedeelte van het integratiedomein  $[0, \infty)$ ! Maar daarom niet getreurd: we zouden de reeks rechts kunnen *definiëren* als de integraal links...

Had Euler gelijk en kan men inderdaad aan iedere (divergente) reeks een zinnige waarde toekennen? Ik denk het niet. Ik ken geen enkele methode die de harmonische reeks

$$H := 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

aankan. Dat zegt niet zo veel, want ik ben geen specialist en al helemaal geen Euler, maar toch: de singulariteit bij  $z = 1$  van  $\sum z^n/n = -\ln(1-z)$  is wel bijzonder akelig. (De functie is meerwaardig bij  $z = 1$ : loop een rondje om  $z = 1$  in het complexe vlak en de functie  $\ln(1-z)$  verschuift over  $2\pi i$ .) Het is overigens enigzins paradoxaal dat de methoden van Cesàro, Abel en Euler falen bij een mild divergente reeks als de harmonische, terwijl ze reeksen die veel heftiger spartelen er wel onder krijgen. Behalve de sommatie van de harmonische reeks is er nog een ander probleem: er zijn situaties waarin verschillende methodes verschillende antwoorden geven (zie bijvoorbeeld bladzijde 16 in [3]).

## 5 Hogere wis- en natuurkunde

Een van de mooiste voorspellingen van de snaartheorie is de dimensie van het universum waarin we leven. Daarin onderscheidt deze theorie zich in positieve zin van Newtoniaanse en relativistische mechanica, die ook in andere dimensies werken. Ergens in die berekening komt de volgende reeks voor:

$$S := 1 + 2 + 3 + 4 + 5 + \dots$$

Een vergelijkbare divergente reeks komt voor in het natuurkundige Casimir-effect, waarover later meer. Hoe groot is  $S$ ?

Het saaie antwoord is  $S = \infty$ , maar er zijn veel intrigerender antwoorden:

$$S = \frac{1}{0^2} - \frac{1}{12}$$

en zelfs

$$S = -\frac{1}{12}.$$

Dit zijn de antwoorden waarmee de natuurkundigen werken! Ik geef drie verklaringen voor deze verrassende antwoorden: een deftige wiskundige verklaring, een houtje-touwtje berekening en een natuurkundige verklaring. Ik vind de eerste verklaring de beste, de tweede de gemakkelijkste, en de derde de mooiste. De eerder besproken methoden van Cesàro, Abel en Euler laten het overigens afweten, want  $f(x) = \sum nx^n = x/(1-x)^2$  heeft een singulariteit bij  $x = 1$ .

## 5.1 Verklaring met behulp van de Riemann-zeta-functie

Hier is de deftige variant. Je bekijkt de Riemann-zeta-functie

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s}.$$

De reeks convergeert, net als de integraal  $\int_1^{\infty} x^{-s} dx$ , voor  $s > 1$ . Dankzij Euler weten we dat  $\zeta(2) = \pi^2/6$  en dankzij Oresme weten we dat  $\zeta(1) = \infty$ . We zijn geïnteresseerd in  $\zeta(-1)$ , maar bij  $s = -1$  divergeert de reeks natuurlijk. Toch kan een wiskundige met goed geweten zeggen dat

$$\zeta(-1) = -\frac{1}{12}.$$

Daarvoor moeten we de parameter van de veilige zone  $s > 1$  deformeren naar  $s = -1$ , waarbij we de Oresme-singulariteit bij  $s = 1$  moeten vermijden. Dat lukt niet als we  $s$  beperken tot reële getallen, maar het lukt wel als we  $s$  door het complexe vlak bewegen. Op die manier – precies gemaakt in de theorie van analytische voortzetting van complexe functies – krijg je een functie  $\zeta$  die voor alle  $s \in \mathbf{C}$  behalve  $s = 1$  gedefinieerd is. Die functie die zo ontstaat heeft een symmetrie die  $\zeta(s)$  met  $\zeta(1-s)$  in verband brengt:

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{1}{2}\pi s\right) \Gamma(1-s) \zeta(1-s)$$

Voor  $s = -1$  krijgen we dus een verband tussen  $S = \zeta(-1)$  en  $\zeta(2) = \pi^2/6$ . We vullen  $s = -1$  in:

$$\zeta(-1) = \frac{1}{2} \frac{1}{\pi^2} \sin\left(-\frac{1}{2}\pi\right) \Gamma(2) \zeta(2) = \frac{1}{2} \frac{1}{\pi^2} (-1)(2-1)! \frac{\pi^2}{6} = -\frac{1}{12}.$$

De Riemann-zeta functie is bekend vanwege het verband met de verdeling van priemgetallen, maar we zien nu dat deze functie ook iets zegt over het Casimir-effect in de natuurkunde en het aantal dimensies van het universum!

## 5.2 Houtje-touwtje variant

Hier is een dubieuze afleiding van de dubieuze bewering  $\sum_{n=1}^{\infty} n = \frac{1}{0^2} - \frac{1}{12}$ . We beginnen met de methode van Abel en Euler en bekijken de machtreeks  $f(x) = \sum nx^n$ . Deze reeks lijkt op de meetkundige reeks  $\sum_{n=1}^{\infty} x^n = (1-x)^{-1}$ . Links en rechts differentiëren geeft:  $\sum_{n=1}^{\infty} nx^{n-1} = (1-x)^{-2}$ , dus  $\sum_{n=1}^{\infty} nx^n = x/(1-x)^2$ . Zoals eerder opgemerkt, zien we dat de reeks

$1 + 2 + 3 + 4 + \dots$  dus niet te temmen is met de methode van Euler of Abel, laat staan met die van Cesàro.

Ik volg nu hoofdstuk I.8 in [5] en substitueer  $x = e^{-a}$ :

$$\sum_{n=1}^{\infty} ne^{-an} = \frac{e^{-a}}{(1 - e^{-a})^2} = \frac{e^a}{(e^a - 1)^2}$$

De limiet  $x \uparrow 1$  correspondeert met de limiet  $a \downarrow 0$ . We berekenen

$$\begin{aligned} \frac{e^a}{(e^a - 1)^2} &= \frac{1 + a + \frac{1}{2}a^2 + \dots}{a^2(1 + \frac{1}{2}a + \frac{1}{6}a^2 + \dots)^2} \\ &= \frac{1}{a^2} \cdot \frac{1 + a + \frac{1}{2}a^2 + \dots}{1 + a + \frac{7}{12}a^2 + \dots} \\ &= \frac{1}{a^2} \cdot (1 + a + \frac{1}{2}a^2 + \dots) (1 - a + \frac{5}{12}a^2 + \dots) \\ &= \frac{1}{a^2} - \frac{1}{12} + \dots \end{aligned}$$

Ten slotte nemen we de limiet  $a \downarrow 0$ , zodat  $x^n = e^{-an} \uparrow 1$ . In deze zin geldt

$$\sum_{n=1}^{\infty} n = \frac{1}{0^2} - \frac{1}{12}.$$

Waarschijnlijk vindt u dit kolder, maar Casimir dacht daar anders over. Daarover gaat de volgende paragraaf.

### 5.3 Polderkolder?

De Nederlandse natuurkundigen Hendrik Casimir (1909 – 2000) en Dirk Polder (1919 – 2001) beredeneerden in 1948 dat tussen

twee parallelle, grote metalen platen een kleine aantrekkende kracht zou moeten bestaan van de grootte

$$|F| = \frac{\pi^2 \hbar c}{240} \frac{A}{d^4} \sim \frac{A}{d^4},$$

waarbij  $d$  de afstand tussen de platen is,  $A$  de oppervlakte en  $\hbar c = 3 \times 10^{-26}$  Jm. Als  $A = 1 \text{ cm}^2$  and  $d = 1 \text{ }\mu\text{m}$ , dan vinden we  $F = 1 \times 10^{-7}$  N. Dit extreem subtiele effect is experimenteel bevestigd.

In de berekening wordt gekeken naar electro-magnetische golfjes tussen de platen die nulpunten hebben op de platen. Als we de platen op  $x = 0$  en  $x = d$  zetten, kijken we dus naar functies van de vorm  $\sin(n\pi x/d)$  met  $n = 1, 2, 3, \dots$ . De energie in zo'n golfje blijkt evenredig met  $n\pi/d$  te zijn. De totale hoeveelheid energie in alle golfjes samen is dus evenredig met

$$\sum_{n=1}^{\infty} \frac{n\pi}{d} = \frac{\pi}{d} \sum_{n=1}^{\infty} n.$$

Casimir liet zich door eventuele wiskundige bezwaren niet van de wijs brengen: hij gebruikte  $\sum_{n=1}^{\infty} n = \frac{1}{0^2} - \frac{1}{12}$ , wist in de loop van zijn berekening de ongewenste term  $\frac{1}{0^2}$  weg te werken en vond daarmee de bovenstaande uitdrukking voor  $F$ . (Eigenlijk heb je in 3 dimensies  $\zeta(-3) = \sum n^3$  nodig in plaats van  $\zeta(-1) = \sum n$ , maar het idee is hetzelfde.) In zekere zin is de rare uitdrukking  $\sum n = \frac{1}{0^2} - \frac{1}{12}$  dus experimenteel bevestigd in een natuurkundig laboratorium!



## 5.4 $\infty - \infty = 0.001159652181$

Moderne natuurkundigen zijn heel sterk in het krijgen van interessante antwoorden uit  $\infty - \infty$ . Het beroemdste is

$$\infty - \infty = 0.001159652181$$

voor het zogenaamde anomale magnetische dipoolmoment van een electron. (Een electron is een klein magneetje en dit getal zegt iets over de sterkte van dat magneetje. Voor ons als wiskundigen is alleen van belang dat het een getal is dat voorspeld wordt door natuurkundige theorie en dat experimenteel zeer nauwkeurig kan worden gemeten.) Ook hier geldt dat dubieuze wiskunde tot subtiele natuurkundige voorspellingen heeft geleid, die vervolgens experimenteel bevestigd zijn. Het betreft zelfs de nauwkeurigste theoretische voorspelling (zo'n 10 decimalen!) die theoretisch bevestigd is!

De technieken om eindige antwoorden te krijgen uit  $\infty - \infty$  vat men samen onder de noemer *renormalisatie*. Onze landgenoten Gerard 't Hooft en Martinus Veltman kregen er in 1999 de Nobelprijs voor de natuurkunde voor. In deze bewieroking van natuurkundigen uit de Nederlandse polder mag Johannes Diederik van der Waals (1837 - 1923) niet ontbreken. Niet alleen omdat hij een oom van mijn oom zou zijn<sup>1</sup>, maar vooral omdat in 1910 eveneens een Nobelprijs voor natuurkunde kreeg en wel voor zijn onderzoek naar de naar hem genoemde kracht tussen moleculen. Die Van-der-Waals-krachten berusten op hetzelfde principe als het Casimir-effect. . . .

Terug naar  $\infty - \infty = 0.001159652181$ . De berekening van 0.001159652181 is erg ingewikkeld, maar uiteindelijk komt het

---

<sup>1</sup>koude kant, helaas

neer op een machtreeksontwikkeling

$$\sum_{n=1}^{\infty} a_n \alpha^n,$$

waarbij  $\alpha \approx 1/137.036$  een natuurconstante is (de beroemde fijnstructuurconstante) en waarbij  $a_n$  coëfficiënten zijn die berekend kunnen worden met behulp van Feynman-diagrammen. Er zijn slechts een paar coëfficiënten numeriek benaderd, dus in de praktijk worden maar een paar termen gebruikt om het anomale magnetische dipoolmoment van een electron te bepalen. Die paar termen geven dus al een zeer nauwkeurige benadering.

Je zou kunnen denken dat meer en meer termen tot steeds nauwkeuriger theoretische voorspellingen leiden, maar hoogstwaarschijnlijk is dat niet waar. Er zijn goede natuurkundige redenen om te vermoeden dat de reeks  $\sum a_n \alpha^n$  divergeert. Men verwacht zelfs dat de machtreeks  $\sum a_n x^n$  alleen voor  $x = 0$  convergeert! Op den duur zullen de termen  $a_n \alpha^n$  groot worden en zal de benadering door toevoeging van zo'n nieuwe term niet beter, maar juist slechter worden. Dit verschijnsel heet *asymptotische benadering*. In de volgende paragraaf geven we een eenvoudig wiskundig voorbeeld van dit verschijnsel.

## 6 Asymptotische ontwikkeling

Hierboven hebben we

$$0! - 1! + 2! - 3! + \dots = \int_0^{\infty} \frac{e^{-t}}{1+t} dt = 0.596347\dots$$

gezien. Dezelfde redenering geeft

$$\int_0^\infty \frac{e^{-t}}{1+tx} dt = 0! - 1! \cdot x + 2! \cdot x^2 - 3! \cdot x^3 + \dots$$

De reeks heeft convergentiestraal 0, net als de Dyson-reeks  $\sum a_n \alpha^n$  uit de vorige paragraaf. Toch kunnen we de reeks voor kleine  $x$  gebruiken om de integraal te benaderen. Als voorbeeld nemen we  $x = \frac{1}{10}$ . De integraal is

$$\int_0^\infty \frac{e^{-t}}{1 + \frac{1}{10}t} dt = 0.915633 \dots$$

We benaderen deze integraal door de deelsommen

$$S_m := \sum_{n=0}^m (-1)^n \frac{n!}{10^n}.$$

We vinden het beste resultaat bij  $m = 9$ :  $S_9 = 0.915456 \dots$  met een fout van  $0.000177 \dots$ . De fout bij  $S_{10}$  is net ietsje groter en bij nog grotere  $m$  neemt de fout steeds verder toe. Bijvoorbeeld is de fout bij  $m = 20$  al ongeveer 0.02, bij  $m = 30$  is de fout bijna 200, bij  $m = 40$  is de fout  $6 \cdot 10^7$ , en bij  $m = 100$  is de fout ongeveer  $10^{58}$ ! Het is gemakkelijk om in te zien dat de fout naar oneindig gaat voor  $m \rightarrow \infty$ .

## 7 Leibniz en de decimalen van $\pi$

### 7.1 Observatie

Leibniz (1646 – 1716) wist al dat

$$4 \cdot \left( 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots \right) = \pi.$$

(Dit volgt uit de Taylorreeksontwikkeling van  $\arctan(x)$ .) De reeks convergeert echter heel langzaam. Als we bijvoorbeeld doorrekenen tot 1 000 000, dan vinden we

$$S := 4 \cdot \left( 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots - \frac{1}{999\,999} \right) = 3.141590 \dots$$

terwijl  $\pi = 3.151592 \dots$ . De zesde decimaal is dus al fout. Dat is niet verwonderlijk, want we verwachten dat de fout ongeveer  $4 \cdot \frac{1}{2} \cdot 1/1\,000\,000$  is.

Maar nu komt het. We kijken naar de decimalen  $na$  de foutieve zesde decimaal:

$$S = 3.14159 \mathbf{06535} 89793 \mathbf{24046} 26433 83269 50288 4197 \dots$$

Vergelijk dit met

$$\pi = 3.14159 26535 89793 23846 26433 83279 50288 4197 \dots$$

We zien dat  $na$  de foutieve zesde decimaal weer 10 correcte decimalen volgen! Daarna komen 2 foutieve decimalen, dan weer 11 correcte, een foute, en weer (minstens) 10 correcte decimalen... Wat hier aan de hand?

## 7.2 Verklaring

Voor  $n$  deelbaar door 4, bijvoorbeeld voor  $n = 1\,000\,000$ , definiëren we

$$R_n := \pi - 4 \left( 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots - \frac{1}{n-1} \right).$$

Borwein, Borwein en Dilcher hebben in [1] bewezen dat

$$R_n = \frac{2}{n} - \frac{2}{n^3} + \frac{10}{n^5} - \frac{122}{n^7} + \dots$$

Vullen we  $n = 10^6$  in, dan vinden we

$$R_{1000000} = 2 \cdot 10^{-6} - 2 \cdot 10^{-18} + 1 \cdot 10^{29} - 1.22 \cdot 10^{40} + \dots$$

Dit verklaart waarom de benadering  $S_{1000000}$  bij de eerste 39 decimalen klopte behalve bij decimalen nummer 6, 17, 18 en 29.

De reeksontwikkeling van Borwein, Borwein en Dilcher voor  $R_n$  blijkt ook een asymptotische te zijn: hoe groot  $n$  ook is, op den duur gaan de termen naar oneindig. De beste benadering van  $R_n$  krijg je door ongeveer  $n$  termen te nemen. (De fout in de BBD-benadering van de fout  $R_n$  gaat naar oneindig als het aantal termen in de BBD-benadering naar oneindig gaat, net als in het voorbeeld  $\int_0^\infty e^{-t}/(1+xt) dt = \sum n!(-x)^n$  dat we eerder hebben bekeken.) Om het vreemde verschijnsel van de vele correcte decimalen te begrijpen, is een handjevol termen gelukkig al voldoende.

## 8 Waaron $\ln(-1) = 0$ zo gek niet is

Wat vindt u van het volgende argument? We beginnen met de bekende alternerende reeks

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

voor  $-1 < x \leq 1$ . We negeren convergentieproblemen en vullen vrolijk  $x = -2$  in:

$$\ln(-1) = -\left(2 + \frac{1}{2} \cdot 2^2 + \frac{1}{3} \cdot 2^3 + \frac{1}{4} \cdot 2^4 + \dots\right).$$

We zijn nu niet meer te houden en zondigen nog meer:

$$2 \ln(-1) = \ln((-1)^2) = \ln(1) = 0.$$

We concluderen dat

$$2 + \frac{1}{2} \cdot 2^2 + \frac{1}{3} \cdot 2^3 + \frac{1}{4} \cdot 2^4 + \dots = 0.$$

Waarschijnlijk vindt u deze afleiding onzin en het resultaat ook; de reeks aan de linkerkant divergeert immers?

Toch is het bovenstaande correct, als we het op een andere manier interpreteren. Voor die interpretatie moeten we ons begrip van *afstand* tussen twee rationale getallen herdefiniëren. De gebruikelijke definitie van de afstand  $d(a, b)$  tussen twee rationale getallen  $a, b \in \mathbf{Q}$  is

$$d(a, b) := |a - b|.$$

Dit afstandsbebegrip ligt ten grondslag aan het begrip convergentie: een rij  $s_n$  van rationale getallen is convergent als  $|s_n - s_m|$  willekeurig klein wordt voor grote  $n$  en  $m$ . Preciezer: de rij  $s_n$  is convergent als voor iedere rationale  $\epsilon > 0$  een natuurlijk getal  $N$  bestaat, zodat voor alle  $n, m \geq N$  geldt:  $|s_n - s_m| < \epsilon$ . Dit is het bekende criterium van Cauchy. Deze definitie van convergentie van rijen rationale getallen is equivalent met de bekendere variant: er bestaat een reëel getal  $\ell$  zodat  $|\ell - s_n|$  willekeurig klein wordt voor voldoende grote  $n$ . Het grote voordeel van de variant van Cauchy is dat de limiet  $\ell$ , die in het algemeen niet in  $\mathbf{Q}$  ligt, niet in de definitie van convergentie voorkomt. In de definitie worden alleen rationale getallen en een afstandsbebegrip op  $\mathbf{Q}$  gebruikt. We kunnen die definitie daarom ook gebruiken bij andere afstandsbebegrppen op  $\mathbf{Q}$ .

Hier is een belangrijk alternatief afstandsbegrip. We kunnen ieder rationaal getal  $x \neq 0$  schrijven als  $x = 2^k p/q$  waarbij  $k$  geheel is (positief of negatief of nul) en  $p$  en  $q$  oneven gehele getallen zijn. We zeggen dan dat  $x$  deelbaar is door  $2^k$ . We definiëren dan

$$\left| 2^k \frac{p}{q} \right|_2 := 2^{-k}$$

en

$$d_2(a, b) := |a - b|_2.$$

Verder definiëren we natuurlijk  $|0|_2 = 0$ . Bijvoorbeeld geldt

$$d_2(123, 23) = |100|_2 = |2^2 \cdot 25|_2 = \frac{1}{4} \quad \text{en} \quad \left| \frac{7}{64} \right|_2 = 64.$$

Een rij  $s_n$  convergeert dus naar 0 als  $s_n$  voor grote  $n$  deelbaar is door grote machten van 2. Preciezer: voor iedere  $k > 0$  is er een  $N$  zodat voor alle  $n \geq N$  geldt dat  $s_n$  deelbaar is door  $2^k$ . Deze vorm van convergentie heet *2-adisch*. In plaats van 2 kan overigens met een willekeurig priemgetal worden gewerkt.

We keren terug naar het resultaat van onze woeste afleiding van  $2 + \frac{1}{2} \cdot 2^2 + \frac{1}{3} \cdot 2^3 + \frac{1}{4} \cdot 2^4 + \dots = 0$ . Het mooie is dat deze afleiding correct is als we alles 2-adisch bekijken. Het argument bewijst dan de volgende niet-triviale stelling:

**Stelling 8.1.** *Voor iedere  $k > 0$  bestaat een  $N > 0$  zodat voor iedere  $n \geq N$  geldt dat  $2 + \frac{1}{2} \cdot 2^2 + \frac{1}{3} \cdot 2^3 + \frac{1}{4} \cdot 2^4 + \dots + \frac{1}{n} 2^n$  deelbaar is door  $2^k$ .*

Ik heb dit geleerd uit het prachtige boek [2]. Kunt u het rechtstreeks bewijzen zonder 2-adische poespas?

Tot slot keren we terug naar de reeks  $S_4 = 1 + 2 + 4 + 8 + \dots$ . Eerder argumenteerden we dat  $S_4 = 1 + 2S_4$ , dus  $S_4 = -1$ . Dus als  $S_4$  sommeerbaar is volgens een sommatiemethode die vol doet aan de drie axioma's uit paragraaf 3, dan geldt  $S_4 = -1$ . We hebben gezien dat Euler slaagde, waar Cesàro en Abel faalden. Dat alles speelde zich af binnen de reële getallen met het gebruikelijke afstandsbegrip. In de 2-adische wereld — waarin wel de rationale getallen voorkomen, maar niet de reële — is  $S_4$  convergent dank zij het 2-adische afstandsbegrip. (In het 2-adische paradijs geldt dat een reeks precies dan convergeert wanneer de termen naar 0 gaan. Het zelfde geldt in de  $p$ -adische wereld voor ieder priemgetal  $p$ . Een reeks als de harmonische, die divergeert hoewel de termen naar 0 gaan, komt in de  $p$ -adische wereld dus niet voor.)  $S_4 = 1 + 2 + 4 + 8 + \dots$  is dus 2-adisch convergent en het bovenstaande argument toont aan dat de limiet  $-1$  is. Hoe kan de som van positieve getallen negatief worden? Wel, in de  $p$ -adische wereld verliest het begrip “positiviteit” zijn betekenis. Algemeener vind je dat  $1 + p + p^2 + p^3 + \dots$  in de  $p$ -adische wereld een convergente meetkundige reeks is met som  $1/(1 - p)$ .

## 9 Slotwoord

De drie citaten aan het begin vormen een heel korte geschiedenis van divergente reeksen: na aanvankelijk optimisme (Euler) volgde een tijd waarin divergente reeksen uit de gratie raakten (Abel). In de twintigste eeuw volgde hun rehabilitatie met het boek [3] van Hardy, de successen uit de natuurkunde (Pauli). De  $p$ -adische getallen zijn eveneens kinderen van die eeuw. Om



Pauli te parafraseren: divergente reeksen zijn zoveel meer dan alleen maar “oneindig”!

## Referenties

- [1] Borwein, J.M., Borwein, P.B., Dilcher, K., *Pi, Euler numbers, and asymptotic expansions*, American Mathematical Monthly **96** (1989), 681 – 687.
- [2] Gouvêa, F.Q., *p-adic Numbers, An Introduction*, Springer, 1997.
- [3] Hardy, G.H., *Divergent Series*, Oxford, 1949.
- [4] K. Jänich, *Funktionentheorie, Eine Einführung*, 3. Auflage, Springer, 1993.
- [5] Zee, A., *Quantum Field Theory in a nutshell*, Princeton University Press, 2003.

# Het benaderen van irrationale getallen door ‘rationale’

Cor Kraaikamp, TU Delft  
e-mail: C.Kraaikamp@tudelft.nl

## Voorwoord

Het is zo alledaags om irrationale getallen te benaderen door rationale, dat we dit nauwelijks meer in de gaten hebben. Wat is  $e$ ? Volgens de Casio die ik bij de HEMA heb gekocht is  $e$  gelijk aan 2.718281828. Maar hoe goed is deze benadering? Meer in het algemeen: hoe goed kunnen we irrationale door rationale getallen benaderen en wat wil ‘goed’ eigenlijk zeggen?

Om deze vragen te kunnen beantwoorden zullen we kettingbreuken voorbij laten komen, maar ook de vertrouwde decimale en binaire ontwikkelingen en generalisaties hiervan zoals

de ontwikkelingen naar niet gehele basis  $\beta > 1$ . Verschillende technieken zullen worden gebruikt; uit de getaltheorie, maar ook uit dynamische systemen en ergodentheorie.

## 1 Warming up

Maar laten we beginnen bij het begin. Iedereen weet dat

$$\frac{1}{3} = 0.33333 \cdots 33333333 \dots,$$

maar wat wil dit ook al weer zeggen? ... en wat bedoelen we ook al weer met

$$\frac{1}{4} = 0.25, \quad \frac{1}{8} = 0.125, \quad \frac{1}{7} = 0.142857142 \dots$$

Wat gebeurt er met deze laatste ontwikkeling? Is deze eindig? oneindig? Zo ja, op wat voor *manier*? Periodiek, 'chaotisch' ... ?

### 1.1 Decimale ontwikkelingen

Als de decimale ontwikkeling van een getal  $x$  tussen 0 en 1 gelijk is aan

$$x = 0.a_1a_2a_3 \dots$$

met digits  $a_n \in \{0, 1, 2, \dots, 9\}$  voor  $n \geq 1$ , dan wil dit zeggen dat

$$x = \sum_{n=1}^{\infty} \frac{a_n}{10^n} = \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{a_3}{10^3} + \dots + \frac{a_n}{10^n} + \dots.$$

Dus

$$\frac{1}{3} = \sum_{n=1}^{\infty} \frac{3}{10^n} = \frac{3}{10} + \frac{3}{10^2} + \frac{3}{10^3} + \cdots + \frac{3}{10^n} + \cdots,$$

terwijl

$$\frac{1}{4} = \frac{2}{10} + \frac{5}{10^2} \quad \text{en} \quad \frac{1}{8} = \frac{1}{10} + \frac{2}{10^2} + \frac{5}{10^3}.$$

Tenslotte is

$$\begin{aligned} \frac{1}{7} &= \frac{1}{10} + \frac{4}{10^2} + \frac{2}{10^3} + \frac{8}{10^4} + \frac{5}{10^5} + \frac{7}{10^6} \\ &+ \frac{1}{10^7} + \frac{4}{10^8} + \frac{2}{10^9} + \frac{8}{10^{10}} + \frac{5}{10^{11}} + \frac{7}{10^{12}} \\ &+ \frac{1}{10^{13}} + \frac{4}{10^{14}} + \frac{2}{10^{15}} + \frac{8}{10^{16}} + \frac{5}{10^{17}} + \frac{7}{10^{18}} \\ &+ \frac{1}{10^{19}} + \frac{4}{10^{20}} + \frac{2}{10^{21}} + \frac{8}{10^{22}} + \frac{5}{10^{23}} + \frac{7}{10^{24}} \\ &+ \cdots . \end{aligned}$$

**Vragen:** *Hoe komen we aan deze 'ontwikkelingen in basis 10', en waarom zijn ze allemaal periodiek? Als we van basis veranderen, krijgen we dan soortgelijke dingen te zien? Hoe zit 't trouwens met de ontwikkeling van een getal als  $\sqrt{2} - 1$ ? Is de ontwikkeling daarvan dan niet periodiek? Als de ontwikkeling van een getal  $x$  periodiek of eindig is, is  $x$  dan rationaal?*

De antwoorden op deze vragen zijn relatief eenvoudig. Zo krijgen we de decimale ontwikkeling van een 'breuk'  $a/b$  (waarbij  $a$  en  $b \neq 0$  gehele getallen zijn) door een staartdeling te doen. In deze staartdeling krijgen we keer op keer een rest; als deze rest op een zeker moment 0 is, dan is de ontwikkeling eindig

(zoals bij  $1/4$ ). Als we nimmer 0 krijgen als rest, dan zal de ‘decimale ontwikkeling’ die we zo krijgen dus oneindig zijn. Merk op, dat er maar eindig veel mogelijkheden zijn voor de rest; je kan als rest alleen maar  $0, 1, \dots, b - 1$  vinden. Dus na eindig veel stappen moeten we wel een rest krijgen die we eerder hebben gezien, en de staartdeling begint zich te herhalen: de decimale ontwikkeling is *periodiek*.

Als  $x$  een *irrationaal* getal tussen 0 en 1 is kunnen we niet een staartdeling toepassen. Intuïtief is direct duidelijk hoe de decimale ontwikkeling van  $x$  gevonden kan worden: verdeel eerst het interval  $[0, 1)$  in 10 deelintervallen:

$$\left[0, \frac{1}{10}\right), \left[\frac{1}{10}, \frac{2}{10}\right), \left[\frac{2}{10}, \frac{3}{10}\right), \dots, \left[\frac{8}{10}, \frac{9}{10}\right), \left[\frac{9}{10}, 1\right).$$

Kijk in welk deelinterval  $x$  ligt: als dat het  $k$ de deelinterval is (met  $k \in \{1, 2, \dots, 10\}$ ), dan is de eerste digit  $a_1$  van  $x$  gelijk aan  $k - 1$ . De eerste benadering van  $x$  is dan  $a_1/10$ . ‘Zoom’ vervolgens in op het deelinterval waarin  $x$  ligt, en verdeel dat ook weer in 10 deelintervallen van gelijke lengte. Dan is de tweede digit  $a_2$  van  $x$  gelijk aan  $\ell - 1$ , als  $x$  ligt in het  $\ell$ de deelinterval (met natuurlijk  $\ell \in \{1, 2, \dots, 10\}$ ). De tweede benadering van  $x$  is dan  $a_1/10 + a_2/100$ . Et cetera.

