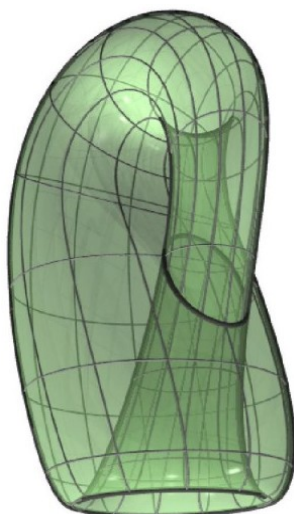


WISKUNDE IN WORDING

Syllabus Vakantiecursus 2013

Eindhoven, 23 en 24 augustus

Amsterdam, 30 en 31 augustus



WISKUNDE IN WORDING

Syllabus Vakantiecursus 2013

(uitgebreide web-versie)

Eindhoven, 23 en 24 augustus

Amsterdam, 30 en 31 augustus

Programmacommissie

prof. dr. Frits Beukers (UU)

prof. dr. Jason Frank (CWI)

drs. Marian Kolleveld (NVvW)

prof. dr. Jan van Maanen (RUG)

prof. dr. ir. Wil Schilders (PWN, TU/e)

dr. ir. Ionica Smeets (ionica.nl)

dr. Jeroen Spandaw (TUD)

dr. Marco Swaen (UvA)

dr. Benne de Weger (TU/e)

prof. dr. Jan Wiegerinck (UvA) (voorzitter)

e-mail: vakantiecursus@platformwiskunde.nl

Platform Wiskunde Nederland

Science Park 123, 1098 XG Amsterdam

Telefoon: 020-592 4006

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

Vakantiecursus 2013

De Vakantiecursus Wiskunde voor leraren in de exacte vakken in HAVO, VWO, 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 en 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).

De presentaties van de sprekers zullen zo veel mogelijk beschikbaar komen op de PWN-website:

http://www.platformwiskunde.nl/onderwijs_vakantiecursus_wiskunde.htm.

Deze uitgebreide web-versie bevat de volgende onderdelen die niet in de gedrukte versie stonden, en op de cursus los zijn uitgereikt:

- de tekst van Jaap Top (hoofdstuk 2),
- de dia's van Harry Buhrman (hoofdstuk 4),
- de opgaven bij het praktikum over het Poincaré-vermoeden (hoofdstuk 7).

Met dank aan

Ondersteuning PWN en CWI: Coby van Vonderen, Sjoukje Talsma.
Ondersteuning TU/e: Anita Klooster, Wil Kortsmit.

Historie

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 stond de wiskunde, op 31 October de didactiek van de wiskunde op de voorgrond. De sprekers waren: Prof.Dr. O. Bottema, "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, "Moelijkheden 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).

Docenten

prof. dr. Harry Buhrman

Centrum voor Wiskunde en Informatica
Science Park 123, 1098 XG Amsterdam
e-mail: buhrman@cw.i.nl

prof. dr. ir. Barry Koren

Faculteit Wiskunde en Informatica, Technische Universiteit Eindhoven
Postbus 513, 5600 MB Eindhoven
e-mail: b.koren@tue.nl

prof. dr. Jaap Top

Faculteit Wiskunde, Universiteit Groningen
Postbus 407, 9700 AK Groningen
e-mail: j.top@rug.nl

dr. Roland van der Veen

Korteweg de Vries Instituut voor Wiskunde, Universiteit van Amsterdam
Postbus 94248, 1090 GE Amsterdam
e-mail: r.i.vanderveen@uva.nl

prof. dr. Ferdinand Verhulst

Mathematisch Instituut, Universiteit Utrecht
Postbus 80010, 3508 TA Utrecht
e-mail: f.verhulst@uu.nl

dr. Benne de Weger

Faculteit Wiskunde en Informatica, Technische Universiteit Eindhoven
Postbus 513, 5600 MB Eindhoven
e-mail: b.m.m.d.weger@tue.nl

prof. dr. Jan Wiegerinck

Korteweg de Vries Instituut voor Wiskunde, Universiteit van Amsterdam
Postbus 94248, 1090 GE Amsterdam
e-mail: j.j.o.o.wiegerinck@uva.nl

Programma

Vrijdag 23 / 30 augustus 2013

15.00–15.25		<i>Ontvangst, koffie</i>
15.25–15.35	Wiegerinck	Introductie “Wiskunde in Wording”
15.35–16.20	de Weger	Het $3n + 1$ -vermoeden
16.20–16.45		<i>Pauze</i>
16.45–17:30	Top	Het vermoeden van Birch en Swinerton-Dyer
17.30–18.30		<i>Diner</i>
18.30–19.15	Koren	Navier-Stokes
19.15–19.45		<i>Pauze</i>
19.45–20.30	Buhrman	De status van het P versus NP probleem

Zaterdag 24 / 31 augustus 2013

Speciale dag over het Poincaré-vermoeden, 10 jaar geleden bewezen door Grigori Perelman

10.00-10.30		<i>Ontvangst, koffie</i>
10.30-11.00	Verhulst	Henri Poincaré en het ontstaan van de topologie
11.00-11.30		Practicum 1: Oppervlakken
11.30-12.00	van der Veen	Het vormen van ruimte: van Poincaré tot Perelman
12.00-13.00		<i>Lunch</i>
13.00-13.30	van der Veen	Vervolg
13.30-14.00		Practicum 2: Ricci-flow op oppervlakken
14.00-14.30		<i>Pauze</i>
14.30-15.00	van der Veen	Afsluiting

Ten geleide

Jan Wiegerinck

Universiteit van Amsterdam

Deze vakantiecursus ziet er wat anders uit dan in voorgaande jaren. De eerste dag is min of meer traditioneel, de tweede grijpt terug op een oude traditie. Er zijn dan naast voordrachten ook practica, waarin u zelf met het aangedragen materiaal aan de slag zult gaan. Zulke practica waren er lang geleden ook, en daar is toen mee gestopt, omdat de toenmalige cursisten er minder belangstelling voor hadden.

Nu horen we van deelnemende docenten dat voordrachten soms wat saai zijn, en dat het leuk zou zijn als zij zelf iets konden doen. We zien ook dat in het algemeen zwaardere vakinhoudelijke eisen aan docenten worden gesteld en er meer aandacht voor de vakinhoud komt. In hoeverre dat in de klas tot uitdrukking komt, kan ik van afstand natuurlijk niet beoordelen, maar ik zie wel dat de aantallen, de kwaliteit en de gedrevenheid van studenten die met wiskunde beginnen erg zijn toegenomen. Reden genoeg om onder leiding van Roland van der Veen en Ferdinand Verhulst een hele dag aan het Poincarévermoeden te wijden. De bedoeling is dat u naar huis gaat met zoveel kennis van dit moderne onderwerp dat het makkelijk is deze in uw onderwijs te gebruiken.

De eerste dag is meer traditioneel. Benne de Weger spreekt over het Collatz- of $3n + 1$ -probleem, makkelijk uit te leggen, moeilijk op te lossen. Een paar jaar terug is dit probleem ook al eens voor het voetlicht gekomen. Het probleem blijft open, maar in de tussentijd is het inzicht gegroeid, en er is weer wat over te vertellen. Zulke open problemen drijven de wiskunde aan. Dat geldt des te meer voor de millenniumproblemen, waarvan het Poincarévermoeden er één is. Drie andere zijn: P versus NP , waarover Harry Buhrman zal spreken, het vermoeden van Birch & Swinnerton-Dyer, waar Jaap Top over zal praten en de vraag naar oplosbaarheid en stabiliteit van de Navier-Stokes vergelijkingen, die door Barry Koren zal worden behandeld.

Deze vier onderwerpen hebben heel verschillende karakters. Het $3n + 1$ -probleem wordt als recreatieve wiskunde beschouwd, vooral omdat er op het eerste gezicht geen link met bestaande centrale wiskundige domeinen lijkt te bestaan. Maar die links zijn er, tot veler verrassing, wel degelijk. Of het daarmee serieuze wiskunde is geworden mag u zelf beslissen. Het vermoeden van Birch & Swinnerton-Dyer is lastig te formuleren, centraal in de algebraïsche meetkunde, en bepaald niet recreatief, maar ik verwacht dat eventuele oplossing de wereld buiten de wiskunde niet op zijn kop zal zetten. De Navier-Stokesvergelijkingen beschrijven stromingen in gassen en vloeistoffen. Ieder beter begrip daarvan heeft onmiddellijk invloed, die zich uitstrekt van vliegtuigbouw tot kunstkleppen voor ons hart. Het $P = NP$ -probleem vraagt of problemen die moeilijk op te lossen lijken, maar waarvan wel eenvoudig te controleren is of een kandidaatoplossing het probleem oplost, eigenlijk eenvoudig kunnen worden opgelost. “Abstract geneuzel” zult u denken, en inderdaad, als blijkt $P \neq NP$ draait de wereld rustig door, maar Harry Buhrman laat u zien hoe een positief antwoord onze maatschappij kan ontwrichten.

Ik hoop en verwacht dat u een leerzame en leuke cursus tegemoet kunt zien!

Inhoudsopgave

1	Het $3n + 1$ -vermoeden Benne de Weger	1
2	Het vermoeden van Birch en Swinnerton-Dyer Jaap Top	31
3	Navier-Stokes Barry Koren	49
4	De status van het P versus NP probleem Harry Buhrman	73
5	Henri Poincaré en het ontstaan van de topologie Ferdinand Verhulst	93
6	Het vormen van ruimte: van Poincaré tot Perelman Roland van der Veen	105
7	Opgaven bij Het vormen van ruimte: van Poincaré tot Perelman Roland van der Veen	151

1 Het $3n + 1$ -vermoeden

Benne de Weger

Technische Universiteit Eindhoven

1.1 Inleiding

1.1.1 Het vermoeden

Neem een natuurlijk getal n . Als het even is, deel het door 2. Als het oneven is, vermenigvuldig het met 3 en tel er 1 bij op. Met de uitkomst doe je hetzelfde, net zolang tot je een patroon herkent.

Omdat voor een oneven n het volgende getal $3n + 1$ altijd even is, is de daaropvolgende stap altijd delen door 2. Deze twee stappen nemen we daarom samen tot één stap. Het hierboven beschreven proces is dan het vanuit startwaarde n *itereren* van de zogeheten $3n + 1$ -*functie*:

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

Dit iteratieproces kunnen we noteren met pijlen, bijvoorbeeld:

$$6 \rightarrow 3 \rightarrow 5 \rightarrow 8 \rightarrow 4 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow \dots$$

Dit voorbeeld laat zien dat je met de getallen 1, 2, 3, 4, 5, 6 en 8 als startwaarde iedere keer op 1 uitkomt, en dan kom je in de *cykel* $1 \rightleftharpoons 2$ terecht waar je nooit meer uitkomt. Een ander voorbeeld:

$$7 \rightarrow 11 \rightarrow 17 \rightarrow 26 \rightarrow 13 \rightarrow 20 \rightarrow 10 \rightarrow 5 \rightarrow \dots$$

laat zien dat je met startwaarde 7 na zeven iteraties op 5 uitkomt, waarvan we al wisten dat je dan vervolgens in de cykel $1 \rightleftharpoons 2$ belandt.

De k -e iteratie van n onder de functie T noteren we als $T^k(n)$. Dus bijvoorbeeld $T^7(7) = 5$ en $T^{11}(7) = 1$.

Het $3n + 1$ -vermoeden zegt nu dat dit iteratieproces bij iedere mogelijke startwaarde altijd een keer bij 1 zal uitkomen, en dus ‘eindigt’ in de cykel $1 \rightleftharpoons 2$. Dat is een wat informele bewering, die eenvoudig formeel te maken is, als volgt:

Het $3n + 1$ -vermoeden:

Voor iedere $n \in \mathbb{N}$ is er een $k \in \mathbb{N}$ zodat $T^k(n) = 1$.

De iteratierij $n \rightarrow T(n) \rightarrow T^2(n) \rightarrow \dots$ heet de *baan* van n . Als een baan weer op z’n beginpunt uitkomt, dus voor zekere $k, n \in \mathbb{N}$ geldt dat $T^k(n) = n$, dan gaat de baan zich vanaf dat moment herhalen: $n \rightarrow T(n) \rightarrow \dots \rightarrow T^{k-1}(n) \rightarrow T^k(n) = n \rightarrow T(n) \rightarrow \dots$. Zo’n herhaald stuk noemen we een *cykel*, en noteren we als $(n, T(n), T^2(n), \dots, T^{k-1}(n))$. We zien $(n, T(n), T^2(n), \dots, T^{k-1}(n))$ en $(T(n), T^2(n), \dots, T^{k-1}(n), n)$ als dezelfde cykel, dus $(1, 2) = (2, 1)$. Een baan noemen we *convergent* als er een cykel in zit (en daar ‘eindigt’ die baan dan ook in), en anders *divergent*. De cykel $1 \rightleftharpoons 2$ heet ook wel de *triviale cykel*.

Het $3n + 1$ -vermoeden kan gezien worden als de combinatie van twee onafhankelijke uitspraken:

Het $3n + 1$ -cykelvermoeden:

De $3n + 1$ -functie kent geen niet-triviale cyclen.

Het $3n + 1$ -convergentievermoeden:

De $3n + 1$ -functie kent geen divergente banen.

Banen kunnen lang divergent lijken en tot grote hoogten stijgen:

$27 \rightarrow 41 \rightarrow 62 \rightarrow 31 \rightarrow 47 \rightarrow 71 \rightarrow 107 \rightarrow 161 \rightarrow 242 \rightarrow 121 \rightarrow 182 \rightarrow 91 \rightarrow 137 \rightarrow 206 \rightarrow 103 \rightarrow 155 \rightarrow 233 \rightarrow 350 \rightarrow 175 \rightarrow 263 \rightarrow 395 \rightarrow 593 \rightarrow 890 \rightarrow 445 \rightarrow 668 \rightarrow 334 \rightarrow 167 \rightarrow 251 \rightarrow 377 \rightarrow 566 \rightarrow 283 \rightarrow 425 \rightarrow 638 \rightarrow 319 \rightarrow 479 \rightarrow 719 \rightarrow 1079 \rightarrow 1619 \rightarrow 2429 \rightarrow 3644 \rightarrow 1822 \rightarrow 911 \rightarrow 1367 \rightarrow 2051 \rightarrow 3077 \rightarrow 4616 \rightarrow \dots$

maar toch convergent blijken en uitkomen in de triviale cykel:

$\dots \rightarrow 4616 \rightarrow 2308 \rightarrow 1154 \rightarrow 577 \rightarrow 866 \rightarrow 433 \rightarrow 650 \rightarrow 325 \rightarrow 488 \rightarrow 244 \rightarrow 122 \rightarrow 61 \rightarrow 92 \rightarrow 46 \rightarrow 23 \rightarrow 35 \rightarrow 53 \rightarrow 80 \rightarrow 40 \rightarrow 20 \rightarrow 10 \rightarrow 5 \rightarrow 8 \rightarrow 4 \rightarrow 2 \rightleftharpoons 1$.

1.1.2 Geschiedenis

Het $3n + 1$ -vermoeden is (naar eigen zeggen) rond 1930 bedacht door Lot-har Collatz, een Duitse wiskundige. In 1950 raakte het vermoeden breder

bekend doordat hij er met verschillende wiskundigen op het International Congress of Mathematicians in Cambridge (Mass.) over sprak. In de jaren 60 verschenen de eerste artikelen over verwante problemen, o.a. over een variant beschreven door de bekende Franse dichter Raymond Queneau, gerelateerd aan rijmschema's in 12e-eeuwse poëzie. In 1971 verschijnt het exacte vermoeden voor het eerst in druk, en in 1972 schrijft de bekende wiskunde-popularisator Martin Gardner erover, waarna het probleem echt beroemd wordt. In de (kennelijk complete) geannoteerde bibliografie die Jeffrey Lagarias bijhoudt [L3, L4] is te zien hoezeer het vermoeden de wiskundigen in de greep heeft gekregen, niet alleen in het aantal publicaties:

decennium:	60-69	70-79	80-89	90-99	00-09
aantal publicaties:	8	34	52	103	134

maar ook in de verscheidenheid van wiskundige deelgebieden waaraan het vermoeden gerelateerd blijkt te kunnen worden. In de literatuur staat het $3n + 1$ -vermoeden ook bekend als o.a. het vermoeden van Collatz, van Hasse, van Kakutani, van Ulam, als het Syracuse-vermoeden en als het ‘hailstone’-vermoeden. De lezer die er meer over wil weten kan worden aangeraden het oude overzichtsartikel [L1] en het nieuwe boek [L2] van Jeffrey Lagarias op te slaan.

In dit artikel willen we de lezer enige argumenten aanreiken waarom hij / zij wel eens in het vermoeden zou kunnen gaan geloven, en een aantal verbanden aangeven met uiteenlopende wiskundige deelterreinen. Diverse van die verbanden leiden tot interessante herformuleringen van het $3n + 1$ -vermoeden in een totaal andere ‘taal’.

Overigens is het vermoeden volstrekt nutteloos. Ionica Smeets [Sm] noemde het in de Vakantiecursus van 2010 een “gênant probleem”, omdat het op feestjes aan leken makkelijk uit te leggen is wat het probleem is, maar niet dat al die slimme wiskundigen zo’n ogenschijnlijk makkelijk probleem nog steeds niet hebben kunnen oplossen.

1.1.3 Grafentaal

Een *graaf* (resp. *gerichte graaf*) is een verzameling punten met lijnen (resp. pijlen) tussen sommige paren punten. Eigenlijk hebben we hierboven al grafen gebruikt om de $3n + 1$ -iteraties te visualiseren. De $3n + 1$ -graaf heeft als punten alle positieve gehele getallen, en voor iedere $n \in \mathbb{N}$ loopt er een pijl van n naar $T(n)$. In Figuur 1.1 is het deel van de $3n + 1$ -graaf getekend met alle getallen t/m 100 en hun banen (in iedere baan

zijn eventuele getallen boven de 100 weggelaten, en dan is telkens tussen rechte haken aangegeven hoeveel getallen zijn weggelaten).

In deze graaf heeft ieder punt één uitgaande pijl en 1 of 2 inkomende pijlen, al naar gelang $n \not\equiv 2 \pmod{3}$ dan wel $n \equiv 2 \pmod{3}$ is. Immers, voor iedere n is er een inkomende pijl uit $2n$, en als er een andere inkomende pijl is moet die wel van de oneven m komen met $n = \frac{1}{2}(3m + 1)$, en dat betekent dat $m = \frac{1}{3}(2n - 1)$, wat precies dan geheel is als $n \equiv 2 \pmod{3}$.

Een graaf bestaat uit *samenhangscomponenten* (kortweg *componenten* genoemd), gedefinieerd door de regel dat twee punten in dezelfde component zitten dan en slechts dan als er een pad is tussen de twee punten (waarbij ook tegen de pijlrichting in gelopen mag worden). Een graaf die uit slechts één component bestaat noemen we *samenhangend*.

Dit geeft meteen een herformulering van het $3n + 1$ -vermoeden:

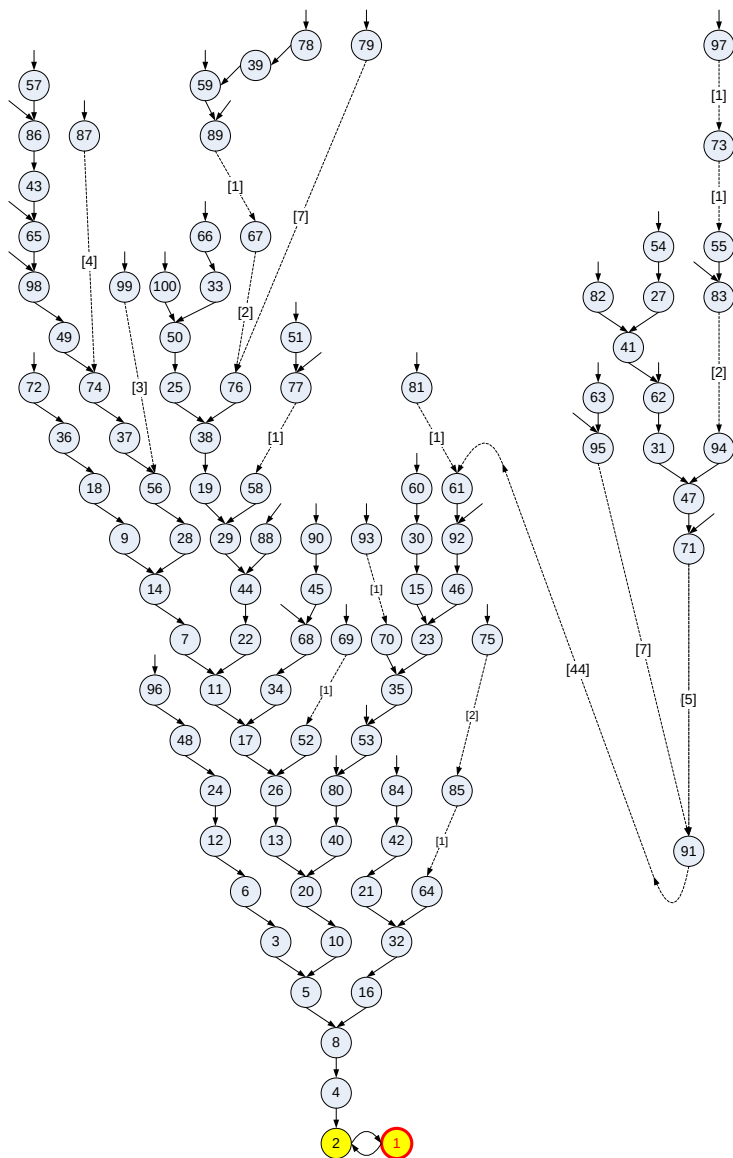
Het $3n + 1$ -vermoeden, herformulering 1:

De $3n + 1$ -graaf is samenhangend.

In zekere zin is er hier niets gebeurd: het is slechts een vertaling van het vermoeden in termen van grafen. Toch geeft dit al wel wat inzicht, vind ik. Onlangs nog was er een serieuze wiskundige [O] die beweerde het vermoeden bewezen te hebben, maar wiens argument gebaseerd was op de veronderstelling dat er tussen ieder tweetal getallen een pad was in de graaf; en dat cruciale punt vergat hij te bewijzen [dW]. Niet verwonderlijk overigens was zijn argument ook overbodig ingewikkeld.

Even er van uitgaande dat we geen idee hebben waarom we het vermoeden zouden moeten geloven, zodat we met alle eventualiteiten rekening willen houden: we kunnen nu in de graaf onderscheiden tussen convergente componenten (die een cykel bevatten) en divergente componenten (die geen cykel bevatten). In een component kan slechts één cykel zitten. Want uit een cykel kun je niet meer ontsnappen door de pijlen te volgen, en als er dan een pad van de ene cykel naar de andere zou moeten lopen moet je vertrekkend vanuit de ene cykel tegen de pijl in lopen, en bij het betreden van de andere cykel met de pijl meelopen, maar dan zou er ergens op het pad een punt moeten zijn met twee uitgaande pijlen, en dat is in tegenspraak met het feit dat elk punt maar één uitgaande pijl heeft.

Zo zien we dus ook herformuleringen van de deelvermoedens:



Figuur 1.1: De $3n + 1$ -graaf.

Het $3n + 1$ -cykelvermoeden, herformulering:

De $3n + 1$ -graaf heeft slechts één convergente component.

Het $3n + 1$ -convergentievermoeden, herformulering:

De $3n + 1$ -graaf heeft geen divergente componenten.

1.1.4 Verwante problemen

Het geeft vaak al wat inzicht in een vermoeden als je verwante problemen bestudeert. Waarom bijvoorbeeld in de functie T niet de getallen 3 en $+1$ vervangen door andere getallen? Dat wil zeggen, we definiëren voor oneven $p > 0$ en oneven q de $pn + q$ -functie:

$$T_{p,q}(n) = \begin{cases} \frac{1}{2}n & \text{als } n \text{ even,} \\ \frac{1}{2}(pn + q) & \text{als } n \text{ oneven.} \end{cases}$$

Dus $T = T_{3,1}$. Het geval $p = 1$ is niet interessant, omdat voor $n > q$ de functie $T_{1,q}$ altijd daalt, en je dus met iedere baan op den duur onder q uitkomt, en daar dan ook blijft. Dus nemen we vanaf nu $p \geq 3$. Het eerste wat wellicht voor de hand ligt te bekijken is de $3n - 1$ -functie $T_{3,-1}$. Maar deze komt neer op de $3n + 1$ -functie op de negatieve getallen. Immers, als we het domein van T uitbreiden tot alle gehele getallen, en hetzelfde functievoorschrift voor T blijven hanteren, dan kunnen we opmerken dat $T_{3,-1}(n) = -T(-n)$. Opgave: toon dit aan.

Figuur 1.2 toont (een stuk van) de $3n - 1$ -graaf. Wat meteen opvalt is dat er drie cycli zijn voor de $3n - 1$ -functie, en dat deze graaf dus (tenminste) drie componenten heeft. Dat betekent in ieder geval al dat het $3n + 1$ -vermoeden niet direct te generaliseren valt. Dat er precies één component zou zijn is kennelijk geen algemene eigenschap van dit soort functies, maar zit vast aan de gekozen parameters. Voor de $3n - 1$ -functie kunnen we het volgende vermoeden formuleren.

Het $3n - 1$ -vermoeden:

Voor iedere $n \in \mathbb{N}$ is er een $k \in \mathbb{N}$ zodat $T^k(n) \in \{1, 5, 17\}$.

In grafen-taal: De $3n - 1$ -graaf heeft precies 3 componenten.

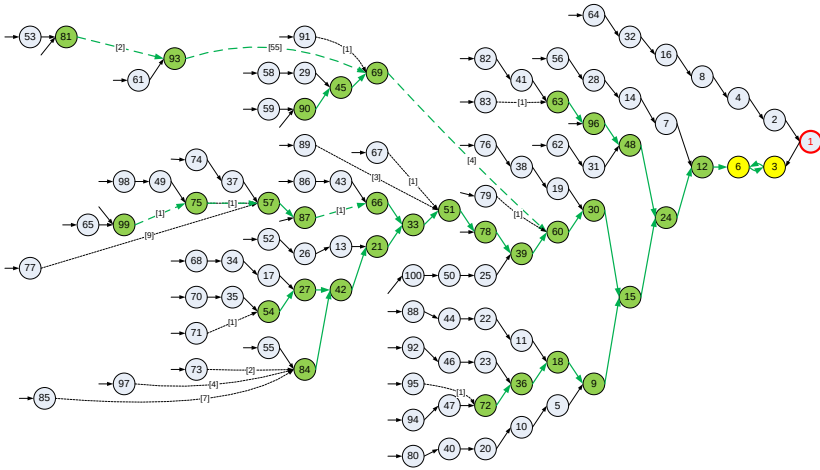
Figuur 1.3 laat een ander interessant geval zien: de $3n + 3$ -graaf. Bij bestudering blijkt de deelgraaf van alle drievouden precies de structuur van de $3n + 1$ -graaf te bezitten, alleen dan met alle getallen vermenigvuldigd met 3. In Figuur 1.3 is dit met een kleur aangegeven. Wat het $3n + 3$ -vermoeden moet worden kunt u nu zelf bedenken. Uit het $3n + 3$ -vermoeden



Een volgend interessant geval is de $5n+1$ -graaf. Deze is getekend in Figuur 1.4. Weer zijn er in ieder geval 3 cycli te ontdekken, en dus tenminste 3 componenten. Maar nu lijken er ook de nodige divergente componenten te bestaan. Het $5n+1$ -vermoeden, waarvoor we in een volgend hoofdstuk een argument zullen aandragen, luidt dan ook als volgt (we geven alleen een formulering in grafen-taal omdat dat nu echt makkelijker is):

De $5n + 1$ -graaf heeft precies 3 convergente componenten, en oneindig veel divergente componenten.

Vermoedelijk zijn alle in Figuur 1.4 getekende componenten inderdaad verschillend. Van de getekende componenten zonder zichtbare cykel is niet bekend of ze verschillend zijn, laat staan dat bekend is of ze wel echt divergent zijn, of niet stiekum ver weg toch nog een cykel blijken te hebben.

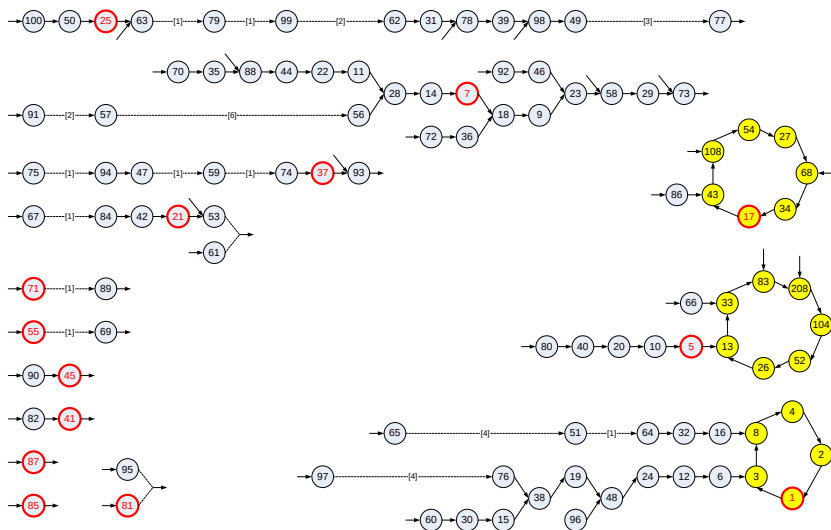


Figuur 1.3: De $3n + 3$ -graaf.

1.1.5 De rationale $3n + 1$ -graaf

Laten we even terugkeren naar de $3n + 3$ -graaf. Door alle getallen door 3 te delen zien we in dat dit in feite een uitbreiding van de $3n + 1$ -graaf is tot de rationale getallen met noemer 1 of 3. Net zo kunnen we inzien dat de $3n + q$ -graaf voor oneven $q > 0$ terug te brengen tot een uitbreiding van de $3n + 1$ -graaf tot de rationale getallen met noemer een deler van q . Maar dan kunnen we ook meteen de rationale $3n + 1$ -graaf bekijken, waarbij we toestaan dat n een willekeurig rationaal getal met oneven noemer is. (Merk op dat even noemers niet interessant zijn, daarvoor kunnen we niet op een goede manier definiëren of deze getallen even of oneven zijn. Een rationaal getal met een oneven noemer kunnen we wel even / oneven noemen, namelijk aan de hand van het even / oneven zijn van de teller.)

Zie Figuur 1.5 voor een deel van de rationale $3n + 1$ -graaf. In feite is deze rationale $3n + 1$ -graaf de vereniging van alle $3n + q$ -grafen in één graaf gestopt. Nu laten we zien dat ieder punt in deze graaf twee inkomende pijlen heeft. Bij een punt in de graaf hoort een rationaal getal $\frac{a}{b}$, met b oneven. Er komt in ieder geval een pijl van $\frac{2a}{b}$, een andere zou moeten komen van $\frac{c}{d}$ met oneven d zodat $\frac{3c}{2d} + \frac{1}{2} = \frac{a}{b}$. Hieruit halen we $\frac{c}{d} =$



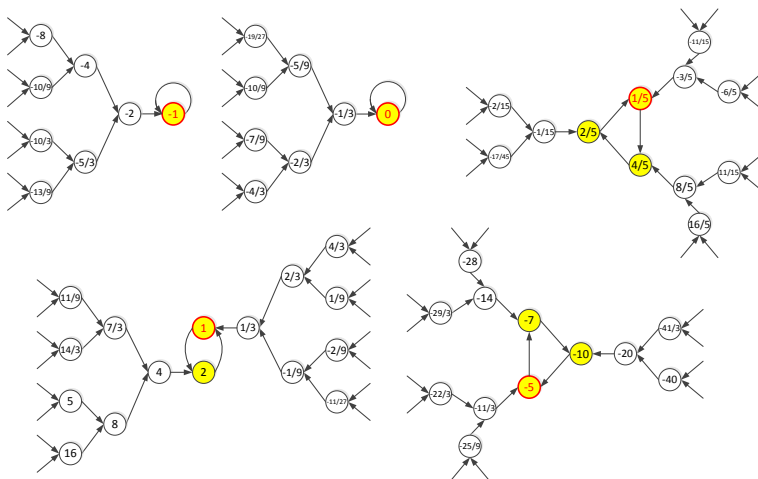
Figuur 1.4: De $5n + 1$ -graaf.

$\frac{2a-b}{3b}$. Als $2a-b$ een 3-voud is, dan volgt $d = b$ en $c = (2a-b)/3$, anders volgt $d = 3b$ en $c = 2a-b$. In alle gevallen is c oneven, en dat toont aan dat deze $\frac{c}{d}$ inderdaad anders is dan $\frac{2a}{b}$. Dus er zijn altijd twee inkomende pijlen bij ieder punt. Terugredenerend betekent dit dat op ieder punt in de graaf een zogeheten *volledige binaire boom* staat.

We gaan nu kijken naar cyclen. Een cykel heeft altijd een zogeheten even/oneven-structuur, die aangeeft of de opeenvolgende getallen in de cykel even (e) dan wel oneven (o) zijn. Een voorbeeld: de cykel $\left(\frac{4}{5}, \frac{2}{5}, \frac{1}{5}\right)$

(dit is inderdaad een cykel want $\frac{3}{2} \frac{1}{5} + \frac{1}{2} = \frac{4}{5}$) heeft de even/oneven-structuur **eeo**. Omdat een cykel een cykel is, is deze structuur identiek aan **eo** en aan **oe**.

Deze even/oneven-structuur schrijft precies voor wat op ieder moment de keuze is in de $3n + 1$ -functie. Dit stelt ons in staat een vergelijking op te stellen voor een cykel met een gegeven structuur. Een voorbeeld: als we uitgaan van de even/oneven-structuur **eeo** en beginnen met het getal x , dan zal x even zijn, en dus $x \rightarrow T(x) = x/2$. Maar op deze tweede



Figuur 1.5: De rationale $3n + 1$ -graaf.

plek in de cykel moet ook een even getal staan, dus $T(x) \rightarrow T^2(x) = T(x)/2 = x/4$. En dit derde getal in de cykel moet nu oneven zijn, dus $T^3(x) = T(x/4) = \frac{3(x/4) + 1}{2} = 3x/8 + 1/2$. Omdat de cykel rond is en lengte 3 heeft is $T^3(x) = x$, dus $3x/8 + 1/2 = x$, en hier staat een eenvoudige eerstegraads-vergelijking in x , die slechts één oplossing heeft. Dat is voor iedere cykel net zo, en die oplossing is altijd een rationaal getal met een oneven noemer (in bovenstaand voorbeeld is dat $x = \frac{4}{5}$). Gevolg is dat bij iedere even/oneven-structuur precies één cykel in de rationale $3n + 1$ -graaf hoort.

Een lijstje van mogelijke even/oneven-structuren (deze staan in de literatuur bekend als *Lyndon-woorden*):

- lengte 1: 2 mogelijkheden: **e**, **o**, horend bij de cyclen $(0), (-1)$;
- lengte 2: 1 mogelijkheid: **eo**, horend bij de triviale cykel $(1, 2)$;
- lengte 3: 2 mogelijkheden: **eeo**, **ooo**, horend bij de cyclen $(\frac{4}{5}, \frac{2}{5}, \frac{1}{5}), (-10, -5, -7)$;
- lengte 4: 3 mogelijkheden: **eeeo**, **eeoo**, **oooo**, horend bij de cyclen

$$\left(\frac{8}{13}, \frac{4}{13}, \frac{2}{13}, \frac{1}{13}\right), \left(\frac{20}{7}, \frac{10}{7}, \frac{5}{7}, \frac{11}{7}\right), \left(-\frac{38}{11}, -\frac{19}{11}, -\frac{23}{11}, -\frac{29}{11}\right);$$

- lengte 5: 6 mogelijkheden: **eeeee**, **eeeee**, **eeeee**, **eeeee**, **eeeee**, **eeeee**, horend bij de cyclen $\left(\frac{16}{29}, \frac{8}{29}, \frac{4}{29}, \frac{2}{29}, \frac{1}{29}\right), \left(\frac{40}{23}, \frac{20}{23}, \frac{10}{23}, \frac{5}{23}, \frac{19}{23}\right), \left(\frac{28}{23}, \frac{14}{23}, \frac{7}{23}, \frac{22}{23}, \frac{11}{23}\right), \left(\frac{76}{5}, \frac{38}{5}, \frac{19}{5}, \frac{31}{5}, \frac{49}{5}\right), \left(\frac{58}{5}, \frac{29}{5}, \frac{46}{5}, \frac{23}{5}, \frac{37}{5}\right), \left(-\frac{130}{49}, -\frac{65}{49}, -\frac{73}{49}, -\frac{85}{49}, -\frac{103}{49}\right),$
- enzovoorts.

Merk op dat even/oneven-structuren waar al een repeterend patroon in zit, zoals **eeeo**, niet meegeteld worden. De bijbehorende vergelijking heeft namelijk exact dezelfde oplossing als die van de even/oneven-structuur **ee**, en dit komt dus gewoon neer op een cykel vaker dan één keer doorlopen.

De rationale $3n+1$ -graaf heeft vermoedelijk geen divergente componenten. Een argument dat daar voor pleit zullen we in het volgende hoofdstuk zien. We hebben zojuist aangetoond dat er oneindig veel convergente componenten zijn, en we hebben een methode gezien om die allemaal systematisch te beschrijven. Zo'n convergente component bestaat uit een cykel, met op ieder punt van de cykel een volledige binaire boom. Ook als er een divergente component zou zijn, zou die een fraaie structuur hebben, namelijk een naar beide zijden onbegrensde volledige binaire boom. Deze rationale $3n+1$ -graaf heeft dus op zich een heel fraaie structuur. Er is dan ook geen rationaal $3n+1$ -cykelvermoeden, omdat we van de cyclen alles al weten. Wat overblijft is alleen het rationale $3n+1$ -convergentievermoeden, het makkelijkst te geven in grafen-taal:

Het rationale $3n+1$ -vermoeden:

De rationale $3n+1$ -graaf heeft geen divergente componenten.

De oorspronkelijke $3n+1$ -graaf is een deelgraaf van de rationale $3n+1$ -graaf. Alleen is volstrekt onduidelijk hoe die er precies inzit.

1.1.6 De amusicale permutatie

Tenslotte noemen we een iets anderssoortige functie, die soms de 'inverse $3n+1$ -functie' genoemd wordt, maar ten onrechte, want het is niet de inverse van de $3n+1$ -functie. John Conway [C2] noemt het de *amusicale permutatie*. Het is de volgende functie:

$$U(n) = \begin{cases} \frac{3n}{2} & \text{als } n \text{ even is,} \\ \frac{3n+1}{4} & \text{als } n \equiv 1 \pmod{4}, \\ \frac{3n-1}{4} & \text{als } n \equiv -1 \pmod{4}. \end{cases}$$

$$U^{-1}(n) = \begin{cases} \frac{2n}{3} & \text{als } n \equiv 0 \pmod{3}, \\ \frac{4n-1}{3} & \text{als } n \equiv 1 \pmod{3}, \\ \frac{4n+1}{3} & \text{als } n \equiv -1 \pmod{3}. \end{cases}$$

Er blijken 4 cycli te bestaan: (1), (2, 3), (4, 6, 9, 7, 5) en (44, 66, 99, 74, 111, 83, 62, 93, 70, 105, 79, 59). De component met 8 erin is vermoedelijk divergent. We kunnen het volgende vermoeden formuleren, weer alleen in grafen-taal.

De amusicale graaf heeft precies 4 convergente componenten, en oneindig veel divergente componenten.

1.2 Argumenten

1.2.1 Experimenteel

Een voor de hand liggende manier om het $3n + 1$ -vermoeden te testen is bruut rekenwerk. Met name Tomás Oliveira e Silva [OS] en Eric Roosendaal [R] hebben zich daar verdienstelijk voor gemaakt.

Stelling (Oliveira e Silva):

Het $3n + 1$ -vermoeden is waar voor alle $n < 5.764 \times 10^{18}$.

Dit is een indrukwekkende prestatie. Natuurlijk passen de rekenaars slimigheden toe om geen overbodig rekenwerk te doen, zoals de volgende:

- je kunt stoppen met de baan van n te berekenen zodra je een getal kleiner dan n bent tegengekomen,
- bepaalde klassen van getallen hoeft je niet te bekijken, zoals getallen van de vorm $128m + 15$, aangezien $T^7(128m + 15) = 81m + 10$, en dat is kleiner dan $128m + 15$,

en vele andere.

Hoewel $n < 5.764 \times 10^{18}$ best een hoge grens is, bereik je met dit soort rekenwerk, hoe ver je ook zou kunnen komen, natuurlijk nooit meer dan 0% van het totale zoekgebied. Toch is het bijzonder nuttig werk, met ook meer implicaties dan alleen maar dat een eventuele tweede cykel of een divergente baan als kleinste getal een getal boven de 5.764×10^{18} zou moeten hebben. Daar komen we in het volgende hoofdstuk op terug.

1.2.2 Probabilistisch

Stel we trekken een random getal n , dus met kans $\frac{1}{2}$ even en met kans $\frac{1}{2}$ oneven. Wat kunnen we nu zeggen over de kans dat $T(n)$ even dan wel oneven is? Daarvoor moeten we kijken naar $n \pmod{4}$:

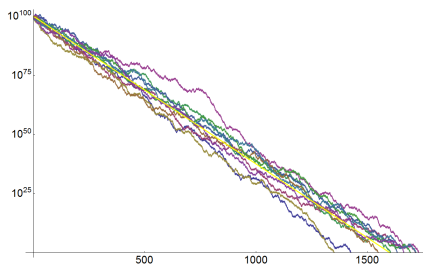
- als $n = 4m$ dan is $T(n) = 2m$ even,
- als $n = 4m + 1$ dan is $T(n) = 6m + 2$ even,
- als $n = 4m + 2$ dan is $T(n) = 2m + 1$ oneven,
- als $n = 4m + 3$ dan is $T(n) = 6m + 5$ oneven,

dus $T(n)$ is ook met kans $\frac{1}{2}$ even en met kans $\frac{1}{2}$ oneven. En het lijkt niet onredelijk te veronderstellen dat de kansen voor n en voor $T(n)$ onafhankelijk zijn.

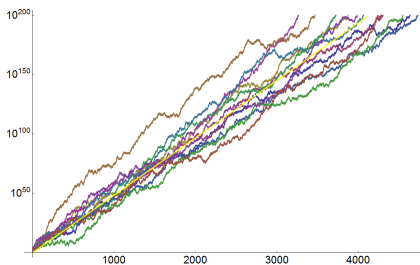
Voor heel grote n gaan we nu de $+1$ in de $\frac{1}{2}(3n+1)$ verwaarlozen, en zien we dat $T(n) = \frac{1}{2}n$ met kans $\frac{1}{2}$, en $T(n) \approx \frac{3}{2}n$ ook met kans $\frac{1}{2}$. Voor de k -e iteratie verwachten we dan dat $\approx \frac{1}{2}k$ maal met $\frac{1}{2}$ vermenigvuldigd wordt, en $\approx \frac{1}{2}k$ maal met $\approx \frac{3}{2}$, zodat

$$T^k(n) \approx \left(\frac{1}{2}\right)^{k/2} \left(\frac{3}{2}\right)^{k/2} n = \left(\frac{1}{2}\sqrt{3}\right)^k n.$$

Merk op dat $\frac{1}{2}\sqrt{3} \approx 0.866 < 1$. Dit laat zien dat itereren van T gemiddeld genomen een exponentieel dalend gedrag zal vertonen. We kunnen nu makkelijk schatten na hoeveel iteratiestappen je verwacht bij 1 uit te komen: $\left(\frac{1}{2}\sqrt{3}\right)^k n = 1 \Leftrightarrow k = \frac{1}{\log(2/\sqrt{3})} \log n \approx 6.95 \log n$. Figuur 1.7 laat zien dat dit experimenteel heel aardig klopt: ik heb daar 10 willekeurige getallen in de buurt van 10^{100} genomen en de banen berekend en geplot; na ongeveer $6.95 \times \log 10^{100} \approx 1600$ stappen zijn alle banen inderdaad op 1 uitgekomen, en nog redelijk op de voorspelde manier ook. Deze figuur heeft op de verticale as een logaritmische schaal; de gele rechte is de lijn die begint bij 10^{100} en richtingscoëfficiënt $\log\left(\frac{1}{2}\sqrt{3}\right)$ heeft.