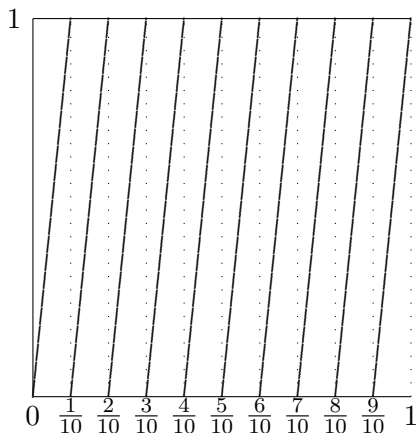
Dit is een beetje moeizame methode, die echter makkelijk te stroomlijnen is. Beschouw de afbeelding  $T_{10} : [0, 1) \rightarrow [0, 1)$ , gedefinieerd door:

$$T_{10}(x) = 10x - [10x].$$

Hierbij is  $[x]$  het grootste gehele getal dat kleiner is, of gelijk aan  $x$ ; zie ook figuur 1.

Als  $x \in [0, 1)$ , zet dan

$$a_n = a_n(x) = [10T_{10}^{n-1}(x)], \quad n \geq 1,$$



Figuur 1: De decimale afbeelding  $T_{10}$

dan zijn de digits  $a_n \in \{0, 1, 2, \dots, 9\}$  en geldt er:

$$T_{10}(x) = 10x - a_1 \quad \Leftrightarrow \quad 10x = a_1 + T_{10}(x).$$

We vinden dat

$$x = \frac{a_1}{10} + \frac{1}{10}T_{10}(x). \quad (4)$$

Als we met  $T^n(x)$  bedoelen dat we de afbeelding  $T$   $n$ -keer herhaald toepassen op  $x$  (en niet  $T(x)$  tot de macht  $n$  nemen), dan geldt er dat

$$T_{10}^2(x) = T_{10}\left(T_{10}(x)\right) = 10T_{10}(x) - a_2,$$

en vinden we dat

$$T_{10}(x) = \frac{a_2}{10} + \frac{1}{10}T_{10}^2(x). \quad (5)$$

Simpelweg invullen van (5) in (4) geeft dan:

$$\begin{aligned}x &= \frac{a_1}{10} + \frac{1}{10}T_{10}(x) \\ &= \frac{a_1}{10} + \frac{1}{10} \left( \frac{a_2}{10} + \frac{1}{10}T_{10}^2(x) \right) \\ &= \frac{a_1}{10} + \frac{a_2}{10^2} + \frac{1}{10^2}T_{10}^2(x)\end{aligned}$$

Zo doorgaand vinden we dus na  $n$  stappen dat

$$x = \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_n}{10^n} + \frac{1}{10^n}T_{10}^n(x). \quad (6)$$

Als we nu in (6) de term  $\frac{1}{10^n}T_{10}^n(x)$  ‘weggooien,’ dan krijgen we dus een rationale benadering  $r_n/s_n$  van  $x$ :

$$\frac{r_n}{s_n} = \frac{a_1}{10} + \frac{a_2}{10^2} + \cdots + \frac{a_n}{10^n} = 0.a_1a_2a_3 \cdots a_n.$$

Omdat  $T_{10}^n(x) \in [0, 1)$  volgt er direct dat

$$0 \leq \frac{1}{10^n}T_{10}^n(x) < \frac{1}{10^n} \quad \text{en} \quad 0 \leq \frac{r_n}{s_n} \leq x,$$

en we zien dat de convergenten  $r_n/s_n$  ‘van onderen’ naar  $x$  convergeren als  $n \rightarrow \infty$ :

$$x = \sum_{n=1}^{\infty} \frac{a_n}{10^n}.$$

De fout die we maken door de  $n$ de convergent  $r_n/s_n$  te nemen is dus ten hoogste  $10^{-n}$ . Merk op, dat een irrationaal getal  $x$  altijd een *unieke* decimale ontwikkeling heeft.

**Vragen:** *Laat  $p/q$  een rationaal getal zijn. Wanneer heeft  $p/q$  een eindige decimale ontwikkeling en wanneer een oneindige? Zijn deze ontwikkelingen uniek? Valt er iets te zeggen over de digits die in de ontwikkeling van  $\sqrt{2}$  voorkomen?*

## 1.2 Binaire ontwikkelingen

Merk op, dat dit het bovenstaande verhaal ook werkt als de basis niet gelijk is aan 10, maar aan 2, of 3, of 12, ... . Het bovenstaande verhaal werkt eigenlijk altijd voor gehele bases  $b \geq 2$ . Een bekend voorbeeld is de keuze van base 2 (dan krijgen we dus *binair ontwikkelingen*). In dat geval hebben we dus alleen maar digits 0 en 1, en kunnen we elk getal  $x \in [0, 1)$  schrijven als

$$x = \sum_{n=1}^{\infty} \frac{b_n}{2^n}, \quad \text{met } b_n \in \{0, 1\} \text{ voor } n \geq 1.$$

**Vraag:** *Laat  $p/q$  een rationaal getal zijn. Wanneer heeft  $p/q$  een eindige binaire ontwikkeling en wanneer een oneindige?*

De binaire ontwikkeling van een getal  $x$  tussen 0 en 1 kan gevonden worden door herhaald de afbeelding  $T_2 : [0, 1) \rightarrow [0, 1)$  te gebruiken. Deze wordt gedefinieerd door

$$T_2(x) = 2x - \lfloor 2x \rfloor = \begin{cases} 2x - 0, & x \in [0, \frac{1}{2}) \\ 2x - 1, & x \in [\frac{1}{2}, 1). \end{cases}$$

Voor een willekeurige basis  $b \in \mathbb{N}$ ,  $b \geq 2$ , kunnen we op dezelfde manier een functie  $T_b$  vinden ‘die het werk doet.’



Dit begint een beetje saai te worden; laten we daarom eens een basis  $\beta > 1$  bekijken met  $\beta \notin \mathbb{N}$ . Om de gedachten te bepalen nemen we  $\beta$  gelijk aan de *gulden snede*  $G$ , gedefinieerd door

$$G = \frac{1 + \sqrt{5}}{2} \approx 1.618033989 \dots$$

Over de gulden snede (en de bijbehorende rij van Fibonacci getallen) zijn bladzijden vol gekletst; voor een verfrissende blik, zie [12]. De gulden snede heeft een aantal opmerkelijke eigenschappen:

1. Het is een van de oplossingen van de vergelijking  $x^2 - x - 1 = 0$  (de andere oplossing is  $-1/G$ ).
2. Het is te schrijven als een kettingwortel met alleen maar 1-nen, want  $G^2 = G + 1$ :

$$G = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \dots}}}}$$

3. Het is te schrijven als een kettingbreuk<sup>2</sup> met alleen maar 1-nen, want  $G = 1 + \frac{1}{G}$ :

$$G = 1 + \frac{1}{1 + \frac{1}{1 + \dots}}$$

Net als de decimale afbeelding  $T_{10}$  en de binaire afbeelding  $T_2$  beschouwen we voor  $\beta > 1$  de afbeelding  $T_\beta : [0, 1) \rightarrow [0, 1)$ , gedefinieerd door

$$T_\beta(x) = \beta x - \lfloor \beta x \rfloor.$$

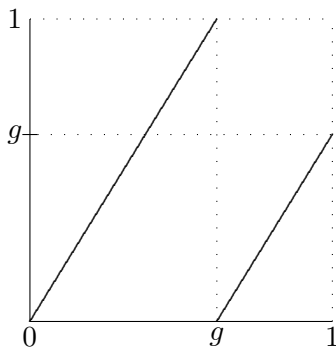
---

<sup>2</sup>We komen zo terug op kettingbreuken.

Voor  $\beta = G$  geeft dit (zie ook figuur 2):

$$T_G(x) = \begin{cases} Gx - 0, & x \in [0, g) \\ Gx - 1, & x \in [g, 1). \end{cases}$$

Hierbij is  $g = 1/G = \frac{1}{2}(\sqrt{5} - 1) = 0.618033\dots$  het ‘andere’ gulden snede getal.



Figuur 2: De afbeelding  $T_\beta$  met  $\beta$  gelijk aan de gulden snede  $G$

Net als in het binaire geval kunnen we elke  $x$  tussen 0 en 1 nu als een oneindige som ontwikkelen met alleen maar digits 0 en 1:

$$x = \sum_{n=1}^{\infty} \frac{c_n}{G^n}, \quad \text{met } c_n \in \{0, 1\} \text{ voor } n \geq 1. \quad (7)$$

Eén blik op figuur 2 leert ons dat in rijtjes digits  $c_1, c_2, \dots$  die de afbeelding  $T_G$  ‘voortbrengt’ we **nimmer** twee keer achterelkaar een digit 1 zullen zien! In het binaire geval heb je weinig kennis van kansrekening nodig om in te zien dat voor een willekeurig gekozen getal de helft van de digits een 0 is en de andere

helft een 1. De getallen  $x$  waarvoor dit niet geldt vormen een *nulverzameling*: een verzameling van (Lebesgue)maat ('lengte') 0. Maar nu is dat overduidelijk niet meer zo! Maar hoe vaak komt digit 1 dan wel voor? Helaas kunnen we hier niet op deze vraag ingaan; zie hoofdstuk 3 (en met name de secties 3.3 en 3.4) in [5] of bekijk eens het artikel [3].

Een andere belangrijke vraag is, of de ontwikkeling (7) uniek is. Het antwoord hierop is verrassend en spectaculair: op 0 en  $G$  na hebben alle getallen  $x$  tussen 0 en  $G$  **overaftelbaar** veel ontwikkelingen van de vorm (7). We zullen nu aantonen dat dit geldt voor alle  $\beta \in (1, G]$  (voor het algemene<sup>3</sup> geval  $\beta > 1$ ,  $\beta \notin \mathbb{N}$  geldt het voor bijna alle  $x$ , dus op een nulverzameling na voor alle getallen).

Laat dus  $\beta \in (1, G]$  en beschouw alle ontwikkelingen van de vorm

$$x = \sum_{n=1}^{\infty} \frac{d_n}{\beta^n}, \quad \text{met } d_n \in \{0, 1\} \text{ voor } n \geq 1. \quad (8)$$

Als we in (8) alle digits  $d_n$  gelijk aan 1 kiezen, dan krijgen we het getal

$$\mathcal{M} = \sum_{n=1}^{\infty} \frac{1}{\beta^n} = \frac{1}{\beta - 1}.$$

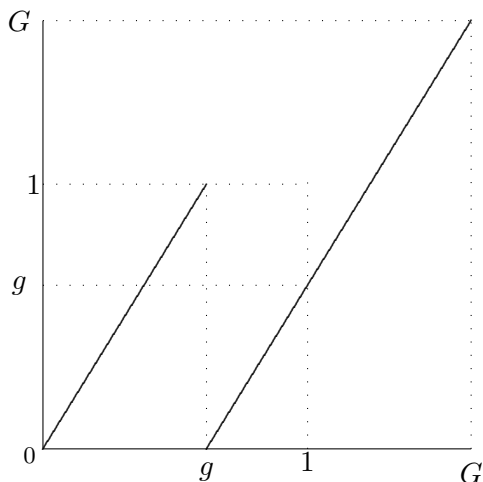
Dat elk getal tussen 0 en  $\mathcal{M}$  tenminste één ontwikkeling van de vorm (8) heeft kunnen we eenvoudig inzien door de definitie

---

<sup>3</sup>Het speciale geval dat we hier gaan bewijzen werd in 1990 verkregen door door Erdős, Joo en Komornik ([8]). Het bewijs in [8] is totaal verschillend van het bewijs dat hier gegeven wordt. Voor het algemene geval, zie [7, 13].

van  $T_\beta$  van  $[0, 1)$  naar  $[0, \mathcal{M}]$  op de voor de hand liggende manier uit te breiden (zie ook figuur 3):

$$T_\beta(x) = \begin{cases} Gx - 0, & x \in [0, 1/\beta) \\ Gx - 1, & x \in [1/\beta, \mathcal{M}]. \end{cases}$$



Figuur 3: De afbeelding  $T_\beta$  voor  $\beta = G$  op het grotere gebied  $[0, G]$

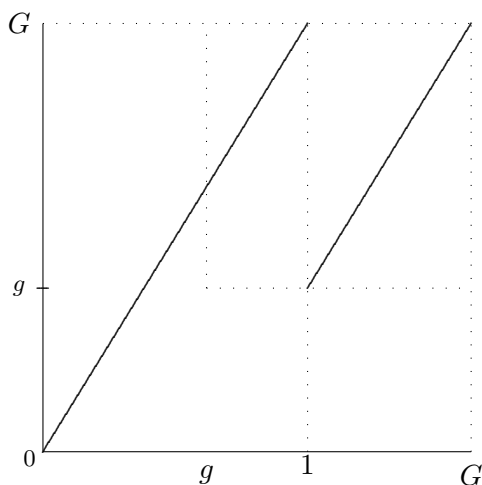
Merk op, dat het interval  $[0, 1)$  een soort ‘aantrekker’ (attractor) is; onder  $T_\beta$  ‘valt’ elke  $x \in [0, \mathcal{M})$  na eindig veel stappen in  $[0, 1)$ . Eenmaal in  $[0, 1)$  komt de baan van  $x$  onder  $T_\beta$  niet meer uit  $[0, 1)$ . Een ontwikkeling van  $x \in [0, \mathcal{M}]$  noemen we een *greedy* ontwikkeling van  $x$ : als het maar even kan kiezen

we  $d_n$  gelijk aan 1, om maar zo snel als het maar kan een ontwikkeling van  $x$  te vinden die ‘dichtbij’  $x$  is.

Als je naar figuur 3 kijkt zie je dat het ook anders zou kunnen. Definieer de *lazy* afbeelding  $S_\beta : [0, \mathcal{M}] \rightarrow [0, \mathcal{M}]$  door

$$S_\beta(x) = \begin{cases} \beta x - 0, & x \in [0, 1/(\beta(\beta - 1))] \\ Gx - 1, & x \in [1/(\beta(\beta - 1)), \mathcal{M}]; \end{cases}$$

zie ook figuur 4 voor het geval  $\beta = G$ . Deze afbeelding  $S_\beta$  is



Figuur 4: De *lazy* afbeelding  $S_G$  en haar ‘attractor’  $[g, G]$

juist zo ‘sloom’ als het maar kan; pas als het niet anders kan kiest zij  $d_n$  gelijk aan 1, anders gelijk aan 0; voor een meer concrete uitleg, zie [5, 8].

In feite hebben we op het interval  $\mathcal{S} = [1/\beta, 1/(\beta(\beta - 1))]$  de keuze uit twee afbeeldingen: de *greedy* afbeelding  $T_\beta(x) = \beta x - 1$  of de *lazy* afbeelding  $S_\beta(x) = \beta x - 0$ . Buiten het interval  $\mathcal{S}$  zijn de *greedy* afbeelding  $T_\beta$  en de *lazy* afbeelding  $S_\beta$  gelijk aan elkaar! Merk op, dat juist omdat  $\beta \in (1, G]$  er voor elke  $x \in (0, \mathcal{M}) \setminus \mathcal{S}$  er geldt dat  $T_\beta^n(x) = S_\beta^n(x) \in \mathcal{S}$ . We kunnen nu ‘linksaf gaan’ (door  $T_\beta$  te gebruiken), of ‘rechtsaf gaan’ (door  $S_\beta$  toe te passen); in beide gevallen komen we weer in  $(0, \mathcal{M}) \setminus \mathcal{S}$  terecht, zodat we na eindig vaak toepassen van  $T_\beta = S_\beta$  weer in  $\mathcal{S}$  terecht komen, en we weer een keuze moment hebben:  $T_\beta$  gebruiken, of toch liever  $S_\beta$ . Et cetera. Elke  $x \in (0, \mathcal{M})$  heeft dus overaftelbaar veel ontwikkelingen van de vorm (8).

Het voordeel van een niet-gehele basis  $\beta > 1$  boven een gehele basis  $b > 1$  is dus, dat bijna elke  $x$  heel veel ontwikkelingen heeft in die basis. Gaat de kwaliteit van de benadering ook echt omhoog? Afkappen van de ontwikkeling (8) op ‘tijdstip’  $n$  geeft een fout tussen 0 en  $\beta^{-n}$ . Dat het ook beter kan gaan we in de volgende sectie zien.

## 2 De reguliere kettingbeuk

Volgens sommigen gaan kettingbreuken terug tot de oude Grieken . . . geen idee of dit waar is, maar het is een feit dat Christiaan Huygens kettingbreuken heeft gebruikt in de constructie van een planetarium om de beste rationale benaderingen van irrationale getallen te vinden. Dit planetarium is in 1682 gemaakt. Het is tegenwoordig in het Boerhave museum te bewonderen.

## 2.1 Het algoritme van Euclides

Een van de belangrijkste en oudste algoritmen is het *algoritme van Euclides* om de grootste gemene deler (ggd) van twee gehele getallen te bepalen. Laat  $a, b \in \mathbb{Z}$  en veronderstel gemakshalve dat  $a > b > 0$ . Zet

$$r_0 = a, \quad r_1 = b,$$

en kies gehele getallen  $a_1 \geq 1, r_2 \geq 0$ , zo dat

$$r_0 = a_1 r_1 + r_2,$$

waarbij  $0 \leq r_2 < r_1$ . In het geval dat  $r_2 \neq 0$  kunnen we deze procedure herhalen. Het is duidelijk dat we na ten hoogste  $r_1$  stappen ‘tot stilstand’ komen. Er bestaat een positief geheel getal  $n$  met  $r_n \neq 0$ ,

$$r_k = a_{k+1} r_{k+1} + r_{k+2} \quad \text{voor} \quad k \leq n-1$$

en

$$0 = r_{n+1} < r_n < \dots < r_1.$$

Er geldt dan dat de *grootste gehele deler*  $\text{ggd}(a, b)$  van  $a$  en  $b$  gelijk is aan  $r_n$ .

We hebben dan, dat

$$a_1 = \left\lfloor \frac{r_0}{r_1} \right\rfloor, a_2 = \left\lfloor \frac{r_1}{r_2} \right\rfloor, \dots, a_n = \left\lfloor \frac{r_{n-1}}{r_n} \right\rfloor.$$

Definieer nu  $t_1, t_2, \dots, t_{n-1}$  door

$$x = \frac{b}{a} = \frac{r_1}{r_0}, \quad t_1 = \frac{r_2}{r_1}, \quad t_2 = \frac{r_3}{r_2}, \dots, \quad t_{n-1} = \frac{r_n}{r_{n-1}},$$

dan geldt er dat

$$\begin{aligned} \frac{1}{x} &= a_1 + t_1 \\ \frac{1}{t_1} &= a_2 + t_2 \\ &\vdots \\ \frac{1}{t_{n-2}} &= a_{n-1} + t_{n-1} \\ \frac{1}{t_{n-1}} &= a_n + 0. \end{aligned}$$

We vinden dat

$$x = \frac{1}{a_1 + t_1} = \frac{1}{a_1 + \frac{1}{a_2 + t_2}} = \dots = \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n}}}. \quad (9)$$

Een uitdrukking zoals in (9) noemen we een eindige *reguliere kettingbreuk*. Een direct gevolg van het algoritme van Euclides is dus dat elke  $x = p/q \in \mathbb{Q}$  geschreven kan worden als een eindige reguliere kettingbreuk

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n}}}, \quad (10)$$

met  $a_0 \in \mathbb{Z}$  zodat  $x - a_0 \in [0, 1)$ . We korten de kettingbreuk in (10) af door

$$[a_0; a_1, a_2, \dots, a_{n-1}, a_n].$$



Uit het algoritme van Euclides volgt bovendien dat  $a_n \geq 2$ . Maar dan heeft elk rationaal getal  $x$  dus twee eindige kettingbreuk-ontwikkelingen:

$$[a_0; a_1, a_2, \dots, a_{n-1}, a_n] = [a_0; a_1, a_2, \dots, a_{n-1}, a_n - 1, 1].$$

## 2.2 De Gauss-afbeelding $T$

Om de kettingbreuk-ontwikkeling van een willekeurig getal  $x \in [0, 1)$  te bepalen, gebruiken we de *Gauss-afbeelding*  $T : [0, 1) \rightarrow [0, 1)$ , gedefinieerd door

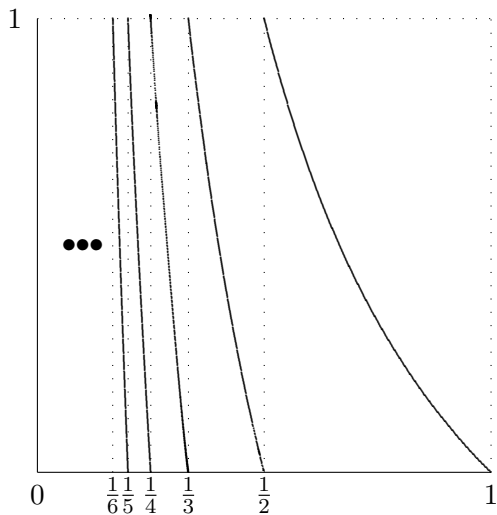
$$T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor, \quad x \neq 0; \quad T(0) = 0;$$

zie ook figuur 5. Definieer voor  $x \in [0, 1) \setminus \mathbb{Q}$  nu de digits  $a_n$  (die nu ook wel *wijzergetallen* worden genoemd) door

$$a_n = a_n(x) = \left\lfloor \frac{1}{T_{n-1}} \right\rfloor, \quad n \geq 1.$$

Dan geldt er dat

$$\begin{aligned} x &= \frac{1}{a_1 + T(x)} = \frac{1}{a_1 + \frac{1}{a_2 + T^2(x)}} = \dots = \\ &= \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n + T^n(x)}}} \quad (11) \\ &= [0; a_1, a_2, \dots, a_{n-1}, a_n + T^n(x)], \quad n \geq 1. \end{aligned}$$



Figuur 5: De (reguliere) kettingbreuk afbeelding  $T$

Merk op, dat als we in (11) de ‘restterm’  $T^n(x)$  weglaten, we een rationale benadering

$$\frac{p_n}{q_n} = [0; a_1, a_2, \dots, a_{n-1}, a_n]$$

van  $x$  vinden. Voor deze *kettingbreuk-convergenten*  $p_n/q_n$  zullen we zo laten zien dat

$$p_{-1} := 1; \quad p_0 := a_0; \quad p_n = a_n p_{n-1} + p_{n-2}, \quad n \geq 1 \quad (12)$$

$$q_{-1} := 0; \quad q_0 := 1; \quad q_n = a_n q_{n-1} + q_{n-2}, \quad n \geq 1, \quad (13)$$

dat

$$p_{n-1}q_n - p_nq_{n-1} = (-1)^n \quad (\text{en dus dat } \text{ggd}(p_n, q_n) = 1),$$

en dat (als we zetten dat  $V_n = q_{n-1}/q_n$ )

$$q_n^2 \left| x - \frac{p_n}{q_n} \right| = \frac{T^n(x)}{1 + T^n(x)V_n}, \quad n \geq 1. \quad (14)$$

Omdat  $(T^n(x), V_n) \in [0, 1] \times [0, 1]$  voor  $n \geq 0$ , volgt er nu direct uit (14) dat

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}, \quad n \geq 1, \quad (15)$$

zodat we met (13) zien dat de convergenten  $p_n/q_n$  ook echt naar  $x$  convergeren ... zelfs akelig snel!

## 2.3 Möbius transformaties

Er zijn verschillende manieren om de hierboven genoemde drie resultaten te bewijzen. Eén van de meer elegante manieren is om *Möbius transformaties* te gebruiken.

Laat  $a, b, c, d \in \mathbb{Z}$ , met  $ad - bc = \pm 1$ , en definieer de afbeelding  $A : \mathbb{R} \cup \{\infty\} \rightarrow \mathbb{R} \cup \{\infty\}$  door

$$A(x) = \frac{ax + b}{cx + d}, \quad x \in \mathbb{R} \cup \{\infty\}.$$

Als we  $A$  nu met de matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$$

identificeren, dan beschrijven we de afbeelding  $A$  (op een beetje illegale wijze) door

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} (x) = \frac{ax + b}{cx + d}, \quad x \in \mathbb{R} \cup \{\infty\}.$$

Verskillende eenvoudige eigenschappen kunnen nu afgeleid worden. Zo geldt er bijvoorbeeld voor matrices  $A, B \in \text{SL}_2(\mathbb{Z})$ , dat

$$(AB)(x) = A(B(x)),$$

waarbij  $AB$  het gebruikelijke matrix product van  $A$  en  $B$  is.

Laat nu  $x \in \mathbb{R} \setminus \mathbb{Q}$  een irrationaal getal zijn, met wijzergetallen  $a_0 \in \mathbb{Z}$  (met  $x - a_0 \in [0, 1)$ ) en  $a_n \in \mathbb{N}$  voor  $n \geq 1$ . Definieer voor  $n \geq 0$  de matrices  $A_n$  en  $M_n$  door

$$A_0 := \begin{pmatrix} 1 & a_0 \\ 0 & 1 \end{pmatrix}, \quad A_n := \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}, \quad M_n := A_0 A_1 \cdots A_n, \quad n \geq 1.$$

Dan geldt er dat

$$M_n(0) = [a_0; a_1, \dots, a_n].$$

Als we nu zetten dat

$$M_n = \begin{pmatrix} r_n & p_n \\ s_n & q_n \end{pmatrix}, \quad n \geq 0,$$

dan volgt er uit het uitschrijven van  $M_n = M_{n-1}A_n$  dat

$$M_n = \begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix}, \quad n \geq 0,$$

en dus dat

$$\begin{aligned} p_{n-1}q_n - p_nq_{n-1} &= \det(M_n) = (-1)^n, \quad n \geq 1, \quad \text{en} \\ \frac{p_n}{q_n} &= M_n(0) = [a_0; a_1, \dots, a_n], \quad n \geq 1. \end{aligned}$$

Bovendien volgen de recurrente relaties (12) en (13) ook direct uit  $M_n = M_{n-1}A_n$ :

$$\begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} = \begin{pmatrix} p_{n-2} & p_{n-1} \\ q_{n-2} & q_{n-1} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_n \end{pmatrix}.$$

Definieer tenslotte de matrix  $M_N^*$  door

$$M_n^* = \begin{pmatrix} p_{n-2} & p_{n-1} \\ q_{n-2} & q_{n-1} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & a_n + T^n(x) \end{pmatrix}.$$

Dan geldt er dat (zie ook (11));

$$M_n^*(0) = \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_n + T^n(x)}}} = x,$$

maar ook, als we het matrix-product uitschrijven en de recurrente relaties (12) en (13) gebruiken, dat

$$M_n^* = \begin{pmatrix} p_{n-1} & p_n + p_{n-1}T^n(x) \\ q_{n-1} & q_n + q_{n-1}T^n(x) \end{pmatrix}.$$

Maar dan zien we dus, dat

$$x = M_n^*(0) = \frac{p_n + p_{n-1}T^n(x)}{q_n + q_{n-1}T^n(x)}, \quad n \geq 1,$$

en we vinden nu direct voor  $n \geq 1$  dat

$$x - \frac{p_n}{q_n} = \frac{p_n + p_{n-1}T^n(x)}{q_n + q_{n-1}T^n(x)} - \frac{p_n}{q_n} = \frac{(-1)^n T^n(x)}{q_n(q_n + q_{n-1}T^n(x))}, \quad (16)$$

waaruit (14) direct volgt. In feite was (14) één van de startpunten van de moderne theorie van de metrische eigenschappen van kettingbreuken, die in het begin van de tachtiger jaren van de vorige eeuw begon met de artikelen [11, 4, 9]; zie ook [6] voor meer achtergrondinformatie en bewijzen.

## 2.4 Waarom koos Huygens voor kettingbreuken?

Uit (13) en (15) volgt dat de convergentie van  $p_n/q_n$  naar  $x$  heel snel is; de rij  $(q_n)_{n \geq 1}$  is ‘op z’n langzaamst’ als  $x = g$ , dus wanneer alle wijzergetallen gelijk aan 1 zijn. In dat geval geldt er voor elke  $n \geq 1$  dat  $q_n$  gelijk aan het  $n$ de Fibonacci-getal  $F_n$ . Jammer genoeg hebben we geen tijd om een klein beetje ergoden theorie te doen, want anders zouden we met redelijk eenvoudige middelen kunnen zien zien dat voor bijna alle  $x$  er geldt dat

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log q_n = \frac{\pi^2}{12 \log 2} = 1.18656911 \dots$$

Dit resultaat van Paul Lévy uit 1929 laat zien dat de rij  $(q_n)_{n \geq 1}$  inderdaad zeer snel (exponentieel!) stijgt voor een ‘typische  $x$ .’ Lévy bewees trouwens ook in 1926 dat er voor bijna alle  $x$  er geldt dat

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left| x - \frac{p_n}{q_n} \right| = \frac{-\pi^2}{6 \log 2} = -2.373138221 \dots$$

zie ook hoofdstuk 3 in [5].

Uit de definitie van  $V_n$  en uit (13) volgt er verder dat

$$V_n = \frac{q_{n-1}}{q_n} = \frac{q_{n-1}}{a_n q_{n-1} + q_{n-2}} = \frac{1}{a_n + V_{n-1}}.$$

Omdat er per definitie geldt dat

$$T^n(x) = T\left(T^{n-1}(x)\right) = \frac{1}{T^{n-1}(x)} - a_n,$$

zien we dat

$$T^{n-1}(x) = \frac{1}{a_n + T^n(x)}.$$

Als we dus in (13) het ‘tijdstip’  $n$  door  $n - 1$  vervangen, dan vinden we voor  $n \geq 2$  dat

$$q_{n-1}^2 \left| x - \frac{p_{n-1}}{q_{n-1}} \right| = \frac{T^{n-1}(x)}{1 + T^{n-1}(x)V_{n-1}} = \frac{V^n(x)}{1 + T^n(x)V_n}. \quad (17)$$

Definieer nu voor  $n \geq 1$  de *approximatie coëfficiënten*  $\Theta_n = \Theta_n(x)$  door

$$\Theta_n(x) = \frac{T^n(x)}{1 + T^n(x)V_n}.$$

Herinner dat we eerder hebben gebruikt dat voor  $n \geq 0$  en alle  $x \in [0, 1)$  er geldt dat  $(T^n(x), V_n) \in [0, 1) \times [0, 1]$ . Een korte elementaire berekening geeft dan dat voor  $n \geq 1$  er geldt dat

$$\Theta_{n-1}(x) + \Theta_n(x) = \frac{T^n(x) + V_n}{1 + T^n(x)V_n} \leq 1.$$

We vinden op deze manier een oud resultaat van Vahlen<sup>4</sup> uit 1895: voor elk irrationaal getal  $x$  en voor alle  $n \geq 1$  geldt er dat

$$\min\{\Theta_{n-1}, \Theta_n\} < \frac{1}{2}.$$

Op zich is Vahlen’s resultaat niet zo spectaculair. Er zijn sterkere resultaten, die met het hierboven ontwikkelde instrumentarium makkelijk bewezen kunnen worden. Zo geldt er voor alle irrationale getallen  $x$  en alle  $n \geq 2$  dat  $\min\{\Theta_{n-1}, \Theta_n, \Theta_{n+1}\} <$

---

<sup>4</sup>Vahlen is naast zijn hier genoemde resultaat over approximatie coëfficiënten ook om zijn politieke carrière in Nazi Duitsland bekend geworden; zie bijvoorbeeld de artikelen [14, 15].