Figuur 1.7: Tien $3n + 1$ -banen.



Figuur 1.8: Elf $5n + 1$ -banen.

Voor de $3n + q$ -functie werkt precies hetzelfde argument, wellicht niet voor n die klein zijn ten opzichte van q , maar verreweg de meeste n zijn natuurlijk groot vergeleken met q . Dit is een, wat mij betreft vrij sterk, argument voor het rationale $3n + 1$ -vermoeden, dat zegt dat iedere $3n + q$ -baan in een cykel zal uitkomen.

Er zijn verdergaande stochastische modellen toegepast op de $3n + 1$ -banen, die voorspellingen doen als:

- hoge banen: er bestaan, met kans vrijwel 1, banen die (beginnend bij n) als hoogste punt $\approx n^{2+o(1)}$ bereiken na $7.645 \log n$ stappen; die banen zullen na nog eens $13.9 \log n$ stappen terugvallen tot 1;
- lange banen: er bestaan, met kans vrijwel 1, banen die (beginnend bij n) pas na $41.7 \log n$ stappen op 1 uitkomen.

Ook deze modellen zijn experimenteel geverifieerd.

We kunnen het eenvoudige probabilistische model dat we hierboven op de $3n + 1$ -functie hebben toegepast ook op de $5n + 1$ -functie loslaten. Het argument gaat dan net zo, maar levert dan $T_{5,1}^k(n) \approx \left(\frac{1}{2}\sqrt{5}\right)^k n$ op, waarbij we nu moeten opmerken dat $\frac{1}{2}\sqrt{5} \approx 1.12$. Opvallend is dat deze factor nu boven de 1 uitkomt, en dat laat zien dat de banen voor de $5n+1$ -functie gemiddeld genomen exponentieel snel zullen gaan stijgen. Zie figuur 1.8, waar de banen van de getallen 7, 21, 25, 37, 41, 45, 55, 71, 81, 85 en 87 getekend zijn.

Voor de $pn + q$ -functie wordt deze vermenigvuldigingsfactor $\frac{1}{2}\sqrt{p}$, en die is voor oneven $p > 3$ altijd groter dan 1. De gemiddeld dalende $3n+q$ -functies zijn in de hele familie dus de uitzondering, en niet de regel.

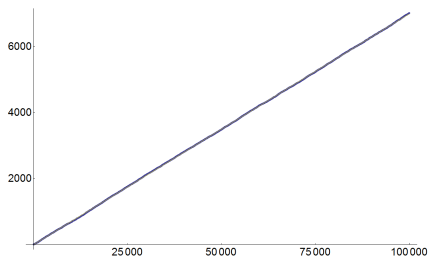
Dat deze vermenigvuldigingsfactor boven de 1 ligt is een sterk argument voor het vermoeden dat de banen van de $5n + 1$ -functie als regel divergent zijn, en slechts bij uitzondering in een cykel uitkomen. Dat het groeigedrag exponentieel is is ook uit te werken tot een argument dat twee banen beginnend in willekeurige startpunten elkaar vrijwel nooit zullen snijden. Vandaar het vermoeden dat er oneindig veel divergente banen zijn.

Dit neemt overigens niet weg dat van geen enkele $5n + 1$ -baan (afgezien van de 3 bekende convergente) bekend is of die divergent is.

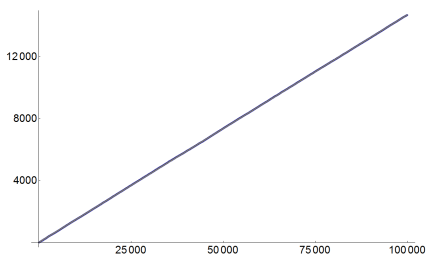
We kunnen proberen voor de $pn + 1$ -graaf (met $p = 5, 7, 9, \dots$) het aantal divergente componenten met een kleinste getal onder n te tellen, en dat aantal als functie van n plotten. Dat aantal blijkt telkens experimenteel vrijwel lineair te zijn. Zie Figuur 1.9-1.11.

De hellingen voor $p = 5, 7, \dots, 49$ zijn geplott in Figuur 1.12. Dit plaatje is vooralsnog (door mij) onbegrepen, maar ik geef direct toe dat ik er niet diep over heb nagedacht.

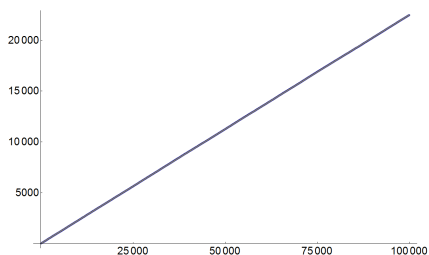
Ook voor de amusicale permutatie is dit eenvoudige stochastische model toepasbaar, zowel op de voorwaartse iteratie $n \rightarrow U(n) \rightarrow U^2(n) \rightarrow \dots$ als op de achterwaartse iteratie $n \rightarrow U^{-1}(n) \rightarrow U^{-2}(n) \rightarrow \dots$. Een opgave voor de lezer is om aan te tonen dat de vermenigvuldigingsfactor voor U



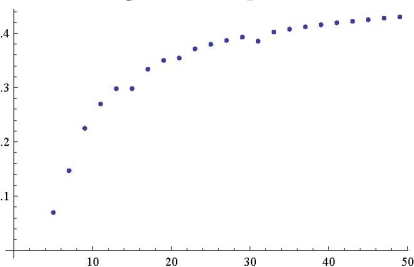
Figuur 1.9: $p = 5$



Figuur 1.10: $p = 7$



Figuur 1.11: $p = 9$



Figuur 1.12: Helling ($p = 5 - 49$)

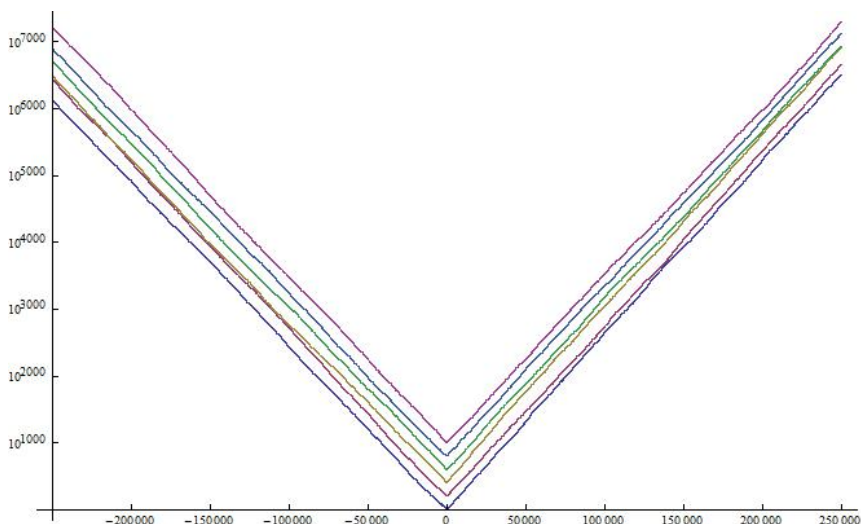
Aantallen divergente banen voor de $pn + 1$ -functie met x als kleinste getal, als functie van x .

dan uitkomt op $\frac{3}{4}\sqrt{2} \approx 1.061$, en die voor U^{-1} op $\frac{2}{3}\sqrt[3]{4} \approx 1.058$. Opvallend is dat beide factoren groter dan 1 zijn, en ook dat ze verschillend zijn. In Figuur 1.13 is van het getal 8 en van willekeurig gekozen getallen met 200, 400, 600, 800 en 1000 cijfers zowel de voorwaartse als de achterwaartse baan getekend. Kennelijk is het zo dat voor een willekeurig gekozen n de kans bijna 1 is dat hij helemaal onderin zijn eigen component zit. Een dergelijk plaatje staat ook in [C2].

1.2.3 Complexiteitstheoretisch

In 1972 heeft John Conway [C1] al laten zien dat er een generalisatie is van de $3n + 1$ -functie waarvan de iteratie een universele Turing-machine simuleert. Zie [C2] voor een concreet voorbeeld. Voor zo'n functie is het beslisprobleem “bereikt een baan een willekeurige macht van 2” onbeslisbaar. In zijn recente artikel [C2] zegt Conway:

It is likely that some simple Collatzian problems (possibly even



Figuur 1.13: Zes amusicale banen

the $3n + 1$ problem itself) will remain forever un settleable.

Dat mag dan misschien zo zijn, en het is zeker een sterke indicatie voor de moeilijkheid van dit soort problemen, maar ik vind dat het ons niet de moed moet ontnemen om toch te blijven proberen ze op te lossen. Ook van de Laatste Stelling van Fermat is wel eens geopperd dat dat misschien wel een onbeslisbaar probleem zou kunnen zijn, en dat is ook goed gekomen.

1.2.4 Diophantisch

Diophantische getaltheorie is de studie van problemen (zoals vergelijkingen en ongelijkheden) met oplossingen in uitsluitend gehele getallen. Een klassiek voorbeeld daarvan is de vraag hoe dicht bij elkaar machten van gehele getallen kunnen liggen, bijvoorbeeld hoe dicht een gehele macht van 2 en een gehele macht van 3 bij elkaar kunnen liggen.

Voor het $3n + 1$ -probleem is dit interessant, omdat, zoals we al hebben gezien, de T -functie als vermenigvuldigingsfactor bij benadering $\frac{3}{2}$ dan wel $\frac{1}{2}$ heeft. Voor een cykel, waarin K maal een oneven getal zit die tot een factor $\approx \frac{3}{2}$ leidt, en L maal een oneven getal met een factor $= \frac{1}{2}$, geldt kennelijk dat $x = T^{K+L}(x) \approx \frac{3^K}{2^{K+L}}x$, met andere woorden: $3^K \approx 2^{K+L}$.

Ray Steiner [St] was in 1977 de eerste die dit idee uitbuitte. Hij bekeek cyclen met de bijzondere eigenschap dat ze bestaan uit eerst K oneven getallen en dan L even getallen, en dan rond zijn. In zo'n geval is eenvoudig aan te tonen (Opgave!) dat het eerste oneven getal x moet voldoen aan $x = a2^K - 1$ voor een oneven a , en dan blijkt $T^K(x) = a3^K - 1$. Dan volgt $T^{K+L}(x) = \frac{a3^K - 1}{2^L}$, en dit moet weer gelijk zijn aan $x = a2^K - 1$ om de cykel rond te maken. Dit geeft een zogeheten *Diophantische vergelijking*

$$a(2^{K+L} - 3^K) = 2^L - 1,$$

die we omschrijven tot een *Diophantische ongelijkheid*

$$0 < \frac{2^{K+L}}{3^K} - 1 = \frac{2^L - 1}{a3^K} = \frac{2^L - 1}{2^L(a2^K - 1) + 1} < \frac{1}{2^K}$$

(waarbij de laatste ongelijkheid met $a \geq 1$ en een beetje gepruts volgt). Inderdaad zien we een macht van 2 en een macht van 3 die dicht bij elkaar liggen. Als we de logaritme nemen dan krijgen we, vanwege $\log(1+x) < x$ voor $x > 0$, dat

$$0 < (K+L)\log 2 - K\log 3 < \frac{2^{K+L}}{3^K} - 1 < \frac{1}{2^K}.$$

De (nogal zware) theorie van Diophantische approximaties (beter: Transcendentietheorie, op kwantitatieve poten gezet in de jaren 60 door Alan Baker) zegt dat dat helemaal niet kan. Preciezer: Georges Rhin (1987) bewees een expliciete ondergrens

$$|(K+L)\log 2 - K\log 3| > 0.00218K^{-13.3}.$$

Boven- en ondergrens vergelijken geeft nu $K \leq 96$. Vervolgens is het niet moeilijk meer om alle oplossingen te vinden, en er blijkt er maar één te zijn: $a = 1, K = 1, L = 1$.

Stelling (Steiner)

Voor de $3n+1$ -functie is de triviale cykel de enige van het even/oneven-type $oo \cdots oee \cdots e$.

Overigens moest Steiner werken met een veel slechtere ondergrens voor $|(K+L)\log 2 - K\log 3|$, die leidde tot $K < 10^{200}$. Het efficiënt afzoeken van alle oplossingen van de hierboven gegeven Diophantische ongelijkheid is, ook voor zo'n hoge bovengrens, goed mogelijk met behulp van de ket-

tingbreuk

$$\frac{\log 3}{\log 2} = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{3 + \frac{1}{1 + \frac{1}{5 + \dots}}}}}}},$$

waar we hier verder niet op kunnen ingaan.

Vervolgens kunnen we kijken of een dergelijk resultaat te generaliseren is tot meerdere cykel-types. Laten we een cykel een m -cykel noemen als er in totaal m overgangen van even naar oneven in zitten (en dus ook m van oneven naar even). Steiner's resultaat gaat dus over 1-cykels.

John Simons (Groningen) bewees in 2004 met een soortgelijke redenering als Steiner (maar ingewikkelder afschattingen waren nodig) dat er geen 2-cykels bestaan. Simons en ik [SdW] breidden dit in 2005–2010 uit:

Stelling (Simons en dW)

Voor de $3n + 1$ -functie is de triviale cykel de enige m -cykel met $m \leq 75$.

De afschattingen die hierbij gebruikt worden hangen af van de experimenteel bepaalde grens waarvoor het vermoeden bekend is, m.a.w. het resultaat van Oliveira e Silva. Dat betekent dat als deze rekenpartij essentieel verder gaat komen, er automatisch grotere m 's vallen. Helaas is de benodigde rekenkracht bijzonder groot. We verwachten niet dat $m = 100$ deze eeuw nog gehaald kan worden.

Door een oude (1978) elementaire afschatting van Richard Crandall [Cr] te combineren met de grens van Oliveira e Silva en de kettingbreuk van $\frac{\log 3}{\log 2}$ ver genoeg uitgerekend, kan worden aangetoond dat een eventuele niet-triviale cykel erg lang moet zijn:

Stelling (Crandall, Oliveira e Silva)

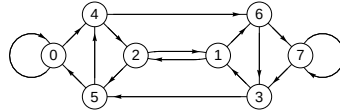
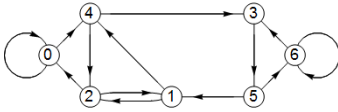
Voor de $3n + 1$ -functie heeft een niet-triviale cykel lengte $> 1.43 \times 10^9$.

Deze resultaten over cykels zijn hard. Geen vermoedens, maar keiharde bewijzen. Ze zeggen niets over het $3n + 1$ -convergentievermoeden, maar geven wel steun aan het $3n + 1$ -cykelvermoeden, in die zin dat het bestaan van een niet-triviale cykel betekent dat er dan kennelijk een zeer uitzonderlijke Diophantische gebeurtenis plaats moet vinden.

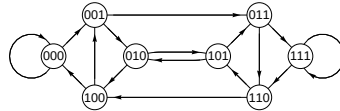
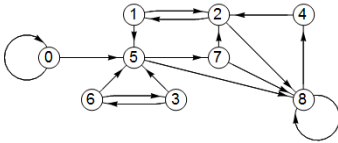
Dergelijke Diophantische resultaten zijn ook op vergelijkbare manieren te vinden voor andere $pn + q$ -functies, maar telkens slechts voor één paar (p, q) tegelijk.

1.3 De Bruijn-Grafen

Een enigszins voor de hand liggend idee is om een modulus m te kiezen, en de $3n + 1$ -functie alleen modulo m te bestuderen. Dit geeft dan eindige grafen om te bestuderen. De modulaire $3n + 1$ -graaf met modulus m bevat de punten $0, 1, 2, \dots, m - 1$, en alle mogelijke pijlen van $n \pmod{m}$ naar $T(n) \pmod{m}$. Zie Figuren 1.14, 1.15, 1.16 voor de $3n + 1$ -graaf modulo 7, 8 en 9.



Figuur 1.14: De $3n + 1$ -graaf mod 7 Figuur 1.15: De $3n + 1$ -graaf mod 8



Figuur 1.17: De binaire De Bruijn-graaf op 8 punten

Figuur 1.16: De $3n + 1$ -graaf mod 9

Dit soort grafen blijkt weinig structuur te hebben, behalve als de modulus een macht van 2 is. Als $m = 2^k$ dan is deze modulaire $3n + 1$ -graaf precies een zogeheten *binaire De Bruijn-graaf*. Deze observatie is gemaakt door Thijs Laarhoven in zijn TU/e-bachelorscriptie uit 2009. Deze De Bruijn-grafen zijn genoemd naar de bekende Eindhovense wiskundige N.G. (Dick) de Bruijn, die onlangs op hoge leeftijd overleden is. Zij ontstaan door alle getallen modulo 2^k in binaire schrijfwijze (bits) $a_0a_1 \dots a_{k-1}$ uit te drukken, en dan pijlen te zetten van $a_0a_1 \dots a_{k-1}$ naar $a_1a_2 \dots a_{k-1}0$ en $a_1a_2 \dots a_{k-1}1$, m.a.w. alle bits schuiven een positie naar links op, de eerste bit verdwijnt, en op de vrijgefallen laatste plaats komen alle mogelijkheden voor. Dit proces heet ook wel de *shiftafbeelding*.

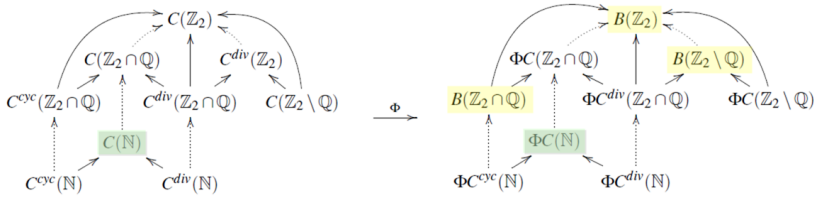
Als je bedenkt dat het $3n + 1$ -proces goed te beschrijven is vanuit de even/oneven-structuur van de baan van n , dan is in te zien dat deze modulaire $3n + 1$ -grafen inderdaad binaire De Bruijn-grafen zijn. Laten we eens kijken naar twee getallen n en n' die modulo 2^k identiek zijn. Zeg $n' = n + b2^k$. In de modulaire $3n + 1$ -graaf modulo $m = 2^k$ zitten die getallen dus op hetzelfde punt. Wat T precies gaat doen met n en n'

wordt bepaald door hun laatste bit, maar die zijn identiek. Dus ofwel n en n' zijn beide even en dan is $T(n') = n'/2 = n/2 + b2^{k-1} \equiv T(n) \pmod{2^{k-1}}$, ofwel ze zijn beide oneven en dan is $T(n') = 3n'/2 + 1/2 = 3n/2 + 3b2^{k-1} + 1/2 = T(n) + 3b2^{k-1} \equiv T(n) \pmod{2^{k-1}}$. Dit argument herhalen laat zien dat de even/oneven-structuur van de banen van n en n' inderdaad op de eerste k posities identiek zijn. Vanaf de $k + 1$ -e positie hangt het af van het even/oneven zijn van b , en die zijn beide mogelijk. Dit laat zien dat modulo 2^k de even/oneven-structuur van de baan van n precies overeenkomt met de shift-afbeelding, en dat de $3n + 1$ -graaf modulo 2^k dus precies de binaire De Bruijn-graaf is op 2^k punten.

Het aardige is nu dat voor de modulaire $pn + q$ -grafen precies hetzelfde argument opgaat. Met andere woorden, de $pn + q$ -graaf modulo 2^k is niets anders dan de binaire De Bruijn-graaf op 2^k punten, *onafhankelijk van p en q* . Alleen de afbeelding van de getallen in de modulaire $pn + q$ -graaf naar de overeenkomende rijtjes bits in de De Bruijn-graaf hangt af van p en q . Deze afbeelding Φ_k heet ook wel de *conjugatie-afbeelding*, en is gegeven door $\Phi_k(n) = a_0a_1 \dots a_{k-1}$ waar a_i telkens het laatste bit van $T^i(n)$ is.

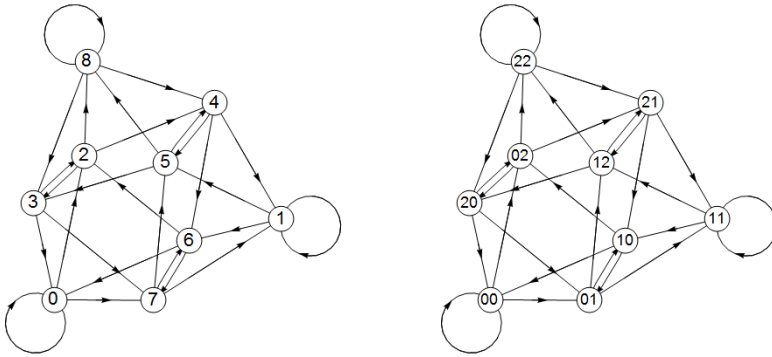
Als we k nu groter en groter nemen, dan kunnen we de overgang naar $k = \infty$ ons voorstellen door de shift-afbeelding te definiëren voor oneindige rijtjes bits: $a_0a_1a_2 \dots$ wordt afgebeeld op $a_1a_2a_3 \dots$. Met andere woorden: het meest linkse bit verdwijnt, maar aan de rechterkant valt niets aan te vullen want daar staan oneindig veel bits die allemaal een plekje naar links opschuiven. Als we al dit soort rijtjes als punten in een enorm grote graaf zetten en de pijlen in de graaf volgens de shiftafbeelding nemen, dan krijgen we de ultieme binaire De Bruijn-graaf, die wel de *2-adische De Bruijn-graaf* genoemd wordt. Ieder punt heeft twee inkomende pijlen en één uitgaande pijl. De rationale $3n + 1$ -graaf zit er compleet in (alle mogelijke cykel-componenten), en is vermoedelijk identiek met de deelgraaf uit de 2-adische De Bruijn-graaf die precies alle convergente componenten (die met cyclen) bevat. Deze 2-adische De Bruijn-graaf heeft daarnaast overaftelbaar veel componenten zonder cyclen, die er allemaal hetzelfde eruit zien: een naar twee kanten oneindig ver uitgebreide volledige binaire boom.

Kortom, het is een monsterachtig ding, maar wel goed begrepen, met bijzonder veel structuur. Alle interessante $pn + q$ -grafen zitten er ergens in verstopt. Zie Figuur 1.18. De geel gemarkeerde grafen zijn goed begrepen. De groen gemarkeerde grafen zijn de eigenlijke $pn + q$ -grafen die we zouden willen begrijpen. Voor een verdere uitleg hiervan, zie [LdW1]. De conjugatieafbeeldingen vertellen hoe precies de interessante grafen in het monster zitten. Alleen begrijpen we erg weinig van deze conjugatieafbeeldingen.



Figuur 1.18: De verschillende $pn + q$ -grafen ingebed in de 2-adische De Bruijn-graaf.

Het is bekend dat ze nergens differentieerbaar zijn. Dat helpt het begrip nou ook weer niet veel verder.



Figuur 1.19: De amusicale graaf modulo 9

Figuur 1.20: De ternaire De Bruijn-graaf op 9 punten

Als we op dezelfde modulo-manier naar de amusicale graaf kijken, dan valt al snel op dat de interessante moduli hier niet machten van 2 zijn, maar machten van 3. Zie Figuur 1.19 voor de amusicale graaf modulo 9. Er zijn nu ook De Bruijn-grafen in het spel, maar dat zijn geen binaire grafen meer, maar ternaire, omdat het gedrag van U (eigenlijk van U^{-1}) bepaald wordt door, verrassend genoeg, $n \pmod{3}$. Nu kijken we dus niet meer naar bits (binair) maar naar trits (ternair, een trit is een 0, 1 of 2). Het idee van de shift-afbeelding is hetzelfde, en die beschrijft nu de modulo-3-structuur van de baan van een getal onder de U -afbeelding. Zie Figuur 1.20 voor de ternaire De Bruijn-graaf op 9 punten.

Het $3n + 1$ -vermoeden, herformulering 2:

Voor $k \rightarrow \infty$ zullen de matrices A^{2k} en A^{2k+1} convergeren naar resp. $(a_0, a_1, 0, 0, \dots)$ en $(a_1, a_0, 0, 0, \dots)$, en $a_0 + a_1 = (1, 1, 1 \dots)$.

1.4.2 Eigenruimtes

[[Deze paragraaf vooronderstelt basiskennis van eigenwaarden, eigenvectoren en eigenruimtes.]]

Merk op dat $Aa_0 = a_1$ en $Aa_1 = a_0$, dus $A(a_0 + a_1) = a_0 + a_1$ en $A(a_0 - a_1) = -(a_0 - a_1)$, dus $a_0 + a_1$ is een eigenvector bij eigenwaarde 1, en $a_0 - a_1$ is een eigenvector bij eigenwaarde -1 . Algemeen geldt: de cyclen van lengte k corresponderen elk precies met een k -tal 1-dimensionale eigenruimtes met k -e-machts eenheidswortels als eigenwaarden.

Bij iedere cykel hoort dus ook precies één 1-dimensionale eigenruimte bij eigenwaarde 1. Ook bij een divergente component hoort precies één 1-dimensionale eigenruimte, namelijk met als voortbrengende eigenvector de vector met 1-tjes op de plekken van de componenten verder nullen.

Op deze manier krijg je alle eigenvectoren bij eigenwaarde 1. Deze observatie kan als volgt geformuleerd worden:

Stelling (Engl, 1982)

De $3n + 1$ -matrix A heeft eigenwaarde 1, en de dimensie van de eigenruimte is gelijk aan het aantal componenten in de $3n + 1$ -graaf.

En dat geeft dan meteen een herformulering van het $3n + 1$ -vermoeden in termen van de dimensie van een eigenruimte:

Het $3n + 1$ -vermoeden, herformulering 3:

De $3n + 1$ -matrix A heeft bij de eigenwaarde 1 een 1-dimensionale eigenruimte.

Dit vind ik wel een heel bijzondere herformulering van het vermoeden. Helaas is de theorie van spectra van oneindige matrices niet ver genoeg ontwikkeld. De verschillende varianten van het $3n + 1$ -vermoeden die we hierboven hebben gegeven laten ook direct zien dat die theorie met veel verschillende verschijnselen rekening zal moeten kunnen houden. Voor de $5n + 1$ -matrix, die visueel sterk lijkt op de $3n + 1$ -matrix, geldt dat de dimensie van die eigenruimte vermoedelijk al oneindig is.

Het complete spectrum van A en van $A^\top$ is inmiddels in kaart gebracht in termen van de componenten van de graaf, zie [LdW2].

1.4.3 Functionaalvergelijkingen

[[Deze paragraaf vooronderstelt basiskennis van de somnotatie $\sum_{n=0}^{\infty}$ en van complexe getallen.]]

Tenslotte: Berg en Meinardus (1994, [BM]) gaan nog een andere interessante kant op. We bekijken de machtreeks

$$f(z) = \sum_{n=0}^{\infty} e_n z^n = e_0 + e_1 z + e_2 z^2 + e_3 z^3 + \dots,$$

en we schrijven voor de coëfficiënten voor dat $e_n = e_{T(n)}$ voor alle n . De vraag is wat dat betekent voor de functie f . We definiëren

$$\begin{aligned} f_0(z) &= \sum_{n=0}^{\infty} e_{3n} z^{3n} = e_0 + e_3 z^3 + e_6 z^6 + \dots, \\ f_1(z) &= \sum_{n=0}^{\infty} e_{3n+1} z^{3n+1} = e_1 z + e_4 z^4 + e_7 z^7 + \dots, \\ f_2(z) &= \sum_{n=0}^{\infty} e_{3n+2} z^{3n+2} = e_2 z^2 + e_5 z^5 + e_8 z^8 + \dots \end{aligned}$$

Laat nu $\omega = -\frac{1}{2} + \frac{1}{2}i\sqrt{3}$, zodat $\omega^3 = 1$ en $\omega^2 = \bar{\omega} = -1 - \omega$. Nu is

$$\begin{aligned} f(z) &= f_0(z) + f_1(z) + f_2(z), \\ f(\omega z) &= f_0(z) + \omega f_1(z) + \bar{\omega} f_2(z), \\ f(\bar{\omega} z) &= f_0(z) + \bar{\omega} f_1(z) + \omega f_2(z), \end{aligned}$$

Met een beetje rekenwerk kunnen we f_2 nu uitdrukken in f :

$$f_2(z) = \frac{1}{3} (f(z) + \omega f(\omega z) + \bar{\omega} f(\bar{\omega} z)).$$

Nu gaan we de eis die we aan f hebben opgelegd, en die met de $3n+1$ -

functie te maken heeft, uitbuiten:

$$\begin{aligned}
 f(z) &= \sum_{n=0}^{\infty} e_{2n} z^{2n} + \sum_{n=0}^{\infty} e_{2n+1} z^{2n+1} \\
 &= \sum_{n=0}^{\infty} e_{T(2n)} z^{2n} + \sum_{n=0}^{\infty} e_{T(2n+1)} z^{2n+1} \\
 &= \sum_{n=0}^{\infty} e_n z^{2n} + \sum_{n=0}^{\infty} e_{3n+2} z^{2n+1} \\
 &= \sum_{n=0}^{\infty} e_n (z^2)^n + z^{-1/3} \sum_{n=0}^{\infty} e_{3n+2} (z^{2/3})^{3n+2} \\
 &= f(z^2) + z^{-1/3} f_2(z^{2/3}).
 \end{aligned}$$

En we hebben net gezien dat we f_2 kunnen terugvertalen naar f . Dat betekent dat we nu een zogenaamde functionaalvergelijking voor f hebben. Om gebroken machten van z te vermijden vervangen we z door z^3 , en de vergelijking kunnen we dan zo opschrijven:

$$3z(f(z^3) - f(z^6)) = f(z^2) + \omega f(\omega z^2) + \bar{\omega} f(\bar{\omega} z^2).$$

Deze functionaalvergelijking, die we de $3n+1$ -functionaalvergelijking noemen, is lineair. Er is direct een verband aan te geven met de $3n+1$ -graaf: bij een samenhangscomponent in die graaf hoort een functie f door $e_n = 1$ te nemen voor alle n in die component, en $e_n = 0$ voor alle n buiten die component. Twee oplossingen zijn $f(z) = 1$ (die hoort bij $n = 0$ die zijn eigen component is) en $f(z) = \frac{z}{1-z} = z + z^2 + z^3 + \dots$. Deze tweede ontstaat door de functie f bij de complete $3n+1$ -graaf te nemen.

Het verband tussen componenten in de $3n+1$ -graaf en onafhankelijke oplossingen van deze functionaalvergelijking (in machtreeksen) is weer één op één. Dit leidt meteen tot een herformulering van het $3n+1$ -vermoeden, namelijk dat er behalve de lineaire combinaties van 1 en $\frac{z}{1-z}$ geen andere oplossingen zijn, althans niet in machtreeksfuncties.

Het $3n+1$ -vermoeden, herformulering 4:

De oplossingsruimte van de $3n+1$ -functionaalvergelijking voor machtreeksfuncties f is 2-dimensionaal.

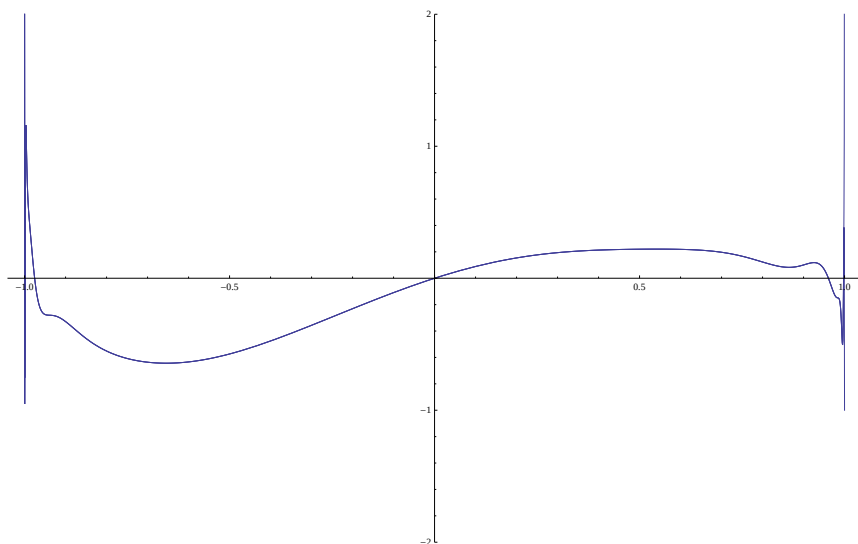
Als we de eis $e_n = e_{T(n)}$ vervangen door $e_n = -e_{T(n)}$, dan kunnen we net zoals zojuist gedaan opnieuw een functionaalvergelijking opstellen. Die verschilt slechts in een minteken:

$$3z(-f(z^3) - f(z^6)) = f(z^2) + \omega f(\omega z^2) + \bar{\omega} f(\bar{\omega} z^2).$$

Merk op dat nu geldt dat $e_n = e_{T^2(n)}$. Dit past precies op een component in de $3n + 1$ -graaf met een cykel van lengte 2. Deze vergelijking heeft dan ook een andere interessante oplossing, met $e_n = 1$ als n in een even aantal stappen op 1 uitkomt, en $e_n = -1$ als n in een oneven aantal stappen op 1 uitkomt (en $e_n = 0$ als n nooit op 1 uitkomt). Deze oplossing heeft dus de machtreeksvoorstelling

$$f(z) = z - z^2 - z^3 + z^4 + z^5 + z^6 - z^7 - \dots$$

Het zou interessant zijn deze functie beter te begrijpen. De reeks convergeert alleen op $(-1, 1)$. In Figuur 1.21 ziet u een grafiek.



Figuur 1.21: Een interessante functie die vermoedelijk alle informatie over het $3n + 1$ -vermoeden bevat.

Tenslotte doen we hetzelfde voor de amusicale permutatie. De functionaalvergelijking kan net zo als bij het $3n + 1$ -geval worden afgeleid:

$$\begin{aligned} 2z(2f(z^4) - f(z^6) - f(-z^6)) &= \\ &= (z^2 + 1)(f(z^3) - f(-z^3)) - i(z^2 - 1)(f(iz^3) - f(-iz^3)) \end{aligned}$$

(voor de doorzetters is dit een redelijke opgave).

Omdat de amusicale functie een permutatie is, zijn de convergente samenhangscomponenten (de cykels) in de amusicale graaf eindig. Dat betekent dat deze amusicale functionaalvergelijking oplossingen heeft in eindige machtreeksen. Maar dat zijn gewoon polynomen. Dat vind ik dan wel weer bijzonder, dat zo'n functionaalvergelijking plotseling blijkt polynomen als oplossingen te hebben. Omdat we vier cykels gevonden hadden, kunnen we zo een viertal polynomen opschrijven: $f(z) = z$, $f(z) = z^2 + z^3$, $f(z) = z^4 + z^5 + z^6 + z^7 + z^9$ en, als klap op de vuurpijl:

$$f(z) = z^{44} + z^{59} + z^{62} + z^{66} + z^{70} + z^{74} + z^{79} + z^{83} + z^{93} + z^{99} + z^{105} + z^{111}.$$

Ga maar na. Het klopt echt.

Het vermoeden is dat er niet meer polynomiale oplossingen zijn dan alle mogelijke lineaire combinaties van deze vier en $f(z) = 1$. Wel zijn er andere (vermoedelijk oneindig veel onafhankelijke, vermoedelijk oneindige) machtreeksoplossingen, waarvan de eerste is

$$f(z) = z^8 + z^{10} + z^{11} + z^{12} + z^{13} + z^{15} + z^{17} + z^{18} + z^{20} + \dots$$

1.5 Conclusie

Wat zijn we nou opgeschoten met het begrijpen van het $3n+1$ -vermoeden? Eigenlijk niet zo heel veel.

We hebben een aantal argumenten er voor gezien: experimentele (flink zoeken maar met een grote batterij computers), probabilistische (getallen lijken zich volgens de wetten van de statistiek te gedragen, en die zeggen dat het vermoeden op z'n minst niet onwaarschijnlijk is), en deelresultaten (bepaalde typen cykels bestaan niet, als er niet-triviale cykels bestaan dan zijn er getallen met onwaarschijnlijke eigenschappen). We hebben geprobeerd te zien hoe de $3n+1$ -graaf ingebed ligt in een veel grotere, wel goed begrepen graaf, maar daar liepen we in vast. En we hebben interessante herformuleringen gezien in de taal van matrixproducten, van dimensies van eigenruimtes, en van dimensies van oplossingsruimten van functionaalvergelijkingen.

Het lijkt er niet op dat het bewijs van het vermoeden (of een tegenbewijs) dichterbij is gekomen. Maar het is op z'n minst interessant te noemen dat het vermoeden vertakkingen blijkt te hebben in heel verschillende deelgebieden van de wiskunde. Je weet maar nooit of er op één van die gebieden ineens een doorbraak komt. We houden dus goede hoop.

Bibliografie

- [BM] LOTHAR BERG EN GÜNTER MEINARDUS, “Functional equations connected with the Collatz problem”, *Results in Mathematics* **25** [1994], 1–12.
- [C1] JOHN H. CONWAY, “Unpredictable Iterations”, Proceedings of the Number Theory Conference (University of Colorado, Boulder, CO, 1972), pp. 49-52. Opgenomen in [L2].
- [C2] JOHN H. CONWAY, “On Unsettleable Arithmetical Problems”, *American Mathematical Monthly* **120** [2013], 192–198.
- [Cr] RICHARD E. CRANDALL, “On the $3x+1$ problem”, *Mathematics of Computation* **32** [1978], 1281–1292.
- [E] H.W. ENGL, “An analytic representation for self-maps of a countably infinite set and its cycles”, *Aequationes Math.* **25** [1982], 90–96.
- [L1] JEFFREY C. LAGARIAS, “The $3x+1$ problem and its generalizations”, *American Mathematical Monthly* 92 [1985], 3–23, <http://www.cecm.sfu.ca/organics/papers/lagarias/>. Opgenomen in [L2].
- [L2] JEFFREY C. LAGARIAS (ED.), *The Ultimate Challenge: The $3x+1$ Problem*, American Mathematical Society, 2010.
- [L3] JEFFREY C. LAGARIAS, “The $3x+1$ problem: An annotated bibliography (1963–1999) (sorted by author)”, arXiv:math/0309224 [math.NT], 2011, <http://arxiv.org/abs/math/0309224v13>.
- [L4] JEFFREY C. LAGARIAS, “The $3x+1$ Problem: An Annotated Bibliography, II (2000-2009)”, arXiv:math/0608208 [math.NT], 2012, <http://arxiv.org/abs/math/0608208v6>.
- [LdW1] THIJS LAARHOVEN EN BENNE DE WEGER, “The Collatz conjecture and De Bruijn graphs”, *Indagationes Mathematicae* [2013], <http://www.sciencedirect.com/science/article/pii/S001935771300027X>.
- [LdW2] THIJS LAARHOVEN EN BENNE DE WEGER, werk in uitvoering, 2013.
- [O] GERHARD OPFER, “An analytic approach to the Collatz $3n +$

1 Problem”, Hamburger Beiträge zur Angewandten Mathematik, Preprint 2011-09, May 2011, <http://preprint.math.uni-hamburg.de/public/papers/hbam/hbam2011-09.pdf>.

- [OS] TOMÁS OLIVEIRA E SILVA, “Empirical Verification of the $3x + 1$ and Related Conjectures”, In [L2], pp. 189–207.
- [R] ERIC ROOSENDAAAL, “On the $3x + 1$ Problem”, <http://www.ericr.nl/wondrous/>.
- [SdW] JOHN SIMONS AND BENNE DE WEGER, “Theoretical and computational bounds for m -cycles of the $3n + 1$ -problem”, *Acta Arithmetica* **117** [2005], 51–70. Updated version 1.44 [2010]: [http://deweger.xs4all.nl/papers/\[35a\]SidW-3n+1-v1.44\[2010\].pdf](http://deweger.xs4all.nl/papers/[35a]SidW-3n+1-v1.44[2010].pdf).
- [Sm] IONICA SMEETS, “Gênante Problemen”, in: *Wiskunde: de uitdaging*, Syllabus Vakantiecursus 2010, CWI, Amsterdam, 2010, pp. 75–82.
- [St] RAY P. STEINER, “A Theorem on the Syracuse Problem”, Proc. 7th Manitoba Conference on Numerical Mathematics 1977, Winnipeg 1978, pp. 553–559.
- [dW] BENNE DE WEGER, “Comments on Opfer’s alleged proof of the $3n + 1$ Conjecture”, <http://www.win.tue.nl/~bdeweger/downloads/opfer-comments%20v0.2.pdf>.

2 Het vermoeden van Birch en Swinnerton-Dyer

Jaap Top

Rijksuniversiteit Groningen

2.1 Inleiding

Precies 50 jaar geleden, in 1963, publiceerden Bryan Birch en Peter Swinnerton-Dyer in het *Journal für die reine und angewandte Mathematik* het eerste gedeelte [B-SD63] van hun artikel *Notes on Elliptic Curves*. De tweede helft [B-SD65] verscheen in 1965 in hetzelfde tijdschrift. In de artikelen doen beide wiskundigen uit Cambridge (Engeland) verslag van hun jarenlange onderzoek betreffende elliptische krommen over de rationale getallen $\mathbb{Q}$. Daarbij was een voor hun tijd nieuw middel ingezet: de EDSAC 2 [W92], een computer die de universiteit van Cambridge in 1958 had aangeschaft en die tot 1965 dienst deed.

Een (over)versimpelde versie van de vraag die Birch en Swinnerton-Dyer probeerden te beantwoorden is de volgende. Begin met een polynoom

$$F(x) := ax^3 + bx^2 + cx + d$$

waarbij $a \neq 0$ en b, c, d rationale getallen zijn. Bestaan er rationale getallen r zodat $F(r)$ een kwadraat is (dat wil zeggen, $F(r) = s^2$ voor zeker rationaal getal s). En zoja, zijn er eindig veel, of zelfs oneindig veel zulke r ?

Fermat (1601–1665) en ook Euler (1707–1783) hadden al een methode aangegeven waarmee uit een oplossing $x = r$ in veel gevallen een andere oplossing gevonden kon worden. We komen hier straks op terug. Op p. 171 van een artikel [P01] van Henri Poincaré (1854–1912) staat een uitspraak die in feite betekent, dat uit een eindig aantal oplossingen $x = r_1, x = r_2, \dots, x = r_q$ door herhaald toepassen van methoden als die van Fermat

en Euler, *alle* oplossingen geconstrueerd kunnen worden. Poincaré (over hem zal in deze Vakantiecursus nog veel meer gezegd worden) verzuimt echter, een bewijs of zelfs maar een schets van een argumentatie bij zijn uitspraak te geven. L.J. Mordell (1888–1972) toonde overigens in 1922 aan, dat de uitspraak van Poincaré correct was, zie [M22].

Wiskundig gezien is het nieuwe idee van Birch en Swinnerton-Dyer vooral, dat ze het polynoom $F(x)$ modulo zoveel mogelijk priemgetallen p gaan bekijken. Voor zo'n priemgetal p kan je tellen hoeveel verschillende r mod p er zijn met de eigenschap, dat $F(r) \equiv s^2 \pmod{p}$ voor een zekere gehele s . Een heel naïeve en niet erg precieze vorm van het vermoeden van Birch en Swinnerton-Dyer, helemaal in de stijl van de inleiding van hun eerste artikel, luidt dan:

als er voor de meeste priemgetallen p onverwacht veel verschillende $r \pmod{p}$ bestaan waarvoor $F(r) \pmod{p}$ een kwadraat modulo p is, dan neemt $F(x)$ voor oneindig veel rationale waarden $x = r$ een kwadraat als waarde aan.

De verdienste van de twee artikelen van Birch en Swinnerton-Dyer is, dat dit is omgesmeed tot een precieze formulering, waarvoor ze bovendien een enorme numerieke evidentie gegeven hebben. In de afgelopen 50 jaar is flink wat vooruitgang geboekt, maar het vermoeden is nog altijd niet bewezen. In het jaar 2000 werd het vermoeden gekozen als een van de problemen in de beroemde Clay-lijst. Andrew Wiles schreef voor die gelegenheid een korte en heel leesbare introductie [W00].