$1/\sqrt{5}$ . Maar Vahlen's resultaat geeft een zekere urgentie aan het volgende resultaat van Legendre uit 1808: als  $x$  een irrationaal getal is,  $p/q$  een rationaal getal met  $q > 0$  en  $\text{ggd}(p, q) = 1$ , dan volgt er uit

$$q^2 \left| x - \frac{p}{q} \right| < \frac{1}{2}$$

dat  $p/q$  een reguliere kettingbreuk convergent is van  $x$ ; er bestaat een  $n \geq 1$  zo dat  $p = p_n$  en  $q = q_n$  (zie [2] voor het bewijs van een iets scherpere versie van Legendre's resultaat).

Het resultaat van Legendre zegt dus, dat als je een irrationaal getal  $x$  echt goed benaderd door een rationaal getal  $p/q$ , er geldt dat  $p/q$  een kettingbreuk convergent van  $x$  is. Vahlen zegt vervolgens dat er heel veel van dergelijke goede benaderingen zijn. Immers, minstens de helft van alle approximatie coëfficiënten  $\Theta_n$  zijn kleiner dan  $1/2$ . Met behulp van ergoden theorie kan dit nog verder aangescherpt worden: voor bijna alle  $x$  geldt er dat

$$\lim_{n \rightarrow \infty} \frac{1}{n} |\{1 \leq j \leq n : \Theta_j(x) < \frac{1}{2}\}| = \frac{1}{2 \log 2} = 0.72134752 \dots$$

(zie sectie 5.3.2 in [5]), dus voor een 'typische'  $x$  is iets meer dan 72% van de approximatie coëfficiënten kleiner dan  $1/2$ .

We sluiten af met een eigenschap, die Huygens nodig had in de constructie van zijn planetarium: de kettingbreuk convergenten zijn de *beste benaderingsbreuken*. Hierbij is het rationale getal  $a/b$  een beste benaderingsbreuk van  $x \in \mathbb{R} \setminus \mathbb{Q}$  is, als  $b > 0$  en als voor alle rationale getallen  $c/d \neq a/b$  met  $0 < d \leq b$  er geldt dat  $|dx - c| > |ax - b|$ ; zie ook sectie 6 in [10].



Merk eerst op, dat uit (16) volgt dat de convergenten  $p_n/q_n$  van  $x$  ‘om-en-om’  $x$  springen:

$$a_0 = \frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \dots < x < \dots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}. \quad (18)$$

Veronderstel eerst, dat  $a/b$  (met  $b > 0$ ) een beste benaderingsbreuk is. We willen dus aantonen dat  $a/b$  een convergent is. Dan moet er wel gelden dat  $a/b \geq a_0$ , want als  $a/b < a_0$  vinden we dat

$$\left| 1 \cdot x - a_0 \right| < \left| x - \frac{a}{b} \right| \leq b \left| x - \frac{a}{b} \right| = \left| bx - a \right|,$$

en we krijgen een tegenspraak, omdat  $a/b$  dan **geen** beste benaderingsbreuk is. Stel nu, dat  $a/b$  geen convergent is. Dan is wegens (18) òf  $a/b > p_1/q_1$  òf bestaat er wegens (18) een  $n \geq 1$  zo dat  $a/b$  tussen  $p_{n-1}/q_{n-1}$  en  $p_n/q_n$  ligt.

In het *eerste geval* ( $a/b > p_1/q_1$ ) geldt er dat

$$\left| x - \frac{a}{b} \right| > \left| \frac{p_1}{q_1} - \frac{a}{b} \right| = \frac{|bp_1 - aq_1|}{bq_1} \geq \frac{1}{bq_1},$$

(hierbij gebruiken we dat  $bp_1 - aq_1$  een geheel getal verschillend van 0 is), en dus dat

$$|bx - a| > \frac{1}{q_1} = \frac{1}{a_1}.$$

Omdat

$$\left| 1 \cdot x - a_0 \right| < \frac{1}{a_1}$$

volgt er direct dat

$$|bx - a| > |1 \cdot x - a_0|,$$

en we hebben opnieuw een tegenspraak.

In het *tweede geval* ( $a/b$  ligt tussen  $p_{n-1}/q_{n-1}$  en  $p_n/q_n$  voor een  $n \geq 1$ ) volgt er ook een tegenspraak. Immers,

$$\left| \frac{a}{b} - \frac{p_{n-1}}{q_{n-1}} \right| = \frac{|aq_{n-1} - bp_{n-1}|}{bq_{n-1}} \geq \frac{1}{bq_{n-1}}$$

en

$$\left| \frac{a}{b} - \frac{p_{n-1}}{q_{n-1}} \right| < \left| \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} \right| = \frac{1}{q_n q_{n-1}},$$

zodat we vinden dat  $b > q_n$ . Verder geldt er, dat

$$\left| x - \frac{a}{b} \right| > \left| \frac{p_{n+1}}{q_{n+1}} - \frac{a}{b} \right| \geq \frac{1}{bq_{n+1}}.$$

Maar dan geldt er dat

$$|bx - a| > \frac{1}{q_{n+1}}.$$

Uit (16) volgt dat

$$|q_n x - p_n| \leq \frac{1}{q_{n+1}},$$

zodat we dus vinden dat

$$|q_n x - p_n| \leq |bx - a|.$$

Maar dan volgt uit dit en  $b > q_n$  een tegenspraak;  $a/b$  is geen beste benaderingsbreuk.

We zien dus dat elke beste benaderingsbreuk een convergent is. Voor het bewijs dat elke convergent een beste benaderingsbreuk is verwijs ik de lezer naar Stelling 17 in [10].

### 3 Kettingbreuken zijn ‘het beste’?

Er zijn heel veel verschillende manieren om getallen te ontwikkelen: naast de decimale-,  $\beta$ - en (reguliere) kettingbreukontwikkelingen zijn er Lüroth reeksen, Engel reeksen, Bolyai ontwikkelingen ... Kunnen we beter doen dan de reguliere kettingbreuk? Het antwoord is ja! De optimale kettingbreuk! Maar dat is weer een heel ander verhaal . . . . . (voor de liefhebber, lees [1] er eens op na).

### Referenties

- [1] Bosma, Wieb, *Optimal continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 49 (1987), no. 4, 353-379.
- [2] Barbolosi, Dominique; Jager, Hendrik, *On a theorem of Legendre in the theory of continued fractions*, J. Théor. Nombres Bordeaux 6 (1994), no. 1, 81-94.
- [3] Barrionuevo, Jose; Burton, Robert M.; Dajani, Karma; Kraaikamp, Cor, *Ergodic properties of generalized Lroth series*, Acta Arith. 74 (1996), no. 4, 311-327.
- [4] Bosma, W.; Jager, H.; Wiedijk, F., *Some metrical observations on the approximation by continued fractions*, Nederl. Akad. Wetensch. Indag. Math. 45 (1983), no. 3, 281-299.
- [5] Dajani, Karma; Kraaikamp, Cor, *From greedy to lazy expansions and their driving dynamics*, Expo. Math. 20 (2002), no. 4, 315-327.

- [6] Dajani, Karma; Kraaikamp, Cor, *Ergodic theory of numbers*, Carus Mathematical Monographs, 29. Mathematical Association of America, Washington, DC, 2002.
- [7] Dajani, Karma; de Vries, Martijn, *Invariant densities for random  $q$ -expansions*, J. Eur. Math. Soc. (JEMS) 9 (2007), no. 1, 157-176.
- [8] Erdős, Pál; Joó, István; Komornik, Vilmos, *Characterization of the unique expansions  $1 = \sum_{i=1}^{\infty} q^{-n_i}$  and related problems*, Bull. Soc. Math. France 118 (1990), no. 3, 377-390.
- [9] Knuth, Donald E., *The distribution of continued fraction approximations*, J. Number Theory 19 (1984), no. 3, 443-448.
- [10] Khintchine, A. Ya., *Continued Fractions*, P. Noordhoff Ltd., Groningen, 1963.
- [11] Nakada, Hitoshi, *Metrical theory for a class of continued fraction transformations and their natural extensions*, Tokyo J. Math. 4 (1981), no. 2, 399-426.
- [12] Van der Schoot, Albert, *Het is niet alles goud wat er snijdt*, Nieuw Arch. Wiskd. (5) 10 (2009), no. 3, 203-207.
- [13] Sidorov, Nikita, *Almost every number has a continuum of  $\beta$ -expansions*, Amer. Math. Monthly 110 (2003), no. 9, 838-842.
- [14] Siegmund-Schultze, Reinhard, *Einige Probleme der Geschichtsschreibung der Mathematik im faschistischen*

*Deutschland—unter besonderer Berücksichtigung des Lebenslaufes des Greifswalder Mathematikers Theodor Vahlen*, Contributions to the history, philosophy and methodology of mathematics (Greifswald, 1982). Wiss. Z. Greifswald. Ernst-Moritz-Arndt-Univ. Math.-Natur. Reihe 33 (1984), no. 1–2, 51–56.

- [15] Siegmund-Schultze, Reinhard, *Theodor Vahlen—zum Schuldanteil eines deutschen Mathematikers am faschistischen Missbrauch der Wissenschaft*, NTM Schr. Geschichte Natur. Tech. Medizin 21 (1984), no. 1, 17–32.

# Eigenwaarden van random matrices

Arno B.J. Kuijlaars, KU Leuven, België  
e-mail: arno.kuijlaars@wis.kuleuven.be

## 1 Inleiding

Een eigenwaarde van een vierkante matrix  $A$  is een reëel of complex getal  $\lambda$  waarvoor een niet-nul vector  $\vec{x}$  bestaat met

$$A\vec{x} = \lambda\vec{x}.$$

Een  $n \times n$  matrix heeft  $n$  eigenwaarden als we eigenwaarden tellen aan de hand van hun multipliciteit. De verzameling van eigenwaarden heet het spectrum van  $A$ .

Een random matrix (of toevalsmatrix) is een matrix waarvan de elementen niet vastliggen maar door het toeval bepaald worden. De eigenwaarden van zo'n random matrix liggen dan ook niet vast en zijn stochastische veranderlijken. Het blijkt dat

eigenwaarden van grote random matrices een mooie structuur hebben die in veel gevallen expliciet beschreven kan worden.

We zullen dit uitleggen voor matrices waarvan de elementen een normale (of Gaussische) verdeling hebben. De kansdichtheid van een standaardnormale verdeling is

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Dit betekent dat de kans dat het element  $A_{i,j}$  van de matrix  $A$  tot het interval  $[a, b]$  behoort gelijk is aan de integraal

$$\mathbb{P}[A_{i,j} \in [a, b]] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx.$$

De normale verdeling treedt op in de centrale limietstelling die zegt dat de som van een groot aantal onafhankelijke en identiek verdeelde stochastische veranderlijken met eindige variantie bij benadering normaal verdeeld is. De limietverdelingen voor random matrices hebben een soortgelijk universeel karakter.

Een random matrix met normaal verdeelde elementen kan eenvoudig gegenereerd worden met een softwarepakket zoals Matlab of Maple. Een voorbeeld van een  $5 \times 5$  matrix is

$$A = \begin{pmatrix} 1.770 & 0.7135 & -2.204 & 1.777 & -0.1129 \\ 0.4162 & 0.6670 & 1.324 & -0.7980 & 1.230 \\ -0.1726 & 2.297 & -0.4617 & -1.332 & 0.1165 \\ -0.6706 & -0.7177 & -0.9024 & 1.269 & 0.3204 \\ 0.7931 & -0.2685 & 0.5650 & -0.6322 & -1.269 \end{pmatrix}.$$

De eigenwaarden van deze matrix zijn (correct tot op drie decimalen),

$$2.511 \pm 1.031i, \quad -1.554 \pm 0.6403i, \quad 0.594.$$

Er zijn in dit voorbeeld één reële eigenwaarde en vier niet-reële die voorkomen in twee complex geconjugeerde paren.

De getransponeerde matrix  $A^T$  is de matrix die uit  $A$  ontstaat door de rijen en de kolommen te verwisselen. Voor het bovenstaande voorbeeld geldt

$$A^T = \begin{pmatrix} 1.770 & 0.4162 & -0.1726 & -0.6706 & 0.7931 \\ 0.7135 & 0.6670 & 2.297 & -0.7177 & -0.2685 \\ -2.204 & 1.324 & -0.4617 & -0.9024 & 0.5650 \\ 1.777 & -0.7980 & -1.332 & 1.269 & -0.6322 \\ -0.1129 & 1.230 & 0.1165 & 0.3204 & -1.269 \end{pmatrix}.$$

Een matrix is symmetrisch als  $A = A^T$ . De bovenstaande matrix is niet symmetrisch maar het gemiddelde

$$M = \frac{1}{2}(A + A^T) = \begin{pmatrix} 1.770 & 0.565 & -1.188 & 0.554 & 0.340 \\ 0.565 & 0.667 & 1.811 & -0.758 & 0.481 \\ -1.188 & 1.811 & -0.462 & -1.117 & 0.341 \\ 0.554 & -0.758 & -1.117 & 1.269 & -0.156 \\ 0.340 & 0.481 & 0.341 & -0.156 & -1.269 \end{pmatrix}$$

is dat wel. De eigenwaarden van een symmetrische matrix zijn altijd reëel. Inderdaad geldt dat de eigenwaarden van de matrix  $M$  gelijk zijn aan

$$-2.252, \quad -1.395, \quad 0.261, \quad 2.046, \quad 3.314.$$

De matrix  $M$  die we gemaakt hebben is een matrix uit het Gaussisch Orthogonaal Ensemble (GOE) van afmeting  $5 \times 5$ . Als we waren begonnen met een complexe matrix  $A$  met complexwaardige normaal verdeelde elementen die onderling onafhankelijk zijn en we vormen

$$M = \frac{1}{2} \left( A + \overline{A}^T \right) \tag{19}$$



dan krijgen we een Hermitische matrix uit het Gaussisch Unitair Ensemble (GUE). De streep in  $\overline{A}^T$  duidt op het nemen van de complex geconjugeerde van elk element van  $A$ . Het GOE en GUE zijn de twee belangrijke modellen van random matrices.

## 2 Limietgedrag

### 2.1 Globaal limietgedrag

Interessant limietgedrag treedt op als de afmetingen van de matrices naar oneindig gaan. Voor een GOE of GUE matrix van afmeting  $n \times n$  zijn de eigenwaarden reëel. Laten we deze noteren we met

$$\lambda_{1,n} < \lambda_{2,n} < \dots < \lambda_{n,n}. \quad (20)$$

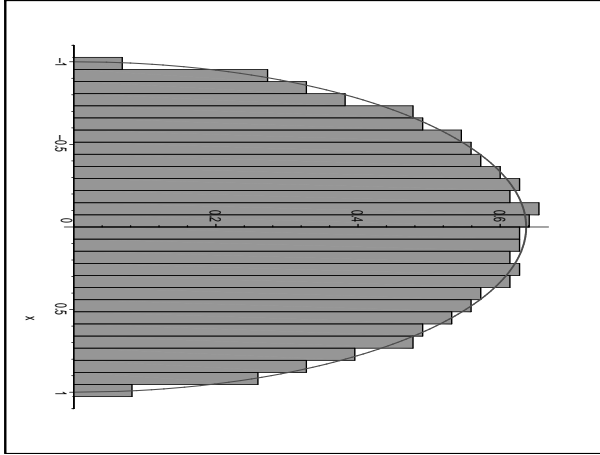
Dit zijn dan  $n$  punten op de reële rechte. Naarmate  $n$  toeneemt krijgen we steeds meer eigenwaarden die zich op een bepaalde manier gaan verdelen over de reële rechte. Het blijkt dat de grootste eigenwaarde  $\lambda_{n,n}$  van de GUE matrix (als we die maken volgens de manier die hierboven beschreven is) grofweg groeit als  $2\sqrt{n}$  en de kleinste eigenwaarde  $\lambda_{1,n}$  daalt als  $-2\sqrt{n}$ . Dat wil zeggen dat alle eigenwaarden zich (met grote kans) bevinden in het interval

$$[-2\sqrt{n}, 2\sqrt{n}].$$

Als we delen door  $2\sqrt{n}$  dan krijgen we herschaalde eigenwaarden in het interval  $[-1, 1]$  die we noteren met

$$x_{k,n} = \frac{\lambda_{k,n}}{2\sqrt{n}}, \quad k = 1, 2, \dots, n. \quad (21)$$

Een histogram van de  $x_{k,n}$ 's ziet er uit als in Figuur 1 voor  $n = 1000$ .



Figuur 6: Histogram van 1000 herschaalde eigenwaarden van een GUE matrix.

We zien in Figuur 1 inderdaad dat vrijwel alle herschaalde eigenwaarden zich bevinden in het interval  $[-1, 1]$ . We zien ook de grafiek van de functie

$$p_{sc}(x) = \frac{2}{\pi} \sqrt{1 - x^2}, \quad -1 < x < 1 \quad (22)$$

die bekend staat als de dichtheid van de halve cirkelwet. De herschaalde eigenwaarden verdelen zich op  $[-1, 1]$  volgens deze halve cirkelwet en de overeenstemming wordt beter naarmate  $n$  groter wordt. Precieser uitgedrukt betekent dit dat met kans

1 geldt dat

$$\lim_{n \rightarrow \infty} \frac{\text{aantal herschaalde eigenwaarden van } M \text{ in } [a, b]}{n} = \frac{2}{\pi} \int_a^b \sqrt{1 - x^2} dx$$

voor elke  $a, b \in [-1, 1]$ . De halve cirkelwet werd het eerst beschreven door de bekende fysicus Eugene Wigner (1902-1995, van oorsprong Hongaar, maar werkzaam in VS, Nobelprijs in 1963).

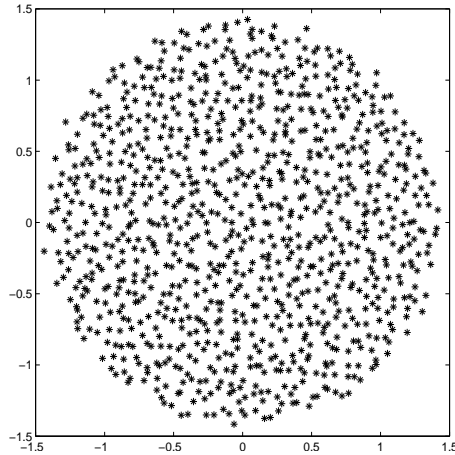
De halve cirkelwet blijkt universeel te zijn. Ze treedt niet alleen op bij symmetrische en Hermitische matrices met normaal verdeelde elementen, maar ook bij andere verdelingen. Belangrijk is wel dat de elementen onderling onafhankelijk zijn, of maar een zwakke afhankelijkheid hebben. Voor random matrices met sterk afhankelijke elementen geldt de halve cirkelwet niet.

Voor een niet-symmetrische matrix  $A$  zijn de eigenwaarden niet noodzakelijk reëel, maar algemeen complex. In dat geval is er ook een opmerkelijk limietgedrag van de eigenwaarden. De eigenwaarden blijken namelijk uniform verdeeld te zijn op een schijf in het complexe vlak zoals in Figuur 2 te zien is. De punten in Figuur 2 zijn weer herschaalde eigenwaarden waarbij de eigenwaarden door een factor  $\sqrt{n}$  gedeeld zijn.

## 2.2 Lokaal limietgedrag

Het hierboven beschreven limietgedrag betreft het globale, collectieve gedrag van alle eigenwaarden tesamen.

Het lokale gedrag gaat over de invloed van naburige eigenwaarden op elkaar. Hierbij gaat het niet om alle eigenwaarden maar alleen om een beperkt aantal eigenwaarden die dicht bij elkaar



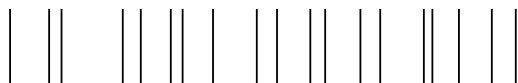
Figuur 7: Plot van herschaalde eigenwaarden van een complexe matrix met onderling onafhankelijke complex normaal verdeelde elementen.

liggen. Bij onafhankelijk getrokken punten uit een zekere verdeling heeft het aanwezig zijn van een punt in een bepaalde positie geen invloed op de positie van andere punten. Dit is precies waarom de punten onafhankelijk zijn. Bij onafhankelijke punten kunnen we een patroon verwachten zoals in Figuur 3.



Figuur 8: Plot van een aantal onderlinge onafhankelijke punten gekozen uit een halve cirkelverdeling. De plot toont niet alle punten maar alleen diegene in een zeker interval rond 0.

We zien in de figuur dat sommige punten heel dicht bij elkaar liggen, terwijl er daarentegen ook relatief grote gaten in het spectrum optreden. Bij eigenwaarden van random matrices treedt dat niet op. Voor eigenwaarden van random matrices is er een afstoting tussen naburige eigenwaarden waardoor de kans op hele kleine afstanden tussen eigenwaarden heel klein is. Dat geeft een meer regelmatig patroon waarbij ook veel grotere afstanden dan normaal niet voorkomen. Na uitvergroting zien de eigenwaarden van een grote GUE matrix rond 0 er als volgt uit



Figuur 9: Plot van een aantal eigenwaarden van een grote GUE matrix. De plot toont niet alle punten maar alleen diegene in een zeker interval rond 0.

Het lokale gedrag van eigenwaarden, zoals te zien is in Figuur 4, is de reden dat random matrices voor het eerst bestudeerd werden in de fysica. In de jaren 50 zocht Wigner naar een model om de energieniveaus van zware atoomkernen te beschrijven. Exacte berekeningen hierrond zijn niet mogelijk en Wigner zocht naar een statistische theorie die een typische situatie zou kunnen beschrijven. Wigner bedacht dat de hoge energieniveaus van zware atoomkernen op een goede manier benaderd zouden kunnen worden met de eigenwaarden van een grote random matrix. Hierbij gaat het niet om het globale gedrag van alle eigenwaarden, maar om het lokale gedrag van naburige eigenwaarden. Deze gedachte bleek heel goed te kloppen. Uit experimentele data van uranium-isotopen bleek dat de ener-

gieniveaus zich inderdaad gedragen volgens Figuur 4 en niet volgens Figuur 3.

Eén van de karakteristieken voor het lokale gedrag is de afstand tussen naburige eigenwaarden. In het Engels heet dit “nearest neighbour spacing”. In een GOE of GUE matrix met eigenwaarden  $\lambda_{1,n} < \lambda_{2,n} < \dots < \lambda_{n,n}$  liggen de middelste eigenwaarden  $\lambda_{\lfloor n/2 \rfloor}$  en  $\lambda_{\lfloor n/2 \rfloor + 1}$  dicht bij 0. Vanwege de halve cirkelwet voor de herschaalde eigenwaarden is de verwachte afstand tussen deze twee eigenwaarden ongeveer

$$\lambda_{\lfloor n/2+1 \rfloor, n} - \lambda_{\lfloor n/2 \rfloor, n} \approx \frac{2\sqrt{n}}{p_{sc}(0)n} = \frac{\pi}{\sqrt{n}}. \quad (23)$$

Na vermenigvuldiging met  $\frac{\sqrt{n}}{\pi}$  krijgen we bijgevolg een verwachte afstand van ongeveer 1. Hierover bestaat een limietstelling.

**Stelling 2.1.** *Zowel voor GOE als voor GUE is er een limietverdeling voor*

$$\frac{\sqrt{n}}{\pi} (\lambda_{\lfloor n/2 \rfloor + 1} - \lambda_{\lfloor n/2 \rfloor})$$

als  $n \rightarrow \infty$ . De limietverdelingen zijn verschillend voor GOE en GUE.

Er zijn dus limietverdelingen  $F_{GOE}(x)$  en  $F_{GUE}(x)$  met de eigenschap dat voor elke  $s > 0$  geldt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \frac{\sqrt{n}}{2\pi} (\lambda_{\lfloor n/2 \rfloor + 1} - \lambda_{\lfloor n/2 \rfloor}) \leq s \right] = \begin{cases} F_{GOE}(s), & \text{voor GOE,} \\ F_{GUE}(s), & \text{voor GUE,} \end{cases}$$

met bijbehorende kansdichtheden  $p_{GOE} = \frac{dF_{GOE}}{ds}$  en  $p_{GUE} = \frac{dF_{GUE}}{ds}$ .

Wigner voorspelde op grond van zekere symmetrieën dat

$$p_{GOE}(s) \approx \frac{\pi}{2} s e^{-\pi s^2/4} \quad \text{en} \quad p_{GUE}(s) \approx \frac{32}{\pi^2} s^2 e^{-4s^2/\pi} \quad (24)$$

een goede beschrijving zou zijn voor de functies  $p_{GOE}$  en  $p_{GUE}$ . Vanwege de factoren  $s$  en  $s^2$  in (24) is de kans klein dat eigenwaarden heel dicht bij elkaar liggen. Dit geeft de lokale afstoting van eigenwaarden en verklaart het regelmatige patroon van eigenwaarden. De afstoting is groter bij GUE dan bij GOE. Deze “Wigner surmise” bleek echter niet juist te zijn. De formules (24) zijn goede benaderingen, maar ze kloppen niet exact. We zullen in (38) en (39) de juist formules zien voor  $F_{GUE}$  en  $p_{GUE}$ . Deze uitdrukkingen zijn wel een flink stuk ingewikkelder.

Merk op dat de constanten in de formules precies zo gekozen zijn dat

$$\int_0^\infty p(x) dx = 1 \quad \text{en} \quad \int_0^\infty xp(x) dx = 1$$

voor zowel  $p(x) = p_{GOE}(x)$  als  $p(x) = p_{GUE}(x)$ .

### 2.3 De Riemann zetafunctie

In de jaren 70 werd door Freeman Dyson (van oorsprong Brits fysicus, werkzaam in Princeton, VS, geboren in 1923) en Hugh Montgomery (Amerikaans wiskundige, geboren in 1944) een verband ontdekt tussen eigenwaarden van random matrices en de nulpunten van de Riemann zetafunctie. We gaan hier niet heel diep op in.

De Riemann zetafunctie is voor  $s > 1$  gedefinieerd door de reeks

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s} \quad (25)$$

Als we ook complexe waarden van  $s$  toelaten, dan is de zetafunctie door de reeks gedefinieerd voor  $\Re s > 1$ . Voor andere waarden is de reeks divergent. Vanwege de unieke ontbinding van een natuurlijk getal als product van priemgetallen geldt

$$\zeta(s) = \prod_{p \text{ priem}} (1 - p^{-s})^{-1}$$

waarbij het product genomen wordt over alle priemgetallen. Dit verband met de priemgetallen verklaart waarom de zetafunctie van belang is in de getaltheorie. Bepaalde eigenschappen van de zetafunctie geven aanleiding tot zekere gedragingen van de priemgetallen.

De Riemann zetafunctie heeft een analytische voortzetting tot het complexe vlak met een enkelvoudige pool in  $s = 1$ . Deze voortzetting heeft nulpunten in  $s = 0, -1, -2, \dots$ . Dit zijn triviale nulpunten. Daarnaast zijn er nog oneindig veel nulpunten in de verticale strip

$$0 < \Re s < 1 \quad (26)$$

in het complexe vlak. Dit zijn de niet-triviale nulpunten van de Riemann zetafunctie.

De Riemannhypothese zegt dat alle niet-triviale nulpunten zich bevinden op de rechte  $\Re s = 1/2$ . Dus volgens deze hypothese liggen de nulpunten allemaal precies in het midden van de strip (26). De Riemannhypothese werd in 1859 geformuleerd door Bernhard Riemann (bekend Duits wiskundige, 1826–1866). De Riemannhypothese is een belangrijk open



probleem in de wiskunde. Het is één van de 7 Clay Millenniumproblemen waarmee een bedrag van 1 miljoen dollar te winnen is (en eeuwige roem) voor diegene die het oplost.

Als we de Riemannhypothese voor waar aannemen dan kunnen we ons in de volgende stap gaan afvragen hoe de nulpunten van  $\zeta(s)$  op de rechte  $\Re s = 1/2$  zich ten opzichte van elkaar gedragen. Het idee van Dyson is dat naburige nulpunten hoog op de kritieke rechte zich gedragen als de eigenwaarden van een GUE random matrix. Dat wil zeggen dat als we een plot maken van een aantal naburige nulpunten  $s = \frac{1}{2} + i\tau$  van  $\zeta(s)$  met groot imaginair deel  $\tau = \text{Im } s$ , en als we op een goede manier herschalen, we een verdeling van punten krijgen die er uit ziet als in Figuur 4 (en zeker niet als in Figuur 3).

De nulpunten van de Riemann zetafunctie zijn met grote precisie bekend. Sinds eind jaren 80 heeft Andrew Odlyzko (van oorsprong Pools wiskundige, werkzaam in VS, geboren in 1949) nulpunten berekend en is gekomen tot maar liefst het  $10^{22}$ ste nulpunt op de kritieke rechte. Een statistische analyse van de afstanden tussen miljoenen opeenvolgende nulpunten rond het  $10^{22}$ ste nulpunten laat het patroon van Figuur 4 zien. Er is bovendien een verbluffend nauwkeurige overeenkomst met de kansverdeling  $p_{GUE}$  die komt uit de random matrix theorie. De overeenstemming is significant beter dan met de benadering (24) van Wigner.