In de voor u liggende tekst gaan we vooral niet proberen, de meest algemene en precieze formulering te geven. Wel hopen we wat handvatten te leveren waarmee u zelf kan rekenen aan en experimenteren met speciale gevallen van het vermoeden, en we geven een kort overzicht van tot nu toe behaalde resultaten.

2.2 Constructies van Fermat en Euler

Gegeven zijn rationale getallen a, b, c, d met $a \neq 0$, en de bijbehorende derdegraads veelterm

$$F(x) := ax^3 + bx^2 + cx + d.$$

Een methode waarmee zowel Fermat als Euler in diverse voorbeelden indrukwekkende rationale getallen r (met tellers en noemers die te groot

zijn voor een moderne standaard grafische rekenmachine) vonden, zodat $F(r)$ een kwadraat is, werkt als volgt. We gaan er vanuit, dat al zo'n oplossing $x = r$ gevonden is, en we zoeken een nieuwe. Daartoe schrijven we $x = \xi + r$, zodat de gegeven oplossing correspondeert met $\xi = 0$. Als veelterm in ξ hebben we dan

$$\tilde{F}(\xi) = F(\xi + r) = a\xi^3 + \tilde{b}\xi^2 + \tilde{c}\xi + \tilde{d},$$

met rationale getallen $a \neq 0$ (dezelfde waarmee we begonnen) en $\tilde{b}, \tilde{c}, \tilde{d}$.

Merk op dat

$$\tilde{d} = \tilde{F}(0) = F(r),$$

en daarvan is gegeven dat het een kwadraat is. Dus we kunnen schrijven $\tilde{d} = e^2$ voor een zeker rationaal getal e . Dit zorgt ervoor, dat we een uit het wat ouderwetse middelbaar onderwijs beproefde rekentechniek kunnen toepassen: "kwadraat afsplitsen". In dit geval houdt dit in, dat rationale getallen α, β gezocht worden zodat geldt

$$\tilde{F}(\xi) = (\alpha\xi^2 + \beta\xi + e)^2 + \gamma\xi^3 + \delta\xi^4$$

voor zekere rationale γ, δ . Door de coëfficiënten van de machten van ξ te vergelijken, zien we dat moet gelden

$$\begin{cases} 2\beta e &= \tilde{c}, \\ 2\alpha e + \beta^2 &= \tilde{b}. \end{cases}$$

Dit levert precies één oplossing (β, α) , behalve in het speciale geval dat $e = 0$ geldt. Dat laatste komt overeen met de voorwaarde dat we waren begonnen met een nulpunt $x = r$ van de veelterm $F(x)$. Gaan we er vanuit dat r geen nulpunt is, dan hebben we nu een schrijfwijze

$$\tilde{F}(\xi) = (\alpha\xi^2 + \beta\xi + e)^2 + \gamma\xi^3 + \delta\xi^4$$

gevonden. Hoe helpt dit alles bij het vinden van een $r' \neq r$ zodat $F(r)$ een kwadraat is, of equivalent, een $r'' \neq 0$ zodat $\tilde{F}(r'')$ een kwadraat is? Wel, in de gevonden uitdrukking voor $\tilde{F}(\xi)$ is de eerste term zonder meer een kwadraat, wat ook voor ξ wordt ingevuld. Dus we hebben een voorbeeld als de resterende termen samen de waarde 0 opleveren. Het resterende stuk laat zich schrijven als

$$\xi^3(\gamma + \delta\xi).$$

Voor $\xi = 0$ levert dit de waarde 0, maar deze ξ hadden we al. Een andere waarde voor ξ die werkt, is te vinden in het geval dat $\delta \neq 0$.

Een voorbeeld: begin met

$$F(x) := x^3 + 2x + 1.$$

Deze is te schrijven als

$$F(x) = \left(-\frac{1}{2}x^2 + x + 1\right)^2 + 2x^3 - \frac{1}{4}x^4.$$

Omdat $2x^3 - \frac{1}{4}x^4$ als nulpunt $x = 8$ heeft, geldt $F(8)$ is een kwadraat. Verder is

$$\begin{aligned} F(x+8) &= x^3 + 24x^2 + 194x + 529 \\ &= \left(-\frac{3287}{24334}x^2 - \frac{97}{23}x - 23\right)^2 - \frac{38998}{279841}x^3 - \frac{10804369}{592143556}x^4. \end{aligned}$$

Een nulpunt van

$$-\frac{38998}{279841}x^3 - \frac{10804369}{592143556}x^4$$

is

$$x = -\frac{82519768}{10804369}$$

en door hier 8 bij op te tellen, volgt dat $F(\frac{3915184}{10804369})$ een kwadraat is. Bij zowel Fermat als Euler komen aanzienlijk omvangrijkere rekenvoorbeelden voor...

Als $\delta = 0$, dan levert de methode geen nieuwe waarde r' zodat $F(r')$ een kwadraat is. Nagaande hoe de $\delta, \gamma, \beta, \alpha$ bepaald worden, blijkt dat $\delta = 0$ equivalent is met $\alpha = 0$, oftewel dat was begonnen met een veelterm

$$\tilde{F}(\xi) = a\xi^3 + \tilde{b}\xi^2 + \tilde{c}\xi + \tilde{d} = \gamma\xi^3 + (\beta\xi + s)^2.$$

Een meer analytische beschrijving van wat hier gebeurt, werkt dan als volgt. De vraag is, waarden voor ξ te vinden zodat $\tilde{F}(\xi)$ een kwadraat is, oftewel zodat $\sqrt{\tilde{F}(\xi)}$ bestaat, en rationaal is. De functie

$$\xi \mapsto \alpha\xi^2 + \beta\xi + s$$

die hierboven bepaald wordt, is precies de tweede orde Taylorbenadering bij $\xi = 0$ van $\pm\sqrt{\tilde{F}(\xi)}$ (met teken $+$ als $s > 0$ en teken $-$ als $s < 0$). En deze tweede orde benadering voldoet aan $\alpha = 0$ (oftewel, heeft slechts graad ≤ 1), precies dan als de functie

$$\xi \mapsto \sqrt{\tilde{F}(\xi)}$$

voor $\xi = 0$ een buigpunt heeft.

Ook als $\delta \neq 0$ kan er iets misgaan: namelijk, we zoeken naar een nulpunt van $\gamma + \delta\xi$, en dat is natuurlijk $\xi = -\gamma/\delta$. Dat zou de nieuwe waarde voor ξ moeten zijn zodat $\tilde{F}(\xi)$ een kwadraat is. Echter, als $\gamma = 0$ dan geeft dit gewoon weer de al bekende waarde $\xi = 0$.

Terugvertaald naar $F(x)$: gegeven een rationaal getal r zodat $F(r)$ een kwadraat is. Dan levert de door Fermat en Euler gebruikte methode een ander rationaal getal $r' \neq r$ waarvoor eveneens geldt $F(r')$ is een kwadraat, behalve in drie gevallen:

1. $F(r) = 0$;
2. voor $x = r$ heeft $\sqrt{F(x)}$ een buigpunt;
3. de methode levert dezelfde r weer op omdat “ $\gamma = 0$ ”.

Het bovenstaande laat zich vrij eenvoudig vertalen in de moderne manier waarop “elliptische krommen” behandeld worden. Daartoe eerst heel kort iets over de theorie. Een goed leesbare, elementaire inleiding in dit fascinerende onderwerp is het boekje van Joe Silverman en John Tate [ST92]. Allereerst eisen we van de veelterm $F(x)$ dat deze geen meervoudige nulpunten heeft, dus $F(x)$ is niet te schrijven in de vorm $a(x - \lambda)^2(x - \mu)$ of zelfs $a(x - \lambda)^3$ (was dat wel het geval, dan is het oorspronkelijk gestelde probleem heel simpel op te lossen, en er zijn dan in alle gevallen oneindig veel verschillende rationale waarden $x = r$ waarvoor $F(r)$ een kwadraat is). Voldoet $F(x)$ aan de bovengenoemde eis, definiëer dan een “elliptische kromme” als de collectie van alle paren (x, y) die voldoen aan

$$y^2 = F(x).$$

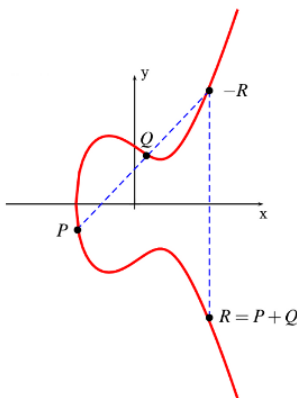
Hier voegt men nog een extra punt aan toe dat O genoemd wordt. Dat is niet een van de paren (x, y) , maar een nieuw punt, waaraan gedacht mag worden als een oneindig ver van de oorsprong verwijderd punt. De aanname dat $F(x)$ geen meervoudige nulpunten heeft, betekent dat in elk punt (α, β) van de kromme een unieke raaklijn aan de kromme is gedefiniëerd, namelijk de lijn met vergelijking

$$2\beta(y - \beta) = F'(\alpha)(x - \alpha).$$

Op de collectie van alle punten op de kromme laat zich dan als volgt een optelling definiëren.

1. $P + O = P$;

2. $P + Q = O$ als P en Q elkaars gespiegelde zijn ten opzichte van de x -as;
3. Als $P \neq Q$ en P en Q zijn niet elkaars gespiegelde ten opzichte van de x -as, dan snijdt de verbindingslijn van P en Q de kromme in drie punten (het derde punt kan samenvallen met P of met Q in het geval dat de verbindingslijn een raaklijn is). Nu geldt $P + Q = R$ waarbij R de gespiegelde in de x -as is van het derde gevonden punt.
4. Als $P = Q$ en dit punt ligt niet op de x -as (in dat geval was de optelling al in 2. hiervoor gedefinieerd), neem dan de raaklijn door P aan de kromme en het derde snijpunt van die raaklijn met de kromme. Dan geldt $P + P = R$ waarbij R de gespiegelde in de x -as is van het derde snijpunt.



De zo gegeven optelling blijkt aan de gebruikelijke regels te voldoen, dus bijvoorbeeld geldt $(P + Q) + R = P + (Q + R)$ voor elk drietal P, Q, R , en elk punt P heeft een unieke tegengestelde Q (namelijk de gespiegelde in de x -as) zodat $P + Q = O$. Die tegengestelde wordt als $-P$ geschreven. Pas op: we zijn misschien een beetje gewend om bij $-(x, y)$ beide coördinaten van het punt van teken te veranderen, maar voor de hier gegeven regel geldt $-(x, y) = (x, -y)$. Tenslotte geldt per definitie dat de gemaakte optelling commutatief is: er geldt $P + Q = Q + P$ voor elk paar punten P, Q .

Als begonnen wordt met punten P, Q die beide coördinaten rationaal hebben, dan geldt hetzelfde voor $P + Q$ en voor $-P$ (voorzover die $\neq O$ is). Dus samen met O is de collectie van al zulke punten gesloten onder de bewerkingen optellen en tegengestelde nemen. Deze collectie wordt in de

artikelen van Birch en Swinnerton-Dyer met de letter Γ aangeduid.

In termen van de optelling blijkt de constructiemethode van Fermat en van Euler het volgende te doen: gegeven een punt $P \in \Gamma$ wordt het ‘nieuwe’ punt

$$R := -3P = -(P + P + P)$$

bepaald. Als begonnen wordt met een P op de x -as, dan is $P + P = O$, en dus $R = P$. Die hadden we al. En als begonnen was met een buigpunt van de grafiek van $\sqrt{F(x)}$, dan geldt $R = O$, dus ook hier krijgen we geen ‘echte’ oplossing. Het derde geval waar geen nieuwe oplossing gekregen wordt is nu ook eenvoudig te verklaren: als geldt $4P = O$, dan is $R = P$, dus dan levert de methode hetzelfde punt op waarmee was begonnen. In termen van de gebruikte variabelen, gebeurt dit als $\gamma = 0$, dus als geldt

$$\tilde{F}(\xi) = (\alpha\xi^2 + \beta\xi + e)^2 - \alpha^2\xi^4 = 2\alpha\beta\xi^3 + (2\alpha e + \beta^2)\xi^2 + 2\beta e\xi + e^2.$$

Deze veelterm heeft graad 3 en geen meervoudige nulpunten als aan de drie eisen

$$\begin{cases} \alpha \neq 0; \\ e \neq 0; \\ \beta^2 \neq 8\alpha e. \end{cases}$$

is voldaan. In dat geval voldoet $P = (0, e)$ aan $4P = O$ op de kromme gegeven door $y^2 = F(\xi)$.

Nog een voorbeeld: begin met

$$F(x) := x^3 + 9x^2 - 16x + 64.$$

Beginnend met $x = 0$ levert de methode van Fermat en Euler dan achtereenvolgend (reken maar na!) de waarden $x = 8$ en $x = -8$ op, gevolgd door (opnieuw) $x = 0$ enzovoort! Dus we krijgen een periodieke baan, met periode 3. Het punt $P = (0, 8)$ op de kromme met vergelijking $y^2 = x^3 + 9x^2 - 16x + 64$ blijkt in dit geval te voldoen aan $7P = O$.

Tenslotte: de methode zoals hier beschreven houdt zoals we zagen nauw verband met de tweede orde Taylorbenadering van de functie $\sqrt{\tilde{F}(\xi)}$ om $\xi = 0$. Maar we hadden ook kunnen volstaan met de eerste orde Taylorbenadering! Dat levert een schrijfwijze

$$\tilde{F}(\xi) = (\beta\xi + e)^2 + a\xi^3 + \epsilon\xi^2.$$

Door ξ zo te kiezen dat $a\xi^3 + \epsilon\xi^2$ de waarde 0 krijgt, wordt weer een oplossing gekregen. In termen van de gegeven optelling betekent dit, dat

uitgaande van een punt P op de kromme, het nieuwe punt

$$R' = -2P = -(P + P)$$

wordt geconstrueerd. Oftewel: neem het derde snijpunt met de kromme van de raaklijn door P aan de kromme. Deze “raaklijnmethode” staat in veel teksten over elliptische krommen vermeld, maar de “paraboolmethode” die door Fermat en Euler veel vaker wordt toegepast, lijkt wat in de vergetelheid geraakt.

2.3 modularekenen

Het probleem van kwadraten als waarden $F(x)$ bij een rationale x te vinden, verandert niet als x met een vast rationaal getal $\neq 0$ wordt vermenigvuldigd, en evenmin als $F(x)$ met een vast kwadraat wordt vermenigvuldigd. Door $F(x) = ax^3 + bx^2 + cx + d$ te vermenigvuldigen met a^2 en vervolgens x te vervangen door $X = ax$, mogen we dus aannemen

$$F(X) = X^3 + bX^2 + cX + d$$

voor zekere (misschien nieuwe) rationale b, c, d . Vermenigvuldigen we dit met n^6 en schrijven we $x := n^2X$ (een nieuwe variabele x), dan staat er

$$x^3 + n^2bx^2 + n^4cx + n^6d.$$

Voor een geschikte keuze van n is dit een veelterm die alle coëfficiënten geheel heeft. Kortom, zonder verlies van algemeenheid kan worden aangenomen dat

$$F(x) = x^3 + bx^2 + cx + d$$

voor zekere gehele getallen b, c, d . Dit stelt ons in staat, de veelterm te reduceren modulo ieder geheel getal, in het bijzonder modulo priemgetallen.

Daartoe eerst kort iets over het rekenen modulo een priemgetal. Kies een willekeurig priemgetal p . De ‘getallen modulo p ’ zijn de uitdrukkingen $\bar{a} = a \bmod p$ waarin a een geheel getal is. Hierbij geldt de regel

$$\bar{a} = \bar{b} \iff p \text{ is een deler van } b - a.$$

Anders gezegd: $\bar{a}$ hangt alleen maar af van de rest die overblijft wanneer a door p gedeeld wordt, niet van het gehele getal a zelf. Er zijn dus in totaal precies p ‘getallen modulo p ’, namelijk

$$\bar{0}, \bar{1}, \bar{2}, \dots, \overline{p-1}.$$

Met ‘getallen modulo p ’ valt net zo te rekenen als met gewone gehele getallen:

$$\bar{a} + \bar{b} := \overline{a + b}$$

en

$$\bar{a} \cdot \bar{b} := \overline{a \cdot b}$$

blijken goed-gedefinieerde bewerkingen op te leveren.

Tot nu toe is niet gebruikt, dat p wordt verondersteld een priemgetal te zijn. Die eigenschap levert iets speciaals voor de vermenigvuldiging: neem een willekeurige $\bar{a} \neq \bar{0}$. Met ook nog twee elementen $\bar{b}, \bar{c}$ geldt dan

$$\begin{aligned} \bar{a} \cdot \bar{b} = \bar{a} \cdot \bar{c} &\Leftrightarrow \overline{ab} = \overline{ac} \\ &\Leftrightarrow p \text{ is een deler van } ab - ac \\ &\Leftrightarrow p \text{ is een deler van } a(b - c) \\ &\Leftrightarrow p \text{ is een deler van } b - c \\ &\Leftrightarrow \bar{b} = \bar{c}. \end{aligned}$$

Immers, $\bar{a} \neq \bar{0}$ oftewel p is geen deler van a . Dus we zien: alle producten $\bar{a} \cdot \bar{b}$ verschillen, als $\bar{b}$ de collectie ‘getallen modulo p ’ doorloopt. Er zijn p zulke producten en ook precies p ‘getallen modulo p ’, dus kennelijk komt elk getal modulo p voor als zo’n product. In het bijzonder bestaat er een unieke $\bar{b}$ waarvoor geldt

$$\bar{a} \cdot \bar{b} = \bar{1},$$

deze $\bar{b}$ heet de ‘inverse’ van $\bar{a}$. Met de getallen modulo p kan men dus optellen, aftrekken, vermenigvuldigen en zelfs delen vergelijkbaar met gewone (rationale of reële) getallen. De collectie getallen modulo p , voorzien van deze optelling en vermenigvuldiging, wordt gewoonlijk genoteerd als $\mathbb{F}_p$.

Wij zijn met name geïnteresseerd in kwadraten in $\mathbb{F}_p$, oftewel elementen van de soort $\bar{a}^2 = \bar{a} \cdot \bar{a}$. Er geldt

$$\begin{aligned} \bar{a}^2 = \bar{b}^2 &\Leftrightarrow p \text{ is een deler van } a^2 - b^2 \\ &\Leftrightarrow p \text{ is een deler van } (a - b)(a + b) \\ &\Leftrightarrow p \text{ deelt } a - b \text{ of } p \text{ deelt } a + b \\ &\Leftrightarrow \bar{a} = \bar{b} \text{ of } \bar{a} = \overline{-b}. \end{aligned}$$

Voor $p = 2$ geldt bovendien $\bar{b} = \overline{-b}$ voor elke b , in dat geval is elk van de twee elementen van $\mathbb{F}_2$ zijn eigen kwadraat. Voor $p > 2$ geldt $\bar{b} = \overline{-b}$ alleen als $\bar{b} = \bar{0}$. Dus in dat geval zijn de $(p - 1)/2$ kwadraten van de paren $\{\pm \bar{a}\}$ met $\bar{a} \neq \bar{0}$ allemaal verschillend, en bovendien verschillend van $\bar{0}^2 = \bar{0}$. Conclusie: voor oneven p zijn er precies $(p + 1)/2$ kwadraten

in $\mathbb{F}_p$ (inclusief $\bar{0}$). Kortom, ongeveer de helft van alle elementen van $\mathbb{F}_p$ is kwadraat.

Voorbeeld: de kwadraten in $\mathbb{F}_{17}$ zijn

$$\{\bar{0}, \bar{1}, \bar{2}, \bar{4}, \bar{8}, \bar{9}, \bar{13}, \bar{15}, \bar{16}\}$$

en die in $\mathbb{F}_{23}$ zijn

$$\{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{6}, \bar{8}, \bar{9}, \bar{12}, \bar{13}, \bar{16}, \bar{18}\}.$$

De veelterm

$$F(x) = x^3 + bx^2 + cx + d$$

met gehele b, c, d levert ook een veelterm met coëfficiënten in $\mathbb{F}_p$, namelijk

$$\bar{F}(x) := x^3 + \bar{b}x^2 + \bar{c}x + \bar{d}.$$

Voor elke $r \in \mathbb{F}_p$ is dan ook $\bar{F}(r) \in \mathbb{F}_p$, en men kan de vraag stellen hoe vaak het gebeurt dat $\bar{F}(r)$ een kwadraat is. Equivalent hiermee: neem de vergelijking

$$s^2 = \bar{F}(r),$$

en schrijf M_p voor het aantal oplossingen $(r, s) \in \mathbb{F}_p \times \mathbb{F}_p$ dat deze vergelijking heeft. Beter nog, omdat in de theorie van elliptische krommen er ook nog dat ene “oneindige” punt O is, tellen we die er ook bij:

$$N_p := 1 + M_p.$$

Een triviale grens is dan $N_p \geq 1$, want minder dan 0 oplossingen kan de vergelijking niet hebben. Evenzo is er de bovengrens $N_p \leq 2p + 1$: immers, die zou bereikt worden als elke $r \in \mathbb{F}_p$ een waarde $\bar{F}(r)$ zou opleveren die een kwadraat $\neq \bar{0}$ is. In 1933 bewees Helmut Hasse [H33] veel sterkere grenzen voor N_p : er geldt

$$p + 1 - 2\sqrt{p} < N_p < p + 1 + 2\sqrt{p}.$$

Deze grenzen laten zien dat voor ieder priemgetal p , het getal N_p vrij dicht bij p ligt, dus voor ruwweg de helft van de mogelijke r zal $\bar{F}(r)$ een kwadraat zijn, voor de andere helft niet.

Als voorbeeld nemen we

$$F(x) := x^3 + x + 1.$$

Dat levert de volgende tabel.

p	2	3	5	7	11	13	17	19	23	29	97	101	211	307	401
N_p	3	4	9	5	14	18	18	21	28	36	97	105	223	301	432

2.4 Het vermoeden

Laat

$$F(x) := x^3 + bx^2 + cx + d$$

een derdegraads veelterm zijn met gehele coëfficiënten b, c, d en zonder meervoudig nulpunt. Zoals in de vorige sectie gezien is, wordt dan voor elk priemgetal p een geheel getal N_p gedefinieerd. Definieer vervolgens, voor elke $P \geq 2$,

$$f(P) := \prod_{p \leq P} \frac{N_p}{p}.$$

Birch en Swinnerton-Dyer doen dan het volgende, eenvoudig geformuleerde vermoeden.

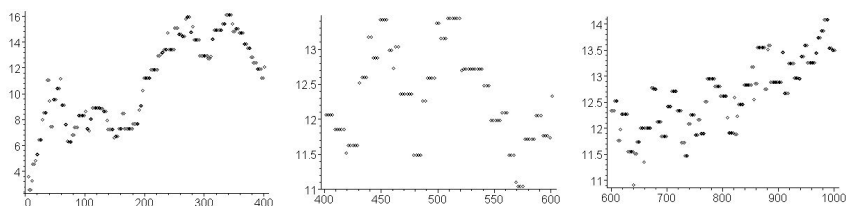
$F(r)$ is voor oneindig veel rationale getallen r een kwadraat, precies dan als de functie $P \mapsto f(P)$ onbegrensd is.

De uitspraak is zelfs preciezer: voor $P \rightarrow \infty$ zou moeten gelden

$$f(P) \sim C \cdot \ln(P)^g,$$

voor een positieve constante C , waarin g de “rang” (het maximale aantal onafhankelijke elementen) van de groep Γ is.

Als voorbeeld nemen we weer $F(x) = x^3 + x + 1$ en we tekenen de grafiek van $P \mapsto f(P)$ op drie verschillende intervallen.



Zelfs in dit kleine voorbeeld blijkt al wel, dat het lastig te voorspellen is of de functie $P \mapsto f(P)$ begrensd is of niet. Om het vermoeden veel preciezer te maken, wordt de zogeheten L -reeks van de elliptische kromme gebruikt. De daartoe gebruikte manipulatie ziet er mogelijk nogal willekeurig uit, maar er zit een omvangrijke theorie achter die we hier laten voor wat het is.

Het polynoom $F(x)$ levert voor bijna alle priemgetallen p een polynoom $\overline{F}(x)$ over $\mathbb{F}_p$ op, dat eveneens geen meervoudige nulpunten heeft. De

eindige verzameling priemgetallen waarvoor er wel een meervoudig nulpunt is, samen met nog het priemgetal 2, noemen we S . Het al eerder genoemde resultaat van Hasse zegt dat als $p \notin S$, dan is N_p ongeveer $p + 1$. We schrijven daarom

$$N_p = p + 1 - a_p$$

waarin dus a_p een geheel getal is dat voldoet aan $|a_p| < 2\sqrt{p}$. Dan is

$$\frac{N_p}{p} = 1 - \frac{a_p}{p} + p^{-1},$$

en dat is wat verkregen wordt door $s = 1$ in te vullen in de uitdrukking

$$1 - a_p p^{-s} + p^{1-2s}.$$

In plaats van kijken of $f(P)$ willekeurig groot wordt als P groeit, kunnen we natuurlijk ook kijken of $\frac{1}{f(P)}$ naar 0 gaat als $P \rightarrow \infty$. Dat betekent, dat we een product nemen van factoren $\frac{p}{N_p}$, oftewel van

$$\frac{1}{1 - a_p p^{-s} + p^{1-2s}}$$

(en dan invullen $s = 1$). Dit leidt ertoe, te kijken naar het product

$$\prod_{p \notin S} \frac{1}{1 - a_p p^{-s} + p^{1-2s}}$$

Voor $s > 3/2$ blijkt dit product te convergeren (maar $s = 1$ is te klein!). De theorie voegt ook voor de resterende $p \in S$ een factor toe, en zo ontstaat een functie $s \mapsto L(s)$ die de “ L -reeks” van de bijbehorende elliptische kromme heet.

Dit leidt tot de volgende versie van het vermoeden. Stel dat het mogelijk is, de functie $s \mapsto L(s)$ op een natuurlijke manier uit te breiden naar een op alle reële getallen gedefinieerde (willekeurig vaak differentieerbare) functie, die we ook $L(s)$ noemen. Zou dan gelden dat $L(1) = 0$, precies dan als er oneindig veel rationale r bestaan waarvoor $F(r)$ een kwadraat is? Beter nog, zou $L(s)$ voor $s = 1$ een nulpunt hebben waarvan de multipliciteit precies de “rang” van de groep Γ is? En zelfs hiermee eindigt het vermoeden niet; als die rang van de groep Γ gelijk is aan g , dan zou de g -de afgeleide van $L(s)$ voor $s = 1$ een door de uiteindelijke versie van het vermoeden voorspelde waarde moeten hebben. Maar het zou te ver voeren om de details daarvan hier te geven.

Rekenen met de functie $L(s)$ is voor iedereen mogelijk, bijvoorbeeld met behulp van programma Magma. Op

<http://magma.maths.usyd.edu.au/calc/>

kan je ermee aan de slag. De regels

```
R<x>:=PolynomialRing(Rationals());
E:=EllipticCurve(x^3+x+1);
Evaluate(LSeries(E),1);
Evaluate(LSeries(E),1 : Derivative:=1);
```

leveren als uitvoer

```
0.00000000000000000000000000000000
1.78580949386920068702005538692
```

Dit geeft dit aan dat waarschijnlijk geldt $L(1) = 0$, dus als het vermoeden in dit geval klopt dan is $r^3 + r + 1$ een kwadraat voor oneindig veel verschillende rationale r . Ook zien we dat de afgeleide $L'(1) \neq 0$. En hadden we **Derivative:=5** geschreven, dan was de vijfde afgeleide in het punt $s = 1$ benaderd. Met het commando

```
MordellWeilShaInformation(E);
```

is overigens te zien dat er inderdaad oneindig veel rationale r zijn waarvoor $r^3 + r + 1$ een kwadraat is: de groep Γ van rationale punten op de kromme met vergelijking $y^2 = x^3 + x + 1$ blijkt te bestaan uit alle nP waarbij $P = (0, 1)$, en $nP \neq mP$ voor $n \neq m$.

En nu we toch met Magma bezig zijn: ook de voorbeelden in de voorgaande secties zijn daarmee uitstekend te verifiëren. Ter illustratie:

```
R<x>:=PolynomialRing(Rationals());
E:=EllipticCurve(x^3+2*x+1);
P:=E![0,1];
P;
(-3)*P;
(-3)^2*P;
```

Dit levert als antwoord de punten $P = (0, 1)$ en $-3P$ en $9P$, en het is gemakkelijk te controleren dat de eerste coördinaat daarvan overeenstemt met wat we al eerder hadden uitgerekend. Ook de eenvoudiger raaklijnmethode (dus $-2P$ en $4P$ en $-8P$ enzovoorts) uitrekenen is zo geen probleem.

2.5 Resultaten

Al is het vermoeden van Birch en Swinnerton-Dyer nog onbewezen, er is in de afgelopen 50 jaar flink wat vooruitgang geboekt. Enkele van de meest spectaculaire ontdekkingen daarvan noemen we hier kort, al is dit overzicht verre van compleet.

Voor speciale $F(x)$, namelijk van de vorm $F(x) = x^3 + ax$ en ook de vorm $F(x) = x^3 + k$, konden Birch en Swinnerton-Dyer in hun artikelen al nagaan dat inderdaad een natuurlijke uitbreiding van de functie $L(s)$ tot alle getallen bestond. Bovendien slaagden ze erin, in deze gevallen een formule voor $L(1)$ te vinden waarmee ze in concrete gevallen konden verifiëren of $L(1) = 0$ of $L(1) \neq 0$. Vrij snel na het bekendmaken van hun vermoeden, werd het verregaand veralgemeniseerd. Allereerst werd (met hulp van Ian Cassels, Jean-Pierre Serre en John Tate) de L -functie in alle gevallen precies beschreven, inclusief factoren voor de “slechte priemenvrijdelers”. Bovendien werd het vermoeden uitgebreid naar andere situaties dan alleen maar de rationale getallen, en ook naar hogere dimensies (niet alleen elliptische krommen).

In 1977 publiceerden John Coates en Andrew Wiles [CW77] (dezelfde die later onder meer de Laatste Stelling van Fermat in samenwerking met Richard Taylor zou bewijzen) een eerste flinke stap vooruit. Ze beperkten zich tot elliptische krommen met een extra eigenschap, namelijk dat ze “complexe vermenigvuldiging” toelaten. Bijvoorbeeld de gevallen $F(x) = x^3 + ax$ en $F(x) = x^3 + k$ waar Birch en Swinnerton-Dyer al diverse voorbeelden van deden, voldoen hieraan, maar ook bijvoorbeeld $F(x) = x^3 + 2kx^2 + 2k^2x$. Dankzij die complexe vermenigvuldiging was bekend dat de bijbehorende functie $L(s)$ inderdaad een natuurlijke uitbreiding tot alle getallen bezit. Het resultaat van Coates en Wiles is in dit geval

$$\Gamma \text{ heeft oneindig veel elementen} \Rightarrow L(1) = 0.$$

In 1982 gaf Dorian Goldfeld [G82] een interpretatie van de constante C die optreedt in de eerste, naïeve versie $f(P) \sim C \ln(P)^g$ van het vermoeden. Ook Keith Conrad [C05] schreef hierover een lang artikel.

Een groot probleem bij het vermoeden in het algemene geval was het uitbreiden van de functie $L(s)$. Dankzij ouder werk van Martin Eichler (1912–1992) en Goro Shimura (geboren in 1930) was dat echter geen probleem voor de “modulaire elliptische krommen over $\mathbb{Q}$ ”. Onderzoek in de jaren ’80 ging daarom met name uit van de voorwaarde dat de elliptische kromme over $\mathbb{Q}$ die beschouwd werd, “modulair” was. In 1986 publiceerden Dick

Gross en Don Zagier een resultaat [GZ86] over zogeheten “Heegner punten” op modulaire elliptische krommen. Ze lieten zien, dat zo’n punt P oneindig veel verschillende veelvouden nP heeft, precies dan als $L(1) = 0$ en $L'(1) \neq 0$. En dus

$$L(1) = 0 \text{ en } L'(1) \neq 0 \Rightarrow \Gamma \text{ heeft oneindig veel elementen.}$$

voor modulaire elliptische krommen.

Kort daarop completeerde de Russische wiskundige Victor Kolyvagin [K88] dit werk van Gross en Zagier, met de twee stellingen

$$L(1) \neq 0 \Rightarrow \Gamma \text{ is eindig}$$

en ook

$$L(1) = 0 \text{ en } L'(1) \neq 0 \Rightarrow \Gamma \text{ heeft rang 1.}$$

Ook hier onder de aanname dat het om modulaire elliptische krommen over $\mathbb{Q}$ gaat.

In 1998 completeerden (de publicatie volgde in 2001) Christophe Breuil, Brian Conrad, Fred Diamond en Richard Taylor [BCDT01] het werk dat een paar jaar eerder was begonnen door Andrew Wiles [W95], [TW95]: het resultaat is, dat *elke* elliptische kromme over $\mathbb{Q}$ “modulair” is. Daarmee gelden dus de stellingen van Gross en Zagier, en die van Kolyvagin, voor alle elliptische krommen over $\mathbb{Q}$. Toch is hiermee het originele vermoeden van Birch en Swinnerton-Dyer nog zeker niet bewezen. De Amerikaan Chris Skinner geeft de afgelopen maanden (zomer 2013) voordrachten over de omkering van Kolyvagin’s stellingen: als Γ eindig is, of rang 1 heeft, hoe zou je dan kunnen aantonen dat $L(1) \neq 0$ resp. $L'(1) \neq 0$? Skinner heeft ideeën hoe je dit wellicht kan aanpakken. En zodra $L(s)$ in $s = 1$ een nulpunt heeft met multipliciteit minstens 2, dan is uit de tot nu toe bewezen stellingen nog niet te concluderen of Γ oneindig is. Ook omgekeerd, bevat Γ oneindig veel elementen, dan volgt dat $L(s)$ in $s = 1$ een nulpunt heeft, maar over de multipliciteit ervan valt nog niks te concluderen. Kortom, het laatste woord over het vermoeden van Birch en Swinnerton-Dyer is nog zeker niet gezegd!

Bibliografie

- [B-SD63] B.J. Birch & H.P.F. Swinnerton-Dyer, *Notes on elliptic curves*. I, J. reine angew. Math., **212** (1963), 7–25.

- [B-SD65] B.J. Birch & H.P.F. Swinnerton-Dyer, *Notes on elliptic curves. II*, J. reine angew. Math., **218** (1965), 79–108.
- [BCDT01] C. Breuil, B. Conrad, F. Diamond, R. Taylor, *On the modularity of elliptic curves over $\mathbf{Q}$: wild 3-adic exercises*, J. Amer. Math. Soc., **14** (2001), 843–939.
- [CW77] J. Coates & A. Wiles, *On the Conjecture of Birch and Swinnerton-Dyer*, Invent. Math., **39** (1977), 223–252.
- [C05] K. Conrad, *Partial Euler products on the critical line*, Canad. J. Math., **57** (2005), 267–297.
- [G82] D. Goldfeld, *Sur les produits eulériens attachés aux courbes elliptiques*, C.R.Acad. Sci. Paris Sér. I Math., **294** (1982), 471–474.
- [GZ86] B.H. Gross & D.B. Zagier, *Heegner points and derivatives of L -series*, Invent. Math., **84** (1986), 225–320.
- [H33] H. Hasse, *Beweis des Analogons der Riemannschen Vermutung für die Artinschen und F.K. Schmidtschen Kongruenzzetafunktionen in gewissen elliptischen Fällen*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen (1933), 253–262.
- [K88] V.A. Kolyvagin, *Finiteness of $E(Q)$ and $\text{III}(E, Q)$ for a subclass of Weil curves* (Russian), Izv. Akad. Nauk SSSR Ser. Mat., **52** (1988), 522–540, 670–671; engelse vertaling in Math. USSR-Izv., **32** (1989), 523–541.
- [M22] L.J. Mordell, *On the rational solutions of the indeterminate equations of the third and fourth degrees*, Proc. Cambridge Phil. Soc., **21** (1922), 179–192.
- [P01] H. Poincaré, *Sur les propriétés arithmétiques des courbes algébriques*, Journ. de math. pures et appl. 5^e série, **7** (1901), 161–234.
- [ST92] J.H. Silverman & J.T. Tate, *Rational Points on Elliptic Curves*. New York etc.: Springer-Verlag, Undergraduate Texts in Mathematics, 1992.
- [W95] A. Wiles, *Modular elliptic curves and Fermat’s last theorem*, Ann. of Math., **141** (1995), 443–551.
- [TW95] R. Taylor & A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math., **141** (1995), 553–572.
- [W00] A. Wiles, *The Birch and Swinnerton-Dyer conjecture* (2000),

www.claymath.org/millennium/Birch_and_Swinnerton-Dyer_Conjecture/birchswin.pdf

- [W92] M.V. Wilkes, *Edsac 2*, IEEE Annals of the History of Computing, **14** (1992), 49–56.

3 Navier-Stokes

Barry Koren

Technische Universiteit Eindhoven

3.1 Stromingsleer

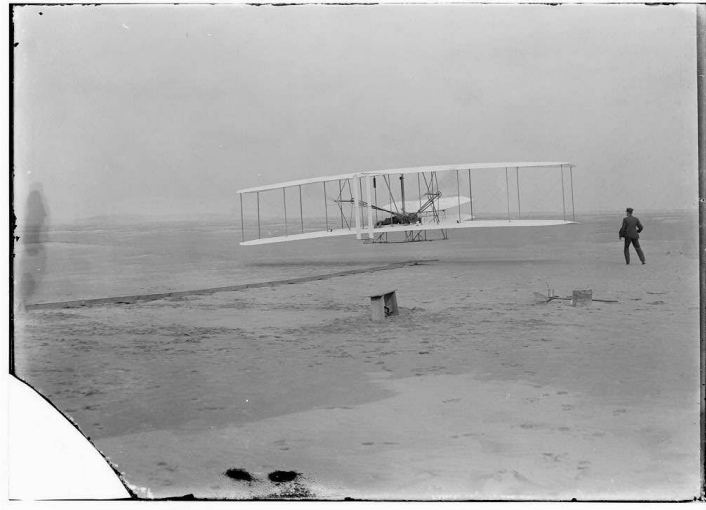
3.1.1 Belang

Overall stromen gassen en vloeistoffen. Lucht stroomt om de aarde, om vliegtuigen, raketten en windmolens, en door onze longen. Water stroomt om de aarde en om schepen, benzine door motoren, en bloed door ons hart. De wetenschap die zich bezighoudt met de bestudering van de stroming van gassen en vloeistoffen heet stromingsleer, in het Engels *fluid dynamics*. Stromingsleer is een onderdeel van de natuurkunde. In het natuurkunde-onderwijs op middelbare scholen wordt in het algemeen geen les gegeven in stromingsleer. Stromingsleer vereist daarvoor te veel voorkennis van vooral wiskunde. Op universiteiten wordt er wel onderwijs in gegeven, op technische universiteiten zoals in Delft, Eindhoven en Twente zelfs heel veel. Wil je bijvoorbeeld een vliegtuig, raket, benzinemotor of kunsthart ontwerpen, dan kun je niet om stromingsleer heen.

Met name voor het ontwerpen van vliegtuigen is kennis en begrip van stromingsleer, aërodynamica in dit geval, van groot belang. Met uitzondering van de gevaarlijke zwaartekracht zijn alle krachten die in de vlucht op een vliegtuig werken: luchtkrachten. Het veilig door de lucht laten bewegen, over onze hoofden heen, van een vliegtuig afgetankt met tonnen brandstof en aan boord een paar honderd passagiers, maakt een precieze kennis, begrip en ook beheersing van deze luchtkrachten van levensbelang. Vliegen moet bovendien niet alleen veilig, maar ook nog eens zuinig en stil. Voor het bouwen van vliegtuigen is stromingsleer echt onontbeerlijk.

3.1.2 Experiment en theorie

Voor het onderzoeken van stromingen staan vandaag de dag zowel experimentele als theoretische middelen ter beschikking. Bij experimentele middelen kan in geval van aërodynamica worden gedacht aan windtunnels. De gebroeders Wright, die de eerste bemande motorvlucht uitvoerden (Figuur 3.1), hadden al de beschikking over een windtunnel, een zelfgemaakte.



Figuur 3.1: Glasplaatfoto van, men zegt, de eerste bemande motorvlucht ooit (vliegafstand: 37 meter, vluchtduur: 12 seconden), Kitty Hawk, North Carolina, 10.35 u., 17 december 1903. Vooroverliggend op de onderste vleugel: Orville Wright. Meelopend om het vliegtuig zo nodig met de hand in evenwicht te houden: Wilbur Wright. Zichtbaar op de voorgrond onder andere: de rail waarop het vliegtuig is gestart en het bankje waarop de vleugel heeft gesteund. Dezelfde dag nog voerden de gebroeders Wright al een vlucht uit van 260 meter en 59 seconden!

Windtunnelwerk geniet vertrouwen omdat het met echte luchtstromingen werkt en niet met denkbeeldige, zoals de theoretische aërodynamica. Theoretische aërodynamica, theoretische stromingsleer in het algemeen, is lang een academische aangelegenheid geweest, met resultaten die sterk afweken van experimentele waarnemingen. Technische toepassingen van stromingsleer ontwikkelden zich dan ook eerst vrijwel onafhankelijk van de theorie.

Theoretische en technologische doorbraken hebben de kloof tussen theorie en experiment echter gedicht en er is nu sprake van een vruchtbare wisselwerking tussen beide. Het vliegtuig heeft hierbij een sterk stimulerende rol gespeeld. Moderne vliegtuigen zijn wonderen van experimentele en theoretische stromingsleertechniek. Tegenwoordig wordt het technologische belang van theoretische stromingsleer alom erkend.

3.1.3 Theoretische stromingsleer

Theoretische stromingsleer heeft een roemrijke geschiedenis, zie bijvoorbeeld ^{1 2 3}. Vele grote geesten hebben in de loop der eeuwen hun tanden gezet in de bestudering van stromingen en daarmee stap voor stap het vak opgebouwd. Theoretische stromingsleer begint al bij Aristoteles, die de begrippen continuüm en luchtweerstand invoerde.

Een mijlpaal in de theoretische stromingsleer is de invoering van partiële differentiaalvergelijkingen voor de beschrijving van de beweging van gas- en vloeistofstromingen op basis van de Tweede Wet van Newton (kracht = massa $\times$ versnelling). Partiële differentiaalvergelijkingen zijn wiskundige formules waarmee het verloop van vele ingewikkelde processen in allerlei vakgebieden – niet alleen exacte wetenschappen en techniek – op heldere en uiterst compacte wijze kan worden beschreven. De eerste partiële differentiaalvergelijkingen voor gas- en vloeistofstromingen werden in 1755 ingevoerd door de wiskundige Leonard Euler⁴. Dat een wiskundige de stromingsleer vooruit helpt zien we wel vaker. De stromingsleer heeft met haar rijkdom aan wiskundeproblemen altijd wiskundigen aangetrokken.

Leonhard Euler (Figuur 3.2) is een van de knapste wiskundigen aller tijden. Legendes vertellen dat Euler, getrouwd en vader van 13 kinderen, veel van zijn wiskundepublicaties schreef bij kaarslicht, met een kind op schoot en andere kinderen om hem heen spelend. De laatste 12 jaar van zijn leven was hij blind, maar bleef hij nog volop wiskundeonderzoek doen en daarover publiceren met de hulp van assistenten. Euler was bovenal wiskundige, maar hij was ook uitvinder. Hij is de uitvinder van onder andere de scheepsschroef. En bovendien had Euler een zeer goed politiek en zakelijk inzicht. De Euler-biografie van Fellmann⁵, in het Engels vertaald

¹T. VON KÁRMÁN, *Aerodynamics*, McGraw-Hill (1963).

²H. ROUSE AND S. INCE, *History of Hydraulics*, Dover (1963).

³J.D. ANDERSON, *A History of Aerodynamics*, Cambridge University Press (1997).

⁴L. EULER, Principes généraux du mouvement des fluides, *Mémoires de l' Académie des Sciences de Berlin*, **11**, 274–315 (1755).