Van dit alles is uiteraard niets strikt bewezen. Ook de Riemannhypothese blijft tot op heden een open probleem. Men heeft de hoop gehad om de Riemannhypothese te bewijzen met een combinatie van getaltheorie en technieken uit de random matrix theorie, maar ook dit heeft tot nu toe tot niets geleid.

### 3 Wiskundige technieken

We kunnen hier niet ingaan op alle wiskundige details, maar we willen wel een overzicht geven van de belangrijkste stappen en ideeën die leiden tot de limietverdeling  $p_{GUE}$  van (23) voor de afstanden tussen naburige GUE eigenwaarden. Het is een interessante combinatie van kansrekening, combinatoriek en asymptotiek van speciale functies, met ook aspecten van functionaalanalyse en differentiaalvergelijkingen.

#### 3.1 Kansdichtheid voor eigenwaarden

De eerste stap in de wiskundige behandeling van de eigenwaarden van een GOE of GUE random matrix is een opmerkelijke expliciete uitdrukking voor de gezamenlijke kansdichtheid

$$\frac{1}{Z_{n,\beta}} \prod_{1 \leq i < j \leq n} |\lambda_j - \lambda_i|^\beta \prod_{j=1}^n e^{-\frac{1}{2}\lambda_j^2} \quad (27)$$

van de eigenwaarden. Hierin is  $\beta = 1$  voor GOE en  $\beta = 2$  voor GUE. De factor  $Z_{n,\beta}$  is een normalisatieconstante die ervoor zorgt dat het een kansdichtheid is.

Het product  $\prod_{1 \leq i < j \leq n} |\lambda_j - \lambda_i|^\beta$  wordt 0 als  $\lambda_i = \lambda_j$ , hetgeen betekent dat de kans op samenvallende eigenwaarden gelijk aan nul is. Deze factor is verantwoordelijk voor de afstoting tussen naburige eigenwaarden en de afstoting wordt groter naarmate

$\beta$  groter is. Merk op dat

$$\prod_{1 \leq i < j \leq n} (\lambda_j - \lambda_i) = \begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{n-1} \\ 1 & \lambda_3 & \lambda_3^2 & \cdots & \lambda_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{vmatrix}$$

vanwege een bekende formule van de determinant van een Vandermonde matrix. Door gebruik te maken van eigenschappen van determinanten kan men nog meer relevante informatie uit de kansverdeling halen.

## 3.2 Hermitepolynomen

De formules worden het mooist voor  $\beta = 2$ , d.w.z. voor eigenwaarden van een GUE matrix. We beperken ons vanaf nu tot dit geval. Voor  $\beta = 2$  is de gemiddelde karakteristieke polynoom

$$\begin{aligned} P_n(x) &= \mathbb{E}[\det(xI - M)] \\ &= \frac{1}{Z_{n,2}} \int_{\mathbb{R}^n} \prod_{j=1}^n (x - \lambda_j) \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^2 \prod_{j=1}^n e^{-\frac{1}{2}\lambda_j^2} d\lambda_1 \cdots d\lambda_n \end{aligned} \quad (28)$$

een polynoom van graad  $n$  met de bijzondere eigenschap dat

$$\int_{-\infty}^{\infty} P_n(x) P_m(x) e^{-\frac{1}{2}x^2} dx = 0, \quad \text{voor } n \neq m.$$

De polynomen die aan deze eigenschap voldoen staan bekend als Hermitepolynomen die genoemd zijn naar de Franse wiskundige Charles Hermite (1822–1901). Er is een expliciete

formule voor  $P_n$

$$P_n(x) = (-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} \left( e^{-\frac{1}{2}x^2} \right).$$

Alle nulpunten van  $P_n$  bevinden zich in het interval  $[-2\sqrt{n}, 2\sqrt{n}]$ . Na deling door  $2\sqrt{n}$  volgen de nulpunten hetzelfde patroon als de GUE eigenwaarden. Ze zijn ook verdeeld volgens de halve cirkelwet, zie histogram van Figuur 1, en kunnen gezien worden als een typische verdeling van de eigenwaarden van een GUE matrix. Figuur 5 geeft de grafiek van de Hermitefunctie

$$\varphi_k(x) = \frac{1}{(2\pi)^{\frac{1}{4}} k!} e^{-\frac{1}{4}x^2} P_k(x)$$

van graad  $k = 40$ . De constante  $\frac{1}{(2\pi)^{\frac{1}{4}} k!}$  is zodanig gekozen dat

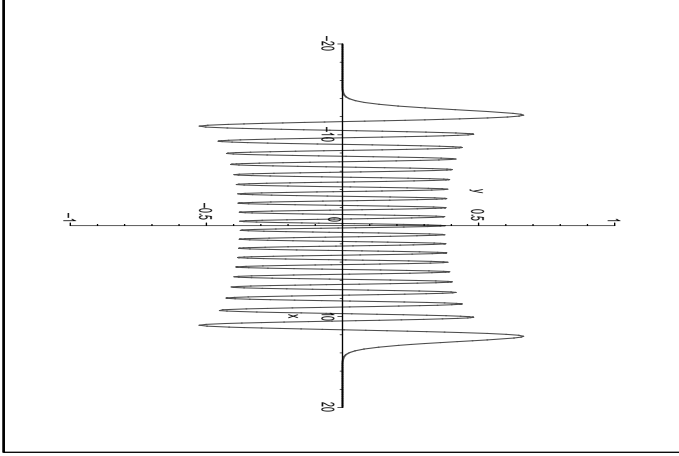
$$\int_{-\infty}^{\infty} \varphi_k(x) \varphi_l(x) dx = \delta_{k,l} = \begin{cases} 0 & \text{als } k \leq l, \\ 1 & \text{als } k = l. \end{cases} \quad (29)$$

De Hermitefuncties zijn een orthonormale basis van de Hilbertruimte van kwadratisch integreerbare functies op  $\mathbb{R}$ . In de quantummechanica treden de Hermitefuncties op als eigenfuncties van de quantum harmonische oscillator.

### 3.3 Correlatiefuncties

De verdere analyse maakt gebruik van de volgende functie van twee veranderlijken

$$K_n(x, y) = \sum_{k=0}^{n-1} \varphi_k(x) \varphi_k(y) \quad (30)$$



Figuur 10: Grafiek van de Hermitefunctie van graad 40. De nulpunten van de Hermitefunctie verdelen zich bij toenemende graad volgens de halve cirkelwet en volgen dus dezelfde verdeling als de eigenwaarden van een GUE matrix.

waarin de Hermitefuncties  $\varphi_k$  optreden. Vanwege eigenschappen van determinanten kan men de gezamenlijke kansdichtheid (27) als een  $n \times n$  determinant

$$\frac{1}{n!} \det [K_n(\lambda_i, \lambda_j)]_{i,j=1}^n$$

schrijven. Dit is de kansdichtheid voor alle eigenwaarden tezamen.

Hieruit krijgt men marginale dichtheden door een aantal veranderlijken uit te integreren. Door gebruik te maken van de orthogonaliteit (29) van de Hermitefuncties kan men laten zien dat de marginale dichtheden ook gegeven worden door deter-

minanten, maar van kleinere afmeting. In statistische fysica spreekt men van correlatiefuncties in plaats van marginale dichtheden (deze twee zijn gelijk op een factor na). De  $k$  punts correlatiefunctie is gelijk aan de  $k \times k$

$$\det [K_n(\lambda_i, \lambda_j)]_{i,j=1}^k \quad (31)$$

met daarin de functie  $K_n$ , die daarom de correlatiekern genoemd wordt.

Voor  $k = 1$  is (31) gelijk aan de correlatiekern op de diagonaal  $K_n(x, x)$  hetgeen de verwachte dichtheid van eigenwaarden geeft. Dat wil zeggen dat

$$\int_a^b K_n(x, x) d\lambda$$

gelijk is aan het verwachte aantal eigenwaarden in het interval  $[a, b]$ . De 2-punts correlatiefunctie is een  $2 \times 2$  determinant

$$\rho_2(\lambda_1, \lambda_2) = \det \begin{pmatrix} K_n(\lambda_1, \lambda_1) & K_n(\lambda_1, \lambda_2) \\ K_n(\lambda_2, \lambda_1) & K_n(\lambda_2, \lambda_2) \end{pmatrix} = K_n(\lambda_1, \lambda_1)K_n(\lambda_2, \lambda_2) - K_n(\lambda_1, \lambda_2)K_n(\lambda_2, \lambda_1)$$

De dubbele integraal hiervan (we gebruiken  $x_1$  en  $x_2$  als integratieveranderlijken)

$$\int_a^b \int_c^d \rho_2(x_1, x_2) dx_1 dx_2$$

geeft het verwachte aantal paren  $(\lambda, \mu)$  van eigenwaarden met  $\lambda \in [a, b]$  en  $\mu \in [c, d]$ . Hierin is verondersteld dat de intervallen  $[a, b]$  en  $[c, d]$  disjunct zijn.

### 3.4 Gatkansen

Andere relevante grootheden kan men uitdrukken in de correlatiefuncties (31).

Een voorbeeld hiervan zijn de “gatkansen” (in het Engels “gap probabilities”). Voor een interval  $[0, s]$  is dat de kans dat er geen eigenwaarde is in  $[0, s]$ , met andere woorden de kans dat  $[0, s]$  een gat is in het spectrum van de random matrix,

$$E_n([0, s]) = \mathbb{P}(\not\exists \lambda \in [0, s]).$$

Deze kans kan uitgerekend worden met het principe van inclusie-exclusie uit de combinatoriek. Het blijkt dat  $E_n([0, s])$  gelijk is aan een alternerende som

$$\begin{aligned} E_n([0, s]) = & 1 - \int_0^s K_n(x, x) dx + \frac{1}{2!} \int_0^s \int_0^s \det \begin{pmatrix} K_n(x_1, x_1) & K_n(x_1, x_2) \\ K_n(x_2, x_1) & K_n(x_2, x_2) \end{pmatrix} dx_1 dx_2 \\ & - \frac{1}{3!} \int_0^s \int_0^s \int_0^s \det \begin{pmatrix} K_n(x_1, x_1) & K_n(x_1, x_2) & K_n(x_1, x_3) \\ K_n(x_2, x_1) & K_n(x_2, x_2) & K_n(x_2, x_3) \\ K_n(x_3, x_1) & K_n(x_3, x_2) & K_n(x_3, x_3) \end{pmatrix} dx_1 dx_2 dx_3 + \dots \end{aligned} \quad (32)$$

De term  $\int_0^s K_n(x, x) dx$  is gelijk aan het verwachte aantal eigenwaarden in  $[0, s]$ . Als er niet meer dan één eigenwaarde in dit interval kan zijn, dan is dit ook gelijk aan de kans op een interval in  $[0, s]$ , en dan zou  $E_n([0, s])$  gelijk zijn aan  $1 - \int_0^s K_n(x, x) dx$ .

Als er meer dan één eigenwaarde in  $[0, s]$  kan zijn, dan moeten we hiervoor een correctie toepassen. De volgende term in (32) is gelijk aan het verwachte aantal paren van eigenwaarden in  $[0, s]$ . De factor  $\frac{1}{2!}$  staat er omdat we een paar  $(\lambda, \mu)$  en  $(\mu, \lambda)$

maar één keer willen meetellen. Als er niet meer dan twee eigenwaarden in  $[0, s]$  kunnen voorkomen, dan is de gatkans precies gelijk aan de som van de eerste drie termen in (32). Als  $n \geq 3$  dan kunnen er wel drie of meer eigenwaarden voorkomen (hoewel de kans wellicht klein is als  $s$  niet al te groot is) en dan moet er weer een correctie plaatsvinden. Als we zo doorgaan vinden we uiteindelijk de som.

De alternerende som (32) is een Fredholmdeterminant en wordt genoteerd als

$$E_n([0, s]) = \det(I - K_n)_{[0, s]}. \quad (33)$$

Hierbij identificeert men  $K_n$  met een lineaire integraaloperator

$$f \mapsto K_n f, \quad (K_n f)(x) = \int_0^s K_n(x, y) f(y) dy.$$

die beperkt wordt tot functies op  $[0, s]$ . De Fredholmdeterminant is een manier om een determinant van een matrix uit te breiden naar een determinant van een lineaire operator tussen oneindig dimensionale ruimten. Fredholmdeterminanten zijn genoemd naar Erik Ivar Fredholm (Zweeds wiskundige, 1866–1927).

### 3.5 Limiet van correlatiekern

Het interessante gedrag treedt op als we de limiet  $n \rightarrow \infty$  nemen. Zoals we al gezien hebben moeten we dan wel steeds een herschaling uitvoeren. In het algemeen geldt dat als we de eigenwaarden herschalen met een factor  $a_n$ , dat we dan de



correlatiekern  $K_n$  moeten vervangen door

$$\frac{1}{a_n} K_n \left( \frac{x}{a_n}, \frac{y}{a_n} \right).$$

De globale kansdichtheid voor een enkele eigenwaarde is gelijk aan  $\frac{1}{n} K_n(\lambda, \lambda)$ . Vanwege (21) is de schalingfactor gelijk aan  $a_n = \frac{1}{2\sqrt{n}}$ . Na herscaling krijgen we bijgevolg

$$\frac{2}{n^{1/2}} K_n (2\sqrt{n}x, 2\sqrt{n}x)$$

De halve cirkelwet (22) krijgen we als de volgende limiet

$$\lim_{n \rightarrow \infty} \frac{2}{n^{1/2}} K_n (2\sqrt{n}x, 2\sqrt{n}x) = \begin{cases} \frac{2}{\pi} \sqrt{1-x^2}, & \text{als } -1 < x < 1 \\ 0, & \text{anders.} \end{cases}$$

In het lokale regime rond 0 moeten we een schalingsfactor  $a_n = \frac{\sqrt{n}}{\pi}$  hanteren, zie (23), zodat de herschaalde kern gelijk is aan

$$\frac{\pi}{\sqrt{n}} K_n \left( \frac{\pi x}{\sqrt{n}}, \frac{\pi y}{\sqrt{n}} \right).$$

Deze herschaalde kern blijkt de volgende limiet te hebben als  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \frac{\pi}{\sqrt{n}} K_n \left( \frac{\pi x}{\sqrt{n}}, \frac{\pi y}{\sqrt{n}} \right) = \frac{\sin \pi(x-y)}{\pi(x-y)}, \quad (34)$$

hetgeen een niet-triviaal resultaat is over het asymptotische gedrag van de Hermitefuncties  $\varphi_k$  als  $k \rightarrow \infty$ . Het heeft te maken met de gelijkmatige oscillaties van de Hermitefuncties

rond de oorsprong, zoals te zien is in Figuur 5. De limiet die optreedt in (34) noemt men de sinus kern

$$K_{\sin}(x, y) = \frac{\sin \pi(x - y)}{\pi(x - y)}.$$

De sinus kern is kenmerkend voor het lokale limietgedrag van eigenwaarden van Hermitische matrices. Merk op dat

$$K_{\sin}(x, x) = \lim_{y \rightarrow x} K_{\sin}(x, y) = 1.$$

### 3.6 Limiet van gatkansen

Uit (34) volgt het volgende voor de limiet van de gatkansen.

$$\lim_{n \rightarrow \infty} E_n([0, \frac{\pi s}{\sqrt{n}}]) = \det(I - K_{\sin})_{[0, s]} \quad (35)$$

waarin de limiet de Fredholmdeterminant is van de sinus kern

$$\begin{aligned} & \det(I - K_{\sin})_{[0, s]} \\ &= 1 - \int_0^s K_{\sin}(x, x) dx + \frac{1}{2!} \int_0^s \int_0^s \det \begin{pmatrix} K_{\sin}(x_1, x_1) & K_{\sin}(x_1, x_2) \\ K_{\sin}(x_2, x_1) & K_{\sin}(x_2, x_2) \end{pmatrix} dx_1 dx_2 \\ & - \frac{1}{3!} \int_0^s \int_0^s \int_0^s \det \begin{pmatrix} K_{\sin}(x_1, x_1) & K_{\sin}(x_1, x_2) & K_{\sin}(x_1, x_3) \\ K_{\sin}(x_2, x_1) & K_{\sin}(x_2, x_2) & K_{\sin}(x_2, x_3) \\ K_{\sin}(x_3, x_1) & K_{\sin}(x_3, x_2) & K_{\sin}(x_3, x_3) \end{pmatrix} dx_1 dx_2 dx_3 \\ & \quad + \dots \quad (36) \end{aligned}$$

In tegenstelling tot (32) is (36) wel een oneindige som.

### 3.7 Verdeling van afstanden

Het is nog een stap om over te gaan van de gatkansen naar de afstanden tussen opeenvolgende eigenwaarden. Om dat laatste te berekenen nemen we aan dat er een eigenwaarde is in 0 en we zijn geïnteresseerd in de verdeling van de eerstvolgende eigenwaarde. Hier gebeurt weer iets moois. Bij conditioneren op een eigenwaarde in 0 blijft de determinantale structuur behouden. We moeten alleen overgaan naar een nieuwe correlatiekern. In het limietgeval  $n \rightarrow \infty$  betekent dat we overgaan van de sinuskern (34) naar

$$L_{\sin}(x, y) = \det \begin{pmatrix} K_{\sin}(x, y) & K_{\sin}(x, 0) \\ K_{\sin}(0, y) & K_{\sin}(0, 0) \end{pmatrix} = \frac{\sin \pi(x-y)}{\pi(x-y)} - \frac{\sin \pi x}{\pi x} \cdot \frac{\sin \pi y}{\pi y}. \quad (37)$$

Voor  $y = x$  krijgen we

$$L_{\sin}(x, x) = 1 - \left( \frac{\sin \pi x}{\pi x} \right)^2.$$

De Fredholmdeterminant  $\det(I - L_{\sin})_{[0,s]}$  geeft nu de voorwaardelijke kans dat er (na herscaling van eigenwaarden en in de limiet  $n \rightarrow \infty$ ) geen verdere eigenwaarde is in het interval  $[0, s]$  onder de aanname dat er een eigenwaarde is in 0. De verdelingsfunctie voor de afstand tussen naburige eigenwaarden

is dan

$$\begin{aligned}
 F_{GUE}(s) &= 1 - \det(I - L_{\sin})_{[0,s]} \\
 &= \int_0^s L_{\sin}(x, x) dx - \frac{1}{2!} \int_0^s \int_0^s \det \begin{pmatrix} L_{\sin}(x_1, x_1) & L_{\sin}(x_1, x_2) \\ L_{\sin}(x_2, x_1) & L_{\sin}(x_2, x_2) \end{pmatrix} dx_1 dx_2 \\
 &+ \frac{1}{3!} \int_0^s \int_0^s \int_0^s \det \begin{pmatrix} L_{\sin}(x_1, x_1) & L_{\sin}(x_1, x_2) & L_{\sin}(x_1, x_3) \\ L_{\sin}(x_2, x_1) & L_{\sin}(x_2, x_2) & L_{\sin}(x_2, x_3) \\ L_{\sin}(x_3, x_1) & L_{\sin}(x_3, x_2) & L_{\sin}(x_3, x_3) \end{pmatrix} dx_1 dx_2 dx_3 \\
 &\quad - \dots \quad (38)
 \end{aligned}$$

met bijbehorende kansdichtheid

$$p_{GUE}(s) = \frac{dF_{GUE}(s)}{ds}. \quad (39)$$

De formules (38)–(39) zijn de correcte uitdrukkingen voor de “Wigner surmise” (24).

### 3.8 Painlevé voorstelling

De (38) ziet er nogal overweldigend uit. Toch zit er nog een bijzondere structuur achter de Fredholmdeterminant die een meer expliciete voorstelling van  $p_{GUE}$  mogelijk maakt. Het blijkt namelijk dat (36) en (38) als functie van  $s$  aan speciale niet-lineaire differentiaalvergelijkingen voldoen. Dit zijn differentiaalvergelijkingen die bekend staan als Painlevévergelijkingen.

We zullen hier niet diep op in gaan. Maar één van de resultaten is dat de  $p_{GUE}$  kansdichtheid (39) ook gegeven wordt door

$$p_{GUE}(s) = \frac{\pi^2}{3} s^2 \exp \left( \int_0^{2\pi s} \frac{v(t)}{t} dt \right)$$

waarin  $v$  een oplossing is van de niet-lineaire differentiaalvergelijking

$$(sv'')^2 + (v - sv')(v - sv' + 4 - 4(v')^2) - 16(v')^2 = 0 \quad (40)$$

die voldoet aan de voorwaarde

$$v(s) = -\frac{1}{15}s^2 + \dots \quad \text{als } s \rightarrow 0.$$

De differentiaalvergelijking (40) is een vorm van de derde Painlevévergelijking  $P_{III}$ .

De Painlevévergelijkingen zijn genoemd naar de Franse wiskundige en politicus Paul Painlevé (1863–1933). Painlevé was in 1917 en in 1925 twee keer kort premier van Frankrijk.

# Traitor Tracing

**Boris Škorić, Technische Universiteit Eindhoven**

**e-mail: `b.skoric@tue.nl`**

## 1 Forensisch watermerken en coalitie-aanvallen

Het is in de informatica een bekend verschijnsel dat een algoritme soms veel efficiënter gemaakt kan worden als je een (kleine) kans is op fouten toestaat. Een recent voorbeeld is te vinden in het vakgebied van het ‘forensisch watermerken’, ook bekend onder de naam ‘traitor tracing’.

Distributeurs van films willen graag in staat zijn om te traceren waar de bron ligt van ongeautoriseerde her-distributie, bijv. op peer-to-peer netwerken. Een van de technologieën die hiervoor ingezet kan worden is digitaal watermerken: bij iedere geautoriseerde ontvanger van de film wordt er een unieke code verstoppt in de film, zodanig dat de beeldkwaliteit er niet onder lijdt. Zo’n watermerk is niet eenvoudig te verwij-

deren of te veranderen zonder kennis over de geheime sleutels die gebruikt zijn. Echter, als een aantal mensen samen (een ‘coalitie’) hun exemplaren van de film vergelijken dan wordt het een stuk makkelijker. Het ontwerpen en analyseren van coalitie-resistente codes is een niet-triviaal deelgebied van de coderingstheorie.

Voor het gemak wordt meestal aangenomen dat een film  $m$  segmenten bevat waarin een  $q$ -air symbool verstopt kan worden, en dat de coalitie alleen iets kan uitrichten daar waar ze een verschil waarnemen. Verder kiest men vaak voor een aanvalsmodel waarin de coalitie slechts mag kiezen uit de symbolen die zij heeft ontvangen.

Aanvaller #1	a	a	c	c
#2	b	c	c	a
#3	b	a	c	a
#4	a	b	c	a
Toegestaan	a/b	a/b/c	c	a/c

Figuur 11: *Het gebruikelijke model voor coalitie-aanvallen op watermerken. Voorbeeld met vier aanvallers en alfabet  $\{a, b, c\}$ .*

Deterministische algoritmes, waarbij de code een algebraïsche structuur heeft en gepoogd wordt de coalitie met 100% zekerheid te identificeren, presteren bijzonder slecht: ze kunnen slechts kleine coalities weerstaan, hebben zeer lange codes nodig, of vereisen een groot alfabet. Ter illustratie beschouwen we een binaire code en twee aanvallers. In dit geval is het *onmogelijk* om een code te verzinnen waarmee je met 100%

zekerheid 1 van de aanvallers kunt aanwijzen. We kijken naar een sequentie van drie bits. Stel drie mensen ontvangen als watermerk  $(1, 0, 0)$ ,  $(0, 1, 0)$  respectievelijk  $(0, 0, 1)$ , en het watermerk na de aanval is  $\vec{y} = (0, 0, 0)$ . Dan zijn alle combinaties van twee van die drie personen is staat om  $\vec{y}$  te produceren; niemand kan met zekerheid aangewezen worden als dader.

De best bekende deterministische code tegen  $c$  aanvallers (die niet absurd lang is) heeft  $q \geq c^2$  nodig. Let wel, in de context van video-watermerken wordt  $q$  groter dan ongeveer 16 reeds als onrealistisch beschouwd! De lengte van deze code is  $m = \mathcal{O}(c^2 \ln n)$ , waar  $n$  het aantal personen is. Met andere woorden, als je  $m$  of meer segmenten tot je beschikking hebt om te watermerken, dan kan je minstens 1 van de  $c$  aanvallers traceren.

Probabilistische codes zijn in staat coalitie-resistentie te bieden bij een soortgelijke code-lengte, maar met veel kleiner alfabet, bijv. binair. We zullen nu de best bekende code presenteren tegen algemene  $c$ : de code van Tardos [6] voor algemeen alfabet  $\mathcal{Q}$  [7] (met  $|\mathcal{Q}| = q$ ).

## 2 De Tardos code

De codewoorden worden gegenereerd d.m.v. twee gerandomiseerde stappen. Eerst wordt voor ieder segment  $j \in \{1, \dots, m\}$  een ‘voorkeursvector’  $\vec{p}_j \in [0, 1]^q$  getrokken volgens een Dirichletkansverdeling  $F$ ,

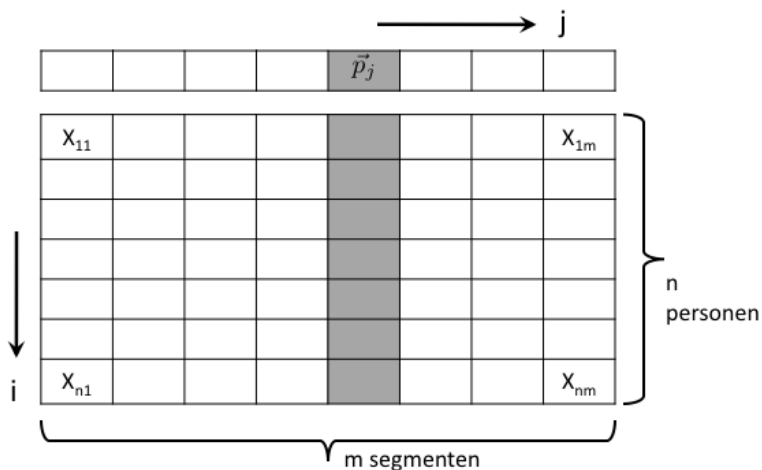
$$F(\vec{p}) \propto \prod_{\alpha \in \mathcal{Q}} p_{\alpha}^{-1+\kappa}, \quad (41)$$



met  $\kappa > 0$  een constante en  $\sum_{\alpha} p_{\alpha} = 1$ . Daarna wordt voor elke persoon  $i \in \{1, \dots, n\}$  het  $q$ -aire symbool in segment  $j$  ( $X_{ij} \in \mathcal{Q}$ ) getrokken volgens de kansverdeling

$$\mathbb{P}[X_{ij} = \alpha] = (p_j)_{\alpha}. \quad (42)$$

Hier staat  $(p_j)_{\alpha}$  voor de  $\alpha$ -component van de vector  $\vec{p}_j$ . We kunnen de codewoorden zien als rijvectoren, en de codewoorden van alle personen onder elkaar zetten; dan krijgen we een matrix  $X$  (zie Fig. 12). Merk op dat de kans in (42) niet van de persoon-index  $i$  afhangt, zodat de  $j$ 'de kolom van  $X$  geheel gevuld wordt met behulp van alleen maar de voorkeursvector  $\vec{p}_j$ .



Figuur 12: Voorkeursvectoren  $\vec{p}_j$  en codewoorden-matrix  $X$ .

Wanneer er een ongeautoriseerde kopie van de film opduikt, met daarin watermerk  $\vec{y} = (y_1, \dots, y_m)$  dan is de volgende

stap om met behulp van de beschikbare informatie, namelijk de  $\vec{p}$  vectoren,  $X$  en  $\vec{y}$ , de aanvallers te traceren. Hiertoe wordt voor iedere persoon  $i$  een score  $S_i$  uitgerekend,

$$S_i = \sum_{j=1}^m S_{ij} \quad ; \quad S_{ij} = \begin{cases} g_1((p_j)_{y_j}) & \text{als } X_{ij} = y_j \\ g_0((p_j)_{y_j}) & \text{als } X_{ij} \neq y_j \end{cases} \quad (43)$$

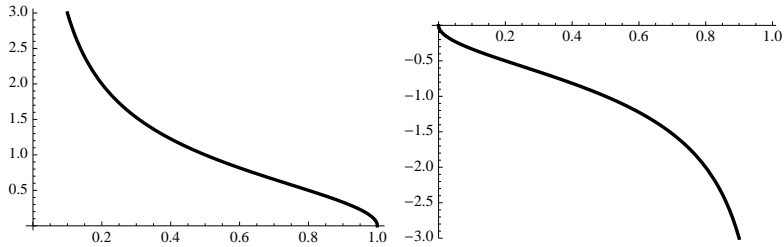
$$g_1(p) = \sqrt{\frac{1-p}{p}} \quad ; \quad g_0(p) = -\sqrt{\frac{p}{1-p}}. \quad (44)$$

Als persoon  $i$  in segment  $j$  hetzelfde symbool  $y_j$  heeft dat in de kopie zit, dan krijgt hij een positieve score; anders een negatieve score. De sommatie in (43) stelt een soort correlatie-som voor tussen  $\vec{y}$  en het codewoord van persoon  $i$ . De  $g_0$  en  $g_1$  functie zorgen voor een speciale weging van de termen:

- In het geval  $X_{ij} = y_j$  is de score  $S_{ij}$  hoger naarmate het symbool  $y_j$  onwaarschijnlijker is.
- In het geval  $X_{ij} \neq y_j$  is de score negatiever naarmate  $y_j$  *waarschijnlijker* is (en daardoor  $X_{ij}$  onwaarschijnlijker).

Op deze manier wordt de score nauwelijks beïnvloed door zeer waarschijnlijke gebeurtenissen en juist sterk beïnvloed door onwaarschijnlijke gebeurtenissen.

Personen van wie de score boven een bepaalde grens  $Z$  uitkomt worden als verdacht beschouwd. Er is een kans dat iemand die onschuldig is toch in de verdachtenlijst belandt. Die kans noemen we  $P_{FP}$  (FP='false positive'). We willen graag dat  $P_{FP} \leq \varepsilon_1$ , met  $n\varepsilon_1 \ll 1$ , zodat de totale kans op valse beschuldigingen klein is.



Figuur 13: De gewichtsfuncties  $g_1(p)$  en  $g_0(p)$ .

Er is ook een kans dat geen van de aanvallers een score heeft groter dan  $Z$ . In dat geval spreken we van een ‘false negative’ fout, waarvoor we de notatie  $P_{\text{FN}}$  gebruiken. We willen dat  $P_{\text{FN}} \leq \varepsilon_2$ , met  $\varepsilon_2 < 1$ . Doorgaans hoeft  $\varepsilon_2$  lang niet zo klein te zijn als  $n\varepsilon_1$ . Zelfs van bijvoorbeeld  $\varepsilon_2 \approx 0.5$  kan een voldoende dreigende werking uitgaan.

### 3 Eigenschappen van de Tardos code

De methode om  $X$  te maken, plus het score-systeem, worden samen ‘de Tardos code’ genoemd. De Tardos code heeft twee eigenschappen waardoor de analyse van het hele systeem sterk vereenvoudigd wordt.

Ten eerste is zowel het genereren van de codewoorden als het bepalen van de scores *segment-symmetrisch*, d.w.z. gebeurtenissen in het ene segment trekken zich niets aan van wat er in de overige segmenten gebeurt. Hierdoor heeft het voor de aan-

vallers geen zin om af te wijken van die symmetrie<sup>5</sup> en kan de analyse van het systeem worden gereduceerd tot een analyse in één segment.

Ten tweede zijn de gewichtsfuncties zo gekozen dat de score  $S_{ij}$  van een onschuldige gemiddeld nul is, met variantie 1,

$$\begin{aligned} \text{Gemiddelde:} \quad & p_y g_1(p_y) + (1 - p_y) g_0(p_y) = 0 \\ \text{Variantie:} \quad & p_y [g_1(p_y)]^2 + (1 - p_y) [g_0(p_y)]^2 = 1. \end{aligned} \quad (45)$$

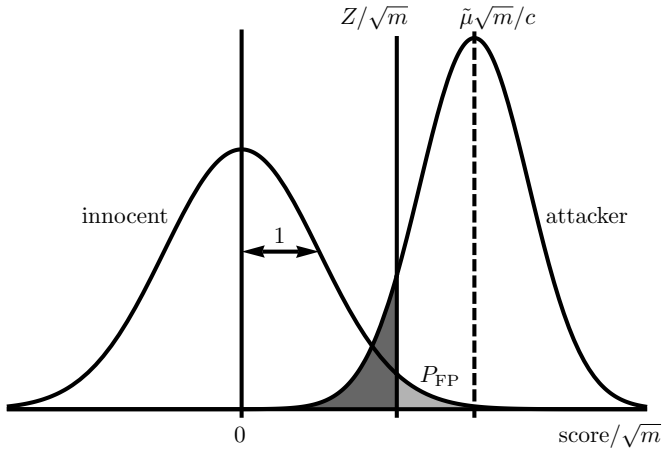
Dit leidt er direct toe dat de stochast  $S_i/\sqrt{m}$ , voor onschuldige  $i$ , verwachtingswaarde nul heeft en variantie 1. We kunnen dan een schets maken als in Fig. 14. Op de horizontale as staat score/ $\sqrt{m}$ . De linkercurve is de kansdichtheid voor de score  $S_i/\sqrt{m}$  van een onschuldige persoon; de rechtercurve hoort bij een aanvaller, en is op een iets ingewikkeldere manier gedefinieerd. We definiëren  $\tilde{\mu}$  als de verwachtingswaarde van de score van de hele coalitie gedeeld door  $m$ , d.w.z. de verwachte score van de hele coalitie in één segment. Dan is  $\tilde{\mu}\sqrt{m}/c$  de verwachte score van één aanvaller, genormeerd met de factor  $1/\sqrt{m}$  van de horizontale as. De  $P_{\text{FP}}$  is gelijk aan het oppervlak onder de linkercurve rechts van  $Z$ ; de  $P_{\text{FN}}$  is ongeveer gelijk aan het oppervlak onder de rechtercurve links van  $Z$ .

Uit Fig. 14 zijn direct enkele belangrijke conclusies te trekken:

1. Het blijkt zo te zijn dat de strategie van de coalitie nauwelijks invloed heeft op de vorm van de linkercurve. (De aanvalsstrategie, d.w.z. de keuze van  $\vec{y}$ , heeft natuurlijk wel een groot effect op de vorm van de rechtercurve.) Verder blijkt de afmeting van de coalitie ( $c$ ) weinig invloed te hebben op het getal  $\tilde{\mu}$ . Wat gebeurt er nu als,

---

<sup>5</sup>Dit is niet eenvoudig te bewijzen [4].



Figuur 14: *Kansdichtheid van de score voor een onschuldige en een aanvaller.*

bij vaste grenswaarde  $Z$ , het aantal aanvallers toeneemt? De linkercurve verandert nauwelijks, en de rechtercurve schuift in zijn geheel naar links. Door  $c$  maar groot genoeg te maken kan  $P_{\text{FN}}$  willekeurig dicht bij de 100% komen. Echter: de  $P_{\text{FP}}$  wordt niet groter! Deze eigenschap, dat een willekeurig grote coalitie geen onschuldigen kan vernachelen, is heel belangrijk en heet ‘**no framing**’. Als de coalitie groter is dan waar de code op berekend was (keuze van  $m$  en  $Z$ ), dan is de lijst met verdachten leeg, wat niet fijn is, maar er verschijnen ook geen onschuldigen in de lijst!

2. Als  $m$  en  $Z$  groter worden gemaakt bij constante ver-

houding  $Z/\sqrt{m}$ , dan verandert de  $P_{\text{FP}}$  nauwelijks. (De vorm van de linkercurve kan een beetje veranderen.) De rechter kansverdeling schuift naar rechts, waardoor de  $P_{\text{FN}}$  kleiner wordt.

3. Definieer de kans in de rechterstaart van de onschuldige als  $P_{\text{FP}} = G_1(Z/\sqrt{m})$ , met  $G_1$  de (complementaire) cumulatieve kansverdeling. Definieer op analoge wijze de kans in de linkerstaart van de schuldige als  $G_2(\frac{1}{\sigma}[Z/\sqrt{m} - \tilde{\mu}\sqrt{m}/c])$ , waarbij  $\sigma^2$  de variantie is van de rechter verdeling. Uit de eis dat deze kansen  $\varepsilon_1$  resp.  $\varepsilon_2$  zijn volgt dan een formule voor  $m$  en  $Z$  als functie van  $\varepsilon_1$  en  $\varepsilon_2$ ,

$$m = \frac{c^2}{\tilde{\mu}^2} [G_1^{\text{inv}}(\varepsilon_1) - \sigma G_2^{\text{inv}}(\varepsilon_2)]^2 \quad ; \quad Z = \sqrt{m} G_1^{\text{inv}}(\varepsilon_1) \quad (46)$$

Merk op dat  $G_2^{\text{inv}}(\varepsilon_2)$  negatief is, en dat typisch de  $G_1^{\text{inv}}(\varepsilon_1)$  veel groter is dan de  $G_2^{\text{inv}}(\varepsilon_2)$  term vanwege  $\varepsilon_1 \ll \varepsilon_2$ . Vergelijking (46) ziet er heel mooi uit, maar er zit een addertje onder het gras: de functies  $G_1$  en  $G_2$  hangen op een ingewikkelde manier van  $m$  af! In de limiet van grote  $m$  (en grote  $c$ ) gebeurt er echter iets moois. Door de Centrale Limiet Stelling worden de twee verdelingen in Fig. 14 Gaussisch. Gebruik makend van de ongelijkheid  $[\text{Erfc}^{\text{inv}}(x/\sqrt{2})]^2 < \ln(\sqrt{\pi}/x)$  krijgen we de volgende ondergrens op  $m$ , benodigd om tegen  $c$  aanvallers opgewassen te zijn, bij foutkansen  $\varepsilon_1, \varepsilon_2$ :

$$m_{\text{min}} \approx \frac{2}{\tilde{\mu}^2} c^2 \ln \frac{1}{\varepsilon_1 \sqrt{2\pi}}. \quad (47)$$

Dit is vergelijkbaar<sup>6</sup> met de lengte van de best bekende deterministische code,  $m = \mathcal{O}(c^2 \ln n)$ , maar wordt hier bereikt met een klein alfabet!

## 4 En er is natuurlijk nog meer te doen

We hebben gezien dat een simpele stochastische tweestaps-opbouw van de codematrix en een handige keuze van de gewichtsfuncties  $g_0, g_1$  leiden tot een goed werkende code: Hij is makkelijk te implementeren, heeft de belangrijke no-framing eigenschap, en is ook nog eens eenvoudig te begrijpen. Het bestaan van gerandomiseerde codewoorden en foutkansen  $\varepsilon_1, \varepsilon_2 > 0$  heeft zoveel ontwerpvrijheid gegeven dat het resultaat beduidend beter is dan deterministische codes.

Om meer inzicht in de Tardos code te krijgen is flink wat werk nodig en veel lastigere bewijstechnieken dan de heuristische argumenten die ik hierboven gaf. Zonder veel uitleg noem ik nog een aantal interessante aspecten van de Tardos code.

Het schaalgedrag in (47),  $m \propto c^2$ , is optimaal. Het bewijs hiervoor werd geleverd door Tardos in dezelfde publicatie [6] waarin hij ook de constructie voor binair alfabet beschreef.

Het getal  $2/\tilde{\mu}^2$  in (47) is gerelateerd aan een zogenaamde *kanaalcapaciteit*. Een kanaalcapaciteit is normaal gesproken een eigenschap van een ruizig communicatiekanaal, namelijk de bovengrens op de hoeveelheid informatie die je zonder fouten door het kanaal kan sturen. In het geval van coalitie-aanvallen

---

<sup>6</sup>De totale kans op een onterechte verdenking is  $\eta = 1 - (1 - \varepsilon_1)^{n-c} \approx n\varepsilon_1$ . De uitdrukking  $\ln \frac{1}{\varepsilon_1}$  is ongeveer gelijk aan  $\ln n + \ln \frac{1}{\eta}$ .

op watermerken kan je de aanval beschouwen als een soort ruis. De kanaalcapaciteit is alleen nog maar bepaald voor de limiet  $c \rightarrow \infty$  [3, 2]. De Tardos code haalt de capaciteit niet, d.w.z. er bestaan codes die beter presteren [1] wat betreft de coëfficiënt die voor de  $c^2$  staat.

De kanaalcapaciteit voor het gebruikelijke aanvalsmodel is een stijgende functie van de alfabetgrootte  $q$ . Het is dus goed om met een zo groot mogelijk alfabet te werken. Wat haalbaar is hangt sterk af van de manier van watermerken.

Met heel veel moeite kan de precieze vorm van de twee kansverdelingen in Fig. 14 uitgedrukt worden [5] in een reeksontwikkeling gerelateerd aan zgn. Edgeworth expansies, die aan de basis liggen van de Centrale Limiet Stelling.

## Referenties

- [1] E. Amiri en G. Tardos. High rate fingerprinting codes and the fingerprinting capacity. In *SODA 2009*, pages 336–345.
- [2] D. Boesten en B. Škorić. Asymptotic fingerprinting capacity for non-binary alphabets. In *Information Hiding 2011*, volume 6958 of *LNCS*, pages 1–13. Springer, 2011.
- [3] Y.-W. Huang en P. Moulin. On the saddle-point solution and the large-coalition asymptotics of fingerprinting games. *IEEE Transactions on Information Forensics and Security*, 7(1):160–175, 2012.
- [4] P. Moulin. Universal fingerprinting: Capacity and random-coding exponents. In *Preprint arXiv:0801.3837v2*, 2008.



- [5] A. Simone en B. Škorić. Accusation probabilities in Tardos codes: beyond the Gaussian approximation. *Designs, Codes and Cryptography*, 63(3):379–412, 2012.
- [6] G. Tardos. Optimal probabilistic fingerprint codes. In *STOC 2003*, pages 116–125.
- [7] B. Škorić, S. Katzenbeisser, en M.U. Celik. Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes. *Designs, Codes and Cryptography*, 46(2):137–166, 2008.

# Wiskunde voor Dichters

**Michiel Doorman, Freudenthal Instituut -  
Universiteit Utrecht**

**e-mail: M.Doorman@fi.uu.nl**

## Inleiding

Veel onderwerpen binnen de wiskunde spreken een breder publiek aan dan alleen béta's. Bekende voorbeelden zijn: het begrip oneindig, de vierde dimensie en de gulden snede. Bovendien komen wiskundige thema's en methoden veelvuldig terug in andere disciplines. Wiskunde is een vakgebied met brede toepasbaarheid en kent een eeuwenlange verwondering die ze bij beoefenaren en andere geïnteresseerden gewekt heeft. Dat is ons romantische beeld van wiskunde.

Erasmus karakteriseert de wiskundige in Lof der Zotheid (1514) als volgt:

Zij zien laag neer op het oningewijde gemeen, als zij drie-

en vierhoeken, cirkels en andere meetkundige figuren, de een over de andere tekenen en als in een doolhof dooreen laten lopen, vervolgens letters als in slagorde scharen, die ze nu eens op deze, dan weer op gene wijze rangschikken, om zo onervarenen zand in de ogen te strooien.

Is deze karakterisering het gevolg van een exacte benadering van de wiskunde die bij Erasmus - en helaas ook bij vele anderen - een beeld van wiskunde en wiskundigen levert dat weinig recht doet aan de romantische doelstellingen?

Met Wiskunde C is op het vwo een mogelijkheid geopend om niet-*bèta*'s te interesseren voor en succeservaringen te laten opdoen met wiskunde. Het nieuwe *cTWO* programma voor wiskunde C is algemeen vormend in de zin dat het leerlingen voorbereidt op de (informatie)maatschappij en vervolgoopleidingen, enerzijds met relevante onderwerpen zoals statistiek en anderzijds door aandacht te geven aan redeneren, argumenteren en reflecteren. Binnen het nieuwe wiskunde C wordt bovendien gewerkt in contexten die passen in het C&M-profiel. Dit betekent dat er minder nadruk ligt op het verwerven en automatiseren van wiskundige technieken en dat er meer aandacht is voor de cultuurhistorische waarde van wiskunde in onze maatschappij.

Op de Universiteit Utrecht bestaat het vak Wiskunde voor Dichters, Denkers en Doeners voor niet-*bèta*'s dat ook een dergelijk doel beoogt: wiskundige vorming in samenhang met de historische en culturele plaats van wiskunde in wetenschap en maatschappij. Wat is mogelijk met een andere benadering van de wiskunde? Hoe reageren leerlingen op experimenten met nieuwe onderwerpen voor Wiskunde C en hoe reageren

studenten op Wiskunde voor Dichters?

Uitgangspunt bij deze initiatieven is dat een benadering van het proces van abstraheren wordt beoogd die voor deze doelgroep minder abstract is dan de benadering die ze eerder in hun school-loopbaan ervaren hebben. Abstracte begrippen en concrete toepassingen worden in samenhang benaderd om hen de gelegenheid te geven de werkplaats van de wiskundige te betreden en te waarderen (Freudenthal, 1967). In dit artikel volgen eerst drie pogingen hiertoe uit het college Wiskunde voor Dichters. Na die voorbeelden volgt een selectie uit het werk van leerlingen en studenten met een reflectie op de oorspronkelijke doelstellingen van wiskunde voor niet- $\beta$ 's.

## **Drie voorbeelden uit het college Wiskunde voor Dichters**

### **1. Ordenen, structureren, abstraheren en de schoonheid van een bewijs**

In het eerste college van Wiskunde voor Dichters staan de oorsprong van wiskundige begrippen en het proces van abstraheren centraal. We bekijken filmpjes van jonge kinderen en telproblemen. Esther Gerritsen observeert een vroege wiskundige activiteit:

“Als ik niet op mijn dochter let en ze ook geen televisie van me mag kijken, begint ze zich na een tijdje enorm te vermaken met mijn spullen te verplaatsen naar plekken waar ze niet horen. (...) Wat ze dan bijvoorbeeld

doet is alle nietjes uit zo'n klein doosje halen, uit elkaar breken en in een theepot stoppen. Of ze opent mijn gereedschapskist, legt alle schroevendraaiers in mijn bed en verdeelt de spijkers en schroeven over heel veel theekopjes.” (VPRO-gids)

Die activiteit van ordenen, het verdelen van voorwerpen over bakjes, zullen vele ouders van jonge kinderen herkennen. Het is fundamenteel voor het latere tellen. Iets wat vaak als triviale activiteit wordt gezien, maar waarbij meer wis-kunde komt kijken dan je denkt: het op een rijtje leggen (ordenen) van de te tellen objecten, het opzeggen van de telrij synchroon met het aanwijzen van die objecten en het inzicht dat met het benoemen van het laatste object ook de omvang van de verzameling wordt vastgesteld (verbinden van ordinaal en kardinaal getalbegrip). Uiteindelijk zijn getallen en telwoorden zelfstandige objecten, onafhankelijk van de aard van de getelde voorwerpen. Zo ook ontstaan wiskundige groepen om structuren van verzamelingen te bestuderen onafhankelijk van de aard van hun elementen. In het onderwijs reflecteren we meestal weinig op het proces van abstraheren naar die wiskundige objecten. In dat proces vormen zich zowel de wiskundige begrippen als de bijbehorende taal en notaties. Inzicht in en misschien ook zelfs het doorlopen van dat proces is belangrijk voor het begrijpen en flexibel kunnen hanteren van die wiskundige begrippen en werkwijzen.

In de werkplaats van de wiskundige is dit proces van abstraheren een zoektocht naar patronen en structuren (van der Blij, 2004). Deze zoektocht is ook in de schilderkunst mooi zichtbaar. Bekende voorbeelden zijn het ontstaan van perspectief

tijdens de renaissance en de opkomst van abstracte schilderkunst binnen de Stijl begin 20e eeuw. Rudi Fuchs beschrijft elementen daarvan in een column over Mondriaans' Studie bomen 1 van 1912. Deze studie betreft een kale boom die Mondriaan in de zomer van 1912 getekend heeft en die dus eigenlijk vol bladeren zat. Fuchs:

“Dus terwijl hij tekende, heeft hij de boomkruin ontbladerd, en zo als melodische vertakking van kronkelige lijnen gezien - omdat het hem ging om de ritmiek en het rijm ervan. (...) Ook in andere werken uit die jaren werd een intuïtie van abstractheid geleidelijk merkbaar - en toen begon hij zo ook te kijken. (...) Daarbij, in zijn zorgvuldige kijken, zien we dat de ruimtes tussen de takken (de doorkijkjes) visueel steeds zelfstandiger worden. (...) De takken van bomen, liet Mondriaan zien, werken als contourlijnen van open vlakken die, in hun onderlinge verhouding, weer een eigen dynamisch patroon vormen. Zo is het, kort gezegd, met de abstractie begonnen. Nu, honderd jaar later, is dat zo kijken een eigen, vruchtbaar idioom geworden.” (De Groene Amsterdammer, 26.01.2012)

Enkele maanden later komt Rudi Fuchs weer terug op dit proces van abstraheren als hij werk van Cassée bespreekt:

“Vanaf het moment dat een vorm hem opvalt, begint hij te tekenen en te abstraheren. Dat laten de schetsboekjes goed zien. Het maken daarna van een prent is een behoedzaam proces van verdere reductie waarbij de vormen ontdaan worden van elk visueel rumoer. (...) De

tijd die het kost om een prent te maken is ook de tijd, eigenlijk, die een kunstenaar heeft om naar de groei van het beeld te kijken. (...) Door langer te blijven kijken begon hij steeds meer te zien. Daarin beleven deze kunstenaars hun avontuur - dat ons meevoert naar wat wij ook nooit zo gezien hebben.” (De Groene Amsterdammer, 14.06.2012)

Vormen ontdoen van visueel rumoer. Het lijkt alsof abstractie de kunst van het weglaten is. Liever spreken we over de kunst van het zien: het identificeren, soms zelfs creëren, van de essentie van dat wat wordt afgebeeld. Daarmee is het abstraheren een constructief en creatief proces. Een creatief proces dat veel gemeenschappelijk heeft met de manier waarop in de wiskunde zich objecten als getallen en groepen zich vormen.

Deze voorbeelden vertellen misschien niet iets over de wiskunde zelf, maar illustreren het proces van abstraheren. Een proces dat met dergelijke analogieën meer betekenis kan krijgen.

In het zoeken naar de kern van een beeld, structuur of patroon zijn er dus overeenkomsten te herkennen tussen wiskundigen en kunstenaars. Hoe zit dat met de esthetische waarde van het eindproduct? Kunnen we vergelijkbare kenmerken hanteren bij het beschouwen van een mooi bewijs en een kunstwerk?

In het college Wiskunde voor Dichters wordt daartoe het volgende probleem behandeld:

*Hoeveel 7's komen voor in de getallen van 0 tot 1000?*

Twee uitwerkingen worden naast elkaar gezet. In de eerste wordt het antwoord gezocht met systematisch tellen van 7's in de achtereenvolgende getallen. Dat levert een patroon waarbij

je 20 7's vindt in de eerste 100 getallen. Dit is te extrapoleren naar de volgende 100-tallen, waarbij je alleen iets speciaals moet doen met de getallen van 700 tot 799. Uiteindelijk vind je zo een totaal van 300 7's.

De tweede uitwerking maakt gebruik van een blikwisseling van getallen naar cijfers: ieder getal tussen 0 en 1000 kan worden gerepresenteerd met 3 cijfers (van 000 tot 999). Uitgaande van een eerlijke verdeling van de cijfers 0-9 over deze drie posities geeft 3000 cijfers, waarvan  $1/10$  een 7 is. En dus volgt verrassend eenvoudig het resultaat: 300 7's.

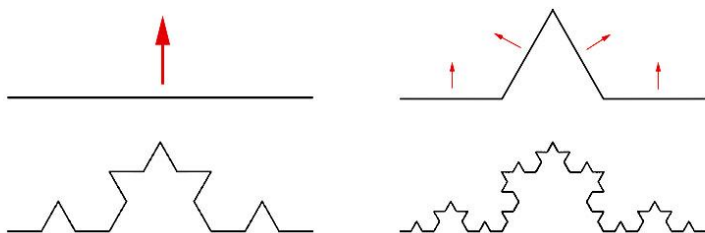
Welke uitwerking vind je mooist en waarom? Bij het eerste bewijs zie je precies wat gebeurt en sluit het zoekproces aan bij je eerste ideeën over een natuurlijke benadering van het probleem. De tweede uitwerking bevat een blikwisseling en laat zich eenvoudig toepassen op variaties in het probleem. Maar het mooie is dat een aantal studenten aarzelden over de tweede uitwerking (ongeacht het identieke antwoord). Hoe weet je zeker dat de cijfers van alle getallen gelijkmatig verdeeld zijn? Het leuke en leerzame van zo'n bespreking is dat zowel esthetische aspecten een rol spelen als inzicht en reikwijdte van de onderliggende redenering ("hoe kom je op die blikwisseling, typisch iets van een wiskundige" verzucht een student). Na het bespreken van een gedicht dat zou kunnen gaan over een wiskundige (de Albatros van Baudelaire die vleugellam is tussen de 'gewone' mensen op het dek van een schip) komen gemeenschappelijke kenmerken van bewijzen en gedichten aan de orde. Overeenkomsten in vormkenmerken (eenvoud, elegantie, symmetrie) en betekenis geven (gebruik van analogie, metaforen, generaliserend, verbindend, kernachtig, inzicht leverend). Is dit zoeken naar overeenkomsten misschien wat vergezocht?



Zoek zelf de overeenkomst tussen wiskunde en Gerrit Komrij's definitie van poëzie: "Poëzie is er voor het intellectuele spel en het verruimde kunstbegrip" (Komrij, 1995).

## 2. Fractals en hun dimensie

In het college komt 'echte' wiskunde aan de orde in onderwerpen waarvoor in het reguliere lesprogramma op de middelbare school vaak weinig gelegenheid is. Hopelijk kunnen uw leerlingen in een praktische opdracht iets over fractals en fractale dimensies verkennen. Hoewel er interactieve lessen nodig zijn om een nieuwe benadering van dimensie met hen te behandelen. Een benadering waarvoor niet meer achtergrond nodig is dan wiskunde uit 3 havo/vwo. Eerst bekeken we een aantal bekende krommen, hun definities en het kenmerk zelfgelijkvormigheid. Hierbij ontbreekt niet de bekende Koch kromme.



Figuur 15: Begin van de constructie van een Koch kromme

Tijd is nodig om te rekenen aan de lengte van de kromme en een bewijs te geven dat die oneindige lengte binnen de grenzen van een A4-tje blijft. Als fractals naast elkaar bekeken worden valt op dat de ene grilliger is dan de ander. Dan rijst

de vraag: kunnen we daar een maat voor vinden? Zo'n maat is geïnspireerd door het meten van de lengte van de kust van Groot Brittannië. Het bleek dat die lengte afhankelijk is van de maat die je hanteert (zie figuur 2). Het is natuurlijk niet

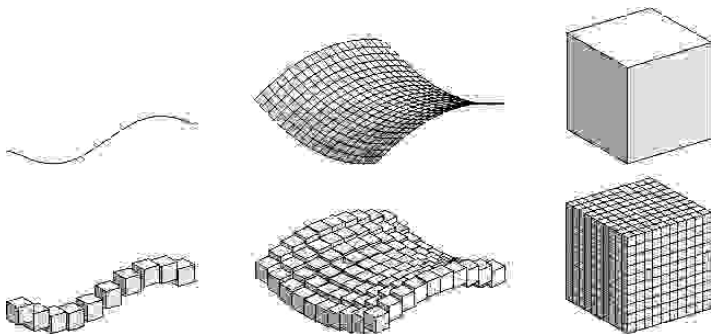


Figuur 16: Het meten van de kustlijn van Groot Brittannië.  
[http://en.wikipedia.org/wiki/How\\_Long\\_Is\\_the\\_Coast\\_of\\_Britain%3F\\_Similarity\\_and\\_Fractional\\_Dimension](http://en.wikipedia.org/wiki/How_Long_Is_the_Coast_of_Britain%3F_Similarity_and_Fractional_Dimension)

merkwaardig dat een kleinere maateenheid meer detail meet en dus een langere lengte levert, maar die lengte bleek exponentieel toe te nemen. Dit was aanleiding voor een alternatieve dimensie-definitie. Bij 1-dimensionale objecten verwacht je dat het halveren van de maateenheid vraagt om twee keer zoveel eenheden om het object te meten. Als het object grillig is dan heb je in het begin misschien last van het kustlijn-effect, maar na voldoende inzoomen is alles netjes glad en geldt deze vuistregel. Uitbreiding naar 2- en 3-dimensionale objecten levert

een verband tussen toename van het aantal maat-eenheden en dimensie. Als voor alle  $n$  geldt dat  $1/n$  keer ribbe vraagt om een overdekking van  $n^d$  meer kubusjes, dan geldt dat de figuur  $d$ -dimensionaal is. Hierbij gaan we er maar van uit dat die kubusjes oneindig dimensionaal zijn (je weet tenslotte niet uit hoeveel dimensies het in te pakken object bestaat). In het Engels wordt dit ook wel de box-dimension genoemd. Een Nederlandse vertaling zou inpak-dimensie kunnen zijn.

In figuur 3 zijn lijn, vlak en kubus ingepakt, waarbij de lengte van de ribbe van de (eigenlijk oneindig-dimensionale) kubusjes met  $1/10$  vermenigvuldigd zijn. De lijn vraagt om  $10^1$  keer zoveel kubusjes, dus de dimensie is 1. Evenzo volgt dat het vlak dimensie 2 heeft en de kubus dimensie 3. Voor ‘normale’ objecten geeft de inpak-dimensie dus precies de waarden die verwachtten. Bij de Koch kromme geldt echter dat een



Figuur 17: Inpak-dimensie van lijn, vlak en kubus (van internet: [http://www.math.sunysb.edu/~scott/Book331/Fractal\\_Dimension.html](http://www.math.sunysb.edu/~scott/Book331/Fractal_Dimension.html))

vermenigvuldiging van de ribbe met  $1/3$  om 4 keer meer kubusjes vraagt, hoe ver je ook inzoomt op de kromme. Dit valt op te maken uit de constructie van deze kromme (figuur 1). Dus hier komen we op de vergelijking  $3^d = 4$ . Nu zou je met logaritmen de dimensie exact kunnen geven, maar benaderen met inklemmen kan ook. Voor de Koch kromme geldt dan dimensie  $d \approx 1,26$ .

De behandeling van zo'n alternatieve definitie van dimensie verdiept het begrip zelf, is een toepassing van het werken met vergrotingsfactoren en levert bijzondere resultaten. Het blijkt dat hiermee het oppervlak van een bloemkool dimensie 2,3 levert en het oppervlak van ons brein en onze longen respectievelijk dimensies van 2,79 en 2,97. Deze getallen zeggen iets over de aard van de grilligheid van die objecten. Overigens heeft de kustlijn van Groot Brittannië zo bij benadering een inpak-dimensie van 1,25.

### 3. De gulden snede

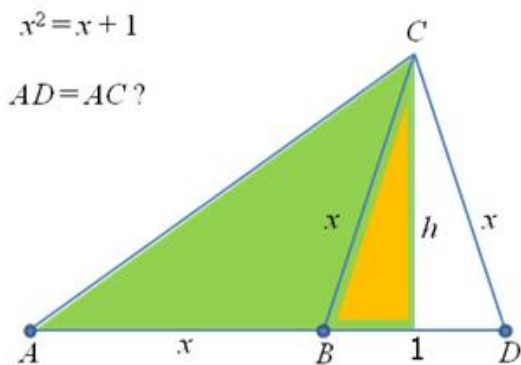
Het derde voorbeeld is de bekende gulden snede. Dit onderwerp is vooral zo geschikt omdat het gebruik en misbruik van wiskunde in de kunst fraai illustreert (lees bijvoorbeeld Ernst & Konings, 2008).

De introductie van de gulden snede in de westerse maatschappij heeft veel te danken aan de Italiaanse wiskundige Leonardo van Pisa (c. 1170 - c. 1250). Hij leefde in een periode waarin Noorditaliaans steden en kleine republieken in voortdurende machtsstrijd verwickeld waren tussen het keizerrijk en het Vaticaan. Pisa was een welvarende stad door de handel met

Noord Afrika en daarmee een kenniscentrum voor de Arabische cultuur. Dat was de tijd van de Fibonacci reeks en aandacht voor Hindoe Arabische getallen en notaties (inclusief de 0). Er werden boeken geschreven voor rekenmeesters, landmeters en handelaren. Sommige wiskunde kreeg een ‘goddelijke’ waarde.