⁵E.A. FELLMANN, *Leonhard Euler*, Birkhäuser (2007).

ter gelegenheid van Eulers 300^e geboortedag, laat dit laatste mooi zien. Zo legde Euler als voorwaarde voor zijn terugkeer naar Sint Petersburg – hij had daar eerder gewerkt – een verlanglijstje neer bij Tsarina Catharina II. Behalve een goed salaris wilde hij: gratis huisvesting, gratis brandstof voor in de winter, vrijwaring van inkwartiering van soldaten, geschikte posities voor zijn drie zonen, plus aanstelling van een bekwaam wetenschapper als vice-president van de Academie van Wetenschappen in Sint Petersburg. Euler wilde wetenschappers op leidinggevende wetenschappelijke posities. (Statutair was al geregeld dat het presidentschap door een edelman moest worden vervuld.) Catharina II willigde Eulers eisen in, met uitzondering toch van de laatste eis. Als extraatje leende ze hem echter een van haar koks uit.



Figuur 3.2: Leonhard Euler (1707–1783).

De door Euler voorgestelde stromingsvergelijkingen waren nog niet geschikt voor de meeste praktische problemen; wrijving die in vloeistofstromingen kan optreden was er nog niet in opgenomen. Een uitbreiding hiertoe werd in de 19^e eeuw verzorgd door – onafhankelijk van elkaar – Claude Navier in Frankrijk en George Stokes in Engeland. Deze meer uitgebreide partiële differentiaalvergelijkingen staan thans bekend als de Navier-Stokes-vergelijkingen, vergelijkingen die nog steeds een prach-

tig werkterrein vormen voor wiskundigen. We berekenen inmiddels allerlei stromingen met de Navier-Stokes-vergelijkingen, luchtstromingen om vliegtuigontwerpen bijvoorbeeld. Vliegtuigen waar we dan vervolgens gewoon instappen als ze gebouwd zijn, ofschoon we niet zeker weten of de bij het ontwerpen van die vliegtuigen met de Navier-Stokes-vergelijkingen berekende luchtstromingen wel echt betrouwbaar zijn!

Als u nog een mooie geldprijs en eeuwige roem wilt verdienen: een van de zes nog openstaande millenniumproblemen betreft het leveren van een bewijs voor het bestaan van betrouwbare oplossingen van de Navier-Stokes-vergelijkingen⁶. Stromingsleer is nog steeds een goudmijn voor wiskundigen.

In het vervolg van dit hoofdstuk leest u meer over de Navier-Stokes-vergelijkingen en over het Navier-Stokes-Millenniumprobleem. In de volgende paragraaf leiden we eerst op beknopte wijze de Navier-Stokes-vergelijkingen af.

3.2 Navier-Stokes-vergelijkingen

3.2.1 Vergelijkingen van Euler

Om de beweging van een vloeistof (het kan ook een gas zijn) in een driedimensionaal, vast, rechthoekig x, y, z -assenstelsel te beschrijven beschouwen we een oneindig kleine kubus van die vloeistof met snelheidscomponenten u, v en w in respectievelijk de x -, y - en z -richting. Deze snelheidscomponenten zijn in het algemeen functies van de drie plaatscoördinaten en de tijd. Noteren we de tijd met t , dan kunnen we dus schrijven voor de snelheidscomponenten van onze oneindig kleine vloeistofkubus: $u = u(x, y, z, t)$, $v = v(x, y, z, t)$, $w = w(x, y, z, t)$. Voor de kleine verandering van bijvoorbeeld de snelheidscomponent u van ons vloeistofkubusje, van tijdstip t naar het iets latere tijdstip $t + dt$, kunnen we schrijven:

$$du = \frac{\partial u}{\partial t} dt + \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial y} dy + \frac{\partial u}{\partial z} dz, \quad (3.1a)$$

waarin

$$dx = u dt, \quad dy = v dt, \quad dz = w dt. \quad (3.1b)$$

⁶ C.L. FEFFERMAN, Existence and smoothness of the Navier-Stokes equations, www.claymath.org/millennium/Navier-Stokes_Equations/

In (3.1b) zijn dx , dy en dz de in het tijdstapje dt door het vloeistofkubusje afgelegde afstanden in respectievelijk x -, y - en z -richting. En in (3.1a) zijn de vier termen $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ en $\frac{\partial u}{\partial z}$ de zogenaamde partiële differentiaal. $\frac{\partial u}{\partial t}$ is de verandering van u bij een oneindig kleine verandering van t met vastgehouden x , y en z , en dat gedeeld door die verandering:

$$\frac{\partial u}{\partial t} = \lim_{dt \downarrow 0} \frac{u(x, y, z, t + dt) - u(x, y, z, t)}{dt}, \quad (3.2)$$

$\frac{\partial u}{\partial x}$ de verandering van u bij en gedeeld door een oneindig kleine verandering van alleen x :

$$\frac{\partial u}{\partial x} = \lim_{dx \downarrow 0} \frac{u(x + dx, y, z, t) - u(x, y, z, t)}{dx}. \quad (3.3)$$

Voor $\frac{\partial u}{\partial y}$ en $\frac{\partial u}{\partial z}$ gelden soortgelijke uitdrukkingen. Invulling van (3.1b) in (3.1a) levert

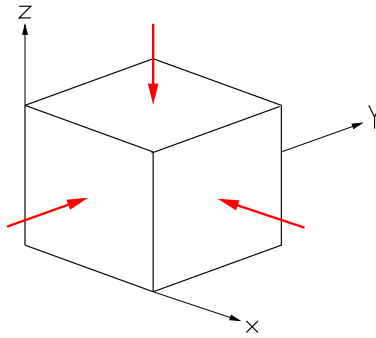
$$\frac{du}{dt} = \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z}. \quad (3.4)$$

Voor $\lim_{dt \downarrow 0}$ is $\frac{du}{dt}$ de versnelling in x -richting van het vloeistofkubusje op tijdstip t . Voor de versnelling in y - en z -richting gelden vergelijkbare uitdrukkingen. Voor het rijtje van drie versnellingscomponenten $(\frac{du}{dt}, \frac{dv}{dt}, \frac{dw}{dt})$, de versnellingsvector van ons vloeistofkubusje, kunnen we dus schrijven:

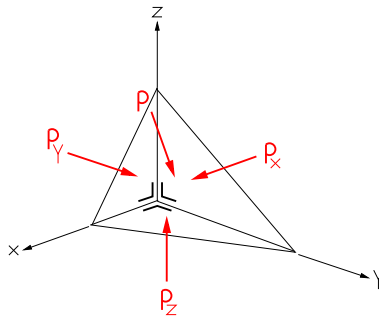
$$\frac{d}{dt} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \left(\frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} \right) \begin{pmatrix} u \\ v \\ w \end{pmatrix}. \quad (3.5)$$

Volgens de Tweede Wet van Newton geldt dat massa $\times$ versnelling van onze oneindig kleine vloeistofkubus de som van de krachten op dat kubusje is. Als notatie voor de dichtheid (massa van de vloeistof per volume-eenheid) voeren we vervolgens ρ in, waarbij, evenals voor de snelheidscomponenten, ook voor de dichtheid geldt: $\rho = \rho(x, y, z, t)$. Vervolgens beschouwen we – in navolging van Euler – als kracht die vanuit de omringende vloeistof op dit vloeistofkubusje kan werken: een oppervlaktekracht *loodrecht* op de zijwandjes van het kubusje, de drukkracht (Figuur 3.3). Verder beschouwen we ook een volumekracht, een kracht die van buitenaf op alle vloeistof *in* het kubusje werkt, bijvoorbeeld de zwaartekracht.

Het bijzondere van de drukkracht is dat deze, in tegenstelling tot de snelheid en de versnelling, in elk punt in alle richtingen hetzelfde is; de druk is isotroop. Wiskundig gezegd: de drukkracht is geen vector maar een scalar.



Figuur 3.3: Oneindig kleine vloeistofkubus met drukkracht op de zijwandjes.



Figuur 3.4: Oneindig klein, rechthoekig vloeistofviervlak met drukkrachten op de zijwandjes. De wet van Pascal zegt dat in elk punt (x, y, z) de druk in alle richtingen even groot is; $p_x = p_y = p_z = p$.

Dit resultaat staat bekend als de Wet van Pascal en kan op zuiver theoretische wijze worden afgeleid, er slechts van uitgaande dat de versnelling van een oneindig klein, rechthoekig vloeistofviervlak (Figuur 3.4) in geen van de drie coördinaatrichtingen oneindig groot mag worden. De volumekracht kan wel een vector zijn (opgebouwd uit bijvoorbeeld de componenten van de zwaartekracht in elk van de drie coördinaatrichtingen).

De (scalaire) drukkracht noteren we met p en de (vectoriële) volumekracht met (f, g, h) , waarbij ook hier druk- en volumekracht functies van de ruimtecoördinaten en de tijd kunnen zijn; $p = p(x, y, z, t)$, enzovoort. De Tweede Wet van Newton kunnen we dan uiteindelijk voor onze oneindig

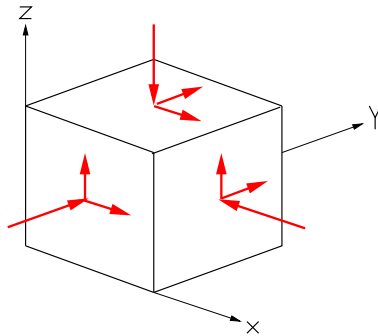
kleine vloeistofkubus schrijven als:

$$\rho \left(\frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} \right) \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \frac{\partial}{\partial x} \begin{pmatrix} -p \\ 0 \\ 0 \end{pmatrix} + \frac{\partial}{\partial y} \begin{pmatrix} 0 \\ -p \\ 0 \end{pmatrix} + \frac{\partial}{\partial z} \begin{pmatrix} 0 \\ 0 \\ -p \end{pmatrix} + \begin{pmatrix} f \\ g \\ h \end{pmatrix}. \quad (3.6)$$

Dit zijn de beroemde Euler-vergelijkingen. De vergelijkingen lieten al snel de juistheid zien van het reeds eerder door Daniel Bernoulli voorgestelde verband tussen druk en snelheid in bepaalde eenvoudige stromingen. Echter, de Euler-vergelijkingen toonden ook de ‘juistheid’ aan van de eveneens eerder gevonden Paradox van d’Alembert, welke zegt dat een lichaam dat in zijn geheel wordt omstroomd door een vloeistof, daarvan geen weerstand ondervindt, hetgeen in strijd is met onze alledaagse ervaring.

3.2.2 Vergelijkingen van Stokes

De Euler-vergelijkingen veronderstellen dat de enige vloeistofkracht die in vloeistofstromingen kan optreden de drukkracht is. Voor praktische toepassingen schieten de Euler-vergelijkingen vaak tekort. In vloeistofstromingen kunnen naast de drukkracht ook schuifkrachten (wrijvingskrachten) optreden; krachten tangentieel aan de zijvlakjes van ons vloeistofkubusje (Figuur 3.5).



Figuur 3.5: Oneindig kleine vloeistofkubus met druk- en schuifkrachten op de zijwandjes.

Het was in 1845 dat in Cambridge de natuur- en wiskundige Stokes (Figuur 3.6), een van de twee helden van dit hoofdstuk, een beter model invoerde dan dat van Euler, een model met schuifkrachten⁷.



Figuur 3.6: Sir George Gabriel Stokes (1819–1903).

George Gabriel Stokes werd op 13 augustus 1819 geboren, in Skreen, een plaatsje in het noorden van Ierland. Hij groeide op in een godsdienstig milieu; zijn vader was predikant, zijn grootvader aan de kant van zijn moeder ook, en drie van zijn broers waren voorganger in de kerk. Stokes koos voor de wetenschap en bleef zijn gehele leven ook belijdend christen. Hij was sterk geïnteresseerd in geloof en wetenschap. Zo was hij president van een vereniging ter bestudering van de relatie tussen het christelijk geloof en wetenschap.

Stokes genoot degelijk onderwijs, als kind thuis van zijn vader en als tiener op het Bristol College in Engeland. Op 18-jarige leeftijd ging hij studeren,

⁷ G.G. STOKES, On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids, *Transactions of the Cambridge Philosophical Society*, **8**, 287 (1845).

aan het Pembroke College aan de Universiteit van Cambridge. Hij moet een uitstekend student zijn geweest; bij zijn afstuderen kreeg hij een beurs aangeboden om als wetenschappelijk medewerker op Pembroke College te blijven. En acht jaar later al werd hij hoogleraar op dezelfde leerstoel waar Newton bijna twee eeuwen daarvoor hoogleraar was geweest, de Lucasian Chair. Stokes was meer theoretisch natuurkundige dan wiskundige. Zijn onderzoeksonderwerp was eigenlijk niet stromingsleer, maar optiek, in het bijzonder de voortplanting van licht door de ether, een denkbeeldige, dunne vloeistof waar de aarde doorheen zou bewegen in het hemelruim. Het doel van Stokes' afleiding van vergelijkingen voor de beschrijving van vloeistofstroming was niet het berekenen van zoiets aards als bijvoorbeeld de stroming van water door een rivier, maar de berekening van de stroming van hemelse ether om de aarde. Als opticus was Stokes zeer succesvol. Zo gaf hij een verklaring voor het verschijnsel van zichtbaar licht uitgestraald door een kleurloze stof als daar onzichtbare ultraviolet-straling op invalt, een verklaring waarin hij het woord fluorescentie invoerde en waarvoor hij in 1852 een prijs ontving van de Royal Society. Hij stelde ook voor om met fluorescentie de moleculaire eigenschappen van stoffen te onderzoeken en legde daarmee mede de basis voor de spectraalanalyse.

Stokes bleef de Lucasian Chair tot aan zijn dood bezetten, tot in 1903, het jaar van de eerste vlucht van de Wright Flyer.

Voor het modelleren van de schuifkrachten maakte Stokes gebruik van stromingsleerwerk van Newton, werk dat in het algemeen veel minder bekend was dan de Tweede Wet van Newton, zij het waarschijnlijk niet bij Stokes als houder van de Lucasian Chair. Stokes postuleerde dat Eulers scalaire druk in feite een zogenaamde krachttensor is:

$$\begin{pmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{pmatrix}, \quad (3.7)$$

en wel een te eenvoudige krachttensor. Stokes breidde Eulers tensor uit tot:

$$\begin{pmatrix} -p + \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & -p + \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & -p + \sigma_{zz} \end{pmatrix}, \quad (3.8)$$

waarin de niet-diagonaaltermen schuifkrachten zijn, en hij drukte alle nieuwe σ -termen uit in partiële differentiaal van u , v en w . Om te beginnen schreef hij voor de nieuwe diagonaaltermen σ_{xx} , σ_{yy} en σ_{zz} :

$$\sigma_{xx} = \lambda \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) + 2\mu \frac{\partial u}{\partial x}, \quad (3.9a)$$

$$\sigma_{yy} = \lambda \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) + 2\mu \frac{\partial v}{\partial y}, \quad (3.9b)$$

$$\sigma_{zz} = \lambda \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) + 2\mu \frac{\partial w}{\partial z}, \quad (3.9c)$$

waarin λ en μ uit metingen bekende eigenschappen zijn van de specifieke beschouwde vloeistof (water, lucht of wat dan ook). Het zijn de viscositeitscoëfficiënten, getallen die de mate van viscositeit (strokerigheid) van de vloeistof kwantificeren. De belangrijkste van de twee is μ , de afschuifviscositeit, de mate van weerstand die een vloeistof biedt tegen afschuiving (uitsmering). Hoe groter μ , hoe strokeriger (minder goed uitsmeerbaar) de vloeistof. Honing heeft bijvoorbeeld een grotere μ dan water. Bij temperatuurverlaging neemt μ in het algemeen toe; honing wordt minder goed uitsmeerbaar als het kouder wordt. De viscositeitscoëfficiënt λ hangt samen met de visceuze weerstand die een vloeistof biedt tegen samendrukking of uitzetting. Het vakgebied waarin vloeistofkrachten worden gemodelleerd, beschreven in termen van vervormingstermen, heet reologie. De door Stokes voorgestelde reologische relaties (3.9a), (3.9b) en (3.9c) lijken misschien ingewikkeld, maar zijn in feite eenvoudig; ze veronderstellen een lineair (en dus eenvoudig) verband tussen krachten en vervormingssnelheidstermen. Voor de niet-diagonaalelementen in (3.8), de schuifkrachten, schreef Stokes nog eenvoudiger lineaire verbanden:

$$\sigma_{xy} = \sigma_{yx} = \mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right), \quad (3.10a)$$

$$\sigma_{yz} = \sigma_{zy} = \mu \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right), \quad (3.10b)$$

$$\sigma_{xz} = \sigma_{zx} = \mu \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right). \quad (3.10c)$$

Het argument dat hij gebruikte voor het symmetrisch laten zijn van de krachtensor ($\sigma_{xy} = \sigma_{yx}$, $\sigma_{yz} = \sigma_{zy}$, $\sigma_{xz} = \sigma_{zx}$) was vergelijkbaar met het argument dat al door Pascal was gebruikt in het isotroop laten zijn van de druk: de hoekversnellingen van ons vloeistofkubusje zijn eindig aldus Stokes. Stokes ging nog verder in zijn vereenvoudiging: hij stelde voor om de zogenaamde bulkviscositeit $\tilde{\mu} \equiv \lambda + \frac{2}{3}\mu$, de visceuze weerstand tegen samendrukking en uitzetting, gelijk aan nul te nemen, dus:

$$\lambda = -\frac{2}{3}\mu. \quad (3.11)$$

Tot zover de bijdrage van Stokes.

3.2.3 Vergelijkingen van Navier en Saint-Venant

Als we aannemen dat $\rho = \rho(x, y, z, t) = \text{constant}$, dat wil zeggen, als we aannemen dat onze vloeistof incompressibel (onsamendrukbaar) is, dan vereenvoudigen de door Stokes afgeleide bewegingsvergelijkingen nog verder, en wel tot:

$$\rho \left(\frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} \right) \begin{pmatrix} u \\ v \\ w \end{pmatrix} = - \left(\frac{\partial}{\partial x} \right) p + \mu \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \begin{pmatrix} u \\ v \\ w \end{pmatrix} + \begin{pmatrix} f \\ g \\ h \end{pmatrix}. \quad (3.12)$$

Stelsel (3.12) is een gekoppeld stelsel van drie partiële differentiaalvergelijkingen, met vier te bepalen oplossingscomponenten: $u = u(x, y, z, t)$, $v = v(x, y, z, t)$, $w = w(x, y, z, t)$, $p = p(x, y, z, t)$. Om het stelsel vergelijkingen sluitend (compleet) te maken wordt de wet van behoud van massa toegevoegd, welke zich voor een incompressibele vloeistofstroming laat schrijven als

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \quad (3.13)$$

Vergelijking (3.13) is al stilzwijgend gebruikt bij het afleiden van (3.12). Stelsel (3.12) met daaraan toegevoegd vergelijking (3.13) vormen tezamen de beroemde, wijd en zijd gebruikte incompressibele Navier-Stokes-vergelijkingen; de vergelijkingen waarvoor een van de zes nog openstaande millenniumproblemen is geformuleerd. Het stelsel vergelijkingen (3.12) – (3.13) is reeds in 1822 afgeleid door de Fransman Navier, de andere held van dit hoofdstuk (Figuur 3.7) ⁸.

Navier was niet uitgegaan van de in Figuur 3.5 geschetste situatie, de situatie met normaal- en tangentialkrachten werkend op de zijvlakjes van de oneindig kleine vloeistofkubus. Naviers uitgangspunt was het aanpassen van de Euler-vergelijkingen voor krachten die tussen vloeistofmoleculen kunnen werken. Navier nam aan dat deze krachten afstotend zijn bij kleine afstanden tussen de moleculen en aantrekkend bij grotere afstanden. Daarnaast nam hij aan dat deze krachten evenredig zijn met de snelheid van moleculen ten opzichte van elkaar. Navier leidde op basis hiervan als stelsel vergelijkingen af: de Euler-vergelijkingen met daaraan toegevoegd

⁸ C.L.M.H. NAVIER, *Mémoire sur les lois du mouvement des fluides*, *Mémoires de l'Académie des Sciences*, **6**, 389–440 (1822).



Figuur 3.7: Claude Louis Marie Henri Navier (1785–1836).

termen voor intermoleculaire krachten; termen die precies hetzelfde waren als die welke later formeel werden afgeleid als zijnde schuifkrachten in incompressibele vloeistofstromingen. Navier kende het begrip schuifkrachten in vloeistofstromingen vermoedelijk niet eens, maar vond deze indirect in feite toch. Naviers afleiding en resultaat waren geen kwestie van geluk. Uit de kinetische gastheorie, de theorie waarin de eigenschappen van gassen worden bestudeerd op basis van het gedrag van hun moleculen, weten we nu dat de viscositeitscoëfficiënt μ te maken heeft met de vrije weglengte tussen de moleculen. Al met al heel knap fundamenteel werk van Navier, die was opgeleid als ingenieur en was gespecialiseerd in het ontwerpen van bruggen.

Claude Louis Marie Henri Navier werd geboren op 10 februari 1785 in Dijon en bracht zijn kinderjaren grotendeels door in Parijs, waar zijn vader juridisch medewerker was in de Nationale Assemblee. Toen de kleine Claude acht jaar oud was overleed zijn vader en werd hij voor zijn opvoeding toevertrouwd aan een oom van zijn moeder, Emiland Gauthey. Oudoom Emiland was een uitstekend ingenieur. Onder zijn invloed ging Claude Navier in 1802 studeren aan de Ecole Polytechnique. In 1806 studeerde hij af, als ingenieur in de weg- en waterbouwkunde. Hij pakte door; hij ging studeren bij de wiskundige Fourier. Fouriers invloed op Navier was

groot en blijvend; Navier werd zijn beschermeling en vriend voor het leven.

Navier werkte zich op tot een van Frankrijks beste ingenieurs. Hij publiceerde ook; hij gaf het werk van zijn oudoom uit en schreef ook zelf. In zijn publicaties legde hij, met zijn bij Fourier verworven analytische inslag, mede de wetenschappelijke basis voor de ingenieurswetenschappen, een basis die tot dan toe tamelijk empirisch was. Het belang van Naviers vernieuwende ingenieursstijl werd erkend. In 1819 werd hij docent aan de Ecole des Ponts et Chaussées. Zijn ster als wetenschapper begon ook te rijzen. In 1826 werd hij lid van de Academie van Wetenschappen in Parijs. Hij ontwikkelde zich steeds meer tot wiskundige, maar bleef ook ingenieur; in 1831 volgde hij de wiskundige Cauchy op als docent aan de prestigieuze Ecole Polytechnique, maar bleef ook nog bruggen ontwerpen.

Navier overleed helaas op betrekkelijk jonge leeftijd, in 1836. Zijn werk aan de beschrijving van vloeistofstromingen was echter niet onopgemerkt gebleven. Op zijn hypothese over de intermoleculaire krachten werd voortgeborduurd door de wiskundigen Cauchy en Poisson. En de ingenieur Adhémar Jean Claude Barré de Saint-Venant (1797-1886) was er ook mee bekend, zoals blijkt uit een boek van hem ⁹. Zeven jaar na Naviers dood leidde Saint-Venant de door Navier gevonden vergelijkingen opnieuw af, nu echter wel uitgaande van schuifkrachten in de vloeistofstroming, en zonder gebruik te maken van Naviers moleculaire aanpak. Hij had het helemaal bij het juiste eind en publiceerde de resultaten van zijn onderzoek ¹⁰, twee jaar voor de publicatie van Stokes! Eigenlijk zouden we van de *Navier-Barré de Saint-Venant-Stokes-vergelijkingen* moeten spreken, maar dat is natuurlijk wel een hele mond vol. De geschiedenis vertelt dat Stokes ook niet van dit werk van Saint-Venant wist. Een geloofwaardig verhaal; wiskundigen en theoretisch natuurkundigen zijn niet altijd goed op de hoogte van fundamenteel wetenschappelijk ingenieurswerk.

3.2.4 Paradox

Met de introductie van de Navier-Stokes-vergelijkingen, nu al bijna twee eeuwen geleden, leek het probleem van het begrijpen en beheersen van een grote klasse stromingen onder handbereik. Het probleem was immers teruggebracht tot het wiskundig oplossen van een handvol partiële differentiaalvergelijkingen.

⁹A.J.C. BARRÉ DE SAINT-VENANT, *Mécanique appliquée de Navier*, Paris (1858).

¹⁰A.J.C. BARRÉ DE SAINT-VENANT, Note à joindre au mémoire sur la dynamique des fluides, *Comptes Rendus des Séances de l'Académie des Sciences*, **17**, 1240 (1843).

Hoewel de Navier-Stokes-vergelijkingen grote vooruitgang brachten, bleek de analytische oplossing van de complete vergelijkingen, de ‘oplossing met potlood en papier’, te hooggegrepen. Het gevolg was het ontstaan van een groot aantal van Navier-Stokes afgeleide, vereenvoudigde vergelijkingen voor speciale gevallen, die *wel* analytisch waren aan te pakken. Bovendien bleef een verschil bestaan tussen experimentele en theoretische stromingsleer. Experimentele stromingsleer ontwikkelde zich sterk tijdens de industriële revolutie. Het was paradoxaal om te zien dat de invoering van de Navier-Stokes-vergelijkingen leidde tot een steeds grotere fragmentatie in verschillende stromingsmodellen, die *allemaal* de stroming beschreven van hetzelfde medium, een theoretisch hoogst ongewenste situatie. Theoretische stromingsleer stagneerde langs een front van niet-lineaire problemen. Het doorbreken daarvan zou uiteindelijk pas in de tweede helft van de 20^e eeuw gebeuren, met numeriek-wiskundige methoden: met getallen, bij benadering, en ten koste van veel, vaak heel veel rekenwerk. Een sleutelrol hierbij zou worden gespeeld door de oorspronkelijk uit Hongarije afkomstige wiskundige John von Neumann. Hoe dit is gegaan illustreren we aan de hand van het probleem van weersvoorspelling.

3.3 Numerieke Stromingsleer

3.3.1 Numerieke weersvoorspelling

Een alledaagse toepassing van de Navier-Stokes-vergelijkingen is weersvoorspelling. Voor de Tweede Wereldoorlog was weersvoorspelling meer een kunst dan een kunde. Elke dag werden weersgegevens van verschillende plaatsen verzameld. Deze gegevens werden op land- en zeekaarten uitgezet, waarna er lijnen van constante druk en temperatuur op deze kaarten werden geschetst. Door vergelijking met oude druk- en temperatuurverdelingen en bijbehorende, bekende weersverlopen werd vervolgens door een weerkundige een weersvoorspelling gedaan. De persoonlijke deskundigheid van deze meteoroloog drukte een zwaar subjectief stempel op de weersvoorspelling. Wat wiskunde betreft werd slechts gebruik gemaakt van elementaire statistiek en wat eenvoudige interpolatie- en extrapolatiemethoden.

In een poging om tot een meer wetenschappelijk gefundeerde weersvoorspelling te komen had de Noor Vilhelm Bjerknes aan het begin van de 20^e eeuw het idee geopperd om het weer te voorspellen op basis van de Navier-

Stokes-vergelijkingen, waar nodig aangevuld met andere vergelijkingen ¹¹. Voor een weersvoorspelling wil je verdelingen weten in ruimte en tijd – de toekomst uiteraard – van onder andere windsnelheid en luchtdruk. Deze verdelingen zitten diep in de Navier-Stokes-vergelijkingen verpakt; ze zijn er niet gemakkelijk uit op te lossen. Exacte oplossing van de Navier-Stokes-vergelijkingen is vandaag de dag nog steeds onmogelijk. Bjerknes kwam niet verder dan zijn idee, wel een belangrijk idee.

Het niet exact, maar bij benadering oplossen van de Navier-Stokes-vergelijkingen is wel te doen. In het geval van weersvoorspelling gaat dat bijvoorbeeld als volgt. In de atmosfeer om de gehele aarde leggen we een grote verzameling netjes verdeelde punten neer (bijvoorbeeld 1000 punten in lengterichting, helemaal rondom de aarde; 500 punten in breedterichting, van pool tot pool; en 100 punten in hoogterichting, vanaf het aardoppervlak tot ergens hoog in de atmosfeer, totaal in dit voorbeeld dus $1000 \times 500 \times 100 = 50$ miljoen punten). De drie-dimensionale verzameling punten noemen we het rooster. Het rooster zit aan de aarde vast. In elk roosterpunt willen we de snelheidsvector en de druk uitrekenen op één of meerdere tijdstippen in de toekomst. We nummeren de roosterpunten in elk van de drie coördinaatrichtingen, bijvoorbeeld met $i = 1, 2, 3, \dots, i_{\max}$, $j = 1, 2, 3, \dots, j_{\max}$, $k = 1, 2, 3, \dots, k_{\max}$. In het getallenvoorbeeld van hiervoor geldt: $i_{\max} = 1000$, $j_{\max} = 500$, $k_{\max} = 100$. Deze discrete (puntsgewijze) benadering van de continue drie-dimensionale ruimte is een voorbeeld van een numeriek-wiskundige benadering. De benadering kan in principe nauwkeuriger worden gemaakt door met kleinere onderlinge afstanden tussen de roosterpunten te werken; door meer roosterpunten te gebruiken. Echter, hoe meer roosterpunten, hoe meer rekenwerk en dus rekentijd. We willen geen ‘weersnaspelling’.

Niet alleen in de ruimte maar ook in de tijd passen we een discrete benadering toe; we vervangen de continue tijd t door de discrete tijd $l\Delta t$, $l = 0, 1, 2, \dots, l_{\max}$, waarbij Δt een te kiezen, kleine stap vooruit in de tijd is, en $l_{\max}$ het nummer van een eindtijdstip tot waaraan toe we willen rekenen. Voor $l = 0$ zitten we op het begintijdstip van *waaraf* we willen gaan rekenen.

De Navier-Stokes-vergelijkingen (3.12) – (3.13) worden nu herschreven op dit discrete ruimte-tijdrooster. We vervangen hiertoe alle partiële differentiaaltermen in de Navier-Stokes-vergelijkingen door zogenaamde eindige differenties. Hier enige voorbeelden (alternatieven zijn mogelijk) van deze eindige

¹¹V. BJERKNES, Das Problem der Wettervorhersage, betrachtet von Standpunkte der Mechanik und der Physik, *Meteorologische Zeitschrift*, **21**, 1–7 (1904).

differenties, waarbij een viervoudig subscript aangeeft in welk punt we ons bevinden in ruimte en tijd, en waarbij we de drie ruimtecoördinaten voor het gemak weer noteren met x, y en z :

$$\left(\frac{\partial u}{\partial t}\right)_{i,j,k,l} \rightarrow \frac{u_{i,j,k,l+1} - u_{i,j,k,l}}{\Delta t}, \quad (3.14)$$

$$\left(\frac{\partial u}{\partial x}\right)_{i,j,k,l} \rightarrow \frac{u_{i+1,j,k,l} - u_{i-1,j,k,l}}{x_{i+1,j,k} - x_{i-1,j,k}}, \quad (3.15a)$$

$$\left(\frac{\partial u}{\partial y}\right)_{i,j,k,l} \rightarrow \frac{u_{i,j+1,k,l} - u_{i,j-1,k,l}}{y_{i,j+1,k} - y_{i,j-1,k}}, \quad (3.15b)$$

$$\left(\frac{\partial u}{\partial z}\right)_{i,j,k,l} \rightarrow \frac{u_{i,j,k+1,l} - u_{i,j,k-1,l}}{z_{i,j,k+1} - z_{i,j,k-1}}, \quad (3.15c)$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j,k,l} \rightarrow \frac{\frac{u_{i+1,j,k,l} - u_{i,j,k,l}}{x_{i+1,j,k} - x_{i,j,k}} - \frac{u_{i,j,k,l} - u_{i-1,j,k,l}}{x_{i,j,k} - x_{i-1,j,k}}}{\frac{1}{2}(x_{i+1,j,k} - x_{i-1,j,k})}, \quad (3.16a)$$

$$\left(\frac{\partial^2 u}{\partial y^2}\right)_{i,j,k,l} \rightarrow \frac{\frac{u_{i,j+1,k,l} - u_{i,j,k,l}}{y_{i,j+1,k} - y_{i,j,k}} - \frac{u_{i,j,k,l} - u_{i,j-1,k,l}}{y_{i,j,k} - y_{i,j-1,k}}}{\frac{1}{2}(y_{i,j+1,k} - y_{i,j-1,k})}, \quad (3.16b)$$

$$\left(\frac{\partial^2 u}{\partial z^2}\right)_{i,j,k,l} \rightarrow \frac{\frac{u_{i,j,k+1,l} - u_{i,j,k,l}}{z_{i,j,k+1} - z_{i,j,k}} - \frac{u_{i,j,k,l} - u_{i,j,k-1,l}}{z_{i,j,k} - z_{i,j,k-1}}}{\frac{1}{2}(z_{i,j,k+1} - z_{i,j,k-1})}, \quad (3.16c)$$

enzovoort. (Tweede-orde differentiaal worden dus herschreven als differenties van differenties.) Op deze manier vervangen we het stelsel van vier partiële differentiaalvergelijkingen door een stelsel van, in ons voorbeeld, 4×50 miljoen eindige-differentievergelijkingen. Dus geen onoplosbare partiële differentiaalvergelijkingen meer, maar relatief gemakkelijke algebraïsche vergelijkingen. De algebraïsche vergelijkingen zijn wel groot in aantal, niet-lineair en bovendien gekoppeld. Het laatste omdat elk roosterpunt is gekoppeld aan buurpunten.

Het rekenwerk kan nu beginnen. Op basis van meetgegevens van het weer over de gehele aarde op een bepaald tijdstip kunnen we een beginoplossing voor onze numerieke weersvoorspelling definiëren: in elke roosterpunt de daar ter plaatse *gemeten* waarden van windsnelheid en luchtdruk. Voor roosterpunten waar meetgegevens ontbreken moet de beginoplossing door interpolatie of extrapolatie vanuit de meetgegevens worden aangeemaakt. Deze beginoplossing kunnen we vervolgens invullen in het stelsel eindige-differentievergelijkingen, en uit dat stelsel kunnen we dan in alle

roosterpunten de oplossing op het nieuwe tijdsniveau $l = 1$ uitrekenen, dus voor alle i, j, k de numerieke oplossing op tijdstip $t = \Delta t$: $u_{i,j,k,l=1}$, $v_{i,j,k,l=1}$, $w_{i,j,k,l=1}$, $p_{i,j,k,l=1}$, $i = 1, 2, 3, \dots, i_{\max}$, $j = 1, 2, 3, \dots, j_{\max}$, $k = 1, 2, 3, \dots, k_{\max}$. Een complicatie is dat in het stelsel (3.12) – (3.13) geen term $\frac{\partial p}{\partial t}$ voorkomt. Met de druk kunnen we dus niet rechthoekig voortaan vooruit stappen in de tijd. Voor wie wil weten hoe dit probleem is opgelost, zij verwezen naar bijvoorbeeld het klassieke numerieke stromingsleerboek van de Nederlandse wiskundige Pieter Wesseling ¹². Een subtiliteit tenslotte is de keuze van de tijdstap Δt . Deze kan gemakkelijk te groot worden genomen, met numerieke instabiliteit tot gevolg.

We hebben hier in grote lijnen al een numeriek algoritme te pakken. We zagen: hoe fijnmaziger het rooster, hoe nauwkeuriger de ruimtelijke weergave van het weer. Hetzelfde geldt voor het verloop van het weer in de tijd; hoe kleiner de stap Δt vooruit in de tijd, hoe nauwkeuriger de voorspellingen, maar ook hier: hoe meer rekenwerk tot we bij de gewenste eindtijd zijn. De hoeveelheid rekenwerk voor betrouwbare weersvoorspelling over een redelijk lange termijn is gigantisch.

De Britse meteoroloog Lewis Fry Richardson was de eerste die aan echte numerieke weersvoorspelling ging doen ¹³. Richardson deed in de jaren twintig van de vorige eeuw een poging om het reeds voorbijgevoerde van 20 mei 1910 over een tijdsduur van zes uur te ‘voorspellen’. Hij koos die dag omdat het een internationale ballonvaartdag was geweest, een dag waarvoor veel weersgegevens voor zijn beginoplossing beschikbaar waren. Richardsons numerieke weersvoorspelling werd een grote teleurstelling, in twee opzichten. De voorspelling was helemaal fout en kostte bovendien veel te veel tijd: zes weken hard rekenen voor zes uur ‘weersvoorspelling’. Alle rekenwerk was nog met de hand gedaan, in teamverband, dat wel. Bjerknæs’ idee van numerieke weersvoorspelling op basis van de Navier-Stokes-vergelijkingen ging weer de mottenballen in, tot de Tweede Wereldoorlog.

Een goede weersvoorspelling is van groot militair-strategisch belang. Toen de Verenigde Staten in de Tweede Wereldoorlog werden betrokken besloot men daar tot opleiding van maar liefst 5000 extra weerambtenaren. En in 1942 pakte de Hongaars-Amerikaanse wiskundige John von Neumann (Figuur 3.8) er de handschoenen van numerieke weersvoorspelling weer op. Voor numerieke berekeningen waarin heel veel keer dezelfde soort sommen worden herhaald is numerieke stabiliteit, het niet buitensporig groeien van

¹²P. WESSELING, *Principles of Computational Fluid Dynamics*, Springer (2000).

¹³L.F. RICHARDSON, *Weather Prediction by Numerical Process*, Cambridge University Press (1922).

kleine foutjes (afrondfoutjes bijvoorbeeld), van cruciaal belang. Von Neumann bedacht een goede methode voor het vooraf analyseren (en zo nodig verbeteren) van de stabiliteit van numerieke methoden. De eerder door Richardson gebruikte numerieke methode voor weersvoorspelling bleek altijd instabiel te zijn volgens Von Neumanns stabiliteitsanalyse, dat wil zeggen voor elke positieve tijdstap Δt . Dat verklaart mogelijk ook waarom Richardsons resultaten niet klopten. Von Neumann ging verder. Hij zag de ontwikkeling van een automatische, elektronische, digitale rekenmachine met interne programma-opslag – de computer zoals we die nu kennen – als de manier om onder de te lange rekentijden uit te komen. Hij stortte zich ook op de ontwikkeling van zo'n machine, in al zijn facetten: van abstracte computerarchitectuur tot concrete invoer- en uitvoerapparatuur, en van algemene richtlijnen voor programmeren tot details over getallenrepresentatie. Von Neumann had veel succes met al dit werk. Hij is de grondlegger van het vak numerieke stromingsleer. Numerieke weersvoorspelling is de manier van weersvoorspellen van vandaag de dag. Een gedetailleerd overzicht van Von Neumanns werk aan numerieke weersvoorspelling, aan numerieke wiskunde en informatica in het algemeen, is te vinden in ¹⁴.



Figuur 3.8: John von Neumann (1903–1957).

¹⁴ W. ASPRAY, *John von Neumann and the Origins of Modern Computing*, MIT Press (1990).

3.3.2 Numerieke wiskunde

Met het werk van Von Neumann kwam de numerieke stromingsleer en daarmee numerieke wiskunde tot grote groei en bloei. Er kwam financiering voor onderwijs en onderzoek in de numerieke wiskunde. Wiskundig talent voelde zich steeds meer aangetrokken tot het vak. Allerlei elegante en krachtige numerieke algoritmen werden en worden nog steeds ontwikkeld. Numerieke wiskunde is een breed vakgebied geworden met daarbinnen allerlei specialismen.

Het weer is slechts een van de vele gevallen waarin we tegenwoordig met behulp van numerieke wiskunde en computers gedetailleerde simulaties kunnen doen. De simulaties kunnen hun oorsprong hebben in allerlei vakgebieden, in de ingenieurswetenschappen, natuurkunde, scheikunde, sterrenkunde, biologie, medische wetenschappen, economie, noem maar op. Veel vakgebieden hebben vandaag de dag hun computersimulatie-versie. We hebben tegenwoordig computational physics, computational chemistry, computational biology, computational finance, enzovoort. Computersimulatie is van groot belang voor de samenleving van vandaag en morgen. Het maakt de voorspelling mogelijk van veel van wat niet met echte experimenten kan worden onderzocht. Experimenten kunnen te gevaarlijk, te duur, onethisch of gewoonweg technisch onmogelijk zijn. Bovendien, in tegenstelling tot echte experimenten, maken computers ook ontwerp en optimalisatie mogelijk: invers rekenen. Gegeven de nog steeds bestaande groei in computersnelheden en -geheugens zijn de verwachtingen over de toekomst van computersimulatie nog steeds hooggestemd.

Numerieke algoritmen speciaal ontwikkeld voor de Navier-Stokes-vergelijkingen zijn misschien wel het verst ontwikkeld van alle numerieke algoritmen. Nog steeds echter – ondanks het zeer wijdverbreide numeriek-wiskundige gebruik van de Navier-Stokes-vergelijkingen – is het niet met zekerheid bekend of er nog wel een oplossing bestaat in het limietgeval waarin roosterpuntafstanden en tijdstapgrootte naar nul gaan (de existentievraag). En indien zij bestaat, is het de vraag of deze oplossing wel altijd en overal voldoende differentieerbaar is.

3.4 Millenniumprobleem

3.4.1 Turbulentie

Oplossing van het Navier-Stokes-Millenniumprobleem is niet alleen vanuit wetenschappelijk maar ook vanuit praktisch oogpunt van groot belang. Binnen de stromingsleer is bijvoorbeeld turbulentie een uitgebreid, reeds lang bestaand onderzoeksterrein. Turbulentie is het (ogenschijnlijk?) wanordelijk door elkaar heen bewegen van wervels in stromingen, wervels van allerlei lengteschalen, van groot tot zeer klein. Leonardo Da Vinci had dat al gezien, waarbij hij ook coherente structuren meende te zien in de wirwar van wervels (Figuur 3.9).



Figuur 3.9: Tekening van turbulente stroming aan de voet van een waterval, Leonardo da Vinci (1508).

Turbulentie treedt op in bijna alle vloeistofstromingen, ten goede of ten kwade. Turbulentie is goed bij bijvoorbeeld het roeren van melk door koffie. U maakt met een lepeltje één korte beweging in uw kopje waardoor er één of twee grote wervels in de koffie ontstaan. De grote wervels vervallen snel in vele, steeds kleinere en steeds chaotischer bewegende wervels. De vele kleine wervels zorgen voor een prima menging van de melk door

uw koffie; turbulentie als nuttig proces. Turbulentie treedt echter ook op in bijvoorbeeld de torusvormige vaten waarin, in een magneetveld en bij extreem hoge temperatuur een plasma van deuterium en tritium wordt opgesloten ten behoeve van de opwekking van energie uit de fusie van hun atoomkernen. Hierbij is turbulentie juist ongewenst; het deel van het plasma waarin kernfusie optreedt mag namelijk niet te veel mengen met koeler plasma; de kernfusie zou daardoor immers kunnen stilvallen.

Turbulentie is een onderzoeksonderwerp met nog veel onbeantwoorde vragen.

3.4.2 Existentie

Het Navier-Stokes-Millenniumprobleem betreft de bewegingsvergelijkingen (3.12) plus massabehoudsvergelijking (3.13). Om stelsels partiële differentiaalvergelijkingen te kunnen oplossen moet de oplossing (of differentialen daarvan) op het begintijdstip en op de ruimtelijke randen van het rekengebied worden voorgeschreven; er moeten begin- en randvoorwaarden worden gespecificeerd. Begin- en randvoorwaarden zijn meestal probleemspecifiek; ze beschrijven bijvoorbeeld de initiële stromingstoestand (de beginvoorwaarde) van de koffie in het koffiekopje van hiervoor, alsmede de geometrie van het kopje en de stromingstoestand van de koffie op de binnenkant van het kopje (de randvoorwaarden).

In de beschrijving van het Navier-Stokes-Millenniumprobleem zijn ook beginvoorwaarden gegeven. Deze zijn echter niet probleemspecifiek, maar zeer algemeen. Aan de beginoplossing wordt ‘slechts’ de voorwaarde opgelegd dat deze oneindig veel differentialen heeft; dat deze oneindig vaak differentieerbaar is. En voor wat betreft de randvoorwaarden: deze zijn er in feite niet; de randen mogen worden beschouwd als aan elkaar vastgeplakt; de rechter rand op de x -as is vastgeplakt aan de linker rand op die as, idem voor de randen op de y - en z -as: zogenaamde periodieke randvoorwaarden. Van de volumekracht tenslotte wordt geëist dat deze oneindig vaak differentieerbaar is op alle tijdstippen.