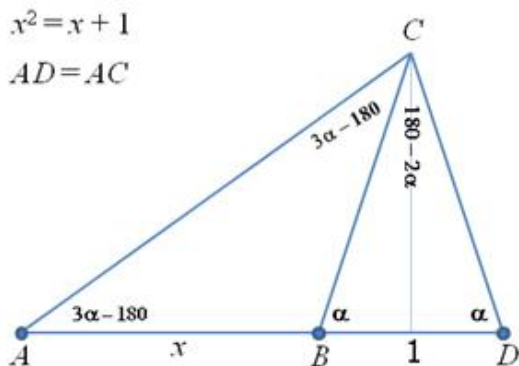
De gulden snede is een verhouding die kan worden berekend met de vergelijking  $x^2 = x + 1$ . De achtergrond blijft hier buiten beschouwing. Alleen de constructie van de regelmatige vijfhoek lopen we langs. Deze constructie illustreert weer hoe met onderbouw-wiskunde toch fraaie klassieke resultaten bereikt kunnen worden.

De constructie begint met het tekenen van de lijnstukken  $AB$  en  $BD$  met verhouding  $x : 1$ , waarbij  $x^2 = x + 1$ .  $C$  wordt geconstrueerd via gelijkbenige driehoek  $BCD$  (figuur 4). Nu is te bewijzen dat  $AD = AC$ . Waarom is dat nodig? Hoe gaat dat helpen? Dat blijkt uit de volgende stap. Door twee keer



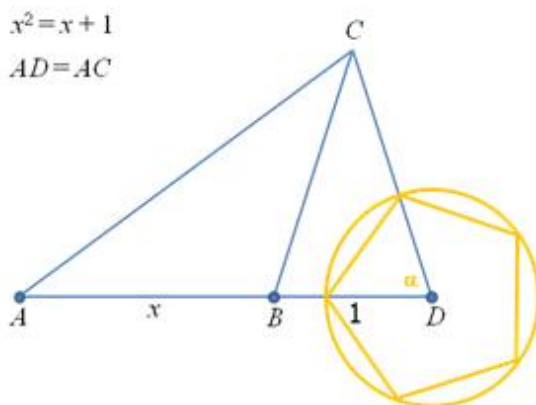
Figuur 18: De basisconstructie voor een regelmatige vijfhoek

de stelling van Pythagoras toe te passen volgt dat  $AC = AD$  en dat driehoek  $ACD$  dus ook gelijkbenig is. Dat betekent voor de hoeken dat hoek  $ACB$  gelijk is aan  $3\alpha - 180^\circ$ . Maar ook dat die hoek gelijk moet zijn aan  $180^\circ - 2\alpha$  (figuur 5). Uit



Figuur 19: Redeneren met hoeken

die twee gegevens volgt dat  $\alpha = 72^\circ$ , en dat is precies wat we nodig hebben voor de constructie van de vijfhoek (figuur 6). Met dit onderwerp kunnen eenvoudig vele lessen gevuld worden. Bovenstaande constructie was een tamelijk technische uitweiding en geeft inzicht in de werkplaats van de wiskundige kunstenaar. De vraag is natuurlijk: Hoe kom je op het idee om met deze driehoek te beginnen? “It is like a rabbit pulled out of a hat” (Polya, 1965, p. 64). Enerzijds is fraai om te zien hoe plotseling die  $72^\circ$  uit de afleiding rolt. Anderzijds is door deze elegante constructie het inzicht in het zoekproces - de werkwijze van de kunstenaar - verloren gegaan. Een beeld van die zoektocht is te krijgen in één van de *Wat-te-bewijzen-*



Figuur 20: Redeneren met hoeken

*is-columns* van Martin Kindt (Kindt, 2003).

Dergelijke excursies omvatten technieken die de studenten niet geautomatiseerd hadden en die af en toe een zijweg veroorzaakten. Dat gaat echter prima. Af en toe terugvallen op de bordjesmethode is geen probleem als je de tijd ervoor neemt (en hebt). Bij deze studenten is het inslijpen van automatismen niet gelukt, en je moet dus ook niet pretenderen dat dit nu wel even te repareren is. Liever leveren we een basis van waaruit ze afleidingen kunnen volgen en eventueel zelf kunnen reproduceren. Tenslotte moeten de technieken niet ten koste gaan van historische achtergronden en de vele voorbeelden en non-voorbeelden (waaronder het Parthenon in Athene), waarbij een filmpje van Vi Hart (zie Figuur 7) niet mag ontbreken.



Figuur 21: Doodling in Math: Spirals, Fibonacci, and Being a Plant <http://www.youtube.com/watch?v=ahXIMUkSXX0>

## Werk van leerlingen en studenten

Al weer enkele jaren geleden experimenteerde Bart Zevenhek (docent aan het Barlaeus gymnasium te Amsterdam) met nieuwe inhoud voor wiskunde C. Hij leest met zijn wiskunde A1-groep *De ontstelling van Pythagoras - Over de geschiedenis van de goddelijke proportie* van Albert van der Schoot. Iedere week presenteert een leerling een hoofdstuk uit dit boek. Voorafgaand aan die presentatie moet een samenvatting worden ingeleverd. Een van de presentaties betreft de rol van Pacioli in het ontstaan van de mythe rond de gulden snede. De leerling heeft een kopie van een schilderij van Pacioli op het bord gehangen (zie figuur 8). De leerling beschrijft hoe Luca Pacioli, een Franciscaner monnik (1445 - 1517), het boek *Divina Proportioni* schreef met illustraties van Leonardo da Vinci. Het boek slechtte een brug tussen kunst en wetenschap. Een



fraai voorbeeld van het positioneren van wiskunde in maatschappelijke context door leerlingen. In het college Wiskunde



Figuur 22: Een portret, Luca Pacioli (Museo di Capodimonte, Napels)

voor Dichters presenteren studenten over een zelfgekozen onderwerp. Een groepje heeft als onderwerp de Möbiusband. Dit is natuurlijk ook geen verrassend onderwerp, maar toch is aardig om te zien hoe ze een systematisch onderzoek hebben uitgevoerd naar resultaten van het doorknippen van banden met meerdere halve draaiingen. Daarnaast laten ze zien hoe je een Möbiusband vanaf het begin in een keer kunt breien met een rondbreinaald. Door de steken op een speciale manier op te zetten, creëer je een halve draai in het breiwerk. Tot slot laten ze een krabcanon van J.S. Bach horen. Door de notenbalk op een Möbiusband te projecteren krijg je een

oneindige uitvoering van dit stuk (zie Figuur 9). Dit laatste



Figuur 23: Möbiussjaal en krabcanon

voorbeeld laat nog eens zien dat zowel het componeren van muziek als het luisteren ernaar een vorm van constructie omvat waarbij patronen worden herkend en bewerkingen op die patronen kunnen worden gevolgd (Doorman, 1999).

Tot slot een voorbeeld van een student (Jasper Leuven) die nog eens de relatie tussen poëzie en wiskunde analyseert. Hij schrijft eerst over de overeenkomsten en benadrukt vervolgens een verschil:

“Hoewel wiskundige bewijzen soms verrassend kunnen zijn, denk ik toch dat gedichten vaker een verborgen boodschap of dubbele bodem hebben: een soort ongeschreven betekenis. Hierover filosoferend heb ik geprobeerd een nieuwe versie te schrijven van *Does it matter?* met uitsluitend getallen en wiskundige termen. (...) Getallen kunnen een betekenis met zich meedragen, maar het blijft lastig om met uitsluitend getallen een verhaal te vertellen.”

## Does It Matter?

*(Siegfried Sassoon, 1918)*

DOES it matter? - losing your  
legs? ...

For people will always be kind,  
And you need not show that  
you mind

When the others come in after  
hunting

To gobble their muffins and  
eggs.

Does it matter? - losing your  
sight? ...

There's such splendid work for  
the blind;

And people will always be  
kind,

As you sit on the terrace  
remembering

And turning your face to the  
light.

Do they matter? - those  
dreams from the pit? ...

You can drink and forget and  
be glad,

And people won't say that  
you're mad;

For they'll know you've fought  
for your country

And no one will worry a bit. <sup>110</sup>

## Maakt het wat uit?

*(Jasper Leuven, 2012)*

Maakt het wat uit? 1918 min  
4?

$1914 + 4 = 1918$

22477500 en 16403000

Want plus, keer, min of delen  
door,

dat gaat allemaal nog steeds  
door.

Maakt het wat uit? Zonder  
irrationele getallen?

Er zijn zo'n mooie rationele  
en ook daarmee kun je rekenen

-  $1918 - 1914 =$  slechts 4 jaren  
zelfs bewijzen geven licht.

Wat maakt het ook uit? De  
getallen die dromen?

Zolang wij ze kunnen gebrui-  
ken voor

Zolang een 3 een 3 blijft en een  
5 een 5.

Ze zijn ons van dienst geweest  
en wat geeft het

dat ze zelf niks meer kunnen.

## Slot

Deze voorbeelden laten een andere benadering van wiskunde in het onderwijs zien. Een benadering die verder gaat dan een leuke les en die recht doet aan de ruimte en het brede perspectief die contexten bieden. Een benadering die mogelijk is als het curriculum niet al te grote eisen stelt aan het beoogde beheersingsniveau van technische wiskundige vaardigheden. Overigens verliepen niet alle colleges Wiskunde voor Dichters vlekkeloos. De poging om in twee bijeenkomsten een idee van chaostheorie te leveren mislukte jammerlijk. Vanuit lineaire en exponentiële groei werd de formule voor logistische groei opgebouwd en vervolgens leverde het draaien aan parameters chaotisch gedrag (gevoelige afhankelijkheid van de beginvoorwaarde). Het beoogde onderwijsleergesprek, waarin formulewerk omzeild werd door gebruik van Excel resulteerde snel in een eenzijdige one-man show waar niemand van genoot. Deze ervaring maakte nog eens duidelijk hoe het proces van abstractie te snel kan gaan als toehoorders te weinig betrokken zijn bij ontwikkeling van notaties en onderliggende begrippen. Eerste experimenten met recurrente betrekkingen bij wiskunde A2 lieten ook zien dat nieuwe notaties en denkwijzen veel vragen van de leerlingen.

Deze ervaringen met wiskunde C en het college wiskunde voor dichters tonen enkel mogelijkheden van een wiskundige benadering van kunst. Vanwege haar exactheid en het constructieve karakter kan wiskunde een beter begrip van artistieke activiteiten ondersteunen. In sommige opzichten is deze benadering misschien minder exact en minder wiskundig dan we gewend zijn, maar het biedt leerlingen en studenten wel een ander

beeld, minder geleid door sommetjes uit het boek en minder gericht op technische vaardigheden. Hopelijk is het gelukt hen daarmee geen zand in hun ogen te strooien en een rijker beeld van de wiskunde te geven. Een beeld dat meer recht doet aan de wiskunde dan Erasmus ervan weergeeft. Een beeld gevormd door ervaringen zoals Belle van Zuylen die beschrijft in een brief aan haar vriend Constant d’Hermenches. Wiskunde is volgens hem geen bezigheid voor een vrouw, het vernauwt de verbeeldingskracht, verdort de geest en brengt schade aan het gevoelsleven. Belle van Zuylen antwoordt hem in een brief (25 februari - 5 maart 1764):

“ik ben nog niet gaan merken dat mijn geest zich vernauwt en dat mijn verbeelding onvruchtbaar wordt, maar wel dat weet ik, dat een of twee uur wiskunde mijn geest vrijer maakt en mijn hart vrolijker. Ik heb het gevoel dat ik beter slaap en beter eet wanneer ik evidente en onweerlegbare waarheden heb gezien”.

## Referenties

- Van der Blij, F. (2004). Abstractie in kunst en Wiskunde. Een denkbeeldige wandeling door een museum van gedachten. *Euclides* 79(4), 180-183.
- Doorman, M. (2007). Wiskunde C: daar komt muziek in. *Nieuwe Wiskrant* 27(1), 31-34.
- Doorman, S.J. (1999). Wiskunde en culturele vorming. *Nieuwe Wiskrant* 19(2), 13-16.

Erasmus, D. (1944). *Lof der Zotheid*. Wereldbibliotheek N.V. (eerste druk 1517).

Ernst, B. & T. Konings, 2008. *Kunst en Wiskunde*. Epsilon Uitgaven.

Freudenthal, H. (1967). *Wiskunde in wetenschap en dagelijks leven*. De Haan/Meulenhof.

Kindt, M. (2003). Wat te bewijzen is (21). *Nieuwe Wiskrant* 22(4). 28-29.

([http://www.fisme.science.uu.nl/wiskrant/artikelen/224/224juni\\_wtbi21.pdf](http://www.fisme.science.uu.nl/wiskrant/artikelen/224/224juni_wtbi21.pdf))

Komrij, G. (1995). In Liefde Bloeyende. In: *NRC Handelsblad*, 20 juli 1995.

Polya, G. (1965). *Mathematical Discovery: On understanding, learning and teaching problem solving. Volume 1*. John Wiley and Sons.

Sassoon, S. (1918). *Counter-Attack and Other Poems*. E. P. Dutton & Company. (Op internet geraadpleegd: <http://www.bartleby.com/136/14.html> op 8 juli 2012)

# Een GeoGebra- ondersteunde benadering van sinus en cosinus

André Heck, Universiteit van Amsterdam  
e-mail:a.j.p.heck@uva.nl

## 1 Verschillende definities van sinus en cosinus

Sinus en cosinus zijn goniometrische functies die essentieel zijn in een wiskundige beschrijving van periodieke verschijnselen. Ze kunnen op vele manieren worden gedefinieerd.

In de meetkundige definitie, die zijn oorsprong heeft in astronomie en landmeetkunde, zijn sinus en cosinus verhoudin-

gen van bepaalde zijden in een rechthoekige driehoek. Ze zijn dan in eerste instantie te beschouwen als afbeeldingen van een scherpe hoek naar het interval  $(0, 1)$ . Een uitbreiding naar hoekfuncties die ook voor stompe hoeken gedefinieerd zijn is voor de hand liggend. De volgende stappen zijn: (i) een herdefinitie voor hoeken tussen  $0^\circ$  en  $360^\circ$ ; en (ii) een periodieke voortzetting tot willekeurige hoeken. Het meetkundige begrip hoek speelt in deze definitie dus een belangrijke rol en landmeetkunde kan als realistische toepassingscontext fungeren (cf., Mazziotta, 1949). Het gaat hier eigenlijk niet om functies gedefinieerd over de reële getallen, maar om functies gedefinieerd op hoeken en dus om functies gedefinieerd op fysieke grootheden met een maateenheid (graad).

Projectie van een eenparige beweging langs de eenheidscirkel op de assen van een cartesisch coördinatenstelsel in een plat vlak is een modernere definitie, die meer aansluit bij de wiskundige modellering van periodieke processen en de overgang naar functies op reële getallen vergemakkelijkt. Ook hierin speelt in eerste instantie het begrip hoek nog een belangrijke rol, maar in dit geval als draaiingshoek bij rotatie van het punt  $(1, 0)$  om de oorsprong. Een positieve draaiingshoek correspondeert met een draaiing tegen de klok in, een negatieve hoek hoort bij een draaiing met de klok mee. De sinus en cosinus van elke hoek, ongeacht zijn grootte, zijn gedefinieerd en allerlei eigenschappen van sinus en cosinus zijn eenvoudig af te leiden met deze definitie als uitgangspunt. Denk hierbij aan eigenschappen als “hoeken die elkaars tegengestelde zijn hebben dezelfde cosinussen en tegengestelde sinussen” of “de sinus van een hoek en de cosinus van de hoek waarmee hij samen  $90^\circ$  vormt (het complement) zijn gelijk. Als maat voor



de grootte van een draaiingshoek kan ook de afgelegde afstand langs de eenheidscirkel bij draaiing van het punt  $(1, 0)$  om de oorsprong gehanteerd worden. De natuurlijke eenheid voor een hoek is dan de radiaal, die correspondeert met een cirkelboog met standaardlengte. De belangrijkste stap richting goniometrische functies gedefinieerd op  $\mathbb{R}$  is dan gezet: alleen de maateenheid (radiaal) hoeft nog maar genegeerd worden.

Wie de theorie van gewone differentiaalvergelijkingen machtig is, kan sinus en cosinus definiëren als de functies  $s$  en  $c$  die de unieke oplossingen zijn het beginwaardenprobleem  $s' = c, c' = -s, s(0) = 0, c(0) = 1$ . Andere analytische definities van goniometrische functies zijn gebaseerd op machtreeksontwikkelingen, functionaalvergelijkingen, integratie van functies, relaties tussen goniometrische functies en hypergeometrische functies, en de relatie tussen sinus, cosinus en de exponentiële functie via de formule van Euler ( $e^{i\theta} = \cos \theta + i \sin \theta$ ). De geïnteresseerde lezer verwijst ik naar publicaties van Bram van Asch en Frederik van der Blij (1992, 1997, 2002).

## 2 Vakdidactische kanttekeningen bij een traditionele aanpak

Bovengenoemde analytische definities vallen buiten bereik van het voortgezet onderwijs. De meetkundige aanpak en de eenheidscirkelmethode worden wel overal ter wereld gebruikt om sinus en cosinus te introduceren in het wiskundeonderwijs. Universeel is ook de ervaring dat leerlingen (inclusief wiskundeleraars in opleiding) moeite hebben met hoeken en goniometrische functies, of het nu gaat om een hoekmaat, negatieve

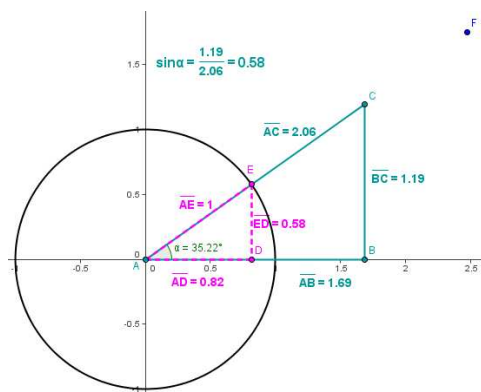
hoeken of hoeken groter dan  $360^\circ$ , de definitie van sinus en cosinus, goniometrische relaties, het oplossen van goniometrische vergelijkingen of om het werken met de goniometrische functies in een reële context. Leerlingen maken allerlei fouten en hebben diverse alternatieve concepties (cf., Akkoç, H, 2008; Blackett, N., & Tall, 1991; Brown, 2005; Fi, 2003; Gür, 2009; Weber, 2005). Ik geef in deze sectie een overzicht van de belangrijkste bevindingen uit vakdidactisch onderzoek. Gek genoeg is er weinig onderzoek naar het leren van goniometrie en het werken met goniometrische functies gedaan; de literatuur bij dit artikel is redelijk compleet en slechts vijf proefschriften (Brown, 2005; Challenger, 2009; Delice, 2003; Fi, 2003; Moore, 2010) heb ik over dit onderwerp kunnen traceren.

In de meetkundige aanpak helpen ezelsbruggetjes leerlingen om de drie basisverhoudingen uit de goniometrie te onthouden. De acroniem *soscastoa* staat bijvoorbeeld voor sinus = overstaande zijde gedeeld door schuine zijde, cosinus = aanliggende zijde gedeeld door schuine zijde en tangens = overstaande zijde gedeeld door aanliggende zijde. Maar het is bekend (cf., Hart, 1981; Behr, et al., 1994; Lamon, 2007) dat werken met verhoudingen niet echt gemakkelijk is voor leerlingen, zeker niet als ook verbanden gelegd moeten worden tussen een meetkundige figuur en numerieke relaties. Tevens komen opgaven in tekstboeken vaak neer op het oplossen van algebraïsche vergelijkingen en dan struikelen veel leerlingen als gevolg van gebrekkige algebraïsche vaardigheden. In de meetkundige aanpak kunnen leerlingen moeite hebben en houden met het inschatten van functiewaarden of zelfs het bepalen van het teken van een sinus of cosinus bij een gegeven hoek. Ook geeft deze aanpak leerlingen weinig houvast bij het bepalen wanneer

de sinus- en cosinusfunctie stijgend en dalend zijn. Volgens Orhun (2010) komt dit vooral omdat leerlingen tijdens en na deze meetkundige aanpak sinus en cosinus blijven associëren met een verband tussen een hoek en de lengtes van zijden van een rechthoekige driehoek. Aan het functiebegrip wordt onvoldoende aandacht besteed in de lessen, met als gevolg dat leerlingen functiewaarden niet goed kunnen inschatten als er niet expliciet een hoekmaat bij vermeld is. Menig leerling kan de verleiding niet weerstaan om  $\sin(3)$  op te vatten als  $\sin(3^\circ)$ . Als tweede oorzaak van gebrekkige kennis en vaardigheden noemt Orhun het onvoldoende oefenen met radiaal als hoekmaat, waardoor de laatste stap naar goniometrische functies gedefinieerd op  $\mathbb{R}$  niet goed uit de verf komt. Eigenlijk duikt voor veel leerlingen de maat radiaal uit het niets op, zien ze niet goed in hoe deze past bij de keuze van booglengte als maat voor een draaiingshoek, en leren ze dan maar omrekeningsformules tussen graden en radialen uit het hoofd. Sinus en cosinus zijn voor de leerlingen op deze manier geen functies gedefinieerd op de verzameling van reële getallen, maar blijvend gekoppeld aan lengteverhoudingen in driehoeken bij een gegeven hoek. Orhun's advies en dat van andere onderzoekers (cf., Brown, 2005; Challenger, 2010; Martínez-Sierra, 2008a,b; Moore, 2010) is om leerlingen expliciet te helpen met het ontwikkelen van een duidelijke link tussen de meetkundige wereld en de wereld van functies en grafieken. Dit gaat niet vanzelf, maar vereist een uitgebalanceerde instructie met volop uitleg, voorbeelden en oefeningen.

Als de eenheidscirkelmethode gebruikt wordt om sinus en cosinus te introduceren, dan kan men goniometriesommen over driehoeken maken door een gegeven rechthoekige driehoek in

een opdracht, zeg  $\triangle ABC$ , te vergelijken met een gelijkvormige referentiedriehoek, zeg  $\triangle ADE$ , met de oorsprong  $A$  als hoekpunt, met een horizontaal en een loodrecht hierop staand lijnsegment als twee zijden van de driehoek en met een derde hoekpunt  $E$  op de eenheidscirkel; zie Figuur 24, die met het dynamische wiskundepakket GeoGebra gemaakt is.



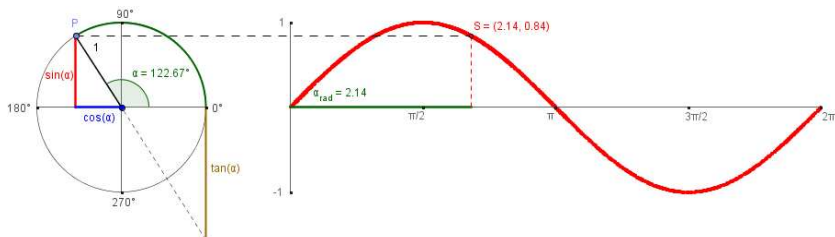
Figuur 24: Schermafdruk van een GeoGebra activiteit voor sinusberekening via lengteverhoudingen in een driehoek en door gebruik te maken van een referentiedriehoek binnen de eenheidscirkel. Door het punt  $F$  te verslepen kan de hoek  $\alpha$  ingesteld worden. Door het punt  $C$  te verslepen kan een gelijkvormige driehoek gemaakt worden.

Het numeriek benaderen van sinus en cosinus kan in de eenheidscirkelmethode uitgevoerd worden door bij een gegeven hoek een halflijn te tekenen vanuit de oorsprong die samen met de positieve  $x$ -as de gegeven hoek vormt en dan de coördinaten van het snijpunt te bepalen van deze lijn met de eenheidscirkel. De coördinaten van het snijpunt horen bij sinus en cosinus.

Weber (2005, 2008) constateerde evenwel dat leerlingen in de schoolpraktijk deze procedure weinig met pen en papier uitvoeren. In zijn optiek gaan tekstboekschrijvers en docenten er te gemakkelijk van uit dat leerlingen het proces kunnen begrijpen zonder veel praktijkervaring, ook al is uit onderzoeksliteratuur (cf., Tall et al., 2000) bekend dat persoonlijke, fysieke ervaring een haast onmisbare stap in het leerproces is. Hier worden kansen gemist, zoals ook bleek in het master research project van Tompogliou (2007) waarin hij een RME-aanpak m.b.v. de eenheidscirkel in een Grieks lyceum beproefde, met een reuzenrad als realistische context. Tompoglou vond dat de zestien- en zeventienjarige leerlingen via de eenheidscirkelmethode een beter begrip van eigenschappen van goniometrische functies verwierven dan met een traditionele meetkundige aanpak. In de meetkundige aanpak leunden leerlingen sterk op het inprenten en onthouden van allerlei eigenschappen. In de eenheidscirkelmethode konden zij veel meer eigenschappen zelf afleiden. Ook concludeerde hij dat een sterke link tussen de meetkundige aanpak en de eenheidscirkelmethode goed te realiseren is, maar dat leerlingen meer moeite hebben met de vervolgstap van het werken met de eenheidscirkel naar de wereld van functies en grafieken (met eigenschappen als stijgende en dalende functie, maxima en minima, symmetrie, etc.).

Wel is in de loop van de tijd duidelijk geworden dat ICT bij kan dragen aan een beter begrip van sinus en cosinus door leerlingen, of dit nu via grafische rekenmachine, specifieke computerprogramma's of dynamische applets gebeurt (cf., Blackett & Tall, 1991; Johari et al., 2010; Johnson & Walker, 2011; Kissane & Kemp, 2009; Moore, 2009; Ross et al., 2011; Takači et al., 2005; Zengin et al., 2012). De belangrijkste winstpunten

van ICT-gebruik bij het leren werken met goniometrische functies zijn: (i) leerlingen en docenten kunnen snel en gemakkelijk grafieken van deze functies tekenen en hiermee verder aan de slag gaan; (ii) leerlingen hebben meer mogelijkheden om transformaties van goniometrische functies te exploreren; (iii) dynamische user-interfaces ondersteunen de bestudering van de link tussen de eenparige cirkelbeweging van een punt en goniometrische functies (Figuur 25); (iv) meetgegevens van periodieke verschijnselen kunnen m.b.v. ICT gemakkelijker vergeleken worden met resultaten van wiskundige modellen gebaseerd op goniometrische functies.



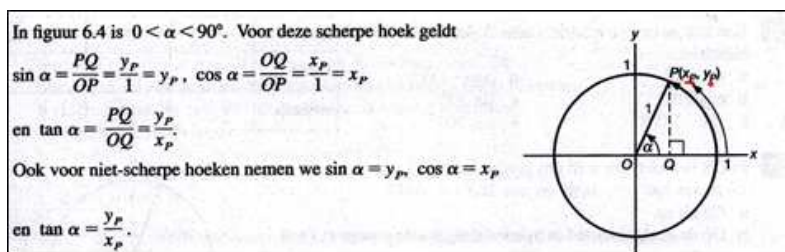
Figuur 25: Schermafbeelding van GeoGebra activiteit voor sinusberekening via de eenheidscirkelmethode. Door het punt  $P$  te verslepen kan de draaiingshoek  $\alpha$  ingesteld worden en het bijpassende punt  $S$  op de sinusgrafiek gevisualiseerd worden.

Hoe dan ook, zowel de meetkundige aanpak met driehoeken als de methode die gebruik maakt van de eenheidscirkel kan als uitgangspunt gekozen worden om sinus en cosinus te introduceren. Maar het is belangrijk zich te realiseren dat, hoewel beide aspecten van sinus en cosinus in onderwijs aan bod moeten komen, er een keuzemogelijkheid is voor de leerroute. Vak-

didactisch onderzoek heeft nog geen uitsluitsel gegeven welke route het meest effectief is. Kendall en Stacey (1996, 1997) rapporteerden dat de meetkundige aanpak beperkt tot scherpe hoeken in hun studie op een Australische school met tien- en elfjarige leerlingen tot betere leeropbrengsten leidde dan de eenheidscirkelmethode. Dit is eigenlijk niet zo vreemd, want als men zich beperkt tot het scherpe hoeken, dan heeft de eenheidscirkelmethode weinig of geen meerwaarde en is deze waarschijnlijk te abstract voor de tien- en elfjarigen. Weber (2005, 2008) vond in zijn studie met oudere studenten op een Amerikaans college juist het tegenovergestelde: studenten die hij onderwees via de eenheidscirkelmethode, met een nadruk op het daadwerkelijk uitvoeren van het proces van numeriek benaderen van sinus- en cosinuswaarden, presteerden beter dan de studenten die hij via de meetkundige aanpak onderwees. In een vergelijkingsstudie over goniometrische kennis en vaardigheden van Engelse en Turkse leerlingen vonden Delice en Roper (2006) dat Turkse leerlingen beter overweg konden met algebraïsche aspecten van goniometrie dan hun Engelse collega's, en dat Engelse leerlingen op hun beurt beter presteerden in toepassingen van goniometrie in contexten uit de echte wereld. Zij noemden de verschillen in curricula en schoolpraktijk in beide landen als voornaamste oorzaak en verklaring van de verschillen. Challenger (2009) formuleerde dit als volgt: "What you get is what you teach." Iets soortgelijks geldt mijns inziens voor de onderzoeken van Weber (2005) en van Kendall en Stacey (1997): de onderzoeksresultaten hangen af van de mate van overeenstemming tussen het onderzoeksinstrument (voor- en natoetsen) en de toegepaste instructiemethode.

Nederlandse wiskundeboeken combineren de bovenstaande be-

nadering van goniometrie. In de onderbouw wordt de meetkundige aanpak voor sinus en cosinus van scherpe hoeken in een rechthoekige driehoek behandeld. Hierna stapt men in de bovenbouw over op de eenheidscirkelmethode, waarin sinus en cosinus gedefinieerd worden als de verticale respectievelijk horizontale coördinaat van een punt verkregen na draaiing van het punt  $(1,0)$  rondom de oorsprong over een gegeven hoek (zie Figuur 26).



Figuur 26: Fragment uit een wiskundeboek waarin de overgang van driehoek naar eenheidscirkel toegelicht wordt.

Maar in deze aanpak zijn sinus en cosinus eigenlijk nog steeds niet als functies van reële getallen gedefinieerd. Dit wordt min of meer afgerond door de introductie van radialen en het simpelweg negeren van deze maateenheid (zie Figuur 27).

De overgang van functies op hoeken naar goniometrische functies is overigens lastiger dan op het eerste gezicht lijkt. De connectie tussen graden en radialen als hoekmaat is en blijft dominant aanwezig en eenieder kan er zich op betrappen dat hij of zij bijvoorbeeld bij de vraag wat de waarde van  $\sin \frac{\pi}{3}$  is toch denkt aan een draaiingshoek van  $\frac{\pi}{3}$  radialen of aan een driehoek met een hoek van 60 graden. De connectie tussen



**Theorie A**

**De functie  $f(x) = \sin(x)$**   
 Het sinusbegrip wordt niet alleen bij hoeken, maar ook bij getallen gebruikt. De sinus van het getal 2 is de sinus van een hoek van 2 radialen.

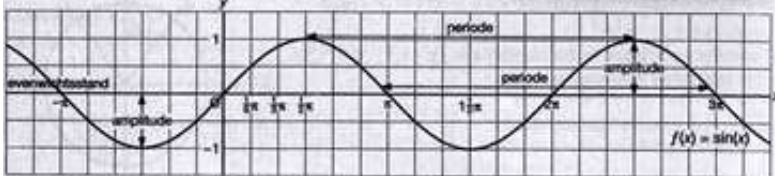
We definiëren de functie  $f(x) = \sin(x)$  als de functie die aan elk getal  $x$  de sinus van  $x$  radialen toevoegt.

Hieronder is de grafiek van de goniometrische functie  $f(x) = \sin(x)$  getekend. Op de horizontale as is als eenheid  $\frac{1}{3}\pi$  genomen.

hoek

$\sin(x) = \sin(x \text{ rad})$

getal



**figuur 6.15** Op de  $x$ -as is als eenheid  $\frac{1}{3}\pi$  genomen. Merk op dat  $\frac{1}{3}\pi \approx 1,047 \approx 1$ .

De grafiek is periodiek met periode  $2\pi$ .  
 De evenwichtsstand is 0 en de amplitude is 1.  
 De nulpunten zijn  $\dots, -2\pi, -\pi, 0, \pi, 2\pi, 3\pi, \dots$

Een nulpunt van een functie  $f$  is een  $x$ -waarde waarvoor geldt  $f(x) = 0$ .

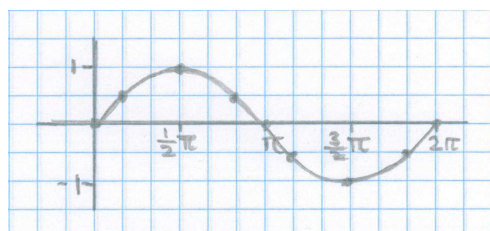
Een nulpunt is geen punt maar een getal.

Figuur 27: Fragment uit een wiskundeboek waarin de overgang van functies op hoeken naar goniometrische functies toegelicht wordt.

$\pi$  en  $180^\circ$  is ook buitengewoon sterk en in ieder geval sterker dan de link tussen radiaal en booglengte op een eenheidscirkel (Akkoç, 2008; Akkoç en Gül, 2010; Orhun, 2010). Dit hindert leerlingen bij het beschouwen van  $\pi$  als een reëel getal in de buurt van 3,14.

In bovenstaande getrapte benadering van goniometrie zijn de sinus- en cosinusfuncties weliswaar geïntroduceerd, maar hun

grafieken zijn toch nog een raadsel voor leerlingen en eigenlijk niet meer dan door grafische rekenmachine of wiskundige software geproduceerde objecten. Leerlingen leren nog wel de exacte waarden van enkele bijzondere hoeken, maar gebruiken deze resultaten haast niet meer om een grafiek zoals in Figuur 28 te schetsen.



Figuur 28: Schets van de sinusgrafiek op ruitjespapier volgens een standaardmethode.

Probleem is en blijft dat leerlingen geconfronteerd worden met functies waarvan de functiewaarden niet meer via rekenkundige bewerkingen berekend kunnen worden; ze moeten afgelezen worden uit gegeven grafieken of uitgerekend worden m.b.v. een rekenmachine of wiskunde software. Werden functies tot dan toe in de schoolloopbaan als machientjes met in- en uitvoer van getallen geïntroduceerd, nu moeten leerlingen ineens een veranderlijke grootte (hoek) via een grafisch gerepresenteerde constructie koppelen aan twee andere grootheden (coördinaten van een punt op de eenheidscirkel). Dit proces is van een hoger abstractieniveau dan leerlingen eerder bij veeltermfuncties en rationale functies tegen zijn gekomen. Het vereist een sterk ontwikkeld functiebegrip en voldoende rijpheid om te kunnen werken op een tamelijk hoog abstractieniveau. Maar lang niet

alle leerlingen zijn al zo ver als het onderwerp sinus- en cosinusfunctie behandeld wordt. Bijkomend probleem is dat leerlingen tegelijkertijd geconfronteerd worden met twee functies die niet los van elkaar geconstrueerd worden, maar direct aan elkaar gekoppeld zijn:  $\sin x = \cos(\frac{\pi}{2} - x)$  voor alle  $x$ . Er wordt dus als het ware in één keer een familie van wiskundige functies ingevoerd. Dit is de eerste keer dat leerlingen zoiets mee maken en houdt een risico in dat ze de boel door elkaar gaan halen en uit nood maar allerlei zaken uit hun hoofd gaan leren i.p.v. de wiskundige essentie proberen te doorgronden.

### **3 Een alternatieve introductie van goniometrische functies**

Uit de eerder besproken vakliteratuur en uit de onderwijspraktijk komt het beeld naar voren dat de notie van wiskundige functie niet goed genoeg uit de verf komt bij de gangbare introductie van sinus en cosinus. Leerlingen blijven te veel hangen in de meetkundige context van hoeken en beschouwen sinus en cosinus niet als functies op  $\mathbb{R}$ . Redenen te over om eens uit te zoeken of er geen alternatief is.

Ik bespreek hier een idee dat ik vijftientig jaar geleden al had, maar dat eigenlijk pas sinds de komst van dynamische wiskunde software in voortgezet onderwijs gerealiseerd kan worden. Basisidee is om een vroege introductie van radialen als hoekmaat via de eenheidscirkelmethode uit de weg te gaan, maar in plaats daarvan de zogenaamde opwindfunctie die de reële lijn op de eenheidscirkel afbeeldt centraal te stellen en daarmee goniometrische functies via het begrip boog lengte

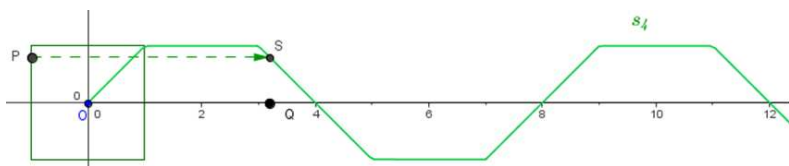
te introduceren. Essentieel onderdeel in deze aanpak is dat leerlingen eerst concreet kunnen oefenen en ervaring opdoen met deze tamelijk compliceerde constructie via opwindfunctie en coördinaatfuncties. Ik laat hen namelijk eerst met pen en papier werken met opwindfuncties gedefinieerd op regelmatige veelhoeken die de eenheidscirkel omschrijven en waarvoor functiewaarden uit te rekenen zijn d.m.v. rekenkundige bewerkingen, alvorens hen te laten onderzoeken wat er gebeurt bij een benadering van de eenheidscirkel met een regelmatige  $n$ -hoek voor grote  $n$ . Voor deze laatste stap heb ik ICT, in het bijzonder dynamische wiskunde applets gemaakt met GeoGebra, nodig.

In mijn aanpak komen diverse eigenschappen van sinus en cosinus eerst aan bod voor corresponderende functies gedefinieerd m.b.v. regelmatige veelhoeken die de eenheidscirkel omschrijven. De grafieken en eigenschappen van sinus en cosinus komen dan niet meer zo maar uit de lucht vallen, maar zijn limietgevallen van grafieken van corresponderende functies gedefinieerd m.b.v. regelmatige veelhoeken. Sowieso is de hoop en verwachting dat leerlingen sinus en cosinus meer leren appreciëren als reële functies door hun grafieken meer op de voorgrond te plaatsen bij het begin van de behandeling van goniometrische functies en niet slechts als toegift aan het einde van de leerroute te geven.

### **3.1 De opwindfunctie bij een regelmatige vierhoek**

De bedoeling in de instructiemethode is om leerlingen zelf de sinus- en cosinusfunctie te laten construeren door uit te gaan

van de grafiek van een periodieke beweging langs de rand van een regelmatig meetkundig object. Om dit proces te vereenvoudigen start ik met een regelmatige vierhoek die in het cartesisch vlak zodanig geïntendeerd wordt dat het punt  $(1, 0)$  zich halverwege op een verticale zijde bevindt en een punt  $P$  dat eenparig tegen de wijzers van de klok in beweegt langs de vierhoek vertrekkend vanuit  $(1, 0)$ ; zie Figuur 29, waarin een schermafbeelding van een bijpassende GeoGebra activiteit staat.



Figuur 29: Schermafbeelding van een GeoGebra activiteit waarin de grafiek van een sinusachtige functie getekend is en de connectie tussen het punt  $P$  op de regelmatige vierhoek en het punt  $S$  op de grafiek van de verticale positie van  $P$  uitgezet tegen afgelegde afstand van het punt  $P$  gevisualiseerd wordt. Het punt  $Q$  op de horizontale as correspondeert met de afgelegde weg. De grafiek hoort bij een functie aangeduid met  $s_4$ .

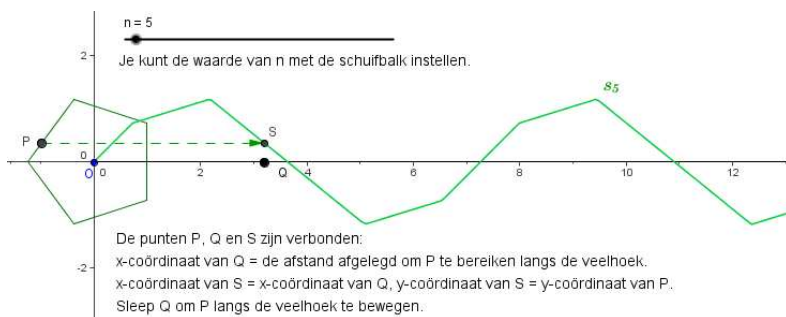
Leerlingen beginnen niet gelijk met zo'n dynamische applet, maar dokteren eerst met pen en papier uit wat de vorm van de grafiek van de verticale positie van het bewegende punt  $P$  uitgezet tegen de afgelegde afstand langs de regelmatige vierhoek precies is. Dit helpt hen naar verwachting met het goed begrijpen van de constructie: de coördinaten van een punt langs de vierhoek hangen af van de afgelegde afstand van het

bewegende punt. De sinusachtige functie  $s_4$  is op deze manier een samenstelling van de opwindfunctie, waarin elk positief reëel getal  $z$  afgebeeld wordt op het punt op de vierhoek dat bereikt wordt na een wandeling tegen de wijzers van de klok in langs de vierhoek over een afstand  $z$  met vertrekpunt in  $(1, 0)$ , de afbeelding die een punt in het vlak projecteert op de verticale as, en de afbeelding van de verticale as naar de reële getallen door de verticale coördinaat van elk punt te nemen. De cosinusachtige functie  $c_4$  is analoog als samenstelling van afbeeldingen te definiëren.

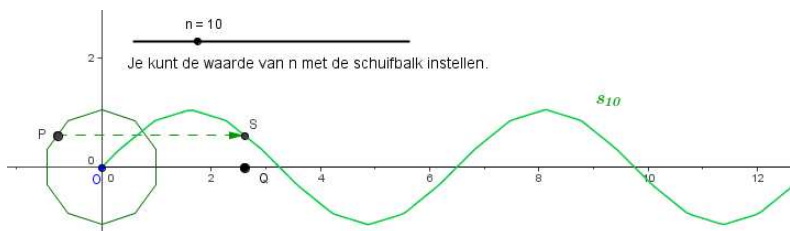
De volgende stap is om de grafiek in Figuur 27 uit te breiden over de negatieve horizontale as. Leerlingen moeten inzien dat deze uitbreiding correspondeert met wandelingen met de richting van de klok mee langs de regelmatige vierhoek. Dit is van belang als later in de leerroute de eenheidscirkelmethode behandeld wordt. Aan de hand van de grafiek van de sinusachtige functie  $s_4$  kunnen leerlingen allerlei eigenschappen bestuderen, zoals de periodiciteit (met periode 8), de eigenschappen van  $s_4(-x) = -s_4(x)$  en  $s_4(x + 4) = -s_4(x)$  voor alle  $x$ , en  $s_4(x) = x$  voor kleine waarden van  $x$ . Door ook de grafiek van de cosinusachtige functie  $c_4$  te bestuderen kunnen leerlingen ook inzien dat  $c_4(x) = s_4(4 - x)$  voor alle  $x$ . Kortom, leerlingen kunnen hun aandacht richten op functies gedefinieerd door de constructie m.b.v. een regelmatige vierhoek die de eenheidscirkel omschrijft en kunnen allerlei nuttige eigenschappen van deze functies zelf uitdokteren.

### 3.2 De opwindfunctie bij een regelmatige veelhoek

De eerder besproken constructie van een opwindfunctie bij een regelmatige vierhoek laat zich uitbreiden tot regelmatige  $n$ -hoeken die de eenheidscirkel omschrijven en die in het cartesisch vlak zodanig geïnterpreteerd worden dat het punt  $(1, 0)$  zich halverwege op een verticale zijde bevindt. Een punt  $P$  beweegt eenparig tegen de wijzers van de klok in langs de veelhoek vertrekkend vanuit  $(1, 0)$ . Door ook een beweging met de wijzers van de klok mee toe te staan worden sinus- en cosinusachtige functies  $s_n$  en  $c_n$  gedefinieerd op alle reële getallen. De grafieken van deze functies zijn voor een regelmatige 8-hoek door leerlingen met enige moeite nog wel met pen en papier te achterhalen, maar voor andere veelhoeken biedt dynamische wiskunde software uitkomst; zie de schermafdrucken van GeoGebra activiteiten in de Figuren 30 en 31.



Figuur 30: Grafiek van de sinusachtige functie  $s_5$  horende bij een regelmatige 5-hoek.



Figuur 31: Grafiek van de sinusachtige functie  $s_{10}$  bij een regelmatige 10-hoek.

Bij de grafiek van  $s_{10}$  horende bij een tienhoek ontstaat al het beeld van een grafiek van een sinusfunctie. De periode van de functie  $s_n$  komt voor grote  $n$  ook steeds dichterbij  $2\pi$ : de periode van de functies  $s_n$  horende voor bij een regelmatige  $n$ -hoek is namelijk gelijk aan  $2n \sin\left(\frac{\pi}{n}\right)$  en dus geldt

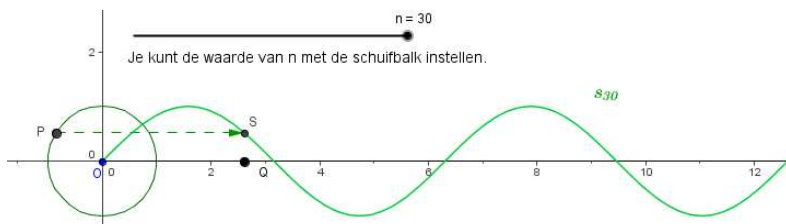
$$\lim_{n \rightarrow \infty} 2n \sin\left(\frac{\pi}{n}\right) = 2\pi \cdot \lim_{y \rightarrow \infty} y \sin\left(\frac{1}{y}\right) = 2\pi \cdot \lim_{x \downarrow 0} \frac{\sin x}{x} = 2\pi.$$

Merk op dat de constructie van de grafiek van de sinusachtige functie  $s_n$  door de speciale oriëntatie van de regelmatige veelhoek automatisch de volgende eigenschap oplevert: voor elke  $n$  geldt dat  $s_n(x) = x$  voor voldoende kleine  $x$ . De eigenschap dat  $s_n(-x) = -s_n(x)$  voor alle  $n$  en  $x$  is ook evident in deze constructie. Dit zijn eigenschappen die voor de echte sinusfunctie ook gelden.



### 3.3 De opwindfunctie bij de eenheidscirkel

Bij een keuze van  $n = 30$  lijkt de grafiek van  $s_{30}$  horende bij een regelmatige 30-hoek erg glad en is met het blote oog bijna geen verschil te zien met de grafiek van de sinusfunctie (Figuur 32)



Figuur 32: Grafiek van de sinusachtige functie  $s_{30}$  bij een regelmatige 30-hoek.

Dit is niet zo vreemd omdat de regelmatige 30-hoek met het blote oog niet goed meer te onderscheiden is van de eenheidscirkel. De functies  $s$  en  $c$  kunnen nu respectievelijk geïntroduceerd worden als limieten:

$$s \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} s_n, \quad c \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} c_n.$$

Uiteraard zijn de functies  $s$  en  $c$  welgedefinieerd en respectievelijk gelijk aan de sinus- en cosinusfunctie, maar de bewijsvoering vereist kennis van analyse op functieruimten. Belangrijker is dat leerlingen kunnen inzien dat een opwindfunctie voor de eenheidscirkel gedefinieerd kan worden op eenzelfde manier als bij regelmatige  $n$ -hoeken en dat het erg aannemelijk is dat de grafieken van  $s_n$  en  $c_n$  voor grote  $n$  erg lijken op de grafie-

ken van de corresponderende functies gedefinieerd m.b.v. de eenheidscirkel.

### 3.4 Van eenheidscirkelmethode naar de meetkundige aanpak van goniometrie

Tot nu toe is de alternatieve aanpak nog geen aandacht besteed aan draaiingshoeken en is er ook nog geen link gelegd met de meetkundige definitie van sinus en cosinus. Er van uitgaand dat leerlingen nog weten dat de omtrek van een eenheidscirkel gelijk is aan  $2\pi$  en dat een punt op de rand van de eenheidscirkel bij een draaiing over  $360^\circ$  op zichzelf afgebeeld wordt, kunnen zij in een paar opdrachten er toe gebracht worden zich te realiseren dat bij een wandeling tegen de wijzers van de klok langs de eenheidscirkel over een afstand  $x$  een draaiing hoort van het punt  $(1, 0)$  over een hoek gelijk aan  $\frac{360^\circ x}{2\pi}$ . Zij moeten hiervoor de link leggen tussen booglengtes en draaiingshoeken en proportioneel redeneren. Door bij een punt op de eenheidscirkel een rechthoekige driehoek binnen de eenheidscirkel te tekenen, zoals in Figuur 24, is door leerlingen in te zien dat de volgende formule geldt:

$$s(x) = \sin\left(\frac{180^\circ x}{\pi}\right).$$

Met deze formule kunnen leerlingen functiewaarden zoals  $\sin(3)$  uitrekenen m.b.v. tabellen met sinuswaarden of m.b.v. een rekenmachine ingesteld op graad als hoekmaat. Speciale functiewaarden als  $s(2\pi)$ ,  $s(\pi)$ ,  $s(\frac{\pi}{2})$ ,  $s(\frac{\pi}{4})$  kunnen hiermee exact berekend worden. De introductie van de radiaal als hoekmaat komt simpelweg neer op het besef dat een cirkelboog van lengte

te 1 correspondeert met een draaiingshoek gelijk aan  $\frac{180^\circ}{\pi}$  en dat dan dus inderdaad geldt

$$s(x) = \sin(x \text{ rad}).$$

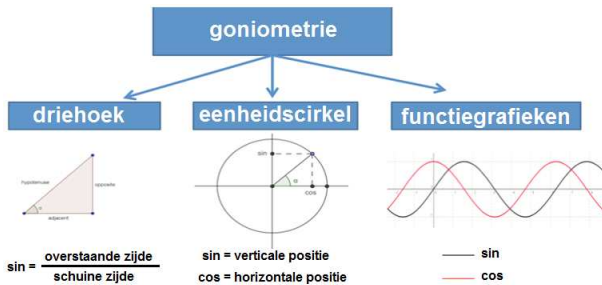
De stap om dan ook maar de functies  $s$  en  $\sin$  eenzelfde naam te geven is nu niet groot meer.

### 3.5 Achterliggende didactische overwegingen

Het moment is aangebroken om nog eens even stil te staan bij didactische principes die ten grondslag liggen aan de alternatieve aanpak en hoe ze in de instructiemethode uitgewerkt zijn.

Op de eerste plaats onderscheid ik drie wiskundige werkvelden die allemaal in min of meer gelijke mate aan bod moeten komen bij de introductie van sinus en cosinus: (i) de meetkunde van driehoeken; (ii) het gebruik van de eenheidscirkel bij goniometrie; (iii) het terrein van goniometrische functies en hun grafieken. In Figuur 33 wordt voor elk werkveld de meest dominante wiskundige representatie getoond.

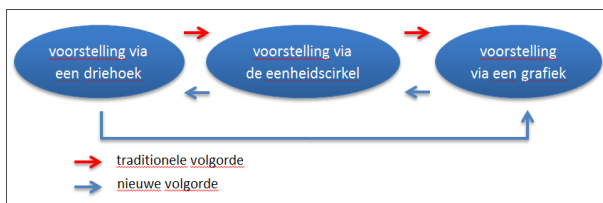
Hoe en in welke volgorde deze wiskundige objecten in onderwijs aan bod komen is een kwestie van traditie, smaak, of een doelbewuste keuze. In Figuur 34 is aangegeven dat men traditioneel met de meetkundige aanpak van driehoeken begint, daarna overstapt op de eenheidscirkel en tenslotte bij goniometrische functies uitkomt. Nadelen zijn dat de eenheidscirkel en de radiaal als hoekmaat een beetje uit de lucht komen vallen en leerlingen hier maar beperkt mee kunnen oefenen, het functiebegrip in deze benadering onvoldoende aandacht krijgt, de



Figuur 33: Verschillende representaties bij goniometrie.

vorm van de grafieken van sinus en cosinus niet op inzichtelijke wijze behandeld wordt en leerlingen niet goed voorbereid worden op belangrijke eigenschappen van goniometrische functies. Figuur 34 illustreert ook een alternatieve aanpak. Opnieuw is de meetkunde van driehoeken het vertrekpunt, maar in plaats van direct doorgaan met de eenheidsccirkel kies ik er voor om eerst maar eens periodieke functies en hun grafieken te behandelen die ontstaan door wandelingen over regelmatige veelhoeken en dan pas, als limietgeval, naar de eenheidsccirkel over te stappen. Omdat ik op deze manier vroegtijdig de aandacht verleg van hoeken naar functies op reële getallen ben ik aan het einde van de leerroute wel genoodzaakt om de link tussen eenheidsccirkel en meetkunde van driehoeken te behandelen; ik keer dus terug naar waar ik in de leerroute begon.

Ik ben van mening dat afstand en afgelegde weg voor leerlingen beter hanteerbare begrippen zijn dan de toch wat complexe notie van hoek. Maar om gelijk met booglengtes te beginnen voert te ver. Liever laat ik leerlingen eerst met pen en papier puzzelen op grafieken van een sinus- en cosinusachtige func-



Figuur 34: Traditionele versus alternatieve aanpak van goniometrie.

ties die ontstaan uit wandelingen op een regelmatige vier- en achthoek. Zo kunnen ze alvast wennen aan de constructie van een periodieke functie op grond van een zogenaamde opwindfunctie. Door m.b.v. het dynamische wiskunde programma GeoGebra te kijken naar regelmatige  $n$ -hoeken met grotere waarden voor  $n$ , ontstaat het beeld van functies die horen bij de eenheidscirkel.

Deze alternatieve aanpak is geworteld in Tall's theorie over drie stadia in het begrijpen van een wiskundige object (Tall et al., 2000). Voor goniometrische functies betekent dit het volgende: eerst moeten leerlingen leren om de berekening van een functiewaarde als een procedure of stap-voor-stap algoritme te doorlopen en alle details in zich opnemen. Dit gebeurt meestal met pen en papier of via concrete objecten. Als de leerlingen de procedures herhaaldelijk hebben toegepast en voldoende in de gelegenheid zijn geweest om op hun handelen te reflecteren, dan kunnen ze hierna de procedure als een proces beschouwen dat op elke regelmatige  $n$ -hoek van toepassing is en in het limietgeval leidt tot een proces dat toepasbaar is voor een eenheidscirkel. Uiteindelijk wordt de eenheidscirkel een zogenaamd procept waarmee leerlingen kunnen anticiperen

ren op resultaten van een proces zonder dit nog daadwerkelijk uit te voeren. Zij kunnen bijvoorbeeld m.b.v. de eenheidscirkel allerlei eigenschappen van goniometrische functie begrijpen en afleiden zonder dit in detail, stap-voor-stap uit te hoeven werken. De behandeling van periodieke functies gedefinieerd m.b.v. regelmatige veelhoeken en hun eigenschappen is een concreet voorproefje op dit begripsniveau.

## 4 Ervaringen in de klas met de nieuwe aanpak

Het bedenken van een alternatieve aanpak om goniometrische functies te introduceren en het theoretisch onderbouwen van het ontwerp is één ding, maar dat betekent nog niet dat de instructie ook in schoolpraktijk goed werkt. Dit laatste moet ook echt onderzocht worden.

De eerste keer is dit in 2008 gebeurd in een vwo-4 klas met leerlingen die voor Wiskunde D gekozen hadden. Twee master studenten, Rafiq Mehdiyev en Georgia Papageorgiou, onderzochten in een kortlopend onderzoeksproject hoe de lessen in praktijk verliepen en hoe de leerlingen deze aanpak ervoeren. Hun conclusies waren dat de methode voor deze groep leerlingen positief uitpakte: leerlingen vonden de opdrachten goed te doen, het ICT gebruik inspirerend en de lessen leerzaam. Hoewel de alternatieve leerroute wat langer is dan in de traditionele aanpak (zie Figuur 34), bleek achteraf dat de extra benodigde tijd volledig gecompenseerd werd door het betere begrip van de leerlingen van de hoekmaat radiaal: toen de leraar van de klas namelijk later de leerlingen sommen uit

het wiskundeboek over omrekening van graden naar radialen en omgekeerd liet maken, zeiden leerlingen tegen haar: “Dit hebben we al gehad.” Vervolgens maakten zij de sommen in sneltreinvaart en foutloos. Bij de nabespreking van de resultaten van het onderzoeksproject opperde de lerares dat zij de alternatieve aanpak ook wel voor een “gewone” vwo-klas bij Wiskunde A of Wiskunde B geschikt en aantrekkelijk vond. Met name de constructie van de grafiek van de sinusfunctie via een limietproces van sinusachtige functies bij regelmatige veelhoeken sprak haar erg aan, omdat het volop aanknopingspunten biedt om de vorm van de sinusgrafiek te bespreken in de klas en om leerlingen beter begrip van de sinusfunctie bij te brengen.

Pas vier jaar later is de stier bij de hoorn gevat door Ozcan Demir in zijn master research project en is de alternatieve aanpak beproefd in twee vwo-4 klassen met leerlingen die Wiskunde B gekozen hebben. Ten tijde van het schrijven van dit artikel worden de onderzoeksresultaten geanalyseerd, maar als voorlopige conclusie mag vermeld worden dat de instructie succesvol was. De nieuwe aanpak was ook toegankelijk voor wie alleen Wiskunde B in zijn of haar vakkenpakket heeft. Uit een diagnostische toets voorafgaand aan het experiment in de klas bleek dat de leerlingen inderdaad last hadden van alternatieve concepties op het gebied van goniometrie zoals eerder beschreven. Uit een toets na afloop van het experiment in de klas bleek dat de alternatieve instructiemethode de leerlingen inderdaad geholpen heeft in het overwinnen of bijstellen van misconcepties en geleid heeft tot een goed begrip van sinus en cosinus als wiskundige functie, inclusief het domein, bereik en periodiciteit van deze functies. De natoets wekte wel de indruk

dat dat de link tussen de eenheidscirkel en de meetkunde van driehoeken voor sommige studenten nog fragiel was, maar uit gestructureerde interviews met verschillende leerlingen achteraf bleek dat ze toch meer hadden opgestoken dan op het eerste gezicht uit de diagnostische toets naar voren leek te komen.

## 5 Conclusie

Uit vakdidactisch onderzoek en praktijkervaring van vwo-docenten betrokken bij dit onderzoek komt naar voren dat de leeropbrengst van de alternatieve aanpak om goniometrische functies te introduceren goed is, geen extra tijd kost en geen drastische wijzigingen in het onderwijs vereist. De experimenten in de klas waren tot nu toe hoopgevend, maar of de alternatieve aanpak ook op grote schaal gaat werken, en ook bij leraren die niet bij voorbaat enthousiast zijn, moet nog worden onderzocht. Voor dit uitproberen in de klas is het bij het onderzoek gebruikte lesmateriaal, zowel in de Engelse als Nederlandse taal, beschikbaar gesteld op de webpagina [www.science.uva.nl/~heck/goniometrie](http://www.science.uva.nl/~heck/goniometrie).

## 6 Referenties

- Akkoç, H. (2008). Pre-service mathematics teachers' concept image of radian. *International Journal of Mathematical Education in Science and Technology*, 39(7), 857–878.
- Akkoç, H., & Akbaş Gül, N. (2010). Analysis of a teaching approach aiming at eliminating student difficulties with radian. *Ankara University, Journal of Faculty of Educational Sciences*, 43(1), 97–129.



Online: <http://dergiler.ankara.edu.tr/dergiler/40/1342/15553.pdf>

Behr, M.J., Harel, G., Post, T., & Lesh, R. (1994). Rational number, ratio and proportion. In: D.A. Grouws (red.), *Handbook of Research on Mathematics Teaching and Learning*, (pp. 296–333). Reston, VA: National Council of Teachers of Mathematics.