Oplossingen van het aldus gespecificeerde algemene begin-randwaardenprobleem kunnen dus in principe slechts in zeer algemene termen worden bepaald, aannemende dat er een oplossing bestaat. Dit laatste is nu een vraag in het Navier-Stokes-Millenniumprobleem: heeft het hiervoor beschreven algemene begin-randwaardenprobleem geen, één of meerdere oplossingen, waarbij van deze oplossingen wordt geëist dat zij oneindig vaak differentieerbaar zijn en blijven, en tevens dat de totale hoeveelheid bewegingsener-

gie die zij bevatten begrensd is op elk tijdstip na het begintijdstip. Voor het Navier-Stokes-Millenniumprobleem hoeft dus geen exacte, algemene oplossing van de Navier-Stokes-vergelijkingen te worden bepaald. ‘Slechts’ het *wel of niet bestaan* van zo’n oplossing hoeft te worden bewezen, een oplossing die voldoende differentieerbaar is en begrensd in bewegingsenergie. Voor het aantonen dat er een oplossing bestaat mag de volumekracht zelfs gelijk aan nul worden genomen; een moeilijkheid minder. Wat dit millenniumprobleem toch nog zo moeilijk blijft maken is dat er in de vergelijkingen (3.12) niet-lineaire termen voorkomen; produkten van snelheidscomponenten en partiële differentialen daarvan, produkten als $u \frac{\partial u}{\partial x}$ bijvoorbeeld.

3.4.3 Onderzoek

Er zijn ongetwijfeld meerdere onderzoekers aan de slag, al dan niet bewust, om het Navier-Stokes-Millenniumprobleem te kraken. Bij de bewuste pogingen gebeurt dat vast met de nodige geheimhouding, ter vermindering van het risico om essentiële stappen weg te geven aan concurrenten. In 2006 was er het gerucht dat een Amerikaans wiskundige in een op internet geplaatst artikel het probleem had opgelost. Kort na plaatsing ontdekte de wiskundige echter een fout in het artikel en haalde het snel van internet.

In de vorige eeuw is al veel onderzoek gedaan aan de existentie van Navier-Stokes-oplossingen. Te denken valt aan het werk van Oseen¹⁵, Leray¹⁶, Hopf¹⁷, Scheffer¹⁸, en Caffarelli en collega’s¹⁹. Onderzoek op het snijvlak van klassieke en numerieke analyse is dat van Temam²⁰. Het is precies dit overgangsgebied in de wiskunde waar mijns inziens vooruitgang is te verwachten met betrekking tot het Navier-Stokes-Millenniumprobleem. Binnen de numerieke analyse wordt onderzoek gedaan aan de existentie van numerieke oplossingen, onderzoek dat niet gericht is op het oplossen van

¹⁵ G.W. OSEEN, Sur les formules de Green généralisées qui se présentent dans l’hydrodynamique et sur quelques-unes de leurs applications, *Acta Mathematica*, **34**, 205–284 (1911).

¹⁶ J. LERAY, Sur le mouvement d’un liquide visqueux emplissant l’espace, *Acta Mathematica*, **63**, 193–248 (1934).

¹⁷ E. HOPF, Über die Anfangswertaufgabe für die hydrodynamischen Grundgleichungen, *Mathematische Nachrichten*, **4**, 213–231 (1951).

¹⁸ V. SCHEFFER, Hausdorff measure and the Navier-Stokes equations, *Communications in Mathematical Physics*, **55**, 97–112 (1977).

¹⁹ L. CAFFARELLI, R. KOHN AND L. NIRENBERG, Partial regularity of suitable weak solutions of the Navier-Stokes equations, *Communications on Pure and Applied Mathematics*, **35**, 771–831 (1982).

²⁰ R. TEMAM, *Navier-Stokes Equations. Theory and Numerical Analysis*, American Mathematical Society (1984).

het Navier-Stokes-Millenniumprobleem, maar dat mogelijk wel van belang kan zijn daarvoor.

Behalve het prestigieuze wetenschappelijke belang van het oplossen van het Navier-Stokes-Millenniumprobleem is er dus ook het minstens even grote praktische belang van het vinden van antwoorden op openstaande vragen over turbulentie. Oplossing van het Navier-Stokes-Millenniumprobleem zou wel eens direct, veel voor de praktijk belangrijke turbulentievragen kunnen oplossen. Over turbulentie, als berucht moeilijk onderwerp, heeft iemand ooit gezegd, als variatie op wat Sir Winston Churchill ooit over een bepaalde politieke stroming heeft gezegd: ‘Als je als onderzoeker in de stromingsleer voor je 30^e niet aan turbulentie hebt gewerkt, dan heb je geen hart. Als je na je 30^e echter nog aan turbulentie werkt dan heb je geen verstand.’ Veel onderzoekers zijn gelukkig nog geen 30 jaar oud. En veel senior-onderzoekers *voelen* zich nog geen 30; wiskundeonderzoek houdt je jong, wiskundeonderwijs ook.

4 De status van het P versus NP probleem

Harry Buhrman

Centrum voor Wiskunde en Informatica

Samenvatting

Het P versus NP probleem is het enige “Clay probleem” waarbij zowel een positieve als een negatieve oplossing een miljoen dollar oplevert. Het is een van de fundamentele wiskundige puzzels van onze tijd en het belang ervan groeit met het steeds krachtiger en goedkoper worden van hardware. Snellere processoren, grotere harde schijven alles aan elkaar gekoppeld via het internet. Computers zijn een essentieel deel geworden van ons dagelijks leven, industrie en de wetenschap. Vooral binnen de wetenschap, maar zeker ook daarbuiten, staat keer op keer de vraag centraal of een bepaald rekenprobleem nu opgelost kan worden of dat we moeten wachten tot de volgende generatie snellere computers een oplossing uitrekent. Er bestaat echter een grote en groeiende groep van belangrijke problemen waarbij een snellere computer niet veel soelaas biedt. Deze problemen vergen dermate veel rekentijd dat een computer die 1000 keer sneller is nog steeds jaren moet rekenen. Een snellere computer is dus geen optie. De enige uitweg is het ontwikkelen van een slimmere methode of algoritme waarmee we het probleem ook op onze huidige computers kunnen oplossen. Maar is dit wel altijd mogelijk? Hiermee geraken we aan de kern van het P versus NP probleem.

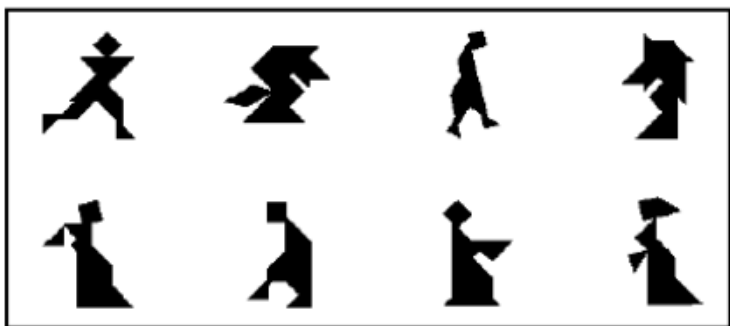
Stel je hebt een landkaart met steden en wegen en afstanden voor de lengte van de wegen en je wilt weten wat de kortste weg van stad A naar stad B is. Je kunt uiteraard alle mogelijke routes van A naar B proberen maar dat loopt bij 30 steden al behoorlijk uit de klauwen. Gelukkig is er een efficiëntere oplossing voor dit probleem, ontwikkeld door Edsger

Dijkstra de enige Nederlander die ooit de Turing award, de belangrijkste internationale informatica prijs, ontving. Je kunt op een slimme manier gericht naar een oplossing zoeken en dat is precies waar Tom Tom gebruik van maakt. Problemen die snel opgelost kunnen worden behoren tot de klasse P (Polynomiale tijd).

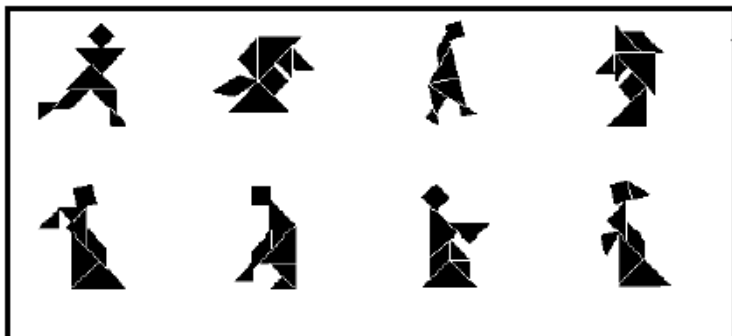
In plaats van de kortste weg had ik ook naar de langste weg van A naar B kunnen vragen, zonder twee keer dezelfde stad aan te doen natuurlijk. Ik had bijvoorbeeld kunnen vragen of er een weg is met lengte minstens 100 km van A naar B. Het vreemde is dat dit probleem veel lastiger is op te lossen dan het eerste probleem. Alle wegen van A naar B langslopen en de langste eruit pikken werkt zeker ook voor dit probleem, maar een wezenlijk betere methode is niet bekend! Als ik echter een weg van A naar B onder ogen krijg dan is gemakkelijk te zien of hij langer dan 100 km is. Goede oplossingen voor het langste pad probleem zijn dus eenvoudig te verifiëren. Zulk soort problemen, waarbij de oplossing snel te controleren valt, behoren tot de klasse NP (Non-deterministisch Polynomiale tijd). De grote vraag is of P gelijk is aan NP. Anders gezegd kunnen we als we een oplossing snel kunnen verifiëren ook snel een oplossing genereren.

Ik zal de huidige stand van zaken schetsen en duidelijk maken wat de verregaande consequenties zijn, voor bijvoorbeeld de cryptografie, quantum computers en de wiskunde, van het positief dan wel negatief oplossen van dit fascinerende probleem.

Een voorbeeld van efficiënt verifiëren (NP) versus efficiënt genereren (P) is Tangram. Het is lastig om de vormen in figuur 1 met de 7 puzzelstukken te maken. In figuur 2 zie je dat gegeven de oplossing het eenvoudig is te verifiëren of die correct is.



Figuur 1: Genereer deze vormen uit de 7 puzzelstukken.



Figuur 2: Controleer of de oplossing correct is.

De dia's

P ? NP

Harry Buhrman

CWI

Centrum Wiskunde &
Informatica

Universiteit van Amsterdam



P versus NP



- One of the seven millennium prize problems
- "In the case of the P versus NP problem and the Navier-Stokes problem, the SAB will consider the award of the Millennium Prize for deciding the question in either direction."
- P not equal NP $\Rightarrow$ 1 million \$
- P equal NP $\Rightarrow$ 6 million \$

Algorithm

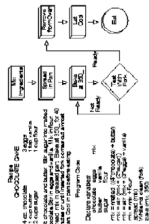
- Algorithm is like a cooking recipe



Main characters of this talk
Algorithms

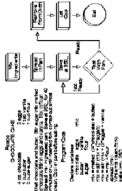
Algorithm

- Algorithm is like a cooking recipe



Algorithm

- algorithm is like a cooking recipe
- input
- computation
- steps (1 time unit)
- output



Slow Algorithm

a=21 b=13

step 1 i=13 13 | 21

step 2 i=12 12 | 21

step 3 i=11 11 | 21

step 4 i=10 10 | 21

step 7 i=7 7 | 13

step 13 i=1

output 1

```

slow-gcd(a, b)
i = min(a,b)
while i % a or i % b
  i := i - 1
output i
  
```

Example Greatest Common Divisor (GCD)

Analysis of Algorithm

Analysis of alg. is preparation time of recipe



Chef

CARL SAGAN'S APPLE PIE

Ingredients:

- 1 universe
- 2 cups all-purpose flour
- 1/2 cup butter
- 1/2 cup brown sugar
- 5 cups sliced apples
- 1/8 cup nutmeg
- 1/4 cup cinnamon
- 1/2 cup raisins

Preparation time: 10 minutes

Servings: 8

Remember:

- Place apples in a large bowl. In a smaller bowl, mix 1/2 cup butter, 1/2 cup brown sugar, and 1/4 cup cinnamon. Sprinkle mixture over apples. Toss until evenly coated.
- Spoon mixture into pie shell.
- Preheat oven to 375 F. Make the universe as usual.
- By now, you should have a large bowl of apples and a smaller bowl of mixture. Add brown sugar and nutmeg to the apples. Mix well.
- Fill the pie shell with the mixture. Cover loosely with aluminum foil.
- Bake for 45 minutes. Let the pie cool for 10 minutes. Cut and serve.

Slow Algorithm

a=21 b=13

step 1 i=13 13 | 21
step 2 i=12 12 | 21
step 3 i=11 11 | 21
step 4 i=10 10 | 21
:
step 7 i=7 7 | 13
:
step 13 i=1

output 1

```
slow-gcd(a, b)
i = min(a, b)
while i { a or i | b
  i := i - 1
}
output i
```

```
if gcd(a, b) != 1 then
  algorithm uses min(a, b) steps
```

Euclidean Algorithm

Greatest Common Divisor (GCD)

step 0 a=21 b=13
step 1 a=13 b= 21 mod 13 = 8
step 2 a=8 b= 13 mod 8 = 5
step 3 a=5 b= 8 mod 5 = 3
step 4 a=3 b= 5 mod 3 = 2
step 5 a=2 b= 3 mod 2 = 1
step 6 a=1 b= 2 mod 1 = 0

output 1

```
function gcd(a, b)
  while b != 0
    t := b
    b := a mod b
    a := t
  output a
```

Better Algorithm

Analysis GCD-Algorithm

- worst case number of steps?

Theorem
alg. terminates in
 $2\log(m) + 1$ steps
 $m = \max(a, b)$

```
function gcd(a, b)
  while b ≠ 0
    t := b
    b := a mod b
    a := t
  output a
```

step 0	a=21	b=13
step 1	a=13	b=21 mod 13 = 8
step 2	a=8	b=13 mod 8 = 5
step 3	a=5	b=8 mod 5 = 3
step 4	a=3	b=5 mod 3 = 2
step 5	a=2	b=3 mod 2 = 1
step 6	a=1	b=2 mod 1 = 0

Proof:
every second step a is
at least halved

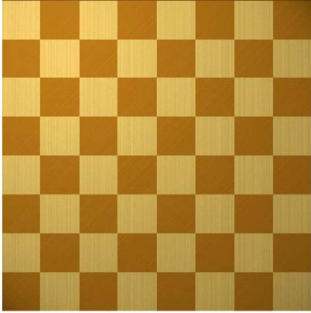
Chess



Complexity

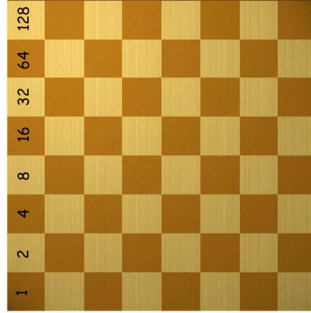
- Euclid: $2\log(m) + 1$ $m = \max(a, b)$
 - Slow: m' $m' = \min(a, b)$
 - length of the input: $\log(a) + \log(b) = n$
- Euclid: $O(n)$ Slow: $2^{O(n)}$
- Euclid exponentially faster than slow!
 - Complexity of computational problem is running time of the best algorithm

A story about exponentials



Reward

- day 1 1 grain of rice
- day 2 2 grains of rice
- day 3 4 grains of rice
- day 4 8 grains of rice
- day 5 16 grains of rice
- day 6 32 grains of rice
- day 7 64 grains of rice
- day 8 128 grains of rice



second week

- day 9 256 grains of rice
- day 10 512 grains of rice
- day 11 1024 grains of rice
- day 12 2048 grains of rice



second week	
• day 9	256 grains of rice
• day 10	512 grains of rice
• day 11	1024 grains of rice
• day 12	2048 grains of rice
• day 13	4096 grains of rice
• day 14	8192 grains of rice
• day 15	16384 grains of rice
• day 16	32768 grains of rice

- | second week | |
|-------------|----------------------|
| • day 9 | 256 grains of rice |
| • day 10 | 512 grains of rice |
| • day 11 | 1024 grains of rice |
| • day 12 | 2048 grains of rice |
| • day 13 | 4096 grains of rice |
| • day 14 | 8192 grains of rice |
| • day 15 | 16384 grains of rice |
| • day 16 | 32768 grains of rice |

day 64

9223372036854775808 grains
92233720368547 kg rice
92233720368 ton rice
92 billion ton rice

world population: 6.5 billion
14 ton rice for everyone
240 years every day rice for everyone

day 64

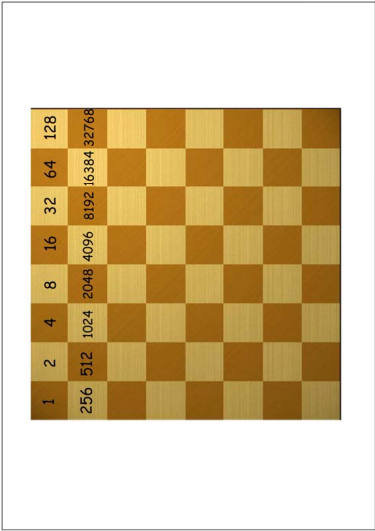
9223372036854775808 grains
92233720368547 kg rice
92233720368 ton rice
92 billion ton rice

world population: 6.5 billion
14 ton rice for everyone
240 years every day rice for everyone

day 64

9223372036854775808 grains
92233720368547 kg rice
92233720368 ton rice
92 billion ton rice

world population: 6.5 billion
14 ton rice for everyone
240 years every day rice for everyone



Complexity

- Euclid: $2 \log(m) + 1$ $m = \max(a, b)$
- Slow: m' $m' = \min(a, b)$
- length of the input: $\log(a) + \log(b) = n$
- Euclid: $O(n)$ Slow: $2^{O(n)}$
- Euclid exponentially faster than slow!
- Complexity of computational problem is running time of the best algorithm

Computation & Complexity

- Computational problem:
 - INPUT $\xrightarrow{\text{computation}}$ OUTPUT
 - Example: a, b output $\gcd(a, b)$
- Complexity:
 - Number of computation steps needed for "best" algorithm
 - function of the input size

Complexity

- Determine the complexity of a computational problem:
 - Upper bound: construct algorithm
 - Lower bound: any algorithm needs this many steps
 - Ideally upper bound = lower bound
- 

Complexity of gcd problem

- Euclid's algorithm runs in $O(n)$ steps
- Can we devise a faster algorithm?
- Not really: any algorithm has to read the whole input: requires n steps
 - Upper Bound: $O(n)$
 - Lower Bound: n
- Complexity of gcd is linear.

Complexity Class P

Feasible Problems: P

- Feasible or efficient algorithms run in **polynomial time**: n^c (some c)
- Complexity Class **P** :
 - All the problems that have feasible algorithms
- Example:
 - Linear Programming
 - Network Flow Problems
 - Shortest Path

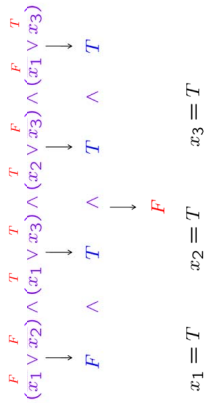
For these problems
upper bound is
linear, lower
bound at best
polynomial for off.

Another problem Satisfiability

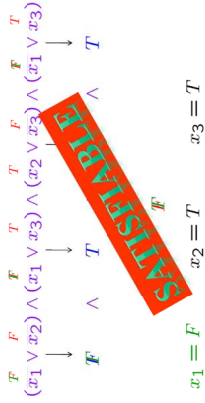
Satisfiability

- variables $x_1 \dots x_n$ $C_l = (x_i \vee x_j \vee \bar{x}_k)$
- Clause $C_1 \dots C_m$ $\phi(x_1 \dots x_n) = C_1 \wedge \dots \wedge C_m$
- formula $\phi(x_1 \dots x_n) = C_1 \wedge \dots \wedge C_m$
- exist $\alpha_1 \dots \alpha_n$ $\alpha_i \in \{T, F\}$
- such that $\phi(x_1 = \alpha_1 \dots x_n = \alpha_n) = T$

Example



Example



Satisfiability

- variables $x_1 \dots x_n$
- Clause $C_1 \dots C_m$ $C_l = (x_i \vee x_j \vee \overline{x_k})$
- formula $\phi(x_1 \dots x_n) = C_1 \wedge \dots \wedge C_m$
- exist $\alpha_1 \dots \alpha_n$ $\alpha_i \in \{T, F\}$
- such that $\phi(x_1 = \alpha_1 \dots x_n = \alpha_n) = T$

SAT = $\{\phi \mid \phi \text{ is satisfiable}\}$

simple algorithm: try all 2^n assignments

Unknown Complexity

- It is hard to determine the complexity of **many** problems
 - Example:
 - Is this formula satisfiable? **SAT**
 - Traveling Salesman Problem. **TSP**
 - Lower Bound: n
 - Upper Bound: 2^n
- Best Known!

Complexity Class NP

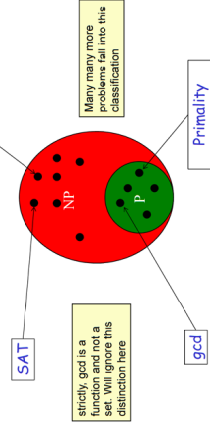
NP

- complexity class NP
 - polynomial time to check solution
 - x in L: exists a polynomial size y :
 - $P(x,y) = \text{True}$
- polynomial time computable
- SAT in NP
- φ is satisfiable
 $\exists \alpha : \varphi(\alpha) = \text{True}$
- in the length of x

P & NP

- complexity class NP
 - easy to check solution
 - polynomial time check
 - easy to check assignment is satisfiable
- complexity class P
 - easy to find solution
 - decide in polynomial time
 - compute in polynomial time $\text{gcd}(a,b)$

$P \subseteq NP$



Reductions & Completeness

reduction

$$A \leq_P^P B$$

compute A in poly-time with B as free subroutine

" A is computationally not harder than B "

"if B in P then A in P "

C is NP-complete

• $C \in NP$

• all $A \in NP$: $A \leq_P^P C$

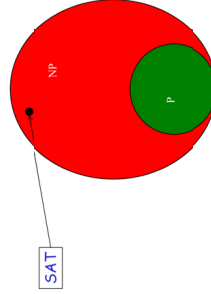
Theorem

• SAT, TSP, many others NP-complete

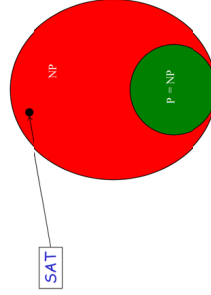
• SAT in $P \Leftrightarrow P=NP$

P versus NP

$P \neq NP$



$P = NP$



P versus NP Question

- $P = NP$?
- widely believed that $P \neq NP$
- how to show this is true?
 - Prove better lower bounds for existing problems like SAT
 - Construct problem in NP with super polynomial lower bound

Lower Bounds

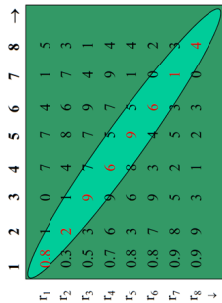
- Construct $D \in NP$
- no poly-time algorithm solves D
 - for every poly time algorithm M exists a string x such that:
 - $M(x) = 1$ & $x \notin D$ or
 - $M(x) = 0$ & $x \in D$ $\Rightarrow D$ not in P
- $D \not\leq_P SAT \Rightarrow SAT$ not in P

Diagonalization

How big are the reals?

- Cantor showed $\mathbb{R}$ not countable
- diagonalization
 - given an enumeration of the reals
 - construct real number d not in the enumeration

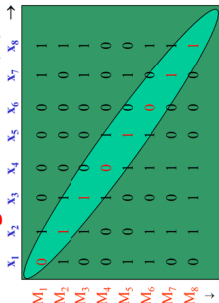
Diagonalization



reals in some enumeration

$d = 0.93070725 \dots$
 i^{th} digit of d is i^{th} entry of diagonal +1

Diagonalization (2)



polynomial time algorithms

$x_1, x_2, x_3, x_4 \in D$ and only if $M_1(x_1) = 0$

Diagonal Language

$$D = \{x_i \mid M_i(x_i) = 0\}$$

i^{th} poly-time algorithm/machine

$D \notin P$, every poly-time machine errs on some input

$D \in NP$??
probably not, but

$D \in \text{time}(n^{\log n})$, quasi polynomial time
with more time can compute more

More Bad News

• Relativization (Oracles):

- Exists oracle A : $P^A = NP^A$
- (Exists oracle B : $P^B \neq NP^B$)



Proof technique should not relativize

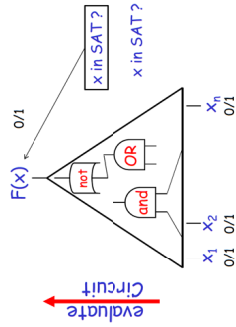
Diagonalization and
most other techniques
we know
relativize

Try something easier

- Study weaker models of computation and develop new lower bound techniques
 - Circuits with small depth
 - Monotone circuits
 - Decision Trees
 - Branching Programs
- The weaker the model the better the lower bounds!

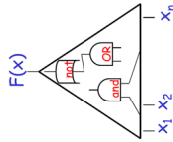
Simple model:
Circuits

Circuit Model of Computation



Size of the Circuit

1. most important:
number of gates
parallel time of computation
2. Depth of the circuit
time of computation



Constant Depth

depth is constant
size is polynomial

AC^0

compute parity:

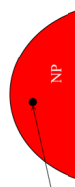
$$F(x) = x_1 + x_2 + \dots + x_n \bmod 2$$

Theorem

parity requires $2^{n^{1/d}}$ size circuits
of depth d

Note: $d = \log n$ bound is meaningless

$NP = AC^1?$



P is poly size
poly depth

Parity

AC^0 poly size
constant depth

AC^1 poly size
 $\log n$ depth

natural proofs another hurdle?

- proof technique that shows parity not in AC^0 likely won't work to separate P from NP
- these proofs fit in a framework called **natural proofs**

Theorem

if one-way functions exist then
natural proofs can't separate P and NP

Approaches

- Structural approach using eg. autoreducibility
- Combinatorial approach
- Algebraic, degrees of multivariate polynomials
- Geometric Complexity
 - algebraic geometry
 - representation theory
- Communication complexity

P vs NP & Cryptography

- computational hardness guarantees security of cryptographic protocols
 - factoring, discrete logarithm
 - lattice problems
 - learning problems
- one-way functions
 - compute $f(x)$ quickly
 - hard to invert
- if $P=NP$ then no cryptography

efficient on
quantum computer



$P=NP$

- $P=NP$, but the proof does not give us an algorithm
- $P=NP$, but algorithm for SAT runs in time $n^{1000000}$
- $P=NP$, but algorithm for SAT runs in time 2^{100n}
- $P=NP$, and algorithm for SAT runs in time n^2

n^2 algorithm for SAT

- Wonderful!!!
 - computing ground states of Hamiltonians
 - protein folding problem solved
 - artificial Intelligence takes really off
 - optimal scheduling
 - computational learning theory
 - weather prediction improves

n^2 algorithm for SAT

- for mathematics
 - can find proofs to theorems, provided they have short proofs
 - can simply ask computer whether theorem/conjecture is true/false
 - mathematics will change dramatically
 - quickly solve the other 5 remaining Clay problems

Summary

- P versus NP central, not just in mathematics and computer science but also in physics, biology, chemistry, cryptography etc.
- not clear how to attack it, several obstacles: relativization, natural proofs, algebraization
- much simpler questions are still way out of reach

5 Henri Poincaré en het ontstaan van de topologie

Ferdinand Verhulst

Universiteit van Utrecht

Voorwoord

Het ontstaan van de moderne topologie is onlosmakelijk verbonden met Henri Poincaré (1854-1912). Hij werd geboren in Nancy, de hoofdstad van Lotharingen, in een familie van ingenieurs, apothekers, medici en juristen. In Nancy bracht hij een gelukkige jeugd door, hij had geen problemen met school (enseignement secondaire en lycée). Wel klaagden zijn leraren soms dat zijn antwoorden niet duidelijk waren, hij vatte zijn oplossingen vaak samen in enkele zinnen omdat naar zijn mening de tussenstappen triviaal waren. Zijn interesses waren breed, pas rond zijn zestiende werd het min of meer duidelijk dat hij zich zou specialiseren in de wiskunde en natuurwetenschappen. Zijn hele leven lang behield hij echter een grote belangstelling voor filosofie, literatuur en andere kunstuitingen.

In 1873 ging Henri naar Parijs om te studeren aan de École Polytechnique, een van de ‘grandes écoles’ van Frankrijk. Na het behalen van het diploma van deze instelling in 1875, vervolgde hij zijn studie aan de École des Mines, eveneens een van de ‘grandes écoles’. Daar voltooide hij zijn studie als mijnningénieur in 1879 terwijl hij tevens in de wiskunde promoveerde aan de universiteit van Parijs (Sorbonne). Dat laatste had hij een beetje stil gehouden, zijn leermeesters aan de École des Mines zouden hebben gedacht dat hij zijn opleiding daar niet serieus genoeg nam.

Poincaré, 25 jaar oud, werkte toen inderdaad een korte tijd als mijnningénieur in Vesoul en Ronchamps, werd echter al snel benoemd tot docent wiskundige analyse aan de universiteit van Caen in Normandië en in 1881 aan de Sorbonne in Parijs. Daar vervulde hij allerlei leeropdrachten, wiskunde, natuurkunde, sterrenkunde, waarschijnlijkheidsrekening, tot zijn



Figuur 5.1: Henri Poincaré, 35 jaar oud.

dood in 1912. Zijn grote talenten en zijn brede belangstelling leverden een groot aantal boeken en artikelen op met fundamentele bijdragen op al deze gebieden en de filosofie. Op het gebied van de grondslagen van de wiskunde kan hij voor een deel worden beschouwd als een voorloper van L.E.J. Brouwer. Hij benadrukte het belang van intuïtie, maar stelde ook dat een wiskundig object bestaat als er geen logische tegenstelling is in de definitie zelf of met bestaande stellingen. Een treffende uitspraak van Poincaré is: “Logica en intuïtie spelen elk een noodzakelijke rol in de wiskunde; de logica is het instrument van de bewijzen, terwijl intuïtie het instrument is van het uitvinden.”

In dit hoofdstuk zullen we schetsen in wat voor problemen Henri Poincaré geïnteresseerd was in de periode 1880-1900 en hoe hij er toe kwam om de topologie te ontwikkelen. Meer details zijn te vinden in de biografie [FV].

5.1 Ideeën van voorlopers

Voorzover we weten was Gottfried Wilhelm Leibniz (1646-1716), zie fig. 5.2 de eerste die de term ‘analysis situs’ gebruikte. Rond 1900 kreeg de topologie die toen werd ontwikkeld in eerste instantie deze naam. Leibniz noemde analysis situs het gebied in de wiskunde waar wordt onderzocht



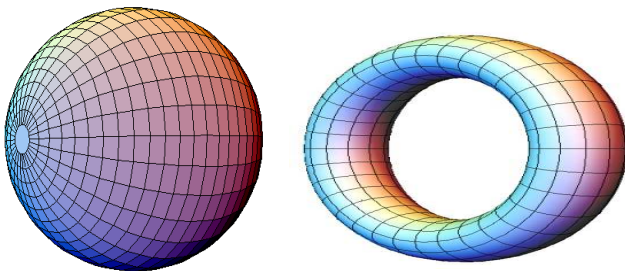
Figuur 5.2: Gottfried Wilhelm Leibniz (1646-1716).

wat meetkundige objecten gemeenschappelijk hebben. Van hem is de uitspraak dat de rechte lijn een kromme is waarvan elk deel lijkt op het geheel.

Het eerste belangrijke resultaat danken we echter aan Leonhard Euler (1707-1783). Dat is de zogenaamde Euler karakteristiek χ (of veelvlak formule van Euler): beschouw een convex maar verder willekeurig veelvlak in $\mathbb{R}^3$ met V het aantal hoekpunten (vertices), E het aantal zijden (edges), F het aantal zijvlakken (faces), dan geldt voor χ dat dit getal een invariant is, namelijk

$$\chi = V - E + F = 2.$$

Na Euler ging men nadenken over eigenschappen van meer algemene gesloten oppervlakken zoals de bol of een torus, zie fig. 5.3. De Italiaan Enrico Betti (1823-1892) streefde er naar om eigenschappen van oppervlakken te karakteriseren in drie en meer dimensies. Zo kan een verschil tussen bol en torus worden aangegeven met het Betti getal P_1 . Elke gesloten kromme op een bol separeert de bol in twee delen, dat wil zeggen dat je niet van een punt in het ene deel over de bol naar een punt in het andere deel kunt gaan zonder die gesloten kromme te snijden. Bij een torus kun je een gesloten kromme door het gat laten lopen, of juist om het gat, die de torus niet separeert. Voeg je zo een tweede kromme toe, dan separeren beide krommen wel. Het maximaal aantal gesloten krommen dat een oppervlak niet separeert is het Betti getal P_1 , bij de bol is $P_1 = 0$. Bij de torus geldt



Figuur 5.3: Bol en torus; een gesloten kromme op de bol separeert deze in twee gebieden, bij de torus levert een gesloten kromme, bijvoorbeeld door het gat, deze separatie niet.

$P_1 = 1$; zie fig. 5.3.

Betti dacht na over de begrippen samenhang en rand van een oppervlak, hij generaliseerde deze begrippen en de Betti getallen voor meer dimensies. Wat de theorie van oppervlakken betreft was dat het zo ongeveer toen Poincaré in 1892 zijn eerste artikel over topologie publiceerde.

5.2 Poincaré en de meetkunde

Hoe kwam Poincaré er toe om na te denken over analysis situs (topologie), over “zaken die meetkundige objecten gemeenschappelijk hebben”?

Het eerste artikel van Poincaré in 1874, hij was nog student aan de École Polytechnique, ging over een klassiek meetkundig probleem. Meetkunde was een belangrijk onderwerp in zijn opleiding, maar belangrijker nog was dat zijn supervisors, onder anderen Gaston Darboux (1842-1917), benadrukten dat meetkunde en analyse elkaar aanvullen bij het verkrijgen van wiskundig inzicht. Wiskunde was voor hen geen vakgebied dat opgesplitst is in deelgebieden die omheind zijn door hoge muren. Voor sommige problemen kom je een heel eind met analyse, maar meetkunde kan daar een aanvullende, verhelderende kijk op leveren. Ook de algebra geeft vaak nieuwe inzichten.

We bekijken een aantal problemen waarmee Poincaré zich bezig hield, maar die niet direct de topologie betreffen, in de periode rond 1880 tot 1900.



Figuur 5.4: Gaston Darboux (1842-1917), een van de supervisors bij het proefschrift van Henri Poincaré.

Zijn behandeling van deze problemen wordt al gekenmerkt door aanzetten van topologische concepten.

5.2.1 Tweede orde differentiaalvergelijkingen

Poincaré's proefschrift (1879) had als onderwerp eerste orde partiële differentiaalvergelijkingen. Dat onderzoek leidde tot gewone differentiaalvergelijkingen met singuliere punten, waarden van de onafhankelijke variabele waar de coëfficiënten onbegrensd worden. Bijvoorbeeld $x = 0$ van de vergelijking:

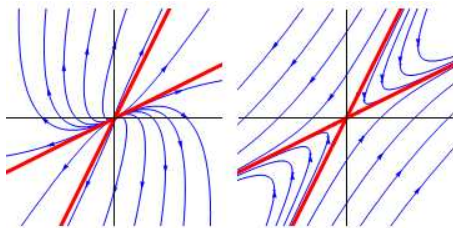
$$\frac{d^2y}{dx^2} + \frac{1}{x} \frac{dy}{dx} + y = 0.$$

In Poincaré's werk over automorfe functies komt dit thema opnieuw terug.

Zijn Mémoire over differentiaalvergelijkingen die in 1881-1882 verscheen is van blijvend belang gebleken. In de Mémoire voert hij de nu algemeen aanvaarde meetkundige klassifikatie in van evenwichtspunten van tweede orde differentiaalvergelijkingen (zadel-, spiraal-, knoop- en centripunt). Beschouw eens het tweede orde stelsel differentiaalvergelijkingen met Euclidische variabelen x, y :

$$\dot{x} = f(x, y), \quad \dot{y} = g(x, y). \quad (5.1)$$

Zoals gebruikelijk staat $\dot{x}$ voor de afgeleide naar de tijd t . De rechterleden zijn voldoende glad (differentieerbaar). Evenwichtso oplossingen (x_0, y_0)



Figuur 5.5: Evenwichten voor vgl. (5.2); links een knoop in $(0,0)$, rechts een zadel.

voldoen aan:

$$f(x_0, y_0) = g(x_0, y_0) = 0.$$

Voer nu de translatie $x - x_0 = u$, $y - y_0 = v$ uit, lineariseer in een omgeving van $u = v = 0$ en noem de variabelen weer x en y . We krijgen het stelsel met constante coëfficiënten:

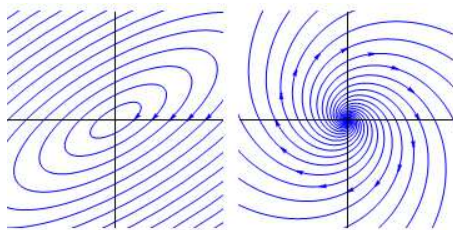
$$\begin{cases} \dot{x} &= ax + by, \\ \dot{y} &= cx + dy. \end{cases} \quad (5.2)$$

Hoe zien de oplossingen van vgl. (5.2) er uit in een omgeving van $(0,0)$? Poincaré classificeerde de oplossingen en het evenwichtspunt $(0,0)$ als volgt. Beschouw de niet-singuliere matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Bereken de eigenwaarden $\lambda_{1,2}$ van de matrix, er zijn dan vier standaard mogelijkheden:

1. λ_1 en λ_2 zijn reëel en met gelijk teken. Het evenwicht $(0,0)$ wordt een ‘knoop’ genoemd; zie fig. 5.5.
2. λ_1 en λ_2 zijn reëel en hebben verschillend teken. Het evenwicht $(0,0)$ wordt een ‘zadel’ genoemd; zie fig. 5.5.
3. λ_1 en λ_2 zijn complex (imaginair deel ongelijk nul) met reëel deel ongelijk nul. Het evenwicht $(0,0)$ wordt een ‘spiraalpunt’ genoemd; zie fig. 5.6.
4. λ_1 en λ_2 zijn complex (imaginair deel ongelijk nul) met reëel deel nul. Het evenwicht $(0,0)$ wordt een ‘centrumpunt’ genoemd; zie fig. 5.6.



Figuur 5.6: Evenwichten voor vgl. (5.2); links een centrumpunt in $(0,0)$, rechts een spiraal.

Als de matrix singulier is, zijn een of twee eigenwaarden nul en hebben we een gedegeneerd evenwichtspunt. Opmerkelijk is dat de dynamica in een omgeving van een evenwichtspunt hiermee behalve analytisch ook meetkundig is gekarakteriseerd. Later bewees Poincaré dat het gedrag van de oplossingen in de buurt van een evenwichtspunt in het geval van knoop, zadel of spiraal, bijna gelijk is (een beetje vervormd) voor de oorspronkelijke, nietlineaire vergelijkingen (5.1).

In de *Mémoire* leidt Poincaré nog meer belangrijke stellingen af, speciaal de indexstelling. Deze zegt het volgende. Beschouw het fasevlak van de twee-dimensionale vergelijking (5.1). Construeer eens een enkelvoudige, maar verder willekeurige gesloten kromme K in het fasevlak. We doorlopen deze gesloten kromme K één keer in positieve richting; het quotiënt $f(x,y)/g(x,y)$ van de rechterleden van vgl. (5.1) zal dan h keer van $-\infty$ naar $+\infty$ springen, k keer van $+\infty$ naar $-\infty$. We noemen I met

$$I = \frac{h - k}{2}$$

de index van de gesloten kromme. Bij het doorlopen van de kromme K zal de vector $(f(x,y), g(x,y))$ over een hoek $2\pi I$ gedraaid zijn. De evenwichtspunten van de vergelijking corresponderen met de nulpunten van $f(x,y), g(x,y)$. Voor een gesloten kromme K waarop geen evenwichtspunt ligt concluderen we:

- Als in het inwendige van K geen evenwichtspunt ligt is zijn index $I = 0$.
- Als er precies één evenwichtspunt in het inwendige van K ligt geldt $I = -1$ als het een zadel is, $I = +1$ als het een knoop, centrum- of spiraalpunt is.

- Als het inwendige van K , N knopen bevat, F spiraalpunten en C zadels, dan is de index van K : $I = N + F - C$.
- De index van een gesloten oplossingskromme (dus niet willekeurig, maar corresponderend met een periodieke oplossing) is $I = +1$.

Uit het laatste resultaat volgt dat binnen een gesloten oplossingskromme altijd minstens één evenwichtspunt moet liggen, maar nooit alleen een zadel. Er geldt verder een relatie tussen het totaal aantal knopen, zadels en spiralen in het fasevlak, maar die relatie is afhankelijk van de aard van het fasevlak. In het boven geschetste geval is het fasevlak Euclidisch, maar het kan ook een bol of torus zijn, afhankelijk van de aard van het probleem. Dit zette Poincaré aan het denken over overeenkomsten en verschillen tussen allerlei soorten oppervlakken.

5.2.2 Automorfe functies

Behalve de theorie van dynamische systemen was er nog een andere aanloop voor topologische gedachten en ideeën bij Poincaré. In 1880, toen hij docent in Caen was, las hij een artikel van de Heidelbergse wiskundige Lazarus Immanuel Fuchs (1833-1902). Deze onderzocht complexwaardige oplossingen in de complexe variable z van de lineaire differentiaalvergelijking:

$$\frac{d^2y}{dz^2} + P(z)\frac{dy}{dz} + Q(z)y = 0.$$

De functies $P(z)$ en $Q(z)$ hebben polen in geïsoleerde punten van het complexe vlak. We zagen eerder zo'n voorbeeld voor reële variabelen met $P = 1/x$, $Q = 1$. Een van de vragen in die tijd was, wat er met de onafhankelijke oplossingen $f(z)$ en $g(z)$ gebeurt als je in een gesloten kromme om een singulariteit heendraait. Bij Fuchs en Poincaré spelen ook de mogelijke invariantie eigenschappen voor transformaties van het quotiënt $f(z)/g(z)$ een belangrijke rol; die details doen er nu niet toe, ze behoren tot het terrein van de automorfe functies. Sinds Riemann (1826-1866) dacht men over singulariteiten (polen) van complexe functies na. Als je bijvoorbeeld de functie $\sqrt{z}$ neemt, in $z = 1$ begint en om $z = 0$ heendraait, krijg je bij terugkeer in $z = 1$ de waarde -1 . Riemann had als oplossing het invoeren van bladen of oppervlakken waarop de functie na zo'n ronddraai gedefinieerd is. Zo vermijdt je de meerduidigheid, maar het is niet zo duidelijk hoe je je dat meetkundig moet voorstellen. Poincaré generaliseerde het werk van Fuchs door complexe functies te bekijken die invariant zijn onder een groep van analytische transformaties. Analytisch en algebraïsch is dat wel

duidelijk, het leidt tot een belangrijke uitbreiding van de speciale functies zoals de elliptische met de zogenaamde hypergeometrische, de Fuchse en zeta-Fuchse functies. De 26-jarige Poincaré liep vooraan in het ontwikkelen van deze resultaten, maar hij voelde zich hier nog steeds ongemakkelijk bij: hoe kon je de structuur van de oplossingen, de speciale functies die door de differentiaalvergelijkingen zijn gedefinieerd, in een omgeving van een singulier punt meetkundig karakteriseren? Poincaré was naast wiskundige ook mijnningenieur en geïnteresseerd in geologie. In dezelfde tijd, in 1880, ging hij met een aantal geologen enkele dagen op expeditie in Normandië, een excursie waarbij hij zich natuurlijk niet met wiskunde kon bezighouden. In Coutances maakte de groep gebruik van een paarden-omnibus; toen Henri Poincaré op de treeplank stapte ervoer hij het volgende:

Juist toen ik op de treeplank stapte kreeg ik het idee, hoewel geen enkele gedachte me daar toen op had voorbereid, dat de transformaties die ik had gebruikt om de Fuchse functies te definiëren, precies die waren die je in de niet-Euclidische meetkunde vindt. Ik kon het niet verifiëren, daar had ik de tijd niet voor, ik ging door met het gesprek dat aan de gang was, maar ik voelde dadelijk absolute zekerheid. Toen ik terug kwam in Caen heb ik het op mijn gemak geverifieerd.

Dit leidde hem geheel onverwacht tot een verband tussen het gedrag van complexe functies in een omgeving van een singulariteit en de hyperbolische meetkunde. In de niet-Euclidische meetkunde van Lobachevsky zijn er transformaties die de afstand tussen objecten constant houden (isometrieën), dezelfde transformaties spelen een rol in de analyse van de Fuchse functies. Het werd de eerste toepassing van de niet-Euclidische meetkunde op een ander gebied van de wiskunde, het leidde bij hem tevens tot een overvloed van andere resultaten. Het is opnieuw een illustratie van de samenhang van analytische, algebraïsche en meetkundige inzichten.

5.2.3 Het Newtonse drielichamenprobleem

De dynamica in het Newtonse drielichamenprobleem wordt beschreven met differentiaalvergelijkingen. De oplossingen van deze vergelijkingen zijn te vinden op een Hamiltoniaans energie-oppervlak in de faseruimte dat singulariteiten vertoont; die singulariteiten corresponderen met botsingen en gedrag op oneindig. De faseruimte van het algemene drielichamenprobleem heeft 18 dimensies, de beweging op het energie-oppervlak is dus nog altijd 17-dimensionaal. Dit reusachtig aantal dimensies wordt verminderd

door bijvoorbeeld aan te nemen dat de bewegingen van alle drie de lichamen in een vlak plaatsvinden, dan is de faseruimte 12-dimensionaal. Elk lichaam heeft in elke dimensie een positie en snelheid (afgeleide in de tijd), dat zijn vier variabelen per lichaam. Daarnaast wordt vaak aangenomen dat een van de lichamen zo weinig massa heeft dat deze de beweging van de andere twee lichamen niet beïnvloedt; denk aan de configuratie Zon-Planeet-Kunstmaan of Zon-Planeet-Asteroïde. Dat is het beperkte drielichamenprobleem.

We noemden 17 dimensies, maar dat is iets te veel. Behalve de energie, zijn er nog meer zulke oppervlakken die de beweging beperken, bijvoorbeeld de drie oppervlakken die corresponderen met de wet van behoud van hoekmoment (zulke behoudswetten worden ook wel ‘integralen van beweging’ genoemd). De oplossingen van het drielichamenprobleem zijn te vinden op de doorsnijdingen van deze oppervlakken met het energie-oppervlak. Om de bewegingen van de deeltjes beter te begrijpen, moet je je dus realiseren dat dit energie-oppervlak weer verder bladert in deeloppervlakken; een belangrijke vraag is dan of het energie-oppervlak helemaal kan worden opgesplitst in zulke oppervlakken en of je die complexe dynamica zo volledig kunt beschrijven. Voor het tweelichamenprobleem kan dit inderdaad, maar Poincaré heeft laten zien dat dit voor het drielichamenprobleem niet het geval is. Daar verstoort chaos de mooie meetkundige structuur.

Over het drielichamenprobleem dacht Poincaré al na in de periode 1885-1888 toen hij werkte aan de beroemde prijsvraag van de Zweedse koning Oscar II. Hij toont daar met berekeningen aan, dat in de omgeving van een instabiele periodieke oplossing er oneindig veel oplossingen zijn die niet op zo’n integraaloppervlak te vangen zijn. In de drie delen van zijn “Méthodes Nouvelles de la Mécanique Céleste” (1892-1899) ontwikkelde hij daarna systematisch de theorie van dynamische systemen waarin steeds meetkundige en analytische methoden elkaar afwisselen en aanvullen. In deel drie beschrijft Poincaré in detail hoe deze complexe dynamica, die we nu chaotisch noemen, tot stand komt. Die beschrijving van oplossingen in het drielichamenprobleem is heel beeldend, het is voor moderne onderzoekers niet te begrijpen hoe hij zonder computer tot dit inzicht kwam.

Meestal beoefende men de meetkunde in het vlak of in de drie-dimensionale ruimte. Het onderzoek aan het drielichamenprobleem maakte het noodzakelijk om na te denken over de structuur van oppervlakken in meerdimensionale ruimtes.

5.3 Poincaré en de Analysis Situs

Tegelijk met zijn werk aan dynamische systemen en automorfe functie, later parallel aan zijn werk in de mathematische fysica en de partiële differentiaalvergelijkingen, publiceert Poincaré in de jaren 1892-1905 zijn artikelen over analysis situs. Het eerste, grote artikel in 1895 beslaat 121 bladzijden in de *Journal de l'École Polytechnique*. Dit artikel en de vijf supplementen zijn in het Engels vertaald en van een inleiding voorzien door John Stillwell [St].

De artikelen zijn in discussiestijl geschreven, waarbij het onderwerp met behulp van allerlei overwegingen wordt ontwikkeld en dan in de latere supplementen wordt verhelderd en gecorrigeerd. Poincaré is in gesprek met de lezer. Dat is bepaald niet de klassieke manier van wiskunde publiceren, maar men moet bedenken dat de topologie hier voor het eerst bedacht wordt. Typerend is de gang van zaken over het Poincaré vermoeden. Losjes geformuleerd zegt dit dat elk enkelvoudig samenhangend, gesloten drie-dimensionaal oppervlak in $\mathbb{R}^4$ dat zonder gaten is, door continue deformatie (homeomorf) kan worden overgevoerd in een bol. Zijn formulering is enigszins anders, namelijk met behulp van de zogenaamde fundamenteaalgroep en later met het begrip homotopie. Poincaré meent eerst dat dit resultaat vanzelf spreekt, pas in het vijfde supplement, in 1905, formuleert hij het als een probleem. In zijn eigen woorden:

Considérons maintenant une variété V a trois dimensions ...
Est-il possible que le groupe fondamental de V se réduise à la substitution identique, et que pourtant V ne soit pas simplement connexe?

Voor een twee-dimensionaal oppervlak in $\mathbb{R}^3$ werd een bewijs snel gevonden, we kunnen dit alleen deformeren naar een bol als het enkelvoudig samenhangend is. Voor $n \geq 5$ werd het in 1961 bewezen door Stephen Smale, voor $n = 4$ in 1982 door Michael Freedman. Het oorspronkelijke Poincaré vermoeden voor $n = 3$ werd in de jaren 2002-2003 bewezen door Grigori Perelman.

Het blijft interessant en wonderlijk dat de problematiek en de bewijzen voor $n = 3$, $n = 4$ en $n \geq 5$ zo verschillend zijn.

Bibliografie

- [St] Henri Poincaré, Papers on topology, analysis situs and its five supplements, translated and introduced by John Stillwell, AMS and London Math. Soc., History of mathematics vol. 37, 2010.
- [FV] Ferdinand Verhulst, Henri Poincaré, impatient genius, Springer, 2012.

6 Het vormen van ruimte: van Poincaré tot Perelman

Roland van der Veen

Universiteit van Amsterdam

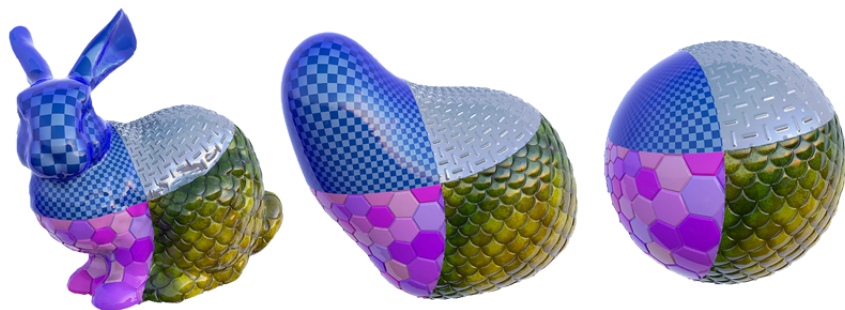
Voorwoord

In deze bijdrage geven we een kort overzicht van recente ontwikkelingen in het vakgebied van de (laagdimensionale) topologie. In 2003 bewees G. Perelman het Poincaré-vermoeden, een doorbraak zowel binnen als buiten de topologie. Aan de hand van oppervlakken geven we een indruk van de belangrijkste ideeën, methoden en toepassingen.

6.1 Inleiding

Aan het einde van de negentiende eeuw ontwikkelde de Franse wiskundige Henri Poincaré een nieuwe tak van de wiskunde: de *topologie*. Dat kwam zo: In zijn bestudering van de bewegingen van de planeten en andere differentiaalvergelijkingen viel het Poincaré op dat de precieze vorm van de vergelijkingen vaak niet zo heel belangrijk is. Kwalitatieve vragen als Loopt het systeem uit de hand of niet? worden vaak al bepaald door de globale vorm van de oplossingen. Hetzelfde fenomeen kwam hij tegen bij het uitrekenen van complexe integralen. De uitkomst van de integraal verandert vaak niet wanneer we het integratiegebied een klein beetje veranderen. Het gaat blijkbaar om de globale vorm van het gebied, niet om de details.

Poincaré zocht naar een nieuwe wiskundige taal om de globale eigenschappen van zijn ruimten mee te beschrijven. Wat blijft er over van meetkunde als je *irrelevante details* zoals hoeken, afstanden en rechte lijnen allemaal



Figuur 6.1: Deze drie objecten zijn topologisch gezien hetzelfde. De vrij ingewikkelde vorm van het konijn links wordt vereenvoudigd tot een perfect ronde bol rechts.

overboord zet? Poincaré ontdekte dat je dan nog steeds een zinnige theorie krijgt en noemde het topologie. Het bekendste deel van de topologie is de grafentheorie van netwerken. In een netwerk gaat het alleen om welke punten verbonden zijn met welke, niet hoe het er precies uitziet. Topologie wordt ook wel rubbermeetkunde genoemd, een object verandert immers topologisch gezien niet wanneer we het uitrekken en vervormen alsof het van rubber was, zie figuur 6.1. Een beroemd voorbeeld is dat een topoloog het verschil niet kan zien tussen zijn koffiekop en de donut die er naast ligt.

De centrale vragen in de topologie zijn: Hoeveel verschillende ruimten zijn er mogelijk? Passen ze in een classificatieschema? Is er een soort periodiek systeem van alle mogelijke ruimten? Poincaré probeerde rond 1900 de eenvoudigste ruimten in kaart te brengen. Hij vermoedde dat dit juist de cirkel, de bol en de hogerdimensionale bollen waren. Deze uitspraak is in de honderd jaar daarna bekend geworden als het *Poincaré-vermoeden*. Een van de beruchtste en belangrijkste problemen in de wiskunde.

Pas in 2003 slaagde de Rus Grigori Perelman er uiteindelijk in om Poincarés vermoeden op een spectaculaire manier te bevestigen. Perelman geeft een volstrekt nieuwe visie op de topologie. Het geeft een effectieve en intuïtieve manier om willekeurige ruimten, hoe ingewikkeld ook, in kaart te brengen. De eenvoudigste ruimten waar Poincaré over nadacht zijn slechts een kleine toepassing. Het zijn inderdaad de bollen zoals Poincaré vermoedde.

De kern van Perelmans visie is dit: **Breng je ruimte in een zo mooi en symmetrisch mogelijke standaardvorm.** Een eenvoudige illustratie hiervan is figuur 6.1. Het vrij ingewikkelde konijnenbeeld¹ links wordt teruggebracht tot een perfect symmetrische bol rechts. Topologisch gezien gebeurt hier helemaal niets maar in de praktijk is de bol veel makkelijker te begrijpen. Het briljante van Perelman is dat hij het gebrek aan vorm in de topologie juist aangrijpt om zo mooi mogelijke vormen te creëren. In deze optimale vorm worden de eigenschappen van de ruimte veel makkelijker zichtbaar. In plaats van abstracte ruimten the classificeren kunnen we dus optimaal symmetrische standaardruimten bestuderen. Zo heeft Perelman de centrale vraag in de topologie teruggebracht tot een eenvoudiger meetkundig probleem. De wiskunde is namelijk veel beter in staat om met mooie symmetrische objecten om te gaan dan met vormeloze ruimten.

Hieronder geven we een korte en niet-technische inleiding in de topologie met nadruk op de visie van Perelman. We illustreren de belangrijkste ideeën aan de hand van heel speciale ruimten: *oppervlakken*. We zullen zien hoe we oppervlakken in kaart kunnen brengen en hoe Perelman deze ruimten in hun optimale vorm brengt. De begrippen *Eulergetal* en *kromming* zullen een hoofdrol spelen. Differentiaalvergelijkingen komen ook aan bod en tot slot heeft de techniek van Perelman onverwachte toepassingen.

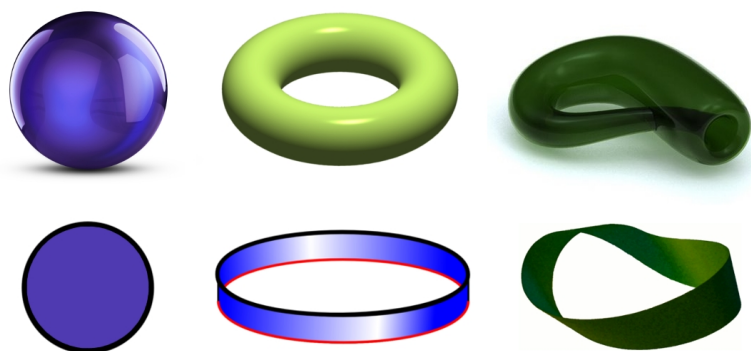
6.2 Topologie en oppervlakken

Topologie gaat over het beschrijven en classificeren van abstracte ruimten. Om het nog abstracter te maken mogen we bovendien niet naar de details van de ruimte kijken, het gaat om de grote lijn, de globale vorm. Maar wat is een ruimte eigenlijk? En wat bedoelen we met de globale vorm?

In plaats van deze vragen in hun algemeenheid te beantwoorden zullen we ons beperken tot een speciaal soort ruimten: *oppervlakken*. In figuur 6.2 hieronder hebben we alvast wat voorbeelden van oppervlakken getekend. Hieronder zullen we precies zeggen wat we bedoelen met een oppervlak en wanneer twee oppervlakken topologisch gezien hetzelfde zijn.

Oppervlakken vormen een ideale proeftuin voor de topologie omdat er veel over bekend is en ze vrij goed te visualiseren zijn. We zullen zien hoe we ze kunnen classificeren en van elkaar kunnen onderscheiden. Zo kunnen we in het volgende hoofdstuk de ideeën van Perelman heel mooi in actie

¹dit konijn is de beroemde *Stanford bunny*. Een van de eerste ingescande 3D-modellen. Het is nog steeds een veel gebruikt voorbeeld in de computergraphics-wereld.



Figuur 6.2: Voorbeelden van oppervlakken. Eerste rij (zonder rand): De bolschil, de torus en de fles van Klein. Tweede rij (met rand): De schijf, de ring en de Möbiusband.

zien. Aan het einde van het hoofdstuk gaan we kort in op driedimensionale ruimten waar Poincaré over nadacht.

6.2.1 Definitie van een oppervlak

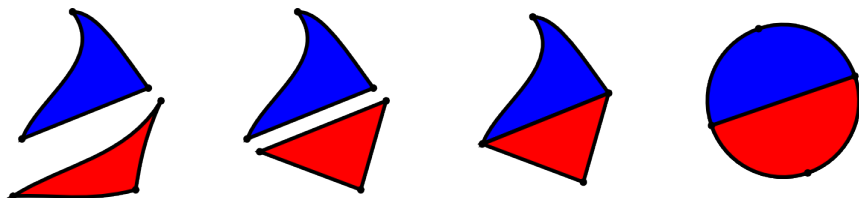
Wat is een oppervlak? Als we nog eens naar figuur 6.2 kijken, dan zijn het een soort flexibele dunne zeepvliezen met gaten erin. Ook kunnen ze een rand hebben. Omdat dit idee concreter te maken brengen we alles terug tot de eenvoudigst mogelijke bouwstenen: driehoeken.

Definitie 6.2.1. *Een oppervlak is een ruimte die je kunt construeren door (eindig veel) driehoeken netjes aan elkaar te plakken.*

Met netjes aan elkaar plakken bedoelen we dat twee driehoeken met de zijden tegen elkaar geplakt worden, zie figuur 6.3. In het resultaat overlappen de driehoeken dus alleen op de zijden en op iedere zijde komen er maximaal twee driehoeken samen.

Het is belangrijk dat de driehoeken geen specifieke vorm of grootte hebben. Ze zijn als het ware van rubber en kunnen naar believen uitgerekt en ingekrompen worden (hoewel we ze voor het gemak meestal recht tekenen). Zolang het patroon waarmee ze op elkaar aansluiten maar niet

verstoord wordt beschouwen we het oppervlak als onveranderd. Het opbouwen van een oppervlak uit driehoeken heet ook wel een *triangulatie* van het oppervlak.

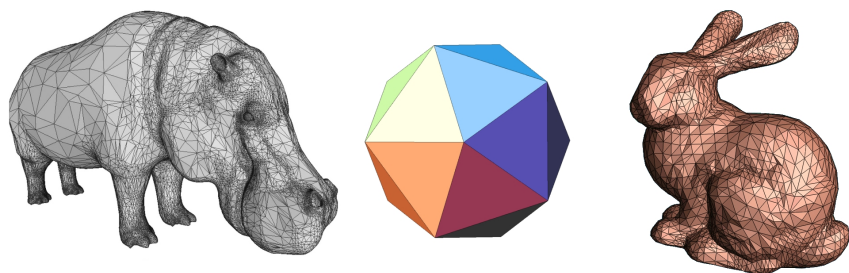


Figuur 6.3: Twee losse driehoeken links worden iets vervormd zodat ze netjes tegen elkaar aan geplakt kunnen worden. De twee plaatjes rechts stellen (topologisch gezien) allebei de schijf voor.

Niet alle zijden hoeven aan een andere zijde vastgeplakt te worden. De overgebleven zijden vormen dan samen de *rand* van ons oppervlak. Dit gebeurde in figuur 6.3 waar we een schijf plakten uit twee driehoeken. Ook de andere oppervlakken op de onderste regel van figuur 6.2 hebben een rand. De ring heeft zelfs twee (de onderste hebben we rood gemaakt). De rand bestaat altijd uit een aantal losse cirkels, de *randcirkels*. Je kunt nagaan dat dit komt doordat er nooit meer dan twee driehoeken een lijn gemeen hebben.

De driehoeken zijn alleen maar een manier om over het oppervlak te kunnen praten. We stellen ons het oppervlak vaak meer als een glad object voor zoals de bol(schil) en de andere voorbeelden uit figuur 6.2. Alles wat je om je heen ziet, dit boek, je hand, je stoel, alles kun je met kleine driehoekjes benaderen. Dit gebeurt bijvoorbeeld dagelijks in de computergraphics toepassingen. Zo zie je in figuur 6.4 bijvoorbeeld een grove triangulatie van het konijn uit figuur 6.1. Daarnaast staan in figuur 6.4 nog twee alternatieve triangulaties van dezelfde bol. Kun je ook triangulaties vinden van de andere oppervlakken in figuur 6.2?

We hebben gezien dat er heel veel verschillende manieren zijn om een oppervlak uit driehoeken op te bouwen, te trianguleren. De kernvraag van de topologie is hoe we kunnen herkennen of twee oppervlakken (topologisch gezien) hetzelfde zijn. Maar wat bedoelen we hier nu precies mee? Intuïtief zijn twee oppervlakken gelijk als ze geleidelijk in elkaar om te vormen zijn, waarbij we zonnodig de driehoeken onderverdelen in kleinere driehoekjes, maar dat is nog erg vaag. Daar hebben we niets aan als we bijvoorbeeld willen begrijpen hoe het kan dat twee Möbiusbanden, langs hun randen aan



Figuur 6.4: Drie triangulaties van de bol. Topologisch gezien is er geen verschil.

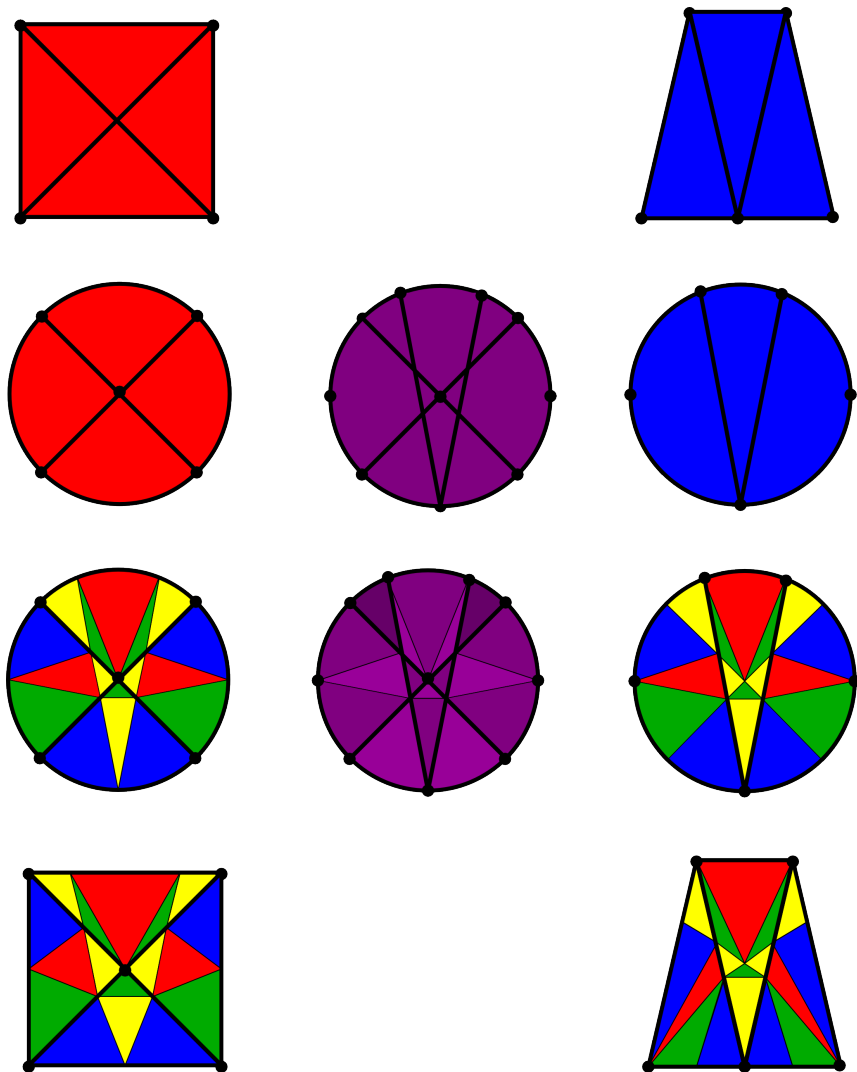
elkaar geplakt, de Kleinse fles vormen (rechtsboven in figuur 6.2). Om in het volgende hoofdstuk Perelmans ideeën te kunnen volgen hebben we een precieze definitie nodig van wanneer we twee oppervlakken als hetzelfde beschouwen:

Definitie 6.2.2. *We zeggen dat oppervlakken A en B topologisch gelijk zijn als we een gemeenschappelijke verfijning van hun triangulaties kunnen vinden.*

Daarmee bedoelen we dat we de driehoeken van A en B kunnen verdelen in kleinere driehoeken zo dat er een één op één verband ontstaat tussen de kleine driehoeken van A en die van B . Bovendien moeten dan twee kleine driehoeken elkaars burens zijn precies wanneer de corresponderende driehoekjes in B dat zijn en omgekeerd.

In figuur 6.5 staat een voorbeeld waar je aan kunt zien hoe dit in de praktijk werkt. We laten zien dat het rode oppervlak linksboven gelijk is aan het blauwe oppervlak rechtsboven. Beide stellen ze de topologische schijf voor. De rode triangulatie uiterst links bestaat uit vier driehoeken, de blauwe rechts uit drie. Volgens de definitie hierboven moeten we een gemeenschappelijke verfijning vinden van de blauwe en de rode driehoeken. Daaruit volgt dan dat ze gelijk zijn. Links- en rechts onder staat met fijne lijntjes de gezochte gemeenschappelijke verfijning aangegeven in de twee triangulaties. De verfijning bestaat uit 17 kleinere driehoeken.

Op de twee regels plaatjes daartussen zien we hoe we aan deze verfijning kwamen. Eerst vervormen we de rode en de blauwe oppervlakken tot ze ‘op elkaar passen’. Dat is gemakkelijk voor elkaar te krijgen door



Figuur 6.5: Boven: twee triangulaties van de schijf. Tweede regel: vervorming tot een ronde schijf en overlap (midden). Derde regel: onderverdeling in driehoeken (fijne lijntjes). Onder: De oorspronkelijke triangulaties met daarin hun gemeenschappelijke verfijning.

beide de vorm van een ronde schijf te geven. Vervolgens leggen we deze twee schijven bovenop elkaar (paars). De lijnen van de twee triangulaties snijden elkaar nu en maken een patroon van (kromme) veelhoeken (vette lijnen). Op de volgende regel voegen we extra lijntjes toe aan dit patroon om er een triangulatie van te maken. Deze triangulatie is onze gezochte gemeenschappelijke verfijning. We kunnen die nu ook ‘terughalen’ naar de oorspronkelijke triangulaties (fijne lijntjes). Anders gezegd krijgen we via het paarse plaatje de gevraagde één op één correspondentie tussen de kleine driehoekjes.

De bovenstaande procedure werkt niet alleen voor ons voorbeeld van de schijven. Voor ieder tweetal triangulaties van hetzelfde oppervlak kunnen we zo een gemeenschappelijke verfijning vinden.

Als een directe toepassing van definitie 6.2.2 kunnen we nu bewijzen dat oppervlakken met rand nooit gelijk kunnen zijn aan oppervlakken zonder rand. We kunnen de rand namelijk nooit wegstreken door een triangulatie te verfijnen of te vervormen. Sterker nog, het aantal randcirkels moet gelijk zijn. Zo heeft de ring twee randcirkels en de schijf maar een (zie figuur 6.2) dus zijn deze oppervlakken ook niet gelijk. Het aantal randcirkels is dus een nuttig getal waarmee we oppervlakken kunnen beschrijven. In de volgende paragraaf komen we nog zo’n getal tegen, het Eulergetal, dat beschrijft wat er binnen het oppervlak zelf gebeurt.

Opgave

Bewijs dat als je twee ringen met hun randcirkels aan elkaar plakt je de torus krijgt (zie figuur 6.2). Laat ook zien dat als je twee Möbiusbanden langs hun randcirkels aan elkaar plakt, je de Kleinse fles krijgt (moeilijker).

6.2.2 Het Eulergetal

De kernvraag in de topologie is hoe je ruimten kunt classificeren en van elkaar kunt onderscheiden. Voor oppervlakken gaat dat heel goed met het zogenaamde *Eulergetal* (naar Leonhard Euler). Het Eulergetal χ (chi) van een oppervlak komt neer op het tellen van het aantal hoekpunten, zijden en driehoeken in een triangulatie van je oppervlak. Traditioneel geven we het aantal punten, lijnen en driehoeken aan met respectievelijk de letters V , E en F .

Definitie 6.2.3. *Het Eulergetal χ van een oppervlak is $\chi = V - E + F$.*

In figuur 6.6 staat een simpel voorbeeld waarin we berekenen dat het Euler-

getal van de schijf gelijk is aan 1 dus $\chi(\text{schijf}) = 1$. De aantallen punten, lijnen en driehoeken, V , E en F staan aangegeven en hun alternerende som is inderdaad 1.

$$\begin{array}{ccccccc} \chi & = & V & - & E & + & F \\ \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} & & \begin{array}{c} \bullet \\ \bullet \end{array} & & \begin{array}{c} \diagup \quad \diagdown \\ \hline \diagdown \quad \diagup \end{array} & & \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} \\ \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} & & \begin{array}{c} \bullet \\ \bullet \end{array} & & \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} & & \begin{array}{c} \bullet \\ \diagup \quad \diagdown \\ \bullet \end{array} \\ 1 & = & 4 & - & 5 & + & 2 \end{array}$$

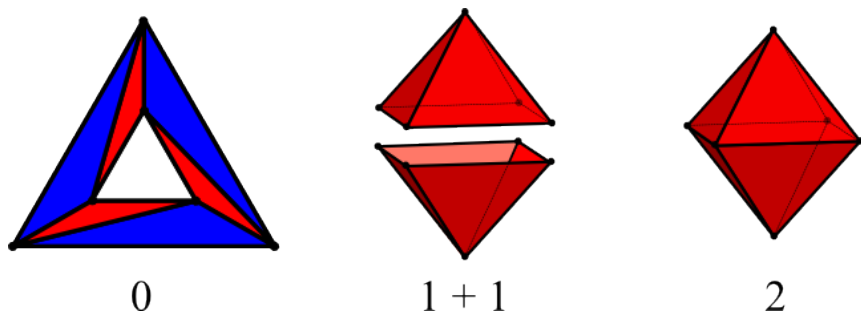
Figuur 6.6: Berekening van het Eulergetal χ van de schijf. $\chi(\text{schijf}) = 1$

We hadden het onszelf ook makkelijker kunnen maken door één enkele driehoek te nemen. Er zijn nu evenveel lijnen als punten en maar één driehoek, dus $\chi = 3 - 3 + 1 = 1$.

In de twee bovenstaande voorbeelden zien we dat het Eulergetal niet af lijkt te hangen van de triangulatie van de schijf die we gebruiken. Dat is geen toeval want volgens definitie 6.2.2 hebben twee triangulaties van hetzelfde oppervlak een gemeenschappelijke verfijning. Je kunt nagaan dat het Eulergetal niet verandert wanneer we een triangulatie verfijnen, daarom leveren beide triangulaties hetzelfde antwoord op.

In het volgende voorbeeld berekenen we de Eulergetallen van de bol en de ring. Figuur 6.7 links is een triangulatie van de ring gegeven met zes driehoeken. In de berekening $\chi = V - E + F$ valt het aantal driehoeken weg tegen de zes inwendige lijnen en de punten tegen de randlijnen. Het antwoord is dus $\chi(\text{ring}) = 0$.

Het Eulergetal van de bol berekenen we uit de octaëder (rechts) in figuur 6.7, een triangulatie met acht driehoeken. In het middelste plaatje hebben we de octaëder langs de evenaar in twee piramidevormige schijven opgedeeld (de piramides hebben geen grondvlak!). We weten al dat $\chi(\text{schijf}) = 1$ dus samen hebben de twee schijven Eulergetal 2. Plakken we de schijven op elkaar dan verdwijnen er evenveel hoekpunten V als lijnen



Figuur 6.7: De Eulergetallen van de ring, twee schijven en de bol op een rijtje.

E en blijft F gelijk. Daarom blijft ook $\chi = V - E + F$ gelijk. Met andere woorden $\chi(\text{bol}) = 1 + 1 = 2$.

De laatste observatie is interessant omdat we hier het Eulergetal van een oppervlak (de bol) berekenden door de Eulergetallen van twee eenvoudigere oppervlakken op te tellen. Die twee eenvoudigere bouwstenen werden langs een van hun randcirkels aan elkaar geplakt. We zagen namelijk in dat het Eulergetal niet verandert wanneer we twee randcirkels van een oppervlak aan elkaar plakken. Andersom mag je dus ook het Eulergetal van je oppervlak berekenen door het langs cirkels in stukjes te snijden. Het Eulergetal van het geheel is niets anders dan de som van de Eulergetallen van de stukjes. Zo kun je bijvoorbeeld de torus uit figuur 6.2 in cylindervormige plakjes snijden. Al die plakjes zijn ringen (met $\chi = 0$) dus het Eulergetal van de torus is ook 0.

Tot nu toe zijn alle Eulergetallen die we hebben gevonden positief of 0 maar dit is misleidend. Snij je namelijk een klein kapje af van de bol dan gaat het Eulergetal 1 omlaag. Dit volgt weer uit de redenering hierboven. Een bol met veel gaten (randcirkels) heeft dus een heel negatief Eulergetal.

Opgave

Zijn er ook oppervlakken zonder rand met negatief Eulergetal?

6.2.3 Classificatie van oppervlakken

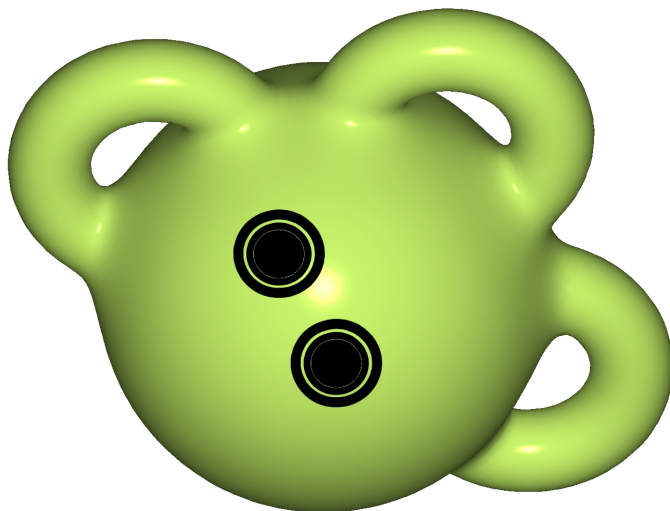
Het Eulergetal is heel praktisch om oppervlakken te onderscheiden, maar kun je hiermee ook werkelijk alle oppervlakken beschrijven? August Möbius (van de band) gaf hierop in 1861 het volgende antwoord. Poincaré was toen 7 jaar oud.

Stelling 6.2.4. *(Classificatiestelling van oppervlakken)*

Een samenhangend oppervlak wordt geheel bepaald door drie parameters:

1. *Het Eulergetal χ*
2. *Het aantal randcirkels r*
3. *Of het oriënteerbaar is of niet.*

Met samenhangend bedoelen we alleen maar dat we er van uit gaan dat het oppervlak niet uiteenvalt in losse stukken. Het begrip oriënteerbaarheid slaat op het verschil tussen de ring en de Möbiusband: op de eerste kun je consistent praten over links en rechts, op de tweede niet. Daarom noemen we de ring oriënteerbaar en de Möbiusband niet.



Figuur 6.8: Een bol met drie handvaten en twee gaten.

Möbius gaf ook een expliciete opsomming van alle mogelijke oppervlakken.

Een oriënteerbaar oppervlak ziet er uit als een bol met handvatten en gaten als in figuur 6.8. Met een gat bedoelen dat er een schijf uit het boloppervlak is weggelaten waarbij een cirkelvormige rand ontstaat. Is het oppervlak niet oriënteerbaar dan is aan een aantal van de randcirkels van de gaten een Möbiusband vastgeplakt. Hoe dit er precies uitziet laten we aan de fantasie van de lezer over.

Een bewijs van deze uitspraken zou te ver voeren hoewel het niet erg moeilijk is en zeker de moeite waard. We verwijzen naar een mooi bewijs van John Conway [FW99]. Ook is er interessant Nederlands lesmateriaal over oppervlakken gemaakt door Jacobien Carstens en Sebas Eliëns [CE10].

Opgave

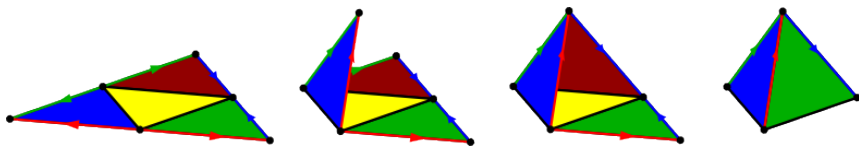
Teken een zo goed mogelijk plaatje van een oppervlak met één handvat en twee gaten waarvan er aan één van de randcirkels een Möbiusband geplakt zit. Is dit de Kleinse fles? Zo nee, hoe kan die dan beschreven worden in termen van de standaardvorm van Möbius?

6.2.4 Intrinsieke oppervlakken

Voor we ons in hogere dimensies begeven kijken we nog even van een hoger standpunt naar onze oppervlakken. We stellen ons oppervlakken graag voor als ruimtelijke objecten in drie dimensies, maar logisch gezien zijn ze dat niet noodzakelijk. Als we terugkijken naar de definities staat er alleen iets over driehoeken waarvan de zijden aan elkaar zitten. Die driehoeken zelf zijn objecten uit het platte vlak dus in zekere zin zijn oppervlakken dat ook. We noemen oppervlakken dan ook wel tweedimensionale topologische ruimten.

Dit standpunt vanwaaruit we oppervlakken zien als opzichzelfstaande ruimten heet wel *intrinsiek*. We praten over oppervlakken zonder te refereren aan hoe (en of!) ze in de ruimte liggen. In zekere zin worden we gedwongen tot dit standpunt omdat sommige oppervlakken nu eenmaal niet in onze gewone $\mathbb{R}^3$ passen. De fles van Klein is hier een voorbeeld van. Iedere poging om hem toch als een ruimtelijk object te tekenen levert een schijnbare zelfdoorsnijding op, zie figuur 6.2.

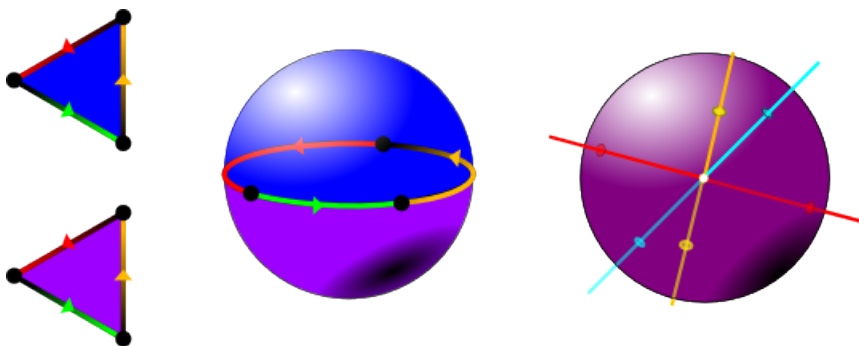
Vanuit het intrinsieke standpunt stellen we ons oppervlakken voor als *bouwplaten*. Van elke driehoek geven we aan waar hij aan vast zit maar zonder te vertellen hoe het geheel eruit ziet. In het geval van de bol is dit een bekende procedure. Hier is bijvoorbeeld een bouwplaat van een



Figuur 6.9: We zetten een uitslag van een tetraëder in elkaar. Hierbij moeten de gekleurde pijltjes op elkaar komen. Linksboven is intrinsiek, rechtsonder extrinsiek.

triangulatie van de bol met vier driehoeken (tetraëder), zie figuur 6.9. De zijden met dezelfde kleur moeten in de richting van de pijltjes aan elkaar geplakt worden.

In het bovenstaande geval kunnen we de bouwplaat van de tetraëder ook daadwerkelijk in elkaar zetten in onze ruimte. Voor de meeste bouwplaten lukt dat niet. Zelfs voor triangulaties van de bol kan het verwarrend zijn. Vergelijk bijvoorbeeld de volgende bouwplaat van een triangulatie van een bol in een noordelijke en zuidelijke driehoek (figuur 6.10). Ze passen netjes op elkaar als je de driehoeken flink wat buigt.



Figuur 6.10: Links een bouwplaat van de bol met twee blauwe driehoeken. Rechts een tekening van het projectieve vlak.

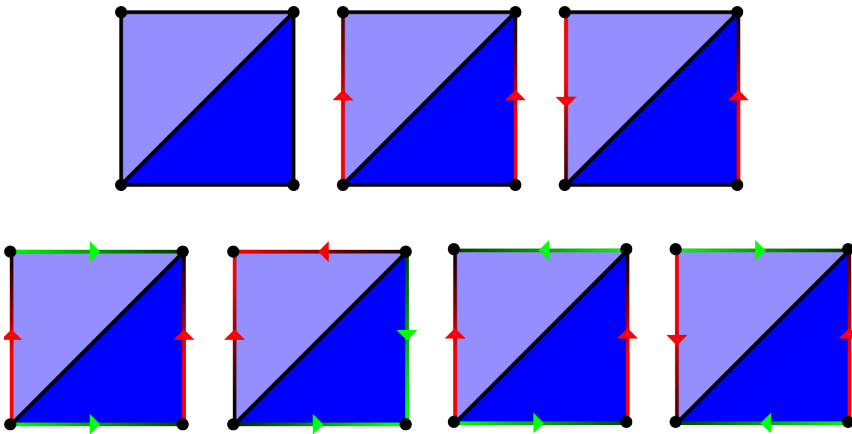
Rechts in dezelfde figuur staat een plaatje van een veel buitenissiger oppervlak: het *projectieve vlak*. Net als de fles van Klein is het gemakkelijk intrinsiek te definiëren maar veel moeilijker te visualiseren. Het projectieve vlak kunnen we beschrijven als de verzameling lijnen in $\mathbb{R}^3$ door de

oorsprong. Deze lijnen zijn als het ware de lichtstralen die ons oog in de oorsprong bereiken.

In figuur 6.10 hebben we drie lijnen door de oorsprong getekend en aangegeven waar ze de eenheidsbol snijden. Zoals gezegd, iedere lijn stelt een punt van het projectieve vlak voor. Omdat iedere lijn de eenheidsbol in twee tegenoverliggende punten snijdt kunnen we de punten van het projectieve vlak ook beschrijven met tegenoverliggende paren punten op de bol. Het is niet direct duidelijk hoe en of dit wel een oppervlak beschrijft want we hebben nog geen bouwplaat bij het oppervlak gemaakt. We komen hier later (en in de opgave) op terug.

Opgave

Hieronder staan zeven uitslagen van zeven oppervlakken met twee driehoeken. Bereken van iedere uitslag het Eulergetal en het aantal randcirkels. Maak ook een zo goed mogelijke schets van het oppervlak. Hebben we nu alle oppervlakken die een triangulatie met maximaal twee driehoeken toestaan in kaart gebracht?

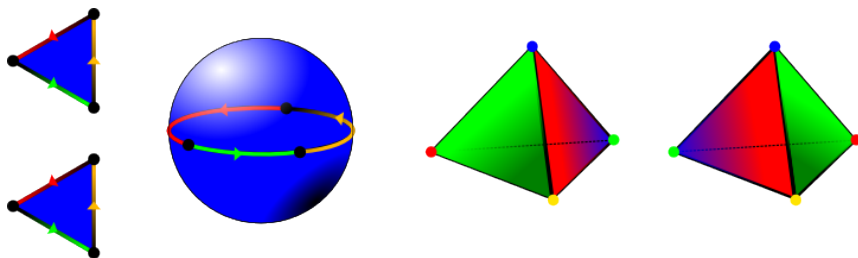


Figuur 6.11: Bouwplaten voor oppervlakken met twee driehoeken. De gekleurde pijltjes moeten met de juiste richting tegen elkaar geplakt worden.

6.2.5 Driedimensionale ruimten

Gewapend met onze kennis over oppervlakken wagen we ons nu aan ingewikkeldere topologische ruimten: driedimensionale ruimten (of korter: *3-ruimten*). Dit zijn de ruimten waar Poincaré en Perelman zich het hoofd over braken. Nog steeds wordt er veel onderzoek gedaan naar deze ruimten. We kunnen helaas alleen een korte indruk geven zonder in al te veel details te treden.

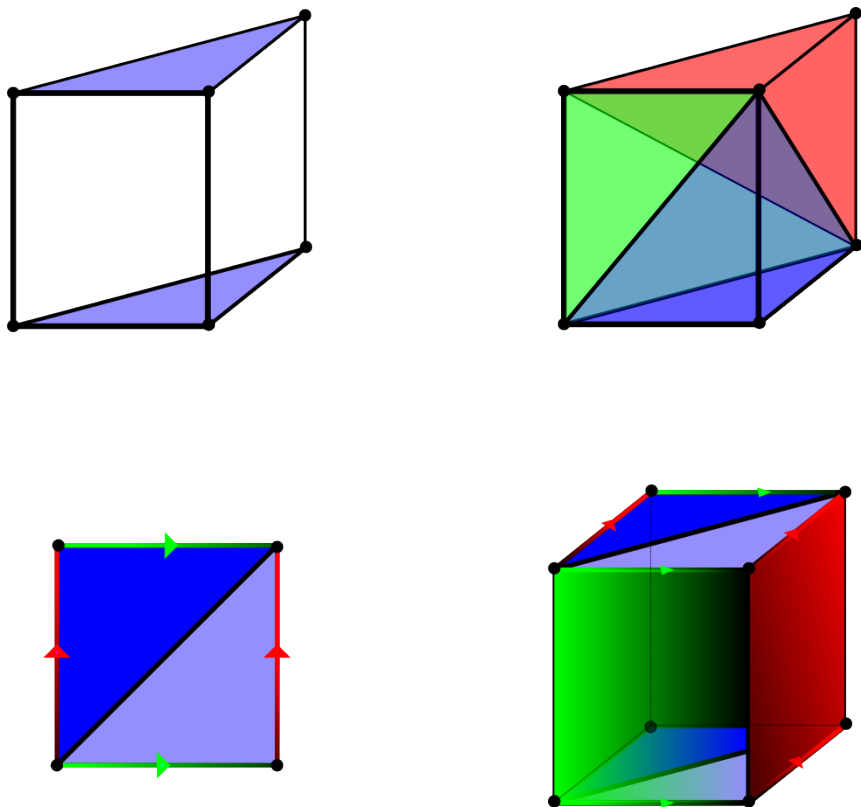
Net zoals oppervlakken uit driehoeken zijn opgebouwd, zo zijn (massieve) tetraëders de bouwstenen voor de 3-ruimten. Volgens Poincaré is de eenvoudigste 3-ruimte de 3-bol. Net als de gewone bolschil (2-bol of kortweg 'bol') kunnen we hem uit twee stukken samenstellen: een noordelijk- en een zuidelijk halfrond, zie figuur 6.12. Het verschil is dat deze twee stukken nu twee tetraëders zijn in plaats van twee driehoeken. Dat neemt niet weg dat de 3-bol wel degelijk ook als ronde bol te realiseren is, maar dan wel als de eenheidsbol in $\mathbb{R}^4$. De intrinsieke bouwplaat van twee tetraëders is vaak praktischer.



Figuur 6.12: Links: de bouwplaat voor de gewone bol uit twee stukken. Rechts: Een driedimensionale bouwplaat voor de 3-bol. De twee (massieve) tetraëders moeten met alle zijvlakken aan elkaar als een soort noordelijk en zuidelijk halfrond. De gekleurde punten komen dan op elkaar te liggen.

Een nog eenvoudiger voorbeeld van een 3-ruimte is natuurlijk een enkele (massieve) tetraëder. Deze heeft een rand en dat is een (topologische) 2-bol. Je kunt nagaan dat de rand, hoe je de tetraëders ook plakt, altijd een oppervlak zal zijn. In figuur 6.13 (boven) hebben we een driehoekig prisma aan elkaar geplakt uit drie tetraëders. Deze 3-ruimte is topologisch gezien volkomen gelijk aan de enkele tetraëder van net. Het wordt anders wanneer we de boven- en ondervlakken van het prisma aan elkaar plakken.

Dan krijgen we een massieve torus. Dat is een 3-ruimte waarvan de rand een torus oppervlak is, zie figuur 6.14 (rechts).



Figuur 6.13: Boven: Een driehoekig prisma is opgebouwd uit drie tetraëders (rood, blauw en groen). Onder: een bouwplaat van een oppervlak wordt een bouwplaat voor een 3-ruimte door op iedere driehoek een prisma te zetten.

Met behulp van de prisma's kunnen we onze kennis van oppervlakken gebruiken om ook 3-ruimten te maken. Dat gaat als volgt, zie figuur 6.13. Begin met een bouwplaat van je favoriete oppervlak (in het voorbeeld is dat bouwplaat van de torus uit twee driehoeken). Op iedere driehoek in de bouwplaat zetten we nu een driehoekig prisma. Vervolgens plakken we de prisma's met hun verticale vlakken aan elkaar volgens de instructies

van de bouwplaat van het oppervlak. De 3-ruimte die we zo krijgen is een ‘verdikking’ van het oorspronkelijke oppervlak. Dit is een 3-ruimte met rand, de boven- en ondervlakken van de prisma’s worden namelijk nergens aan vast geplakt. De bovenvlakken van de prisma’s vormen een kopie van het oppervlak en de ondervlakken ook. De rand bestaat dus uit twee delen.

Tot slot kunnen we dan de boven- en ondervlakken van ieder prisma aan elkaar plakken. In het geval dat ons oppervlak de schijf was (een enkele driehoek) krijgen we zo de massieve torus. Heeft ons oppervlak geen rand dan krijgen we interessante voorbeelden van 3-ruimten zonder rand. In ons voorbeeld waarbij we beginnen met het torus-oppervlak en na de prisma’s er op te zetten en aan elkaar te plakken krijgen we de zogenaamde *3-torus*.

Dit is nog maar een begin van de enorme rijkdom aan verschillende mogelijke 3-ruimten. De meesten zijn ingewikkelder dan die we met de oppervlak-constructie hierboven kunnen maken. Meer hierover is te lezen in het prachtige boek van Jeff Weeks [We01]. Ook het hoofdstuk over het Poincarévermoeden in de bundel [BVCK] gaat hier dieper op in.

Een driedimensionaal Eulergetal

Ons geheime wapen bij het bestuderen van de oppervlakken (ook wel 2-ruimten) was het Eulergetal. Is er ook een Eulergetal voor 3-ruimten? Het Eulergetal $\chi = V - E + F$ was de alternerende som van de aantallen punten, lijnen en driehoeken. Voor de 3-ruimten speelt ook het aantal tetraëders T een rol. Het Eulergetal is daarom gedefiniërd als $\chi = V - E + F - T$.

Zo is het Eulergetal van de 3-bol bijvoorbeeld $4 - 6 + 4 - 2 = 0$, want we konden de 3-bol maken uit twee tetraëders door alle zijvlakken van de een aan alle zijvlakken van de ander te plakken. Dus $T = 2$ en de punten, lijnen en driehoeken worden in paren op elkaar geplakt zodat er nog maar 4 punten, 6 lijnen en 4 vlakken overblijven.

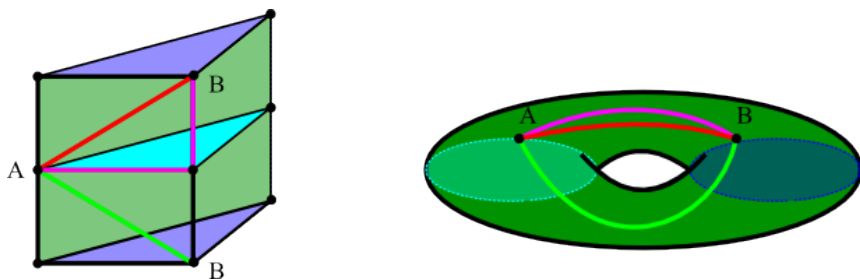
Een ander voorbeeld is de enkele tetraëder. Deze 3-ruimte met rand heeft Eulergetal $\chi = 4 - 6 + 4 - 1 = 1$. Omdat het prisma topologisch equivalent is, moet dit ook Eulergetal 1 hebben (dat klopt ook met de berekening). Plakken we de boven- en ondervlakken aan elkaar om de massieve torus te krijgen dan verliezen we drie punten, drie lijnen en één driehoek dus neemt het Eulergetal 1 af. Het Eulergetal van de massieve torus is dus 0.

Met iets meer moeite kun je berekenen dat het Eulergetal nul is van alle 3-ruimten zonder rand die we hierboven met prisma’s construeerden. Dat is geen toeval want het blijkt dat het Eulergetal van alle (behoorlijke)

3-ruimten zonder rand altijd gelijk is aan 0.

Om 3-ruimten toch te kunnen bestuderen bedacht Poincaré iets anders. Hij keek naar de mogelijke paden door de 3-ruimte. Met een pad bedoelen we een aantal lijntjes van de tetraëders die samen punt A met punt B verbinden. Er zijn een hoop mogelijke paden van A naar B maar de meeste lijken erg op elkaar. Poincaré noemde daarom twee paden equivalent als ze met kleine stapjes in elkaar over kunnen gaan, namelijk door het pad steeds maar in één enkele tetraëder te veranderen.

De eenvoudigste 3-ruimten zijn die waar alle paden van A naar B equivalent zijn (dit heet ook wel enkelvoudig samenhangend). Dat is bijvoorbeeld zo in de tetraëder en ook in de 3-bol, maar niet in de massieve torus en ook niet in de 3-torus. In de massieve torus is goed te zien dat er meerdere niet-equivalente paden tussen A en B mogelijk zijn. Je kunt namelijk van A naar B linksom het gat en rechtsom het gat, zie figuur 6.14.



Figuur 6.14: Twee aanzichten van de massieve torus met daarop drie paden van punt A naar punt B . Het rode pad is equivalent met het paarse pad, maar niet met het groene pad.

Poincaré vroeg zich nu af of de 3-bol de enige 3-ruimte was waar alle paden tussen twee punten equivalent zijn. Dit is het beroemde Poincaré-vermoeden.

Poincarévermoeden: De enige 3-ruimte zonder rand waar alle paden tussen twee punten equivalent zijn is de 3-bol.

Opgave

Is de 2-bol ook het enige oppervlak zonder rand waar alle paden tussen twee punten equivalent zijn?

6.3 Meetkunde op oppervlakken

We hebben nu een hele diertuin aan topologische ruimten opgebouwd. Hoe kunnen we hier orde in aanbrengen? Voor de oppervlakken konden we het Eulergetal gebruiken om ze te classificeren, maar voor 3-ruimten lukte dat niet. De visie van Perelman was om het vormeloze vorm te geven. Tot nu toe zijn de ruimten opgebouwd uit driehoeken of tetraëders zonder specifieke vorm. Ze zijn van rubber want in de topologie maakt de precieze vorm niet uit. Dat was nu juist de kracht van de topologie: om de ingewikkelde vorm van figuren te negeren en alleen naar de hoofdzaak te kijken.

Perelman wilde de vormeloze topologische ruimten vorm geven. Wij zullen dat doen door aan iedere zijde van de driehoeken een lengte toe te kennen. Dat heeft op de topologie geen enkel effect. Perelman liet echter zien dat je ruimten een heel speciale vorm kunt geven die uniek is en als het ware opgelegd wordt door de manier waarop de driehoeken aan elkaar zitten. Deze unieke standaardvorm van Perelman vertelt ons dan wel degelijk iets over de topologische ruimte zelf. Het is een fantastisch classificerend principe.

Om de ingewikkelde machinerie van de Riemannse meetkunde te vermijden werken we met een *discrete* versie van de theorie van Perelman. Dat wil zeggen dat we geen punten toelaten die oneindig dicht bij elkaar liggen. In deze context zullen we precies zeggen wat we bedoelen met de optimale standaardvorm van een oppervlak. Dit doen we in termen van het meetkundige begrip *kromming*. Verrassend is dat het Eulergetal hier weer een belangrijke rol speelt. Het essentiële punt is dat deze methode van Perelman ook voor hogerdimensionale ruimten werkt.

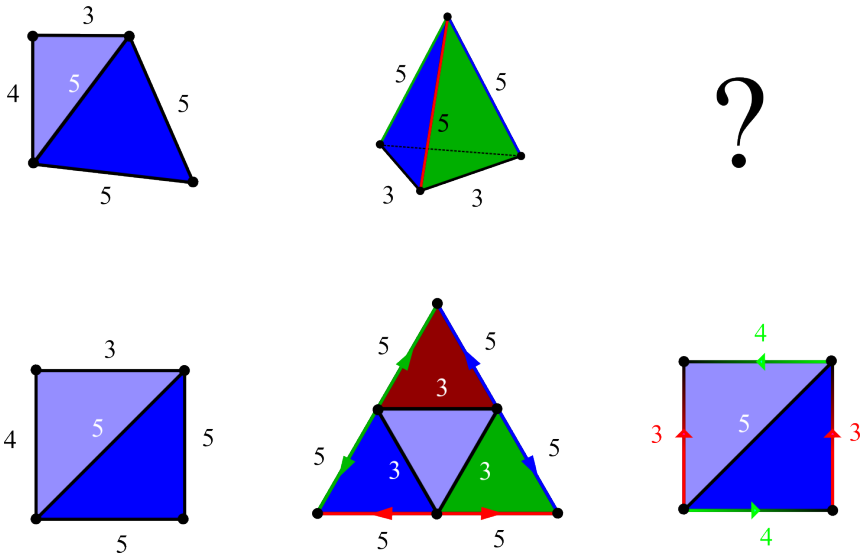
6.3.1 Metrieken op een triangulatie

Om de vormeloze topologische oppervlakken vorm te geven doen we het volgende. We kiezen een triangulatie en geven de zijden van iedere driehoek een specifieke lengte. Intuïtief zijn de driehoeken nu niet meer van rubber maar juist van kristal. Iedere driehoek is nu een specifieke vlakke Euclidische driehoek die geheel is vastgelegd door de lengten van zijn drie zijden. De essentiële meetkundige informatie zit hem in het geven van een positief getal voor iedere zijde, we noemen dit wel een metriek. Om te zorgen dat een drietal zijden ook werkelijk een Euclidische driehoek geeft eisen we dat ze aan de zogenaamde *driehoeksongelijkheid* voldoen. De drie-

hoeksongelijkheid stelt dat in iedere Euclidische driehoek de langste zijde korter is dan de som van de lengtes van de overige twee zijden.

Definitie 6.3.1. Een (discrete) metriek op een getrianguleerd oppervlak is een functie die aan iedere zijde van de triangulatie een positief getal toekent (de lengte). Voor elke driehoek in de triangulatie moeten de lengten van de zijden bovendien aan de bovengenoemde driehoeksongelijkheid voldoen.

Het tekenen van een getrianguleerd oppervlak in $\mathbb{R}^3$ zoals het nijlpaard de icoesaëder en het konijn in figuur 6.4 geeft impliciet een metriek aan. We kunnen dan namelijk de lengten van alle zijden meten. Aan de driehoeksongelijkheid is automatisch voldaan omdat de driehoeken er anders nooit recht uit zouden kunnen zien. Een metriek geven is echter niet hetzelfde als het oppervlak in ruimtelijk realiseren in $\mathbb{R}^3$. Het begrip metriek is namelijk *intrinsiek*, het is niets meer dan een functie $\text{lengte} : E \rightarrow \mathbb{R}_{>0}$ die aan een aantal ongelijkheden voldoet.



Figuur 6.15: Onder: Bouwplaten van de schijf, de bol en de Kleinse fles met metriek. Boven: Zo mogelijk een kloppende tekening van het bijbehorende meetkundige object.

In figuur 6.15 zijn drie bouwplaten van getrianguleerde oppervlakken getekend die de betekenis van de metriek duidelijk maken. Bij iedere lijn staat

de lengte gegeven en boven de bouwplaat hebben we geprobeerd het meetkundige object natuurgetrouw te tekenen. Voor de schijf links lukt dat heel aardig. In het middelste plaatje staat de bouwplaat van de tetraëder uit figuur 6.9. In de gekozen metriek is de top wat uitgerekt.

Helemaal rechts staat een uitslag van de Kleinse fles maar hiervan hebben we niet geprobeerd om een correct plaatje te tekenen. Dat lukt niet in onze ruimte vanwege de zelfdoorsnijdingen. Dat neemt niet weg dat het een prima meetkundig oppervlak is waar we van alles over te weten kunnen komen.

Vergelijking met Riemannse metrieken

Een probleem met onze meetkunde op oppervlakken is dat het werkt met een vaste triangulatie. Het is natuurlijker om, zoals in de *Riemannse meetkunde*, oppervlakken voor te stellen als gladde (differentiërbare) objecten. In dit geval wordt de metriek gegeven door inproduct op de raakruimte in elk punt. De lengte van een pad meten we dan door de lengte van de raakvector te integreren. Onze grovere (discrete) aanpak met platte driehoeken zien we als een benadering van de werkelijke gladde situatie die Perelman beschreef. Naarmate het aantal driehoeken groter en groter wordt beginnen onze resultaten heel goed te lijken op het gladde geval.

Opgave

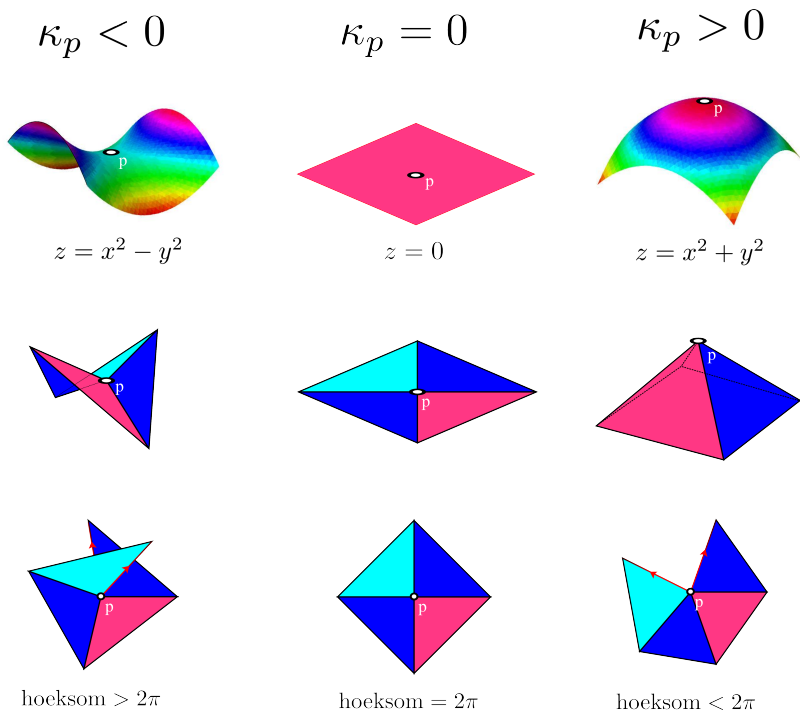
Gegeven een metriek op de tetraëder-triangulatie van de bol. Kunnen we altijd een Euclidische tetraeder in construeren in $\mathbb{R}^3$ waarvan de lengten van zijden overeenkomen met de metriek?

6.3.2 Kromming

In de Riemannse meetkunde beschrijven we gladde oppervlakken allereerst door hun raakvlakken. Zien we een oppervlak als de grafiek van een functie dan is het raakvlak niets anders dan de beste lineaire benadering in het punt. We krijgen nog een beter beeld van het oppervlak door te kijken hoe het wegbuigt van zijn raakvlak. Dat kan in ieder punt p anders zijn en we noemen dit de (Gauss) *kromming* κ_p in punt p . Als we ons oppervlak weer zien als de grafiek van een functie dan beschrijft de kromming κ_p de beste kwadratische benadering van onze functie in het punt p .

Boven in figuur 6.16 zien we een punt p en de drie kwalitatief verschillende

mogelijkheden voor de kromming κ_p in het punt p . Het gaat alleen om hoe het er dichtbij p uit ziet. Is de kromming positief dan is het oppervlak rondom p bolvormig, is κ_p negatief dan lijkt het juist op een zadel dichtbij p . Kromming 0 is heel speciaal, in dat geval is het platte raakvlak zelf juist de beste benadering.



Figuur 6.16: Het effect van het teken van de kromming κ_p in het punt p , op een glad en een getrianguleerd oppervlak en zijn uitslag. Hoeksom rechts slaat op de som van de hoeken rond p .

In onze discrete context kunnen we het begrip kromming heel goed benaderen in termen van de hoeken² van de driehoeken. Dit werkt alleen voor de hoekpunten van de driehoeken. Voor alle andere punten is de kromming 0. Gegeven een metriek op een getrianguleerd oppervlak is de kromming κ_p in een punt p als volgt gedefinieerd.

Definitie 6.3.2. *De kromming κ_p in een punt p is gelijk aan 2π min de*

²We rekenen met radialen in plaats van graden.

som van de hoeken van de driehoeken die samenkomen rond om p . Is p een randpunt dan gebruiken we π min hoekensom.

Rechts in figuur 6.16 staat de discrete kromming in een punt op een getrianguleerd schijfje getekend. Hoewel we hier maar vier driehoeken gebruiken in onze bouwplaten benadert het resultaat het gladde geval verrassend goed.

Kromming is een *intrinsiek* begrip. Het wordt namelijk bepaald door de hoeken van de driehoeken en die kun je met de cosinusregel berekenen. Alle lengten van de zijden zijn namelijk gegeven door de metriek. De driehoeksongelijkheid zorgt er bovendien voor dat de hoeken zich gedragen zoals we dat gewend zijn. Ter herinnering is hier de formule voor een driehoek met lengten a, b, c . De hoek α tegenover a wordt gegeven door:

$$\cos \alpha = \frac{-a^2 + b^2 + c^2}{2bc}$$

Het begrip kromming laat overigens goed zien waarom wereldkaarten nooit helemaal waarheidsgetrouw zijn. Het aardoppervlak is namelijk een bol en heeft dus positieve kromming. We kunnen dit dus nooit correct afbeelden in het platte vlak. Dat heeft namelijk kromming 0.

Als een eenvoudig voorbeeld berekenen we de kromming in de punten van de tetraëder uit figuur 6.15 (midden). We beginnen met de top. Hier komen drie identieke hoeken samen en volgens de cosinusregel is een zo'n hoek gelijk aan $\cos \frac{41}{50}$. De hoeksom is daarom ongeveer 1.83 radiaal en de kromming in het toppunt is dus ongeveer 4.46. De kromming in de drie andere punten is ongeveer 2.70.

Opgave

Bereken de kromming in de hoekpunten van de regelmatige tetraëder, octaëder en de icoëder. Trianguleer ook een regelmatige kubus en bereken de kromming in de hoekpunten daar. Was het nodig om te trianguleren? Bereken voor ieder van deze voorbeelden en het voorbeeld hierboven de som van de krommingen in alle punten samen.

6.3.3 Kromming en het Eulergetal

In de vorige opgave zagen we al dat als we de kromming in alle punten van een bol optellen we altijd op 4π uit lijken te komen. Dit doet denken aan

het Eulergetal uit het vorige hoofdstuk en inderdaad. Achter de schermen speelt het Eulergetal een heel belangrijke rol in de meetkunde. Volgens de beroemde stelling van Gauss-Bonnet is namelijk de totale kromming van het oppervlak gelijk aan 2π maal het Eulergetal. Met totale kromming bedoelen we de som van de kromming in ieder hoekpunt.

Stelling 6.3.3. (*Gauss-Bonnet*):

Voor ieder oppervlak geldt:

$$\sum_p \kappa_p = 2\pi\chi$$

waarbij de som over alle punten van het oppervlak loopt.

Dit is een verbluffend resultaat. Het laat bijvoorbeeld zien dat de totale kromming niet verandert wanneer we de metriek veranderen. Zelfs niet wanneer we een totaal andere triangulatie gebruiken. De totale kromming is blijkbaar net als het Eulergetal een *topologisch* begrip. De stelling van Gauss-Bonnet is een speciaal geval van de belangrijke *index-stelling* van Atiyah-Singer [La04]. Gelukkig is de stelling van Gauss-Bonnet in onze getrianguleerde context veel eenvoudiger te bewijzen.

Bewijs van de Stelling van Gauss-Bonnet.

Voor het gemak zullen we de stelling bewijzen in het speciale geval dat het oppervlak geen rand heeft. We gaan op een handige manier de krommingen van alle punten optellen en zullen dan vanzelf het Eulergetal zien verschijnen.

Om te beginnen levert ieder punt een factor 2π op, dus krijgen we $2\pi V$. (V is het aantal punten). Verder wordt iedere hoek precies een keer afgetrokken. Verzamel de hoeken nu per driehoek waar ze bij horen dan vinden we $-\pi F$. De som van de hoeken van een driehoek is immers π . Het linkerlid is dus gelijk aan $2\pi(V - F/2)$.

Om het bewijs af te maken hoeven we alleen nog te laten zien dat $V - F/2 = V - E + F = \chi$. Iedere driehoek heeft drie zijden en elke zijde wordt gedeeld wordt door twee driehoeken, dus $3F = 2E$. We kunnen dit herschrijven tot $-F/2 = -E + F$, ofwel $V - F/2 = V - E + F$. **Einde bewijs.**

Opgave

Kun je het bewijs uitbreiden naar het geval met rand? Let op de aangepaste definitie van kromming in de randpunten.

6.3.4 Constante kromming: de optimale meetkunde

Na deze inleiding in de meetkunde gaan we terug naar de vragen van Perelman. Gegeven een getrianguleerd oppervlak. Wat is dan de mooiste mogelijke meetkunde (metriek) die we hierop kunnen leggen? Wat is de eenvoudigste? Maar wat bedoelen we eigenlijk met eenvoudig en mooi?

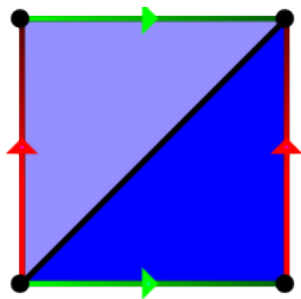
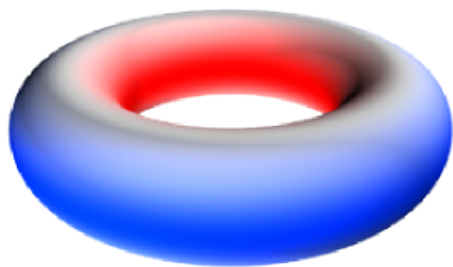
Een goed criterium hiervoor is kromming. Dat bepaalt immers voor een groot deel hoe het oppervlak er uit ziet. Als de kromming in een punt nul is, is het oppervlak daar in de buurt plat. Misschien kunnen we de kromming wel overal 0 maken zodat het oppervlak overal mooi plat is? Dit lijkt een goed idee maar de stelling van Gauss-Bonnet (stelling 6.3.3) maakt dit onmogelijk. De totale kromming is immers altijd $2\pi\chi$ ongeacht de metriek. Als χ niet nul is zoals bij de bol dan moeten er voor iedere metriek altijd een punt zijn met positieve kromming.

Wat we wel kunnen proberen is de kromming zo gelijkmatig mogelijk te verdelen. Dat past ook goed bij de intuïtie dat in de optimale meetkunde ieder punt er hetzelfde uitziet. Dat is waarom de ronde bol zo mooi is: perfecte symmetrie. Met driehoeken wordt het natuurlijk niet helemaal perfect rond maar met een paar miljoen driehoeken begint het er aardig op te lijken.

Hoe zit het met de andere oppervlakken die we hebben leren kennen? Hoe ziet hun optimale vorm eruit? Voor het gemak denken we hier aan een triangulatie met heel veel driehoeken zodat we het oppervlak kunnen tekenen als een glad vlies. Neem bijvoorbeeld de torus. Het gebruikelijke plaatje is verre van optimaal: de punten in de binnenring hebben een negatieve kromming (het is daar net een zadel) terwijl aan de buitenring het meer lijkt op een bol, de kromming is hier positief (zie figuur 6.17). Maar wat is dan wel de optimale torus?

Het Eulergetal geeft ons een hint: de totale kromming is 0, dus als we elk punt hetzelfde willen laten zijn dan moet de kromming overal 0 zijn. Kromming 0 betekent plat dus we zoeken een volstrekt platte torus. Dat is moeilijk voor te stellen maar het dichtst bij komt de bouwplaat zelf: Een vierkantje uit het platte vlak waarvan de tegenoverliggende zijden aan elkaar zouden moeten zitten. De kromming in de punten op de rand van het vierkant lijkt misschien meer dan 0 maar dat komt omdat hier juist twee (of vier) randpunten aan elkaar vast zitten. De hoeken moeten we dan optellen en zo komt de kromming overal netjes op 0.

Het in zijn geheel zien van de optimale meetkundige vorm van de meeste oppervlakken is helaas geen optie. Onze ruimte $\mathbb{R}^3$ is er te klein voor, ze



Figuur 6.17: Links de gebruikelijke, niet optimale torus. De kromming is aangegeven in kleur, blauw is positief, rood negatief. Rechts de optimale platte bouwplaat van de torus.

passen er niet in. Wat we wel goed kunnen zien zijn de optimale bouwplaten. Als het oppervlak er toch overal hetzelfde uitziet is het misschien niet eens zo erg om het hier en daar open te snijden zodat we het beter kunnen zien. Zo kwamen we aan het platte vierkantje voor de torus.

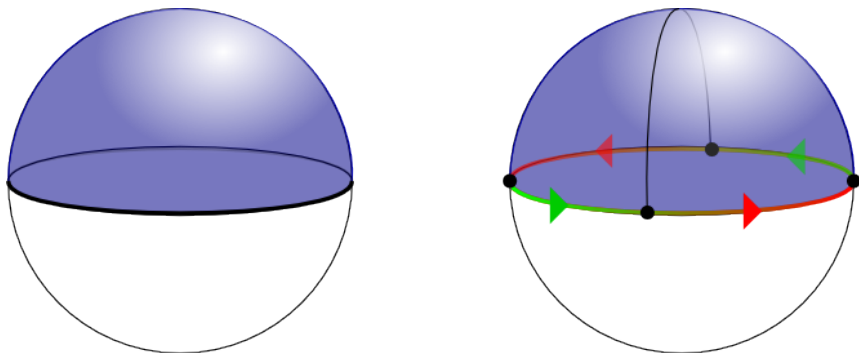
De fles van Klein, de ring en de Möbiusband hebben ook allemaal Eulergetal 0 dus de al bekende platte bouwplaten zijn optimaal.

Voor de schijf en het projectieve vlak ligt de situatie een beetje subtieler: het Eulergetal is voor beide gelijk aan 1 dus volgens onze logica moeten we ze dus overal positief laten krommen, net als de bol. In het geval van de schijf zijn we dus op zoek naar een stukje bol. Alle inwendige punten hebben dan namelijk de goede kromming.

De vraag is alleen welk deel van de bol we het beste kunnen kiezen. Ook hier heeft Gauss-Bonnet een antwoord op. Het Eulergetal van de bol is precies de helft van het Eulergetal van de schijf (zie figuur 6.7). Volgens Gauss-Bonnet is de totale kromming van de optimale schijf dus ook precies de helft van de totale kromming van de bol. Als de kromming gelijk verdeeld is, moeten we dus wel de halve bol kiezen.

Er is alleen wel iets speciaals aan de hand met de randpunten van onze schijf. Daar is de kromming precies half zo groot als in de rest van de punten. De optimale kromming is dus overal constant maar net de helft op de rand. Dat klopt mooi met onze intuïtie dat de totale kromming van het geheel de som is van de totale krommingen van de delen.

Het projectieve vlak is wat lastiger om een optimale bouwplaat voor te vin-



Figuur 6.18: Links de optimale schijf. Rechts de optimale bouwplaat voor het projectieve vlak.

den want tot nu toe hadden we er nog helemaal geen bouwplaat van. Het enige wat we wisten was dat de punten van het projectieve vlak uit paren tegenoverliggende punten op de bol bestaan. Kiezen we van ieder puntenpaar het punt op het noordelijk halfrond dan krijgen we een bouwplaat in de vorm van de optimale schijf. Alleen moeten we op de evenaar de tegenoverliggende punten nog aan elkaar plakken, zie figuur 6.18 (rechts). Met rode pijlen hebben we aangegeven hoe het plakken moet gebeuren. We merken op dat de kromming in de punten op de evenaar aan de rand van de bouwplaat precies de helft is van de kromming van de andere punten. Gelukkig plakken we aan ieder randpunt een ander randpunt vast zodat de kromming inderdaad overal constant en dus optimaal wordt.

En hoe zit het met de oppervlakken met negatieve Eulergetallen? Bijvoorbeeld een bol met drie gaten. Die passen noch in het vlak noch op de bol want in hun optimale meetkunde is de kromming steeds negatief. Is er wel een ruimte waar we bouwplaten mooi op neer kunnen leggen? Het antwoord is het *hyperbolische vlak*. Het is de tegenhanger van de bol en is juist op ieder punt negatief gekromd, als een zadel.

Zo krijgt ieder oppervlak hoe ingewikkeld ook zijn optimale vorm binnen een van de drie fundamentele meetkundes: Bolmeetkunde voor positieve Eulergetallen (positieve kromming), de vlakke meetkunde voor Eulergetal 0 en de wat minder bekende *hyperbolische meetkunde* voor $\chi < 0$. Dit resultaat staat wel bekend als de *Uniformisatiestelling*. Poincaré en Felix Klein (van de fles) bewezen hem onafhankelijk rond 1882, voordat Poincaré zich in met driedimensionale ruimten bezig hield en zijn beroemde

vermoeden uitte.

Het werk van Perelman draait om een 3-dimensionale versie van de uniformisatiestelling. Iedere 3-ruimte krijgt ook een optimale vorm. We komen daar straks op terug.

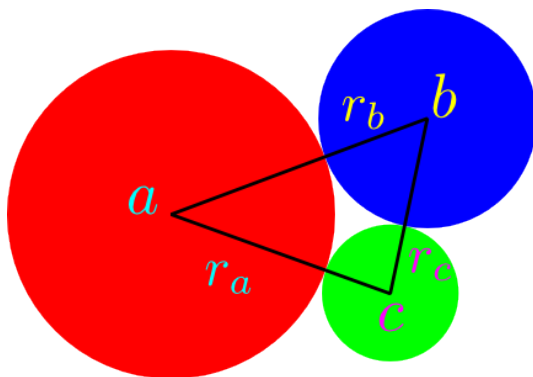
Opgave

Probeer een oppervlak in $\mathbb{R}^3$ te schetsen dat *overall* negatieve kromming heeft.

6.3.5 Cirkelmetrieken

Vanaf nu beperken we ons voor het gemak tot een speciaal soort metrieken, namelijk de zogenaamde *cirkelmetrieken* (ook wel conforme metrieken). Een gewone metriek is niets meer dan het geven van de lengte van iedere lijn van je triangulatie. Bij een cirkelmetriek kennen we juist aan ieder punt van de triangulatie een positief getal (de straal) toe. De lengte van de lijnen zelf is dan gegeven door de som van de stralen, zie figuur 6.19

Definitie 6.3.4. Een cirkelmetriek op een getrianguleerd oppervlak is een functie $r : V \rightarrow \mathbb{R}_{>0}$. Dus aan ieder punt p kennen we een getal r_p toe. En de lengte van zijde pq wordt gegeven door $r_p + r_q$.



Figuur 6.19: Een driehoek gevormd door de cirkelmetriek. De lengten van de zijden zijn $r_a + r_b$, $r_b + r_c$ en $r_c + r_a$.

Cirkelmetrieken zijn veel bijzonderder dan gewone metrieken. Voor cirkelmetrieken hoeven we bijvoorbeeld de driehoeksongelijkheden niet apart te

controleren. Ze volgen automatisch uit de constructie. Dat maakt cirkelmetrieken makkelijker om mee te werken want alle driehoeken zijn altijd automatisch nette rechte Euclidische driehoeken.

Ons hele verhaal draait om de volgende discrete versie van de uniformisatiestelling die we in de vorige paragraaf tegenkwamen. De stelling laat zien dat er optimale cirkelmetrieken bestaan.

Stelling 6.3.5. (*Discrete uniformisatiestelling*) *Voor iedere triangulatie van een oppervlak is er een (op schaling na) unieke cirkelmetriek met constante kromming.*

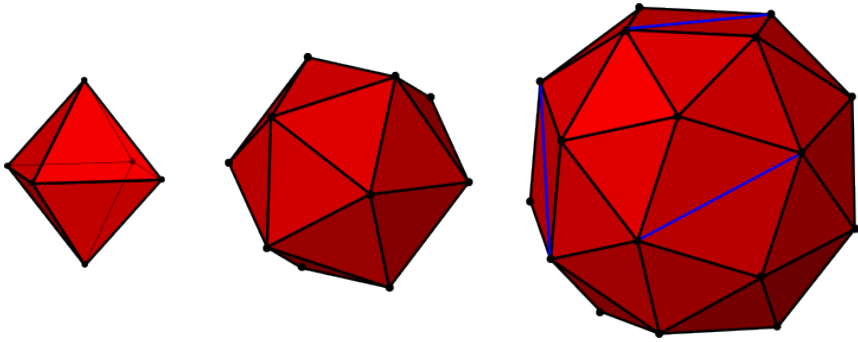
Met schaling bedoelen we dat we alle stralen met dezelfde factor mogen vermenigvuldigen. Er verandert dan immers niets aan de hoeken en dus aan de kromming. Heeft ons oppervlak een rand, dan bedoelen we met de term constante kromming dat de kromming in de randpunten half zo groot is als de kromming in de inwendige punten.

De regelmatige veelvlakken geven eenvoudige voorbeelden van de stelling in het geval van de bol. Ze hebben duidelijk constante kromming door hun symmetrie. De regelmatige octaëder en de icosaeëder in figuur 6.20 zijn inderdaad de cirkelmetrieken, waarbij alle stralen even groot zijn. De afgeknotte kubus rechts is daarentegen wel een voorbeeld van een metriek met constante kromming (weer door symmetrie) maar geen cirkelmetriek. Dat komt omdat we de vierkante vlakken met blauwe lijnen in tweeën moesten verdelen om er een triangulatie van te maken. De lengten van de blauwe lijnen kunnen echter niet komen van een cirkel-metriek. Omdat alle andere zijden dezelfde lengte hebben moeten alle stralen namelijk gelijk zijn, maar dan moeten de blauwe lijnen even lang zijn als de zwarte en dat is niet zo.

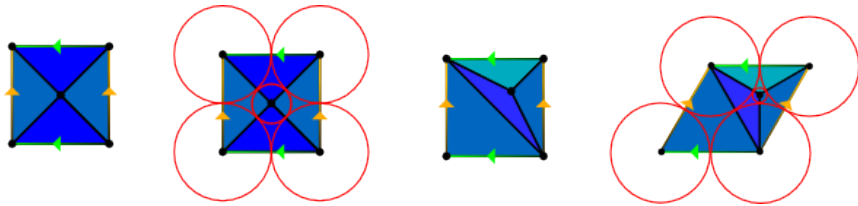
Op de torus heet de cirkelmetriek met constante kromming ook vaak een *circle packing*. Als we rond punt p de cirkel met straal r_p tekenen raken de cirkels elkaar namelijk precies. Het patroon waarin de cirkels elkaar raken wordt gegeven door de triangulatie.

Merk op dat er in beide triangulaties maar twee punten zijn. De vier hoekpunten van het vierkant stellen namelijk hetzelfde punt op de torus voor. Daarom zijn de cirkels in deze vier punten ook even groot. We zouden ze eigenlijk als kwart-cirkels moeten tekenen want samen vormen ze een grote cirkel op de torus.

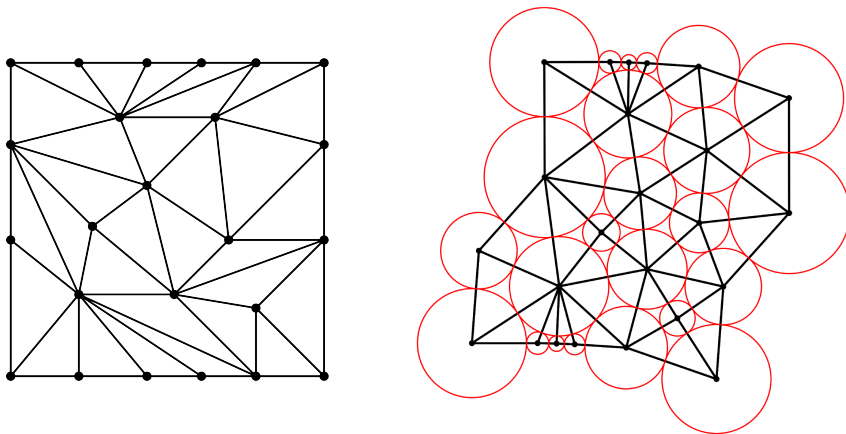
Iets ingewikkelder is het voorbeeld in figuur 6.22, berekend met de Ricci-flow uit het volgende hoofdstuk.



Figuur 6.20: Drie metrieken met constante kromming op de bol. De rechter is geen cirkelmetriek want de lengte van de blauwe zijden zijn onmogelijk.



Figuur 6.21: Twee triangulaties van de torus met er naast de unieke cirkelmetriek met constante kromming 0. De rode cirkels geven de stralen aan.



Figuur 6.22: Een triangulatie van de torus met de bijbehorende kromming 0 cirkelmetriek. Tegenoverliggende cirkels aan de rand van het vierkant zijn even groot.

In het volgende hoofdstuk zullen we een bewijs van stelling 6.3.5 schetsen. Dit hoofdstuk sluiten we af met een korte blik in de drie-dimensionale meetkunde.

Opgave

Zijn er tetraëders in $\mathbb{R}^3$ waarvan de lengten van de zijden nooit van een cirkelmetriek kunnen komen? Wat is de relatie met vier elkaar paarsgewijs rakende bollen? Bereken de optimale cirkelmetriek voor de triangulatie van de schijf met één inwendig punt, drie punten op de rand en drie driehoeken (een tetraëder zonder grondvlak).

6.3.6 Meetkunde in drie dimensies

Ook voor 3-ruimten kunnen we metrieken invoeren, weer door de lengten van de zijden van de tetraëders vast te leggen. Het begrip kromming is hier alleen veel subtieler. Er zijn namelijk veel meer manieren om krom te zijn. Hoe hoger de dimensie, hoe meer getallen er nodig zijn om de kromming in een enkel punt weer te geven. Een manier om de kromming van een 3-ruimte te onderzoeken in een punt p is als volgt. We bekijken alle mogelijke

deeloppervlakken die door zijvlakken van de tetraëders rond p gevormd worden. Ieder oppervlak heeft een kromming die we kunnen berekenen. Al deze krommingen samen geven een idee van de driedimensionale kromming in het punt p .

De uniformisatiestelling voor oppervlakken zegt dat we van ieder oppervlak een bouwplaat maken die constante kromming heeft. Afhankelijk van het Eulergetal χ moeten we hiervoor de bolmeetkunde gebruiken ($\chi > 0$), de vlakke Euclidische meetkunde ($\chi = 0$) of de hyperbolische meetkunde ($\chi < 0$).

Rond 1980 formuleerde de Amerikaanse wiskundige William Thurston een equivalente stelling voor 3-ruimten, het zogenaamde *geometrisatievermoeden*. Naast de driedimensionale bolmeetkunde, Euclidische en hyperbolische meetkunde had Thurston nog vijf andere meetkundes nodig.

De laatste vijf zijn nodig omdat het niet meer altijd mogelijk is om de driedimensionale kromming overal constant te krijgen. Thurstons oplossing was om in die gevallen een voorkeursrichting te kiezen en te eisen dat de kromming van alle oppervlakken loodrecht daarop constant was. Dit fenomeen komt bijvoorbeeld voor bij de 3-ruimten die we maakten uit oppervlakken door prisma's op de driehoeken te leggen. Die prisma's geven zo'n voorkeursrichting en loodrecht daarop bepaalt het oorspronkelijke oppervlak de kromming.

Thurstons geometrisatievermoeden was dat iedere 3-ruimte een bouwplaat had die precies zo kromt als een van de acht optimale meetkonden voorschrijft. Deze bouwplaat stelt dan de optimale standaardvorm van de 3-ruimte voor, zijn blauwdruk. Net als de uniformisatiestelling geeft het een ruwe classificatie van alle 3-ruimten in termen van meetkunde. Het grote voordeel is dat we nu met meetkunde kunnen onderzoeken wat de mogelijkheden zijn voor de bouwplaten.

Uit het geometrisatievermoeden volgt dat het Poincarévermoeden waar is. Hoe dat precies gaat voert te ver maar het komt er op neer dat we ruimten die sterk lijken op de 3-bol volgens het geometrisatievermoeden ook een bouwplaat in de bolmeetkunde moeten hebben. Omdat bolmeetkunde relatief eenvoudig is, zijn er dan geen andere mogelijkheden meer dan de 3-bol zelf.

De grote doorbraak van Perelman was om Thurstons geometrisatievermoeden in zijn geheel te bewijzen. Zijn bewijs is erg gecompliceerd maar in het volgende hoofdstuk zullen we de hoofdpunten van zijn techniek aan de hand van oppervlakken illustreren.

Opgave

Om wat meer gevoel te krijgen voor de meetkunde van de 3-bol en de 3-torus stel je voor dat je in een ruimteschip binnenin een van deze ruimten zou rondlopen/vliegen. Als we aannemen dat licht de kortste weg neemt, wat zou je dan zien? De 3-torus kunnen we maken met de optimale meetkunde door de tegenoverliggende zijden van een kubusvormige kamer aan elkaar te plakken. De optimale meetkunde op de 3-bol komt van de eenheidsbol in $\mathbb{R}^4$.

6.4 Perelmans Ricci-flow

In dit hoofdstuk geven we een inleiding op de Ricci-flow. Dit is de techniek waarmee Perelman het geometrisatievermoeden en daarmee het Poincaré-vermoeden de baas kon. Het woord *flow* slaat op een stroming, een proces waarin we de metriek op onze ruimte geleidelijk laten veranderen. Ricci is de naam van een Italiaanse tijdgenoot van Poincaré die onderzoek deed naar kromming.

Zoals in de vorige hoofdstukken, concentreren we ons hier op oppervlakken en in het bijzonder cirkelmetrieken. Met behulp van de Ricci-flow zullen we laten zien hoe je een willekeurige cirkelmetriek kunt laten veranderen in de unieke cirkelmetriek van constante kromming. Dit geeft een bewijs van de discrete uniformisatiestelling (stelling 6.3.5).

Tot slot zullen we wat opmerkingen maken over de problemen in het drie dimensionale geval spelen en noemen we een aantal onverwachte toepassingen van Ricci-flow.

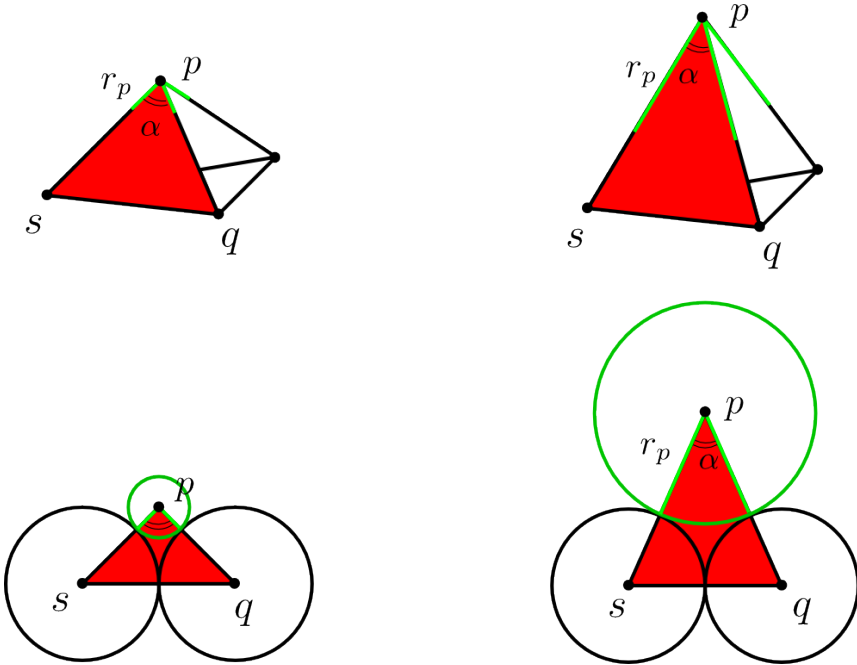
6.4.1 Ricci-flow op oppervlakken

Het basisidee van Ricci-flow is heel eenvoudig. Stel we hebben een getrianguleerd oppervlak en we zoeken er de cirkelmetriek bij die een zo constant mogelijke kromming heeft. Om te beginnen kiezen we een *willekeurige* cirkelmetriek³ en passen die dan stapje voor stapje aan zo dat de kromming steeds beter verdeeld raakt.

Om de kromming overal gelijk te maken aan het gemiddelde gaan we als

³bijvoorbeeld alle stralen gelijk aan 1.

volgt te werk. Is de kromming in een punt p te groot? Dan passen we de straal r_p in het punt p wat aan. Maken we de straal groter, dan wordt ook de kromming κ_p groter en maken we de straal kleiner dan neemt juist de kromming ook af in p . Intuïtief volgt dit al uit figuur 6.16. Een scherpe punt heeft een grote kromming en lange zijden. Maken we de zijden wat korter dan neemt de kromming af.



Figuur 6.23: Vergroten van de straal r_p in punt p verkleint de hoek α en vergroot dus de kromming κ_p in p .

Omdat dit het fundament van de Ricci-flow is gaan we nog wat preciezer na waarom dit zo werkt. De kromming is 2π min de som van de omliggende hoeken. In figuur 6.23 zien we waarom vergroten van de straal in een punt p juist de hoeken rond p verkleint en zo de kromming vergroot. We kunnen deze observatie ook nagaan door met de cosinusregel de hoek α in de figuur uit te rekenen in termen van de stralen. Differentiëren naar r_p geeft dan inderdaad iets negatiefs.

Helaas heeft het veranderen van de straal in p niet alleen effect op de krom-

ming in p maar ook op de kromming in de naburige punten. Wat gebeurt er met die krommingen? Zij gedragen zich precies tegengesteld. Dat moet ook want anders zou de totale kromming niet behouden kunnen blijven onder deze verandering van metriek (Gauss-Bonnet). Ook dit kunnen we vinden door stug door te rekenen met de cosinusregel.

Samenvattend zien we de kromming in een punt p nu als een functie $\kappa_p(r_p, r_q, r_s, \dots)$ van de stralen van p en de omliggende punten $q, s, \dots$. Uit de figuur blijkt dat de kromming heel eenvoudig reageert op veranderingen van de stralen. Dit kunnen we samenvatten in de volgende ongelijkheden

$$\frac{\partial \kappa_p}{\partial r_p} > 0 \quad \text{en} \quad \frac{\partial \kappa_q}{\partial r_p} < 0 \quad (6.1)$$

Uit het behoud van de totale kromming (Gauss-Bonnet) volgt dat er precies even veel kromming bij de buurpunten komt als er afgaat in het punt p als we r_p verkleinen.

Hoewel dit een goed begin is, is het nog niet direct duidelijk dat we zo de kromming ook daadwerkelijk gelijk kunnen verdelen. Neem bijvoorbeeld een tetraëder met twee heel scherpe punten p en q van gelijke kromming. Verkleinen we de kromming in p door de straal r_p te verkleinen, dan vergroten we de kromming κ_q juist. Natuurlijk wordt de kromming in de stompe punten ook iets groter.

In dit eenvoudige geval kunnen we dit oplossen door dan gewoon de κ_q weer kleiner te maken, zo klein als hij eerst was. Dit heeft tot gevolg dat κ_p weer groter wordt maar niet zo groot als hij was want er is wat kromming ‘weggelekt’ naar de andere twee punten.

Toch is het nog steeds niet helemaal duidelijk dat we de kromming ook overall exact gelijk kunnen krijgen op deze manier. Waarom blijft het bijvoorbeeld niet eindeloos op en neer gaan? In het volgende hoofdstuk geven we daarom een wat gestructureerdere versie van dit proces waarbij we dit wel kunnen bewijzen.

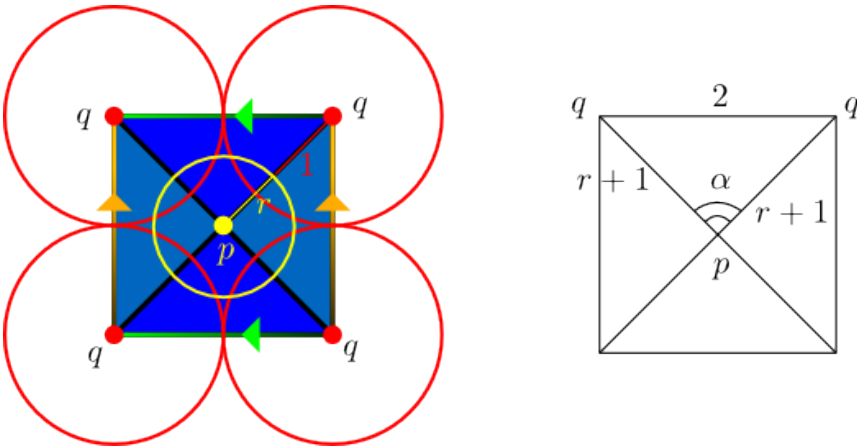
Opgave

Vereenvoudig het voorbeeld van de tetraëder hierboven zo dat de stralen in twee punten gelijk zijn aan x en de stralen in de twee andere punten gelijk zijn aan y . Beredeneer hoe we in dit geval stap voor stap x en y dichter bij elkaar krijgen zo dat de kromming constant wordt.

6.4.2 Differentiaalvergelijkingen

Om meer greep te krijgen op het veranderen van de stralen en de verschillende krommingen schreef Perelman de cirkelmetriek als een functie van de tijd. Voor ieder punt p hebben we dus een bijbehorende straal $r_p(t)$ op tijdstip t . De kromming in p berekenen we steeds uit de stralen dus die wordt zo ook een functie van t . We schrijven $\kappa_p(t)$. In termen van deze functies kon Perelman heel precies voorschrijven hoe de functies moesten veranderen. Deze regels zijn te schrijven in de vorm van een stelsel differentiaalvergelijkingen dat de *Ricci-flow* heet.

We beginnen met een eenvoudig voorbeeld, de triangulatie van de torus met twee punten uit figuur 6.21.



Figuur 6.24: Ricci-flow op de torus. De straal van het middelste punt p (geel) noemen we r , de straal van het andere punt q (rood) zetten we stil op 1.

Voor het gemak stellen we de straal van het punt q gelijk aan 1. Dat kan, want alles is toch bepaald op schalen na. De straal r_p in het inwendige punt p noemen we in de figuur voor het gemak r . Zoals blijkt uit figuur 6.24 is de kromming in het punt p gelijk aan

$$\kappa_p(t) = 2\pi - 4 \arccos \left(1 - \frac{2}{(1 + r_p(t))^2} \right)$$

de hoeken rond p zijn namelijk alle vier gelijk aan α en die vinden we uit

de cosinusregel.

Begin nu met een willekeurige cirkelmetriek op tijdstip $t = 0$. In ons voorbeeld is $r_p(0) \approx 0.8$, getekend als de gele cirkel. Het Ricci-flow proces houdt nu steeds bij hoe ver de kromming op tijdstip t weg is van de gewenste constante kromming κ_G . Het verschil geven we aan met $\bar{\kappa}_p(t) = \kappa_p(t) - \kappa_G$. In ons voorbeeld van de torus is $\kappa_G = 0$ dus $\bar{\kappa}_p(t) = \kappa_p(t)$.

We passen de stralen $r_p(t)$ aan zo dat $\bar{\kappa}_p(t)$ steeds kleiner wordt en hopelijk naar 0 gaat voor t naar oneindig. Dat doen we zoals eerder: Als $\bar{\kappa}_p(t)$ positief is dan moet $r_p(t)$ afnemen dus $\frac{dr_p(t)}{dt} < 0$. Is $\bar{\kappa}_p(t)$ negatief, dan moet de afgeleide juist positief zijn. Dit drukken we uit met behulp van de volgende differentiaalvergelijking (discrete Ricci flow):

$$\frac{dr_p(t)}{dt} = -r_p(t)\bar{\kappa}_p(t) \quad (6.2)$$

De Ricci flow is niets anders dan een precisering van onze eerdere wensen. Om beter te zien hoe dit werkt laten we de tijd in stapjes Δt lopen. Dan kunnen we de afgeleide benaderen met het differentiequotient $\frac{r_p(t+\Delta t) - r_p(t)}{\Delta t}$ en krijgen we:

$$r_p(t + \Delta t) = r_p(t)(1 - \bar{\kappa}_p(t)\Delta t)$$

Dus de straal een tijdje later, $r_p(t + \Delta t)$ is de straal nu, $r_p(t)$ vermenigvuldigd met $(1 - \bar{\kappa}_p(t)\Delta t)$. De factor $(1 - \bar{\kappa}_p(t)\Delta t)$ reguleert of de nieuwe straal groter of kleiner zal worden. Is $\bar{\kappa}_p(t)$ positief dan is de kromming te groot en weten we moet r_p kleiner worden. Dat klopt want dan is de factor juist kleiner dan 1. Anders is de factor juist groter dan 1.

Anders dan in ons intuïtieve voorbeeld met de tetraëder veranderen we volgens de Ricci-flow alle stralen tegelijk. Voor ieder punt p hebben we dezelfde vergelijking (6.2) (de precieze formule voor de kromming is voor iedere p net anders). De Ricci-flow zorgt er op deze manier voor dat we een straal nooit te veel tegelijk aanpassen. We gaan dus recht op ons doel (constante kromming) af zoals inderdaad blijkt uit de volgende stelling.

Stelling 6.4.1. *Voor iedere cirkelmetriek ρ is er een unieke oplossing $r(t)$ van de Ricci-flow vergelijkingen met $r(0) = \rho$. Deze oplossing convergeert naar de unieke cirkelmetriek met constante kromming.*

Aan de hand van ons eenvoudigste geval op de torus schetsen we het bewijs van deze stelling.

Hiervoor gebruiken we de volgende substitutie. We gaan over op de variabele $u = \ln r_p$ dus $\frac{1}{r_p} \frac{dr_p}{dt} = \frac{du}{dt}$. Als we de notatie $k(u) = \kappa_p(e^u)$ gebruiken gaat de Ricci-flow vergelijking (6.2) over in:

$$\frac{du}{dt} = -k(u)$$

Hier kunnen we al veel aan zien maar we maken het nog wat mooier door een primitieve $K(u) = \int_0^u k(v)dv$ in te voeren, dus $K' = k$. De Ricci-flow ziet er dan uit als:

$$\frac{du}{dt} = -\frac{dK}{du} \quad (6.3)$$

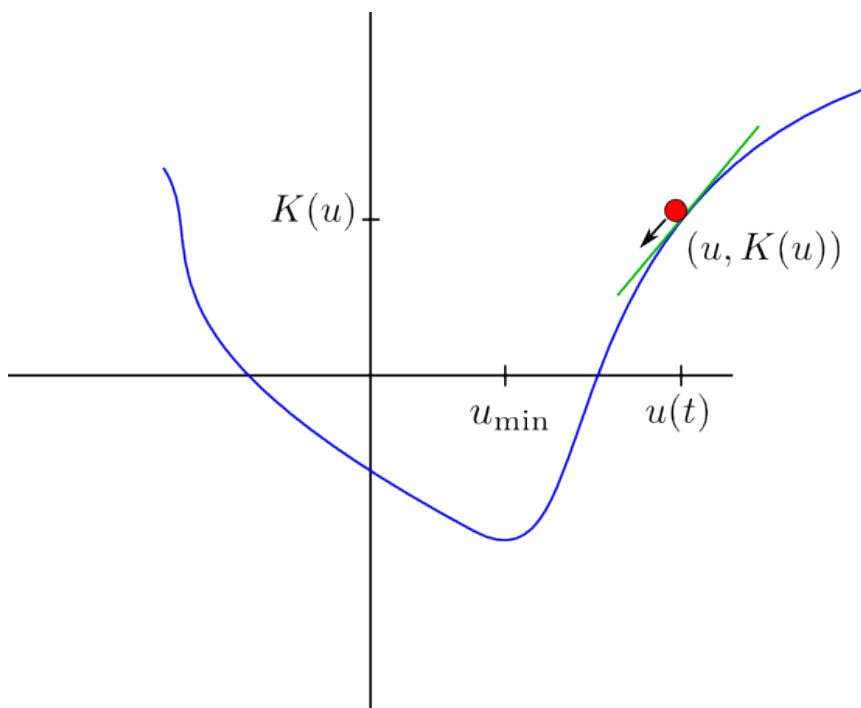
Dit lijkt een nietszeggende verandering maar de vorm van deze vergelijking is heel bijzonder. Het is een zogenaamde *gradient-stroming*. Dat betekent dat we de vergelijking als volgt kunnen lezen. We stellen ons $u(t)$ voor als de projectie op de horizontale as van een balletje dat over de grafiek van $K(u)$ rolt, zoals in figuur 6.25. De positie van het balletje is dus $(u, K(u))$ en u verandert in de tijd volgens de differentiaalvergelijking (6.3). In het linkerlid staat namelijk dat de horizontale snelheid van het balletje $\frac{du}{dt}$ gegeven wordt door de helling $-\frac{dK}{du}$ van de grafiek van K op dat punt. Het minteken zorgt er voor dat het balletje daadwerkelijk omlaag rolt.

Maar als het balletje dan omlaag rolt, waar rolt het dan heen? We weten dat $K' = k$ en $k(u) = \kappa(e^u)$. Bovendien was de afgeleide van $\kappa(r)$ naar r positief, zie vergelijking (6.1). Volgens de kettingregel is $K''(u)$ daarom altijd groter dan 0. Dat betekent dat de grafiek dalvormig is met een uniek minimum bij $u = u_{\min}$. Volgen we het balletje dus maar lang genoeg dan ligt het uiteindelijk op het minimum en daar blijft het stabiel liggen. Zo weten we dus zeker dat $\lim_{t \rightarrow \infty} u(t) = u_{\min}$.

Vertalen we dit terug naar de gewone straal $r = e^u$ dan komen we tot de volgende conclusie: Ongeacht de beginstraal $r(0)$ convergeert de Ricci-flow $r(t)$ naar de waarde $r_{\min} = e^{u_{\min}}$. $r_{\min}$ is bovendien de optimale cirkelmetriek, dus met $\bar{\kappa}(r_{\min}) = 0$. Dat komt omdat als we de definities teruglezen we zien dat:

$$\kappa(r_{\min}) = k(u_{\min}) = K'(u_{\min}) = 0$$

De laatste formule is 0 omdat $u_{\min}$ het minimum van K is. In dit voorbeeld is $r_{\min} = \sqrt{2} - 1$.



Figuur 6.25: De grafiek van K gezien als functie van u . Het balletje rolt langs de grafiek naar beneden met snelheid $-K'(u)$.

Precies dezelfde redenering werkt ook als we meer punten (en dus variabelen) hebben. Voor ieder punt p gaan we dan over op de nieuwe variabele $u_p \ln r_p$. Net als eerder stellen we $k_p(u_p) = \bar{\kappa}_p(e^{u_p})$ en gaat de Ricci-flow vergelijking over in:

$$\frac{du_p}{dt} = -k_p(u_p)$$

Nu gaat de truc van de primitieve werkelijk een rol spelen. We kiezen namelijk één enkele functie $K(u_p, u_q, \dots)$ die afhangt van alle variabelen $u_p, u_q, \dots$ zo dat $\frac{\partial K}{\partial u_p} = k_p$. Net als eerder kunnen we K weer schrijven als een (wat ingewikkeldere) integraal.

De Ricci-flow gaat nu over in de vergelijking

$$\frac{du_p}{dt} = -\frac{dK}{du_p}$$

Hier staat weer de gradient-stroom vergelijking maar nu in meerdere variabelen. We kunnen deze vergelijking nog steeds interpreteren als het rollen van een balletje over de grafiek van de functie $K(u_p, u_q, \dots)$. De functies $u_p(t)$ geven dan de coördinaten aan van de horizontale positie van het balletje.

Door te differentiëren kunnen we weer inzien dat de grafiek van K een dalvorm heeft. Zo concluderen we dat het balletje weer naar het unieke minimum van K rolt en is de stelling bewezen. De Ricci-flow convergeert dus naar de unieke cirkelmetriek waar $\bar{\kappa}_p = 0$ ofwel $\kappa_p = \kappa_G$, de kromming is constant.

Hier en daar hebben we wat details weggelaten maar de hoofdlijn van het bewijs hopelijk duidelijk geworden. Het volledige bewijs is te vinden in [CL03], stelling 1.2.

Opgave

Teken triangulatie op de torus met drie punten en vind hier de optimale cirkelmetriek bij met behulp van de Ricci-flow. Stel een van de stralen in de cirkel metriek gelijk aan 1. De twee andere stralen zij variabelen r en s . Schrijf de kromming in de twee punten in termen van r en s . Ga over op de variabelen $u = \ln r$ en $v = \ln s$ en vind een primitieve voor de krommingsfuncties.

6.4.3 Drie dimensies

In drie dimensies is het nog steeds mogelijk om de vergelijkingen van de Ricci-flow op te schrijven maar er werkelijk iets over zeggen vereist heel veel nieuwe wiskundige techniek. Een groot probleem is dat er geen equivalent van het Eulergetal is die ons vertelt welke kromming we precies kunnen verwachten. Het was van groot belang om in de vergelijkingen van de vorige paragraaf te werken met de afwijking van de kromming van de gemiddelde kromming $\bar{\kappa}_p = \kappa_p - \kappa_G$. Maar wat als we niet weten wat κ_G is? Probeer bijvoorbeeld de Ricci-flow met de verkeerde κ_G op de tetraëder. Nemen we $\kappa_G = 0$ dan proberen we de positief gekromde tetraëder plat te maken maar volgens Gauss-Bonnet kan dat nooit. De tetraëder zal dus terugvechten en zorgen dat de Ricci-flow spaak loopt. Als je het probeert zul je zien dat alle stralen gelijk aan 0 worden. Positieve kromming geeft namelijk kleinere stralen.

In drie dimensies is het zelfs nog erger, verschillende delen van de 3-ruimte kunnen neigen naar een verschillende meetkunde. De optimale bouwplaat bestaat dan uit een aantal losse stukken. Dat openbaart zich dan als het opblazen van de Ricci-flow langs een bepaald oppervlak in de 3-ruimte. Daarlangs moet de 3-ruimte vervolgens worden opengesneden om de Ricci-flow op ieder afzonderlijk stuk opnieuw te starten.

Perelman bewees onder andere dat de Ricci-flow maar eindig vaak opblaast zodat we eindig veel meetkundige stukken overhouden.

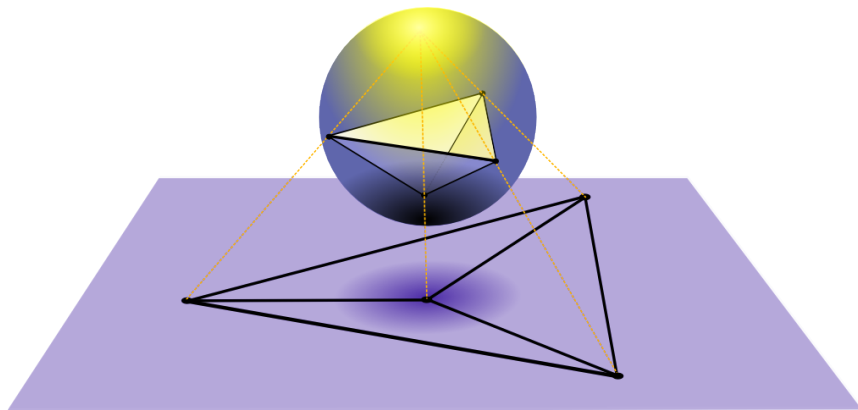
6.4.4 Toepassingen

Tot slot bespreken we kort een aantal eenvoudigere toepassingen van Perelmans Ricci-flow. Belangrijk hierbij is dat onze discrete Ricci-flow niet alleen convergeert naar de optimale metriek, maar dat de convergentie ook uitzonderlijk snel gaat. Daardoor is de techniek populair geworden in de numerieke wiskunde en computer graphics.

Zelfs voor de schijf, het eenvoudigste oppervlak, zijn er interessante toepassingen mogelijk. Zo zegt de stelling van Fáry bijvoorbeeld dat als een graaf in het vlak getekend kan worden, we dit kunnen doen met alleen maar rechte lijnen. Met Ricci-flow kunnen we zo'n tekening eenvoudig vinden. Hieronder beschrijven we kort hoe dat in zijn werk gaat.

Door wat lijntjes toe te voegen maken we van de graaf een triangulatie van de schijf. Op deze triangulatie laten we Ricci-flow los en zo vinden

we de cirkelmetriek met constante (positieve) kromming. De driehoeken kunnen we in $\mathbb{R}^3$ tekenen als rechte Euclidische driehoeken en vormen een mooie benadering van het zuidelijk halfrond van een ronde bol. Projecteren vanuit de noordpool zoals in figuur 6.26 geeft de gezochte triangulatie met rechte lijnen in het vlak. De oorspronkelijke graaf is daar onderdeel van.



Figuur 6.26: Met de stereografische projectie wordt de graaf op het zuidelijk halfrond van de bol geprojecteerd op het vlak.

Een ingewikkeldere stelling langs dezelfde lijnen is de *afbeeldingsstelling van Riemann*. Deze beroemde stelling uit de complexe functietheorie zegt dat ieder gebied in het vlak hoekbehoudend op de eenheidsschijf kan worden afgebeeld. Met andere woorden, we kunnen een één op één verband vinden tussen de punten in het gebied en alle punten in de cirkelschijf. Met hoekbehoudend bedoelen we dat de hoek tussen twee lijnen in het gebied gelijk is aan de hoek tussen de corresponderende kromme lijnen in de schijf. Weer geeft de Ricci-flow een snelle benadering voor deze stelling. Dat gaat als volgt.

Kies een gebied en betegel het vlak er in en er omheen met kleine gelijkzijdige driehoekjes van dezelfde grootte in regelmatig honingraatpatroon. Alle driehoekjes die binnen het gebied vallen vormen samen een oppervlak dat ons gebied benadert. Topologisch gezien is dit oppervlak natuurlijk al een schijf, maar met Ricci-flow kunnen we die schijf in optimale vorm brengen zoals hierboven. De halve bol die we zo krijgen projecteren we weer vanuit de noordpool op het vlak zoals in figuur 6.26. Zo krijgen we een goede benadering van de ronde schijf.

Wat we hier mee bereikt hebben is een één op één verband tussen de gelijkzijdige driehoekjes in het gebied en de driehoekjes in de ronde schijf. Hoe kleiner de driehoekjes waar je mee begint, hoe beter dit gaat lijken op Riemanns afbeelding. Een bewijs van de convergentie en de hoekbehoudendheid heeft wat meer voeten in de aarde. Een goed leesbaar bewijs hiervan is te vinden in het project van Joran van Apeldoorn en Nastasha Wijers, 2e jaars wiskunde aan de Universiteit van Amsterdam [AW13].

Soortgelijke resultaten kunnen natuurlijk ook voor andere oppervlakken worden gevonden en steeds leveren de cirkelmetrieke prachtige cirkelpatronen. De theorie staat daarom ook wel bekend onder de naam *circle-packings*. Zie bijvoorbeeld [St05] en ook de bijbehorende website (met software) www.circlepack.com.

Een andere erg mooie toepassing is het vinden van discrete minimale oppervlakken [Se07]. Hierbij speelt de Ricci-flow voor de bol een grote rol in de vorm van de discrete Gauss-afbeelding. Voor de theorie van discrete differentiaalmeetkunde verwijzen we tot slot naar [BS08].

6.5 Samenvatting

We hebben een overzicht gegeven van hoe topologie, meetkunde en differentiaalvergelijkingen nauw verweven zijn. De topologie vraagt naar de classificatie van alle mogelijke ruimten. Voor de eenvoudige oppervlakken hebben we gezien hoe dit in zijn werk gaat. Hier is een mooie classificatiestelling mogelijk, voornamelijk in termen van het Eulergetal. Voor de iets ingewikkeldere 3-ruimten is de classificatie meteen een ontzettend moeilijk probleem. Het Poincarévermoeden is hier een klein onderdeel van.

De doorbraak van Grigori Perelman was niet zo zeer alleen het bewijzen van het Poincarévermoeden als wel het veel omvangrijkere geometrisatievermoeden. Dit vermoeden hebben we aan de hand van oppervlakken leren kennen als de uniformisatiestelling: ieder oppervlak heeft een optimale vorm (bouwplaat). Dat wil zeggen een bouwplaat met constante kromming. Perelman bewees dat voor 3-ruimten iets dergelijks geldt. Dit geeft gedeeltelijke oplossing van het classificatieprobleem voor 3-ruimten want als we weten dat ze een optimale meetkundige vorm dragen hoeven we alleen nog maar de mogelijke meetkundige vormen in kaart te brengen. Dit is aanmerkelijk makkelijker en leverde meteen de bevestiging van Poincarés vermoeden.

De brug tussen topologie en meetkunde wordt voor oppervlakken gesla-

gen door de stelling van Gauss-Bonnet. Het centrale meetkundige begrip (kromming) en het centrale topologische begrip (Eulergetal) worden hier direct aan elkaar gerelateerd. De totale kromming van het oppervlak is gelijk aan het Eulergetal. Ook in drie dimensies staan topologie en meetkunde allerm minst los van elkaar maar het verband is moeilijker precies te maken. Het in kaart brengen van de precieze relatie tussen topologie en meetkunde van 3-ruimten blijft ook na Perelmans bewijs een actief en veelzijdig onderzoeksgebied.

In het laatste hoofdstuk hebben we een indruk gegeven van de werkelijke bewijstechniek van Perelman: de Ricci-flow. Met dit stelsel differentiaalvergelijkingen geeft hij letterlijk de ruimte vorm en brengt hij hem in zijn optimale gedaante. Voor oppervlakken konden we dit mooi illustreren met de volgende stelling: Iedere cirkelmetriek convergeert met de (discrete) Ricci-flow naar de optimale metriek van constante kromming. Hoewel de theorie van Perelman technisch gezien veel moeilijker is, zijn zo toch alle hoofdzaken aan bod gekomen.

Als bonus vonden we dat Ricci-flow en discrete meetkunde ook voor oppervlakken onverwachte toepassingen en inzichten oplevert. De mooie cirkelpatronen en veelvlakken geven zo een hint van hoe bijzonder de theorie in drie dimensies is.

Bibliografie

- [AW13] J. van Apeldoorn, N. Wijers, *Cirkels naar cirkels: discrete functietheorie en cirkelpakkingen*, project 2e jaar wiskunde Universiteit van Amsterdam, 2013. <http://www.rolandvdv.nl/Files/Cirkels.pdf>
- [BS08] A. Bobenko, Y. Suris, *Discrete Differential Geometry*, GSM 98, American Mathematical Society, 2008.
- [CE10] J. Carstens, S. Eliens, *TopWis Poincaré*, de Praktijk 2010. <http://www.praktijk.nu/lesmateriaal/42/topwis-poincare.html>
- [CL03] B. Chow, F. Luo, *Combinatorial Ricci Flows on Surfaces*, Journal of Differential Geometry, 63 (2003) 97-129.
- [FW99] G. Francis, J. Weeks *Conway's ZIP Proof*, American Mathematical Monthly, 106 5 (1999) 393-399.
- [La04] K. Landsman, *De indexstelling van Atiyah en Singer*, Nieuw Archief voor de wiskunde, 3 (2004) 207-211.

- [Se07] S. Sechelmann *Discrete Minimal Surfaces, Koebe Polyhedra, and Alexandrov's Theorem. Variational Principles, Algorithms, and Implementation*. <http://page.math.tu-berlin.de/~sechel/DiplomaThesis.pdf>
- [St05] K. Stephenson, *Introduction to circle packing*, Cambridge University Press, 2005.
- [We01] J. Weeks, *The shape of space*, Chapman & Hall, 2e druk, 2001.
- [BVCK] A. van den Brandhof, R. van der Veen, J. van de Craats, B. Koren, *De zeven grootste raadsels van de wiskunde*, Bert Bakker 2012.

7 Opgaven bij Het vormen van ruimte: van Poincaré tot Perelman

Roland van der Veen

Universiteit van Amsterdam

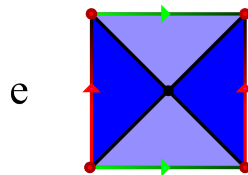
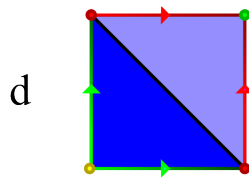
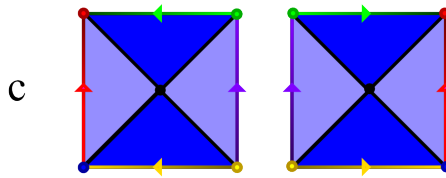
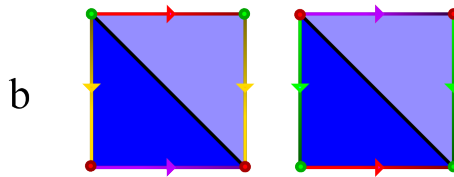
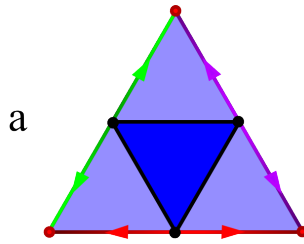
7.1 Topologie

Poincaré dacht na over ruimtelijke objecten, *oppervlakken*, die worden gevormd door driehoeken aan elkaar te plakken. Hierbij ging het hem niet om de precieze details van hoeken en lengten maar om de globale vorm, de *topologie*. Welke oppervlakken lijken op elkaar en welke helemaal niet en hoe bewijst je dat?

Om zulke vragen te beantwoorden gebruikte Poincaré het zogenaamde *Eulergetal*. Het Eulergetal χ (chi) kun je eenvoudig berekenen door het aantal punten V (*vertices*), het aantal zijden E (*edges*) en het aantal driehoeken F (*faces*) te tellen. Het Eulergetal is dan $\chi = V - E + F$.

7.1.1 Opgave bouwplaten

Hieronder staan bouwplaten voor vijf oppervlakken. Iedere bouwplaat bestaat uit een aantal driehoeken waarvan de zijden met dezelfde kleur volgens de pijltjes aan elkaar vast moeten zitten. De punten die dezelfde kleur hebben komen ook op elkaar terecht en vormen dus samen één hoekpunt. Bouwplaat b bestaat bijvoorbeeld uit twee losse stukken van twee driehoeken elk die volgens de instructies uiteindelijk allemaal aan elkaar komen te zitten. Uiteindelijk blijven er twee punten en zes zijden (ribben) over.



Figuur 7.1: Bouwplaten van oppervlakken.

- a. Bereken van elk van de vijf oppervlakken het Eulergetal.
- b. Schets zo goed mogelijk hoe het figuur er uit komt te zien wanneer we de pijltjes daadwerkelijk aan elkaar vast zouden plakken. Hierbij zullen de rechte lijnen en vlakken noodzakelijk vervormd moeten worden. Dat is geen probleem, dat is nu juist topologie!
- c. Zie je een verband tussen het Eulergetal en de globale vorm van je schetsen?

7.2 Meetkunde

Perelman pakte het probleem van het classificeren van oppervlakken heel anders aan: Hij probeerde ze **in een zo regelmatig mogelijke standaardvorm te brengen**. De vorm van een oppervlak wordt bepaald door iedere lijn een specifieke lengte te geven, dit heet ook wel een (discrete) *metriek*. In figuur 7.3 is bij elke bouwplaat uit figuur 7.1 een metriek gegeven. Voor het gemak hebben we eenvoudige getallen gekozen.

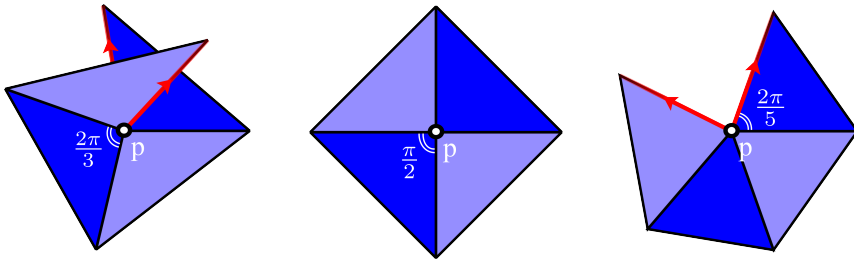
Er zijn een hoop metrieken mogelijk, maar welke is nu de mooiste? Welke metriek is het regelmatigste?

Om hier een antwoord op te vinden gebruikte Perelman het begrip *kromming*. De kromming is een getal dat in ieder hoekpunt de meetkundige vorm van het oppervlak beschrijft. De kromming in een punt p (notatie κ_p) wordt berekend uit de som van de hoeken rond dat punt. Preciezer gezegd is de kromming gedefinieerd door $\kappa_p = 2\pi - \text{hoeksom}$.

7.2.1 Opgave kromming

Hieronder staan drie bouwplaten voor een stukje van een oppervlak. De hoeken van de driehoeken zijn in radialen weergegeven.

- a. Geef voor elk van de drie gevallen aan of de kromming positief, negatief of 0 is in het punt p .
- b. Schets ook hoe het oppervlak ruimtelijk uit zou komen te zien als de rode zijden in de bouwplaat aan elkaar zouden plakken.
- c. Op welke plaatsen op een zwemband denk je dat de kromming positief is? En waar is die juist negatief?



Figuur 7.2: Drie typische mogelijkheden voor de kromming in een punt p .

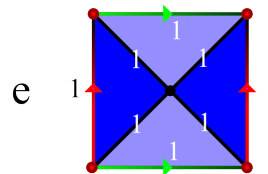
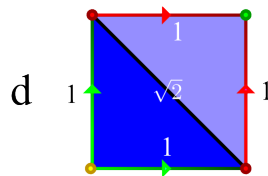
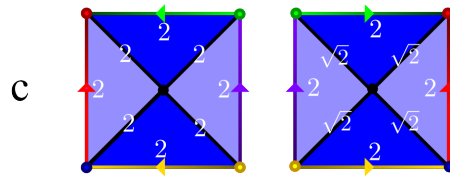
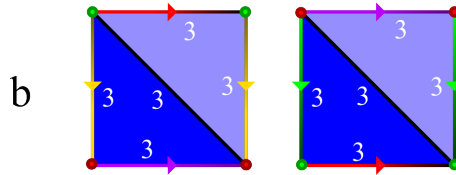
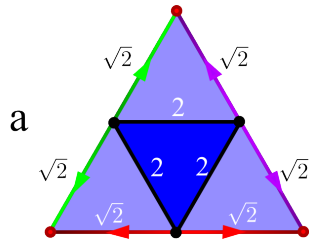
7.2.2 Opgave totale kromming

- Bereken met behulp van de cosinusregel de kromming in ieder punt van de bouwplaten in figuur 7.3. Pas op, op de rand van de bouwplaat komen vaak meerdere punten bij elkaar (die hebben dezelfde kleur) bovendien zijn de bouwplaten niet op schaal getekend. Alle hoeken die bij deze punten samenkomen vormen samen de hoeksom voor één enkel punt.
- Waarom verandert de kromming niet wanneer we de lengten van alle zijden met een positief getal vermenigvuldigen (schalen)?
- Bereken steeds ook de *totale kromming*, de som van de krommingen in alle punten van het oppervlak.
- Zie je een verband met opgave 7.1.1?

7.2.3 Opgave de optimale metriek

De mooiste mogelijke vorm, de optimale metriek van een oppervlak is volgens Perelman die metriek waar de kromming overal constant is. Hij bewees dat deze metriek op schalen na uniek is.

- Ga na dat de metriek die de vier zijden in het midden van bouwplaat e (figuur 7.3) lengte $\sqrt{2}$ geeft en de rest lengte 2, de constante kromming metriek is.
- Zoek ook voor de andere bouwplaten van de bouwplaten uit de vorige opgave de optimale metriek.



Figuur 7.3: De lengte (metriek) van iedere zijde is er bij geschreven.

Na de pauze gaan we dieper in op de vraag hoe je deze metrieken kunt vinden.

7.3 Differentiaalvergelijkingen

Perelman slaagde er in om het Poincarévermoeden op te lossen door iedere ruimte, hoe vreemd ook, in zijn optimale vorm te brengen. Voor ons betekent de optimale vorm de metriek met constante kromming. Perelmans belangrijkste gereedschap was de *Ricci-flow*, een stelsel differentiaalvergelijkingen voor de metriek.

7.3.1 Opgave cirkelmetrieken

Om de Ricci-flow te beschrijven beperken we ons tot een speciaal soort metrieken: *cirkelmetrieken*. Aan ieder punt kennen we een positief getal toe: de straal. De lengte van een zijde wordt nu bepaald door de som van de stralen van begin- en eindpunt te nemen. (Ook als begin en eindpunt gelijk zijn).

- Welke van de metrieken uit figuur 7.3 komt van een cirkelmetriek?
- Wat zijn de bijbehorende stralen?

7.3.2 Opgave straal en kromming

De Ricci-flow is gebaseerd op de antwoorden op volgende vragen:

- Hoe verandert de kromming in een punt als we de straal in dat punt wijzigen?
- Hoe verandert de kromming in een nabijgelegen punt?

Met de cosinusregel kunnen we de kromming in ieder punt altijd expliciet uitdrukken in termen van de stralen (in het onderstaande geval kan het wat eenvoudiger). Om dit te doen voor bouwplaat e in figuur 7.1 noemen we de punten p en q . Het punt p is het punt midden in de bouwplaat. De vier punten op de hoeken van de bouwplaat worden samengeplakt tot het enkele punt q . Er komen dus acht hoeken samen rond q en vier rond p .

- Schrijf de krommingen $\kappa_p(r_p, r_q)$ en $\kappa_q(r_p, r_q)$ als functies van de stralen r_p en r_q .

7.3.3 Opgave Ricci-flow

Het idee van de *Ricci-flow* is om met een willekeurige cirkelmetriek te beginnen en die vervolgens stapje voor stapje aan te passen tot de kromming overall constant is. Voor de pauze hebben we gezien dat de totale kromming niet afhangt van de gekozen metriek. Delen we de totale kromming door het aantal punten dan krijgen we de gemiddelde kromming, notatie $\bar{\kappa}$. Het doel is dus om de kromming in ieder punt gelijk aan $\bar{\kappa}$ te maken.

Om de aanpassingen van de stralen goed bij te houden schrijven we de straal in ieder punt als een functie van de tijd, $r_p(t)$. Op basis van de ideeën uit de vorige opgave stelde Perelman de volgende Ricci-flow vergelijkingen voor om de stralen aan te passen. De Ricci-flow is het volgende stelsel differentiaalvergelijkingen (één vergelijking voor ieder punt p):

$$\frac{dr_p(t)}{dt} = -r_p(t)(\kappa_p(t) - \bar{\kappa})$$

- Laat de tijd in stapjes Δt gaan en vervang de afgeleide in de bovenstaande formule door het differentiequotient.
- Beredeneer hiermee dat de Ricci-flow in iedere stap de kromming gunstig aanpast.
- Schrijf de twee Ricci-flow vergelijkingen op voor de bouwplaat e uit figuur 7.1. Wat is $\bar{\kappa}$ in dit geval?

7.3.4 Opgave convergentie van de Ricci-flow

We gaan dieper in op de Ricci-flow in het geval van bouwplaat e in figuur 7.1. Om het onszelf gemakkelijker te maken schalen we de metrieken zo dat de straal in het punt q steeds gelijk is aan 1. Kies verder een willekeurige beginstraal in het andere punt p . Bijvoorbeeld ook gelijk aan 1. Dus $r_p(0) = 1$.

Het doel van deze opgave is om te laten zien dat de oplossing $r_p(t)$ van de Ricci-flow vergelijking in het punt p convergeert naar de cirkelmetriek met constante kromming 0. De optimale standaardvorm dus. Volgens opgave 7.2.3a betekent dat in ons geval dat

$$\lim_{t \rightarrow \infty} r_p(t) = \sqrt{2} - 1$$

- a. Schrijf de Ricci-flow vergelijking voor het punt p op.
- b. Pas de substitutie $r_p(t) = e^{u(t)}$ toe en herschrijf de vergelijking in de vorm $\frac{du(t)}{dt} = \dots$
- c. Gebruik opgave 7.3.2a (of de formule) om in te zien dat $-\kappa_p(r_p)$ monotoon dalend is in r_p . Leg uit waarom $-\kappa_p(e^{u(t)})$ dus ook dalend is in $u(t)$.
- d. Vind getallen u_1 en u_2 zo dat $-\kappa_p(e^{u_1}) > 0$ en $-\kappa_p(e^{u_2}) < 0$. Leid nu uit deel c af dat $-\kappa_p(e^u)$ een uniek nulpunt u_0 heeft.
- e. Ga na dat de oplossing van de vergelijking uit deel b de eigenschap heeft dat als $-\kappa_p(e^{u(t)})$ positief is, de $u(t)$ toeneemt en dat $u(t)$ afneemt als $-\kappa_p(e^{u(t)})$ negatief is.
- f. Concludeer uit deel e dat de oplossing $u(t)$ convergeert naar het nulpunt u_0 uit deel d.
- g. Concludeer tot slot dat $r_p(t) = e^{u(t)}$ daarom convergeert naar e^{u_0} en dat dit inderdaad de straal is zo dat de kromming in p gelijk is aan 0.
- h. Waarom is de kromming in q dan automatisch ook 0?



Voor wie is PWN interessant?

Beroepswiskundigen

Wiskundeleraren

Bedrijven

Leerlingen en studenten

Breed publiek

Platform Wiskunde Nederland is hét landelijke loket voor alles wat met wiskunde te maken heeft.

PWN behartigt de belangen van, en fungeert als spreekbuis voor, de gehele Nederlandse wiskunde.

Platform Wiskunde Nederland | Science Park 123 | kamer L013 | 1098 XG Amsterdam | 020 592 40 06

Ga voor meer informatie naar:
www.platformwiskunde.nl



platform
wiskunde nederland