Blackett, N., & Tall, D.O. (1991). Gender and the versatile learning of trigonometry using computer software. In: F. Furinghetti (red.), *Proceedings of the 15th Meeting of the International Group for the Psychology of Mathematics Education*, vol. 1 (pp. 144-151). Assisi, Italië.

Brown, S.A. (2005). *The Trigonometric Connection: Students' Understanding of Sine and Cosine*. PhD thesis. Illinois State University, USA.

Challenger, M. (2009). *From Triangles to a Concept: A Phenomenographic Study of A-level Students' Development of the Concept of Trigonometry*. PhD thesis. University of Warwick, UK.

Online: [http://wrap.warwick.ac.uk/1935/1/WRAP\\_THESIS\\_Challenger\\_2009.pdf](http://wrap.warwick.ac.uk/1935/1/WRAP_THESIS_Challenger_2009.pdf)

Delice, A. (2003). *A comparative study of students' understanding of trigonometry in the United Kingdom and the Turkish Republic*. PhD thesis. University of Leeds, UK.

Delice, A., & Roper, T (2006). *Implications of a comparative study for mathematics education in the English education system*. *Teaching Mathematics and its Applications*, 25(2), 64–72.

Fi, C. (2003). *Preservice Secondary School Mathematics Teachers' Knowledge of Trigonometry: Subject Matter Content Knowledge, Pedagogical Content Knowledge and Envisioned Pedagogy*. PhD. thesis. University of Iowa, USA.

- Gür, H. (2009). Trigonometric learning. *New Horizons in Education*, 57(1), 67–88.  
 Online: [www.hktalhk.edu.hk/hkta/NewHorizon/abstract/2009May/6.pdf](http://www.hktalhk.edu.hk/hkta/NewHorizon/abstract/2009May/6.pdf)
- Hart, K.M. (1981). *Children's Understanding of Mathematics*, 11–16. London: John Murray.
- Johari, N.A., Chan, L.O., Ramli, R., & Ahmat, N. (2010). The effect of GSP on students' understanding in the graphs of trigonometric functions. In *Electronic Proceedings of the 15th Asian Technology Conference in Mathematics*.  
 Online: [http://atcm.mathandtech.org/ep2010/regular/3052010\\_18310.pdf](http://atcm.mathandtech.org/ep2010/regular/3052010_18310.pdf)
- Johnson, J., & Walker, M. (2011). Trigonometry students' knowing when to use hand-held CAS technology to make sense of mathematics. *Journal of Mathematical Sciences & Mathematics Education*, 6(2), 17–33.  
 Online: <http://msme.us/2011-2-4.pdf>
- Kendall, M., & Stacey, K. (1996). Trigonometry: Comparing ratio and unit circle methods. In P. Clarkson (red.), *Technology in Mathematics Education: Proceedings of the Nineteenth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 322–329). Melbourne: MERGA.
- Kendall, M., & Stacey, K. (1997). Teaching trigonometry. *Vinculum*, 29(1), 4–8.  
 Online: <http://staff.edfac.unimelb.edu.au/~kayecs/publications/1997/KendalStacey-Trig.pdf>
- Kissane, B., & Kemp, M. (2009). Teaching and learning trigonometry with technology. In *Electronic Proceedings of the 14th Asian Technology Conference in Mathematics*. Beijing, China.  
 Online: [http://atcm.mathandtech.org/EP2009/papers\\_full/2812009\\_17288.pdf](http://atcm.mathandtech.org/EP2009/papers_full/2812009_17288.pdf)

Lamon, S.A. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In: F.K. Lester (red.), *Second Handbook of Research on Mathematics Teaching and Learning*, (pp. 629–667). Charlotte, NC: Information Age Publishing and National Council of Teachers of Mathematics.

Martínez-Sierra, G. (2008a). On the transit from trigonometry to calculus: The case of the conceptual breaks in the construction of the trigonometric functions in school. In *Electronic Proceedings of the 11th International Conference on Mathematics Education*, Mexico. Online: <http://tsg.icme11.org/document/get/667>

Martínez-Sierra, G. (2008b). From the analysis of the articulation of the trigonometric functions to the corpus of Eulerian analysis to the interpretation of the conceptual break present in its scholar structure. In *Electronic Proceedings of the ICME Satellite Meeting “History and Pedagogy of Mathematics”*, Mexico. Online: [www.matedu.cicata.ipn.mx/archivos/Gustavo/HPM.Proceddings\\_Extenso.pdf](http://www.matedu.cicata.ipn.mx/archivos/Gustavo/HPM.Proceddings_Extenso.pdf)

Mazziotta, E. (1949). The basis concept of trigonometry. *Mathematics Magazine*, 22(3), 139–150.

Moore, K.C. (2009). Trigonometry, technology, and didactic objects. In S. Swars, D. Stinson, & S. Lemons-Smith (red.), *Proceedings of 31st Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, vol. 5 (pp. 1480-1488). Atlanta, Georgia. Online: <http://www.pmena.org/2009/proceedings/TECHNOLOGY/techRR369892.pdf>

Moore, K.C. (2010). *The Role of Quantitative Reasoning in Pre-calculus Students Learning Central Concepts of Trigonometry*. PhD thesis. Illinois State University, USA. Online: [www.patthompson.net/PDFversions/Theses/2010Moore.pdf](http://www.patthompson.net/PDFversions/Theses/2010Moore.pdf)

- Orhun, N. (2010). The gap between real numbers and trigonometric relations. *Quaderni di Ricerca in Didattica*, 20. G.R.I.M. (University of Palermo, Italy).  
Online: [http://dipmat.math.unipa.it/~grim/QRDM\\_Orhun\\_20\\_2010.pdf](http://dipmat.math.unipa.it/~grim/QRDM_Orhun_20_2010.pdf)
- Ross, J.A., Bruce, C.D., & Sibbald, T. (2011). Sequencing computer-assisted learning of transformations of trigonometric functions. *Teaching Mathematics and its Applications*, 30(3), 120–137.
- Takači, D., Herceg D., & Stojković, R. (2005). Possibilities and limitations of scientific workplace in studying trigonometric functions. *The Teaching of Mathematics*, 8(2), 61-72.  
Online: <http://elib.mi.sanu.ac.rs/files/journals/tm/15/tm822.pdf>
- Tall, D.O., Thomas, M., Davis, E., Gray, E., & Simpson, A. (2000). What is the object of the encapsulation of a process? *Journal of Mathematical Behavior*, 18(2), 1–19.
- Tompoglou, I. (2007). *An RME Approach to Trigonometry Using a Ferris Wheel and the Unit Circle*. Master thesis. Universiteit van Amsterdam.
- Van Asch, A.G., & Van der Blij, F. (1992). Hoeken en hun maat. *CWI Syllabus* 29, Amsterdam.
- Van Asch, A.G., & Van der Blij, F. (1997). Goniometry between geometry and analysis. *International Journal of Mathematical Education in Science and Technology*, 28(1), 85–96.
- Van Asch, A.G., & Van der Blij, F. (2002). Sinus en cosinus, één functionaalvergelijking? *Euclides*, 78(2), 60–62.
- Weber, K. (2005). Students' understanding of trigonometric functions. *Mathematics Education Research Journal*, 17(3), 94–115.
- Weber, K. (2008). Teaching trigonometric functions: Lessons learned from research. *Mathematics Teacher*, 102(2), 144–150.

Zengin, Y., Furkan, H., & Kutluca, T. (2012). The effect of dynamic mathematics software GeoGebra on student achievement in teaching of trigonometry. *Procedia - Social and Behavioral Sciences*, *31*, 183–187.

# Derdegraads vergelijkingen oplossen

**Frans Keune, Radboud Universiteit Nijmegen**

**e-mail: keune@math.ru.nl**

Voor het oplossen van kwadratische vergelijkingen is er een recept dat al duizenden jaren oud is. In het begin van de zestiende eeuw hebben Tartaglia en Del Ferro een recept gevonden voor vergelijkingen van graad 3 (en iets later deed Ferrari dat voor vergelijkingen van graad 4). Drie eeuwen daarna liet Abel zien dat zoiets niet bestaat voor vergelijkingen van graad 5 en hoger. Echt diep inzicht in het oplossen van vergelijkingen werd bereikt door de op twintigjarige leeftijd in een duel omgekomen Evariste Galois (1811-1832). De theorie van Galois, die pas in 1846 posthuum door Liouville was gepubliceerd, betekende het begin van de ‘moderne’ abstracte algebra.

Vroeger werd meetkunde gebruikt bij het oplossen van verge-

lijkingen. Daar kwamen die vergelijkingen vaak ook vandaan: kwadratische vergelijkingen gingen over oppervlakten en kubische over inhoud. Anders dan nu vonden algebraïsche manipulaties hun rechtvaardiging in de meetkunde. Ook was men niet vertrouwd met negatieve getallen. Er werden dan ook bij bijvoorbeeld kwadratische vergelijkingen meerdere gevallen onderscheiden. Daar gaan we hier niet op in. We gebruiken onze huidige kennis van het getalsysteem met z'n bekende rekenregels.

We zullen met de methode van Tartaglia een 'exacte' uitdrukking bepalen voor het reële getal  $\psi$  dat voldoet aan  $\psi^3 = \psi + 1$ . Dit getal heeft eigenschappen die lijken op die van de gulden snede, zie daarvoor de appendix.

## 1 Vereenvoudiging van de kubische vergelijking

De algemene gedaante van een kubische (= derdegraads) vergelijking is

$$ax^3 + bx^2 + cx + d = 0.$$

Hierbij zijn  $a, b, c, d$  getallen. We laten hier in het midden hoe ruim we het begrip getal opvatten. Van belang is te weten hoe met getallen gerekend wordt. Verder moet gelden  $a \neq 0$ , want anders hebben we geen derdegraads vergelijking. Delen we door  $a$ , dan krijgen we de gedaante

$$x^3 + ex^2 + fx + g = 0,$$

en als we  $x = y - \frac{1}{3}e$  substitueren, dan wordt het linker lid

$$\begin{aligned} & (y - \frac{1}{3}e)^3 + e(y - \frac{1}{3}e)^2 + f(y - \frac{1}{3}e) + g \\ &= y^3 - ey^2 + \frac{1}{3}e^2y - \frac{1}{27}e^3 + e(y^2 - \frac{2}{3}ey + \frac{1}{9}e^2) + f(y - \frac{1}{3}e) + g \\ &= y^3 + (-\frac{1}{3}e^2 + f)y + \frac{2}{27}e^3 - \frac{1}{3}ef + g. \end{aligned}$$

We krijgen zo een vergelijking van de gedaante

$$x^3 + px + q = 0.$$

## 2 De vergelijking $x^3 = 1$

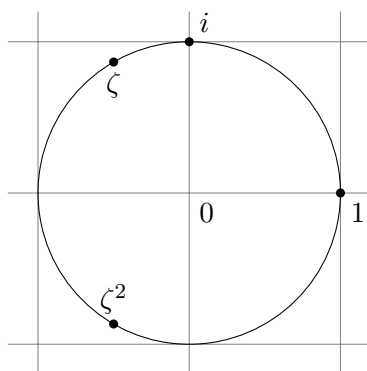
Dit is het speciale geval  $p = 0$  en  $q = -1$ . In het lichaam  $\mathbb{R}$  van de reële getallen is er één oplossing. In het lichaam  $\mathbb{C}$  van de complexe getallen zijn er drie:

$$x^3 - 1 = (x - 1)(x^2 + x + 1).$$

Er zijn nog 2 niet-reële complexe oplossingen:  $x = -\frac{1}{2} + \frac{1}{2}\sqrt{-3} = \zeta$  en  $x = -\frac{1}{2} - \frac{1}{2}\sqrt{-3} = \zeta^2$ . We hebben dan

$$x^3 - 1 = (x - 1)(x - \zeta)(x - \zeta^2).$$

De getallen 1,  $\zeta$  en  $\zeta^2$  zijn de drie mogelijke derdemachts wortels uit 1. De  $n$  oplossingen van  $x^n = 1$ , waarbij  $n$  een natuurlijk getal is, noemt men  $n$ -de eenheidswortels. Hier hebben we dus de drie derde eenheidswortels.



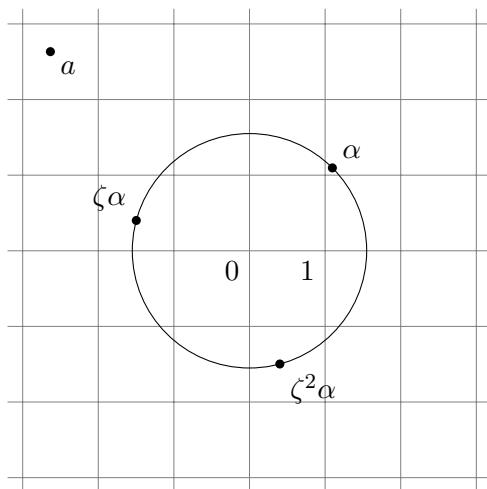


### 3 De vergelijking $x^3 = a$

Het speciale geval  $p = 0$ . In  $\mathbb{C}$  is er voor iedere  $a$  een oplossing en die kunnen we schrijven als een derdemachts wortel:  $x = \sqrt[3]{a}$ . Laten we  $\alpha$  schrijven voor  $\sqrt[3]{a}$ . Dan

$$x^3 - a = x^3 - \alpha^3 = (x - \alpha)(x^2 + \alpha x + \alpha^2) = (x - \alpha)(x - \zeta\alpha)(x - \zeta^2\alpha).$$

De andere twee oplossingen zijn dan  $x = \zeta\alpha$  en  $x = \zeta^2\alpha$ , aangenomen dat  $a \neq 0$ , anders staat er drie keer dezelfde oplossing, namelijk  $x = 0$ . De drie oplossingen liggen gelijkkelijk verdeeld op een cirkel om 0 met straal  $|\alpha| = \sqrt[3]{|a|}$ .



## 4 De vergelijking $x^3 + x - 2 = 0$

Duidelijk is dat  $x = 1$  een oplossing is van de vergelijking  $x^3 + x - 2 = 0$ . Verder

$$\begin{aligned}x^3 + x - 2 &= (x^3 - 1^3) + (x - 1) - (2 - 2) = (x - 1)(x^2 + x + 1) + (x - 1) \\ &= (x - 1)(x^2 + x + 2).\end{aligned}$$

Er zijn nog twee oplossingen, nl. de oplossingen van  $x^2 + x + 2 = 0$ , en dat zijn

$$x = -\frac{1}{2} + \frac{1}{2}\sqrt{-7} \quad \text{en} \quad x = -\frac{1}{2} - \frac{1}{2}\sqrt{-7},$$

niet-reële complexe oplossingen. Doordat we al een oplossing hadden, konden we de het probleem terugbrengen tot het oplossen van een kwadratische vergelijking.

We doen nu net alsof we dat niet door hebben en gaan de vergelijking oplossen volgens de methode van Tartaglia. Zo kunnen we zien hoe het werkt en ook dat het helemaal niet zo gemakkelijk is op deze manier zo'n simpele oplossing als  $x = 1$  terug te vinden. We schrijven de onbekende  $x$  als som van twee nieuwe onbekenden:  $x = u + v$ . De vergelijking wordt dan

$$(u + v)^3 + (u + v) - 2 = 0,$$

of ook

$$u^3 + 3u^2v + 3uv^2 + v^3 + u + v - 2 = 0.$$

We schrijven het een beetje anders

$$u^3 + v^3 + 3uv(u + v) + (u + v) - 2 = 0.$$

We nemen nu  $u$  en  $v$  zo dat  $3uv = -1$ . Hun som is  $x$  en hun product is dus  $-\frac{1}{3}$ . Dan

$$u^3 + v^3 - 2 = 0 \quad \text{en} \quad uv = -\frac{1}{3}.$$

We elimineren  $v$ :

$$u^3 - \frac{1}{27u^3} - 2 = 0, \quad \text{ofwel} \quad 27u^6 - 54u^3 - 1 = 0.$$

We hebben nu een kwadratische vergelijking in  $u^3$ . Hadden we  $u$  geëlimineerd, dan kregen we dezelfde kwadratische vergelijking, maar dan in  $v^3$ . Oplossen geeft

$$u^3 = \frac{9 + \sqrt{21}}{9} \quad \text{en} \quad v^3 = \frac{9 - \sqrt{21}}{9}$$

(of andersom, maar dat doet er hier niet toe). We krijgen

$$x = \sqrt[3]{\frac{9 + \sqrt{21}}{9}} + \sqrt[3]{\frac{9 - \sqrt{21}}{9}}.$$

Zodra de eerste derdemachtswortel gekozen is, ligt de tweede vast:  $uv = -\frac{1}{3}$ . Dit zijn derdemachtswortels uit reële getallen en daar kiezen we gewoonlijk reële getallen voor. Omdat  $x = 1$  de enige reële oplossing van de vergelijking is, staat hier dus 1. En dat is niet meteen duidelijk. Wel zodra je door hebt dat

$$\left(\frac{3 + \sqrt{21}}{6}\right)^3 = \frac{9 + \sqrt{21}}{9},$$

want dan volgt eenvoudig

$$x = \frac{3 + \sqrt{21}}{6} + \frac{3 - \sqrt{21}}{6} = 1.$$

Een andere oplossing krijg je door een andere derdemachtswortel te nemen:

$$\begin{aligned} x &= \frac{-1 + \sqrt{-3}}{2} \cdot \frac{3 + \sqrt{21}}{6} + \frac{-1 - \sqrt{-3}}{2} \cdot \frac{3 - \sqrt{21}}{6} \\ &= \frac{-3 - \sqrt{21} + s\sqrt{-3} + 3\sqrt{-7}}{12} + \frac{-3 + \sqrt{21} - 3\sqrt{-3} + 3\sqrt{-7}}{12} \\ &= \frac{-1 + \sqrt{-7}}{2}. \end{aligned}$$

## 5 De vergelijking $x^3 - 7x + 6 = 0$

Je ziet al snel de drie oplossingen:  $x = 1$ ,  $x = 2$  en  $x = -3$ . De bijbehorende ontbinding is:  $x^3 - 7x + 6 = (x - 1)(x - 2)(x + 3)$ . We doen weer of we dat niet door hebben en gaan het recept van Tartaglia volgen. We schrijven weer  $x = u + v$ . De vergelijking wordt dan

$$u^3 + v^3 + 3uv(u + v) - 7(u + v) + 6 = 0.$$

Nu nemen we  $u$  en  $v$  zo dat  $3uv = 7$ . Net als in de vorige paragraaf leidt dit tot een kwadratische vergelijking in  $u^3$ :

$$27u^6 + 6 \cdot 27u^3 + 7^3 = 0.$$

We kunnen dus nemen  $u^3 = -3 + \frac{10}{9}\sqrt{-3}$  en  $v^3 = 3 - \frac{10}{9}\sqrt{-3}$ . We krijgen

$$x = \sqrt[3]{-3 + \frac{10}{9}\sqrt{-3}} + \sqrt[3]{-3 - \frac{10}{9}\sqrt{-3}}.$$

Afhankelijk van de gekozen derdemachtswortel is dit dus een ingewikkelde manier om 1, 2 of  $-3$  te schrijven. Merk ook

nog op dat er in de formule niet-reële complexe getallen staan. Daar hadden ze in de zestiende eeuw best problemen mee. Ze deden al moeilijk over negatieve getallen.

We hebben dus  $9u^3 = -27 + 10\sqrt{-3}$ , ofwel

$$(\sqrt{-3})^4 u^3 = -(\sqrt{-3})^6 + 10\sqrt{-3}$$

en na delen door  $\sqrt{-3}$ :

$$(\sqrt{-3} \cdot u)^3 = (\sqrt{-3})^5 + 10 = 9\sqrt{-3} + 10 = (-2 + \sqrt{-3})^3.$$

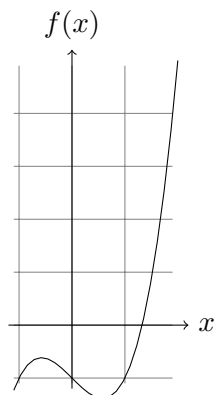
Nemen we  $\sqrt{-3}u = -2 + \sqrt{-3}$ , ofwel  $u = 1 + \frac{2}{3}\sqrt{-3}$  en daarmee  $v = 1 - \frac{2}{3}\sqrt{-3}$ , dan krijgen we

$$x = u + v = 1 + \frac{2}{3}\sqrt{-3} + 1 - \frac{2}{3}\sqrt{-3} = 2.$$

De andere twee oplossingen krijg je weer door de andere derdemachtswortels te nemen.

## 6 De vergelijking $x^3 - x - 1 = 0$

Nu eens een vergelijking zonder voor de hand liggende oplossingen. Hiernaast de grafiek van  $f(x) = x^3 - x - 1$ . Er is een unieke reële oplossing. Dit getal wordt wel het *plastisch* getal genoemd, zie ook appendix A. We zullen het noteren als  $\psi$ . We bepalen deze  $\psi$  ‘exact’.



We schrijven weer  $x = u + v$  en krijgen dan

$$u^3 + v^3 + 3uv(u + v) - (u + v) - 1 = 0.$$

We nemen  $3uv = 1$ . Dat geeft de vergelijking

$$27u^6 - 27u^3 + 1 = 0.$$

We krijgen zo

$$\psi = u + v = \sqrt[3]{\frac{9 + \sqrt{69}}{18}} + \sqrt[3]{\frac{9 - \sqrt{69}}{18}}.$$

## 7 De formule van Cardano

Pas je het recept van Tartaglia toe op de vergelijking  $x^3 + px + q = 0$ , dan is het resultaat de *formule van Cardano*:

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{p}{3}\right)^3 + \left(\frac{q}{2}\right)^2}}.$$

Maak je een keuze voor een van de derdemachts wortels, dan bepaalt dat ook de andere derdemachtswortel: hun product is

gelijk aan  $-\frac{1}{3}p$ . De gevonden uitdrukking voor  $\psi$  in de vorige paragraaf krijg je dus ook door in de formule van Cardano  $p = q = 1$  te nemen.

## 8 Wat is oplossen?

In paragraaf 6 hebben we de vergelijking  $x^3 - x - 1 = 0$  opgelost. Natuurlijk, in  $x^3 - x - 1 = 0$  is  $x$  op impliciete wijze vastgelegd en in de oplossing wordt  $x$  expliciet gegeven. Voor zo'n expliciete beschrijving moet het wel duidelijk zijn wat je daarin mag gebruiken. Traditioneel komt dat neer op: de coëfficiënten van de vergelijking, de bewerkingen optellen etc., d.w.z. de bewerkingen die er in een lichaam zijn, en daarbij ook wortels, in de zin van vierkantswortels, derdemachtswortels, etc. De formule van Cardano is zo'n wortelformule voor kubische vergelijkingen. Je kunt een benadering van de oplossing vinden door de gevonden uitdrukking met wortels in te toetsen in een rekenmachine, maar in de zestiende eeuw kon dit natuurlijk niet. Voor  $\psi$  krijg je

$$\psi = 1,324717957244746\dots$$

De gevonden exacte oplossing vertelt iets over de algebraïsche eigenschappen, maar  $\psi^3 = \psi + 1$  is eigenlijk ook wel een mooie karakterisering van  $\psi$ .

Heeft een derdegraads veelterm met reële coëfficiënten drie reële nulpunten, zoals in paragraaf 5, dan komen er in de formule van Cardano vierkantswortels uit negatieve getallen voor, niet-reële getallen dus. In die tijd was dat een mysterie, maar

omdat het wel werkte was men geneigd zoiets onbestaanbaars te accepteren.

## 9 Vergelijkingen van hogere graad

In de zestiende eeuw is er ook een recept bedacht voor het oplossen van vierdegraads vergelijkingen. Dit is gedaan door Ferrari, een leerling van Cardano. Het heeft drie eeuwen geduurd totdat Abel aantoonde dat er voor vijfdegraads vergelijkingen geen algemene wortelformule bestaat. Galois heeft laten zien hoe je oplosbaarheid van een vergelijking kunt vertalen in groepentheorie. Voor vijfdegraads vergelijkingen komt het erop neer dat je dat kunt zien aan de groep van alle 120 permutaties van vijf elementen.

## 10 Benaderen

Ben je geïnteresseerd in een goede benadering van een oplossing in plaats van een exacte uitdrukking met wortels, dan is de methode van Newton zeer geschikt. Voor het benaderen van een nulpunt van een reële differentieerbare functie  $f(x)$  gaat dat als volgt.

1. Laat  $a \in \mathbb{R}$  zo zijn dat  $f'(a) \neq 0$ . De raaklijn in het punt  $(a, f(a))$  aan de grafiek van  $f(x)$  is de grafiek van de lineaire functie  $f_a(x)$ :

$$f_a(x) = f(a) + f'(a)(x - a).$$



2. Omdat  $f'(a) \neq 0$  heeft  $f_a(x)$  een uniek nulpunt, namelijk

$$a - \frac{f(a)}{f'(a)}.$$

3. We hebben zo een transformatie  $g$  van  $\mathbb{R}$  met de nulpunten van  $f(x)$  weggelaten:

$$g: x \mapsto x - \frac{f(x)}{f'(x)}.$$

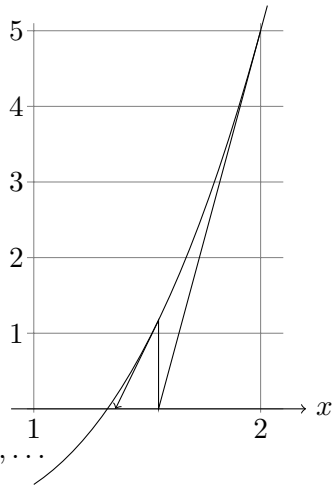
4. Iteratie van  $g$  geeft rijen getallen die naar de nulpunten van  $f$  convergeren. het nulpunt dat zo benaderd wordt is afhankelijk van de startwaarde.

Voor het benaderen van het plastisch getal  $\psi$  kunnen we de functie  $f(x) = x^3 - x - 1$  nemen. Dan wordt de functie  $g$ :

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

Nemen we de startwaarde  $a_0 = 2$ , dan krijgen we de rij  $a_0, a_1, a_2, \dots$  die vast ligt door

$$\begin{cases} a_0 = 2, \\ a_{n+1} = \frac{2a_n^3 + 1}{3a_n^2 - 1} \quad \text{voor } n = 0, 1, 2, \dots \end{cases}$$



Dit is een dalende rij getallen die nadert naar  $\psi$ . Bij  $n = 6$  zijn de eerste 15 decimalen al gevonden:

$$\psi \approx a_6 \approx 1,324717957244746.$$

In de figuur is voor de duidelijkheid de schaal op de  $x$ -as verdrievoudigd.

## A Padovangetallen

In deze paragraaf iets over de achtergrond van het getal  $\psi$ .

De verhouding van opeenvolgende Fibonacci-getallen nadert naar de gulden snede. Ook voor het plastisch getal  $\psi$  hebben zo'n rij: de rij van *Padovan-getallen*, genoemd naar Richard Padovan, die ze toeschreef aan de Nederlandse architect Hans van der Laan:

$$\begin{cases} p_0 = p_1 = p_2 = 1 \\ p_{n+3} = p_n + p_{n+1} \quad \text{voor } n = 0, 1, 2, \dots \end{cases}$$

Dus:

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$p_n$	1	1	1	2	2	3	4	5	7	9	12	16	21	28	37	49	65

Er geldt

$$\lim_{n \rightarrow \infty} \frac{p_{n+1}}{p_n} = \psi.$$

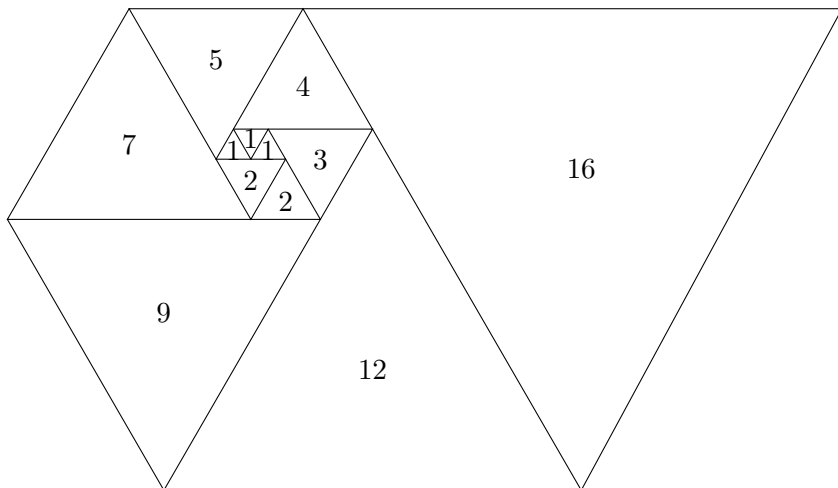
Merk op dat als je al weet dat deze rij van quotiënten een limiet heeft, die limiet  $\psi$  moet zijn:

$$\frac{p_{n+3}}{p_{n+2}} \cdot \frac{p_{n+2}}{p_{n+1}} \cdot \frac{p_{n+1}}{p_n} = 1 + \frac{p_{n+1}}{p_n}.$$

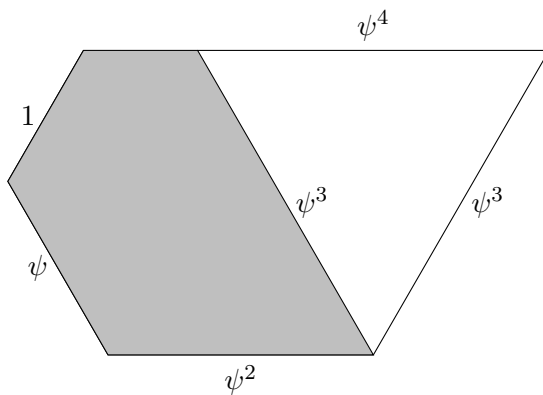
Vanaf  $n = 86$  krijgen we de eerste 15 decimalen van  $\psi$ :

$$\frac{p_{87}}{p_{86}} = \frac{30433364801}{22973467397} \approx 1,324717957244746.$$

Een meetkundige betekenis voor  $p_n$  (kennelijk geldt ook  $p_n - p_{n-1} = p_{n-5}$ ):



Een meetkundige betekenis voor  $\psi$  (het grijze deel is gelijkvormig met het geheel):



En zo is er een spiraal van gelijkzijdige driehoeken:

