Snelle algoritmen in de getaltheorie

Bas Edixhoven
Universiteit Leiden
email: edix@math.leidenuniv.nl

Samenvatting

De auteur van deze tekst is bijzonder blij met de resultaten die hij en zijn medewerkers recentelijk hebben behaald in [Ed-Co], [Bruin] en [Varma], op het gebied van snelle algoritmen in de getaltheorie. Er heeft hier een doorbraak plaatsgevonden die bepaalde objecten toegankelijk heeft gemaakt voor berekeningen. Deze objecten komen uit de Galoistheorie en zijn van groot belang in de getaltheorie. Ze spelen bijvoorbeeld een belangrijke rol in Wiles' bewijs van de Laatste Stelling van Fermat. Men kan deze objecten uitrekenen met computeralgebra (symbolische rekenmachines), maar dat leidt tot exponentiële groei in de rekentijd. Rekenen met benaderingen met voldoend grote precisie leidt tot algoritmen waarvan de rekentijd slechts polynomiaal groeit.

Het doel van deze tekst is de objecten waarom het gaat zo eenvoudig mogelijk te beschrijven, en het geheel in een historische context te plaatsen door te beginnen met een toepassing op het volgende klassieke probleem. Hoe snel kan men uitrekenen op hoeveel manieren een geheel getal n geschreven kan worden als som van d kwadraten van gehele getallen? Beroemde wiskundigen zoals onder andere Diophantus, Fermat, Legendre, Gauss, Jacobi, Eisenstein, en Liouville hebben voor d gelijk aan 2, 4, 6, 8 of 10 formules gegeven waarmee deze aantallen snel berekend kunnen worden. Daarna is zonder succes geprobeerd dit uit te breiden naar grotere even getallen. Met hedendaagse kennis begrijpen we nu dat er dan geen dergelijke formules meer bestaan, maar dat de gezochte aantallen tòch snel kunnen worden uitgerekend.

De bijbehorende voordracht zal veel voorbeelden bevatten, waaronder ook computerberekeningen. Advies aan de lezer: deze tekst bevat vast teveel stof en het gaat over een moeilijk onderwerp, dus probeer niet teveel alles te begrijpen, maar zie het meer als een overdadig buffet en kies eruit wat u lekker vindt.

1 Een klassieke vraag

We beginnen met een klassieke vraag: op hoeveel manieren een geheel getal n geschreven kan worden als som van d kwadraten van gehele getallen. Voor $d = 1$ zijn we snel klaar: als n geen kwadraat is dan kan het niet (dus op 0 manieren), als n wel een kwadraat is (zeg van een geheel getal x) dan kan het op twee manieren ($n = x^2$ en $n = (-x)^2$) als $n \neq 0$ en anders op één manier.

Voor $d = 2$ kunnen we bijvoorbeeld opmerken dat 1 op 4 manieren geschreven kan worden als som van 2 kwadraten: $(\pm 1)^2 + 0^2$ en $0^2 + (\pm 1)^2$. Ook voor 2 is het aantal manieren 4: $(\pm 1)^2 + (\pm 1)^2$.

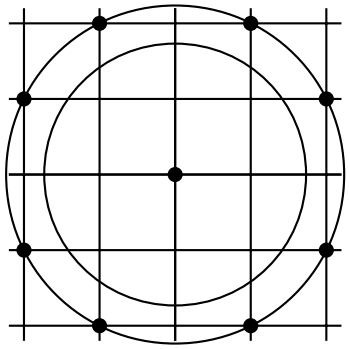
Om wat makkelijk over deze vraag te kunnen praten laten we $r_d(n)$ het gevraagde aantal zijn. In formuletaal:

$$r_d(n) := \#\{x \in \mathbb{Z}^d : x_1^2 + \cdots + x_d^2 = n\}. \tag{1.1}$$

De stelling van Pythagoras geeft ons ook een meetkundige interpretatie van de vraag naar $r_d(n)$:

> *voor $n \geq 0$ is $r_d(n)$ het aantal punten in $\mathbb{Z}^d$ op afstand $\sqrt{n}$ van de oorsprong, met andere woorden, het aantal punten van de doorsnede met $\mathbb{Z}^d$ van de bolschil in $\mathbb{R}^d$ om 0 met straal $\sqrt{n}$.*

In dimensie 2 is deze interpretatie makkelijk te tekenen. Bijvoorbeeld kan 5 op 8 manieren geschreven worden als som van 2 kwadraten: op 4 manieren als $(\pm 2)^2 + (\pm 1)^2$ en op nog eens 4 manieren als $(\pm 1)^2 + (\pm 2)^2$. Maar bijvoorbeeld kan 3 niet geschreven worden als som van 2 kwadraten, want de enige kwadraten kleiner of gelijk aan 3 zijn 0 en 1, en daar gaat het niet mee. Ook 6 en 7 zijn geen som van 2 kwadraten, maar 8 weer wel. In dimensie 3 kunnen we ons ook nog wel voorstellen hoe voor een paar kleine waarden van n de doorsnede van een bolschil van straal $\sqrt{n}$ om de oorsprong met het rooster $\mathbb{Z}^3$ eruit ziet, maar als n groter wordt, en nog meer als de dimensie groter wordt, dan schiet ons voorstellingsvermogen tekort en kan alleen rekenen het antwoord geven.



Figuur 1: $r_2(5) = 8$

2 Oude resultaten en formules in even dimensies $d \leq 12$

De geschiedenis van de vraag naar de getallen $r_d(n)$ gaat ver terug, en bekende wiskundigen hebben zich hiermee bezig gehouden (voor meer detail verwijzen we naar Hoofdstuk XX in [Ha-Wr]). Al in de 3e eeuw gaf Diophantus van Alexandrië de volgende identiteit[1]:

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \tag{2.1}$$


DIOPHANTI
ALEXANDRINI
ARITHMETICORVM
LIBRI SEX.
ET DE NVMERIS MVLTANGVLIS
LIBER VNVS.
Nunc primùm Græcè & Latinè editi, atque absolutissimis Commentariis illustrati.
AVCTORE CLAVDIO GASPARE BACHETO
MEZIRIACO SEBVSIANO, V. C.
LVTETIAE PARISIORVM,
Sumptibus Sebastiani Cramoisy, via Iacobæa, sub Ciconiis.
M. DC. XXI.
CVM PRIVILEGIO REGIS.

Uit deze identiteit volgt dat als n en m beide een som van twee kwadraten zijn, dan zo ook hun product. De vraag welke n dan de som van twee kwadraten zijn is daarmee teruggebracht tot priemgetallen. Wat experimenteren (bekijk bijvoorbeeld alle priemgetallen onder de 100) leidt al snel tot het vermoeden

[1] Vraag aan de lezer: ziet u een verband met complexe getallen?

dat een priemgetal p een som van 2 kwadraten is precies dan als p niet rest 3 geeft na deling door 4. Één van de twee implicaties is makkelijk: een kwadraat van een even getal is deelbaar door 4 (want $(2x)^2 = 4x^2$), en dat van een oneven getal geeft rest 1 na deling door 4 (want $(2x+1)^2 = 4x^2 + 4x + 1$), dus kan een som van twee kwadraten niet rest 3 geven na deling door 4.

De implicatie in de andere richting werd bewezen door Pierre de Fermat (jurist te Toulouse) in de 17e eeuw: ieder priemgetal dat rest 1 geeft na deling door 4 is een som van twee kwadraten.

Algemener bewees Fermat dat voor $n \geq 1$ geldt dat $r_2(n) \neq 0$ precies dan als iedere priemdeler p van n die rest 3 geeft na deling door 4 tot een even macht voorkomt in de ontbinding van n in priemfactoren.

Adrien-Marie Legendre gaf in 1798 een formule voor $r_2(2^a m^2)$. Carl Friedrich Gauss gaf in 1801 een algemene formule voor $r_2(n)$, en zelfs een formule voor $r_3(n)$, maar uit die laatste formule bleek al dat $r_d(n)$ voor oneven d meer gecompliceerd zijn dan voor even d. Vanaf nu bekijken we dan ook alleen nog maar even d.

Een *mooie* formule voor $r_2(n)$ werd in 1829 gegeven door Carl Gustav Jacob Jacobi, die voor $n > 0$ bewees:

$$r_2(n) = 4 \sum_{d|n} \chi(d), \quad \text{waar} \quad \chi(d) = \begin{cases} 0 \text{ als } d \text{ even is,} \\ 1 \text{ als } d = 4r+1, \\ -1 \text{ als } d = 4r+3. \end{cases} \tag{2.2}$$

De notatie $\sum_{d|n}$ betekent dat er gesommeerd wordt over de positieve delers d van n. Bijvoorbeeld zijn de delers van 3 de getallen 1 en 3, en krijgen we $r_2(3) = 4{\cdot}(\chi(1)+\chi(3)) = 4{\cdot}(1+(-1)) = 0$, hetgeen inderdaad klopt. Voor $n = 5$ vinden we: $r_2(5) = 4{\cdot}(\chi(1) + \chi(5)) = 4{\cdot}2 = 8$, wat natuurlijk ook klopt.

Voor $d = 4$ bewees Jacobi een nog mooiere formule:

$$r_4(n) = 8 \sum_{2\nmid d|n} d + 16 \sum_{2\nmid d|(n/2)} d, \tag{2.3}$$

Figuur 2: Jacobi

waarin $\sum_{2\nmid d|n}$ betekent dat er gesommeerd wordt over de oneven delers d van n, en waar we afspreken dat als $n/2$ niet geheel is dat getal ook geen delers heeft en de som $\sum_{2\nmid d|(n/2)} d$ nul is.

Laten we een paar voorbeelden bekijken. Het getal 1 kan op $2{\cdot}4$ manieren geschreven worden als som van 4 kwadraten, want $1 = (\pm1)^2 + 0^2 + 0^2 + 0^2$, en de (±1) kan op 4 verschillende plekken staan. En inderdaad geeft Jacobi's formule $r_4(1) = 8{\cdot}1 = 8$.

We hebben $r_4(2) = \binom{4}{2}{\cdot}2^2 = 6{\cdot}4 = 24$, want $2 = (\pm1)^2 + (\pm1)^2 + 0^2 + 0^2$, met $\binom{4}{2}$ mogelijkheden voor de plaatsen van de twee nullen, en 4 mogelijke tekens. Dat klopt precies met Jacobi's formule: $8 + 16 = 24$. De lezer wordt van harte aanbevolen zelf nog een paar voorbeelden uit te werken. Het is natuurlijk een wonder dat zulke formules bestaan. Verderop zal iets over dit wonder worden geopenbaard, maar eerst gaan we verder omhoog in dimensie.

Uit werk van Jacobi, Ferdinand Eisenstein en Henry Smith volgt dat:

$$\begin{aligned} r_6(n) &= 16\sum_{d|n}\chi(n/d)d^2 - 4\sum_{d|n}\chi(d)d^2, \\ r_8(n) &= 16\sum_{d|n} d^3 - 32\sum_{d|(n/2)} d^3 + 256\sum_{d|(n/4)} d^3. \end{aligned} \tag{2.4}$$

Ook hier wordt de lezer aangeraden een paar kleine gevallen zelf uit te werken.

In 1865 vond Joseph Liouville een formule voor $r_{10}(n)$, in termen van de Gaussische gehele getallen $\mathbb{Z}[i] := \{a + bi \in \mathbb{C} : a \text{ en } b \text{ in } \mathbb{Z}\}$. Deze deelverzameling van $\mathbb{C}$ is gesloten onder optellen en vermenigvuldigen. Voor $d = a + bi$ in $\mathbb{Z}[i]$ schrijven we $|d|$ voor de absolute waarde van d als complex getal, dus $|d|^2 = a^2 + b^2$. Dan luidt de formule van Liouville als volgt:

$$r_{10}(n) = \frac{4}{5}\sum_{d|n}\chi(d)d^4 + \frac{64}{5}\sum_{d|n}\chi(n/d)d^4 + \frac{8}{5}\sum_{d\in\mathbb{Z}[i],\,|d|^2=n} d^4. \tag{2.5}$$

De derde som in Liouville's formule is van een andere soort dan de eerste twee sommen. Voor elke term d^4 van de derde som geldt dat $|d^4| = |d|^4 = (|d|^2)^2 = n^2$, terwijl in de voorafgaande sommen de absolute waarden van de termen tussen 1 en n^4 variëren.

De volgende stap werd gezet door James Whitbread Lee Glaisher in 1907. Hij gaf formules voor $r_d(n)$ voor alle even d van 12 tot en met 18. We geven als voorbeeld de formule voor $d = 12$, zoals geïnterpreteerd door Srinivasa Ramanujan 1916:

$$r_{12}(n) = 8\sum_{d|n} d^5 - 512\sum_{d|(n/4)} d^5 + 16a_n, \tag{2.6}$$

waar:

$$\sum_{n\geq 1} a_n q^n = q\prod_{m\geq 1}(1-q^{2m})^{12} \quad \text{in } \mathbb{Z}[[q]]. \tag{2.7}$$

De notatie $\prod_{m\geq 1}$ betekent dat de factoren $1 - q^{2m}$ vermenigvuldigd moeten worden, over alle gehele $m \geq 1$, en "in $\mathbb{Z}[[q]]$" betekent dat dit product gezien moet worden als een machtreeks in de variabele q waarbij er niet op convergentie wordt gelet: het product wordt dan ook een formele machtreeks genoemd.

Om bijvoorbeeld a_1 tot en met a_5 uit te rekenen merken we op dat alle factoren $(1 - q^{2m})$ met $m > 2$ er niet toe doen, en is het genoeg om het product $q(1 - q^2)^{12}(1 - q^4)^{12}$ uit te werken, en alle termen met q^i waarin $i > 5$ is weg te laten:

$$q(1-q^2)^{12}(1-q^4)^{12} = q - 12q^3 + 54q^5 + O(q^7). \tag{2.8}$$

We zien dus dat $a_1 = 1$, $a_2 = 0$, $a_3 = -12$, $a_4 = 0$, en $a_5 = 54$. Het is duidelijk dat deze formule voor $r_{12}(n)$ weer van een andere soort is dan die voor lagere even d. We gaan nu uitleggen waarom we Glaishers formule echt minder waard vinden dan de voorgaande.

3 Algoritmisch perspectief op het klassieke probleem

In Sectie 2 hebben we formules gezien voor $r_d(n)$ voor even d van 2 tot 12. De vraag die we nu willen stellen is hoe nuttig deze formules zijn voor het *berekenen* van de getallen $r_d(n)$, waarbij we rekening houden met hoe lang die berekeningen dan duren.

We beginnen met het bekijken van het eenvoudigste geval: $d = 1$. In Sectie 1 merkten we al op dat $r_1(n) = 0$ als n geen kwadraat is, $r_1(0) = 1$ en $r_1(n) = 2$ als $n > 0$ en een kwadraat is. Laten we nu eens kijken hoe snel we kunnen uitrekenen of een geheel getal $n > 0$ (willekeurig groot, denk aan minstens honderden cijfers!) een kwadraat is, dat wil zeggen, of er een geheel getal m is zodat $m^2 = n$. Anders gezegd, we moeten beslissen of het reële getal $\sqrt{n}$ geheel is. Een manier om dit te doen is als volgt, waarbij we gebruiken dat de functie $x \mapsto x^2$ van $(0, \infty)$ naar $\mathbb{R}$ strikt stijgend is. We weten dat $0 \leq \sqrt{n} \leq n$. Laat $m = \lfloor n/2 \rfloor$, d.w.z., $n/2$ naar beneden afgerond. We berekenen m^2 en beslissen of $m^2 \geq n$ of $m^2 \leq n$. Als $m^2 \geq n$, dan weten we dat $0 \leq \sqrt{n} \leq m$, en anders dat $m \leq \sqrt{n} \leq n$, in beide gevallen is $\sqrt{m}$ in een interval van lengte hoogstens $(n + 1)/2$ en met gehele eindpunten. We herhalen deze stappen totdat we een interval $[a, b]$ hebben van lengte hoogstens 1 en met gehele eindpunten waarin $\sqrt{n}$ ligt. Dan berekenen we a^2 en b^2 en weten we of n een kwadraat is.

Hoeveel rekentijd kost ons dit? Om hier antwoord op te geven moeten we eerst kijken hoeveel tijd de gebruikte operaties ons kosten. Als we met de methode van de lagere school twee gehele getallen a en b optellen, dan kost ons dit hoogstens r elementaire optellingen die we uit ons hoofd kennen, waarbij r het maximum is van het aantal cijfers van a en b. Uitrekenen van het product ab kost hoogstens ongeveer r^2 elementaire operaties. De deling $\lfloor a/2 \rfloor$ kost hoogstens r operaties, en zo ook $a - b$ en het kijken of $a - b = 0$. Het aantal intervallen dat we construeren tijdens de berekening is hoogstens ongeveer $\log_2(n)$, want (op een klein beetje na) wordt de lengte iedere keer gehalveerd. Het aantal cijfers van de getallen die in onze berekeningen voorkomen is hoogstens $1 + \log_{10} n$. De uitkomst is dan dat de berekening hoogstens ongeveer $3(\log_2 n)(\log_{10} n)^2$ elementaire operaties kost. Iets minder nauwkeurig gezegd: de berekening kost hoogstens $c{\cdot}(\log n)^3$ tijd, voor één of ander getal c. Nog minder nauwkeurig gezegd: de rekentijd is hoogstens polynomiaal in $\log n$. Het is goed om over $\log n$ na te denken als het aantal cijfers van n (op de factor $1/\log 10$ na, dan).

Men zou kunnen denken dat het handig is om de ontbinding van n in priemfactoren te gebruiken, want daaruit volgt meteen of n een kwadraat is of niet. Jammer genoeg hebben de snelst bekende algoritmen om n te factoriseren een rekentijd die veel harder groeit dan polynomiaal in $\log n$ (namelijk $\exp(c{\cdot}(\log n)^{1/3})$); zie "General number field sieve" op Wikipedia. Dat maakt deze methode veel langzamer dan de voorgaande. Er is ook een positieve kant hieraan: in de cryptografie is het RSA protocol gebaseerd op het feit dat we niet zo snel kunnen factoriseren.

Laten we nu eens terugkijken naar de formules voor $r_2(n), \ldots, r_8(n)$. Daarin moet gesommeerd worden over de delers van n, en om dat te doen moeten we de factorisatie van n kennen. Het wordt dan interessanter om de gestelde vraag "Hoe snel kunnen we $r_d(n)$ uitrekenen?" iets aan te passen tot:

> *Hoe snel kunnen we $r_d(n)$ uitrekenen als de ontbinding van n in priemfactoren gegeven is?*

Het antwoord op deze vraag is bevredigend voor d in $\{2,4,6,8,10\}$, namelijk dat dat kan in tijd polynomiaal in $\log n$, maar daarvoor is het wel nodig de formules voor $r_d(n)$ iets anders op te schrijven, want het kost teveel tijd alle delers af te lopen. De functie χ uit (2.2) is multiplicatief: voor alle d_1 en d_2 in $\mathbb{Z}$ geldt $\chi(d_1 d_2) = \chi(d_1)\chi(d_2)$. Laat nu $p_1, \ldots, p_r$ de verschillende priemfactoren van n zijn, en $e_1, \ldots, e_r$ hun exponenten in de ontbinding van n. Dan geldt, bijvoorbeeld:

$$r_2(n) = 4 \sum_{d|n} \chi(d) = 4 \prod_{i=1}^{r} \left(\chi(1) + \chi(p_i) + \cdots + \chi(p_i^{e_i})\right). \tag{3.1}$$

Omdat voor alle i geldt dat $p_i \geq 2$ hebben we $2^r \leq n$, dus $r \leq \log_2 n$, en zo ook $2^{e_i} \leq n$, dus $e_i \leq \log_2 n$. Het berekenen van $r_2(n)$ kost een tijd die hoogstens polynomiaal is in $\log n$. Hetzelfde argument werkt voor d in $\{4,6,8\}$. Voor $d = 10$ kan men iets dergelijks doen met de ontbinding van n in priemfactoren in $\mathbb{Z}[i]$ (bijvoorbeeld $2 = (1+i)(1-i)$, $5 = (2+i)(2-i)$, 3 is priem). De ontbinding in $\mathbb{Z}[i]$ van een priemgetal p in $\mathbb{Z}$ kan berekend worden in tijd polynomiaal in $\log p$, maar daarvoor is wat algebra en elementaire getaltheorie nodig.

Daarentegen is het uitrekenen van de a_n in (2.7) door het uitwerken van het product een kostbare operatie. Om a_n uit te rekenen moet de coëfficiënt van q^n in $q \prod_{m=1}^{n/2} (1 - q^{2m})^{12}$ worden uitgerekend. In deze berekening mogen steeds alle termen waarin q^i met $i > n$ voorkomt weggelaten worden, maar dan moeten nog steeds polynomen van graad n vermenigvuldigd worden en dat betekent dat de tijd minstens lineair in n groeit, dus exponentieel in $\log n$. De conclusie is dat Glaishers formule ons niet in staat stelt om $r_{12}(n)$ uit te rekenen in tijd polynomiaal in $\log n$. Zelfs niet als de factorisatie van n gegeven is (maar dat doet er nu niet toe).

Tot slot van deze sectie merken we op dat het snel uitrekenen van $r_4(n)$ bijna neerkomt op het factoriseren van n. In ieder geval voor getallen $n = pq$ met p en q verschillende priemgetallen > 2 (de n die voor het RSA protocol gebruikt worden zijn van deze vorm), want dan geldt:

$$r_4(n) = 1 + p + q + n, \tag{3.2}$$

en als we $p+q$ en pq kennen, dan weten we hoe we p en q kunnen berekenen via de abc-formule, of via $(p-q)^2 = (p+q)^2 - 4pq$.

4 Een gevolg van het recente onderzoek

Ila Varma, een masterstudente in Leiden in 2009-2010, liet in haar scriptie [Varma] zien dat voor geen enkele even $d > 10$ er een "elementaire formule" is voor $r_d(n)$ zoals we hebben voor $d \leq 10$. Natuurlijk moet het begrip "elementaire formule" precies gedefinieerd worden voordat er sprake is van een wiskundige uitspraak, en een bewijs. Verderop zullen we hierover iets kunnen zeggen.

Aan de positieve kant laten [Ed-Co] en [Bruin] zien dat voor iedere even d de getallen $r_d(n)$ tòch uitgerekend kunnen worden in een tijd die hoogstens polynomiaal is in $\log n$, als de factorisatie van n is gegeven.

De conclusie is dan ook:

Vanuit algoritmisch perspectief is het klassieke probleem van het snel uitrekenen van $r_d(n)$ voor even d en n gegeven met ontbinding is priemfactoren opgelost voor alle *even d. De vraag naar* formules *heeft een negatief antwoord, maar voor* berekenen *maakt dat negatieve antwoord niet uit en hebben we nu een* positief *antwoord.*

De volgende sectie geeft wat inzicht in waar deze vooruitgang vandaan komt, welke wiskundige objecten hieraan bijdragen.

5 Tweedimensionale Galoisrepresentaties

Lieve lezer, schrik alsjeblieft niet teveel bij het lezen van de titel van deze sectie, en stop niet hier met lezen. Het doel is juist om in simpele termen uit te leggen wat deze objecten zijn, en dat zal niet moeilijker zijn dan complexe getallen en 2 bij 2 matrices. We moeten nu dieper inzoomen op de complexe getallen, omdat we die nodig hebben om symmetrie in de getaltheorie te kunnen zien.

We schrijven (zoals hierboven) $\mathbb{C}$ voor de verzameling van complexe getallen. We vatten $\mathbb{C}$ op als het reële vlak, met basis 1 en i. Elementen van $\mathbb{C}$ zijn dan uniek te schrijven als $a+bi$ met a en b in $\mathbb{R}$. De optelling geschiedt coördinaatsgewijs:

$$(a+bi)+(c+di)=(a+c)+(b+d)i, \tag{5.1}$$

en de vermenigvuldiging ligt vast door te eisen dat die lineair is over $\mathbb{R}$ in beide factoren en dat $i^2=-1$:

$$(a+bi)\cdot(c+di)=(ac-bd)+(ad+bc)i. \tag{5.2}$$

De absolute waarde van $a+bi$ is gedefinieerd als:

$$|a+bi|=\sqrt{a^2+b^2}, \tag{5.3}$$

en voor alle complexe getallen z en w geldt dan:

$$|z\cdot w|=|z|\cdot|w| \tag{5.4}$$

We kunnen in $\mathbb{C}$ ook delen (behalve door 0):

$$1/(a+bi)=(a-bi)/(a^2+b^2). \tag{5.5}$$

En complexe getallen kunnen in poolcoördinaten geschreven worden:

$$r\cdot e^{i\phi}=r\cos(\phi)+ir\sin(\phi) \tag{5.6}$$

De e-macht e^z van een complex getal is gegeven door de machtreeks:

$$e^z=\sum_{n\geq 0}\frac{z^n}{n!}=1+z+\frac{z^2}{2}+\frac{z^3}{6}+\cdots. \tag{5.7}$$

Deze machtreeks convergeert, dat wil zeggen, de rij van partiële sommen convergeert, waarbij de afstand tussen twee complexe getallen z en w gelijk is aan $|z-w|$. Voor alle z en w in $\mathbb{C}$ geldt dan dat $e^{z+w} = e^z e^w$, en daaruit volgt dan weer dat vermenigvuldiging van complexe getallen in poolcoördinaten neerkomt op het vermenigvuldigen van de absolute waarden en het optellen van de argumenten.

Voor de symmetrie in de getaltheorie moeten we het afstandsbegrip op $\mathbb{C}$ loslaten, en alleen naar algebraïsche eigenschappen kijken: optelling en vermenigvuldiging. We moeten dus kijken met de bril van de getaltheoreticus. De symmetrieën van $\mathbb{C}$ heten dan *automorfismen*.

Automorfismen van $\mathbb{C}$ zijn afbeeldingen $\sigma\colon \mathbb{C} \to \mathbb{C}$ die voldoen aan:

1. $\sigma(z+w) = \sigma(z) + \sigma(w)$,
2. $\sigma(z{\cdot}w) = \sigma(z){\cdot}\sigma(w)$,
3. σ heeft een inverse afbeelding.

Het volgt dan dat:

$$\sigma(0) = \sigma(0+0) = \sigma(0) + \sigma(0), \quad \text{dus } \sigma(0) = 0. \tag{5.8}$$

En ook:

$$\sigma(1) = \sigma(1{\cdot}1) = \sigma(1){\cdot}\sigma(1), \quad \text{dus } \sigma(1) = 1 \text{ of } \sigma(1) = 0, \text{ dus } \sigma(1) = 1, \tag{5.9}$$

want $\sigma(1)$ is verschillend van $\sigma(0)$.

We kennen twee automorfismen van $\mathbb{C}$: de identiteit $\mathrm{id}_{\mathbb{C}}\colon z \mapsto z$, en de complexe conjugatie: $\sigma(a+bi) = a - bi$, ook wel geschreven als $z \mapsto \overline{z}$. Deze twee automorfismen zijn de enige twee die continu zijn. Maar het is bekend (na 3 colleges algebra op de universiteit) dat de verzameling $\mathrm{Aut}(\mathbb{C})$ van automorfismen van $\mathbb{C}$ heel groot is (even groot als de verzameling van alle deelverzamelingen van $\mathbb{C}$). Toch kennen we van geen van al die andere automorfismen een expliciete beschrijving. Ook van de algebraïsche eigenschappen van bekende complexe getallen weten we niet veel. We weten dat e en π transcendent zijn: ze zijn geen nulpunt van een polynoom $x^n + a_{n-1}x^{n-1} + \cdots + a_o$ met alle a_i in $\mathbb{Q}$. Daaruit volgt dat er een $\sigma \in \mathrm{Aut}(\mathbb{C})$ is met $\sigma(e) = \pi$, en ook dat er een $\sigma \in \mathrm{Aut}(\mathbb{C})$ is met $\sigma(\pi) = e$. Maar we weten niet of e en π algebraïsch onafhankelijk zijn (of er een polynoom $f \neq 0$ in $\mathbb{Q}[x,y]$ is zodat $f(e,\pi) = 0$; iedereen verwacht van niet, maar niemand heeft een bewijs), en dus ook niet of er een $\sigma \in \mathrm{Aut}(\mathbb{C})$ is die e en π verwisselt.

Daarentegen is het opmerkelijk dat $\mathbb{R}$ maar één automorfisme heeft, de identiteit. Dit te bewijzen is een interessante opgave, die niet zo moeilijk is (laat $\sigma \in \mathrm{Aut}(\mathbb{R})$, bewijs dat voor alle $n \in \mathbb{Z}$ geldt dat $\sigma(n) = n$, dan dat voor alle x in $\mathbb{Q}$ geldt dat $\sigma(x) = x$, en dan dat σ de ordening op $\mathbb{R}$ vastlaat: als $x \leq y$ dan is er een z zodat $y = x + z^2$, en dan $\sigma(y) = \sigma(x) + \sigma(z)^2 \ldots$).

Elementen van $\mathrm{Aut}(\mathbb{C})$ kunnen worden samengesteld: voor σ en τ in $\mathrm{Aut}(\mathbb{C})$ is ook de samenstelling $(\sigma \circ \tau)\colon z \mapsto \sigma(\tau(z))$ in $\mathrm{Aut}(\mathbb{C})$. Tenslotte: voor alle p/q in $\mathbb{Q}$ geldt dat $\sigma(p/q) = p/q$. We zijn nu klaar voor de symmetrieën van de getaltheorie.

Laat $f = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ een veelterm zijn met alle a_i in $\mathbb{Q}$. Dan zegt de hoofdstelling van de algebra dat de vergelijking $f(z) = 0$ precies n oplossingen in $\mathbb{C}$, met multipliciteiten

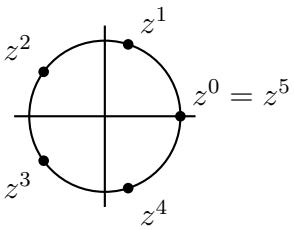
geteld. We schrijven $\mathrm{Wortels}(f)$ voor de verzameling wortels van f. Voor σ in $\mathrm{Aut}(\mathbb{C})$ en z in $\mathrm{Wortels}(f)$ geldt dan:

$$\begin{aligned} 0 = \sigma(0) = \sigma(f(z)) &= \sigma(z^n + \cdots + a_1 z + a_0) \\ &= \sigma(z^n) + \cdots \sigma(a_1 z) + \sigma(a_0) \\ &= \sigma(z)^n + \cdots + \sigma(a_1)\sigma(z) + \sigma(a_0) \\ &= \sigma(z)^n + \cdots + a_1\sigma(z) + a_0 = f(\sigma(z)), \end{aligned} \tag{5.10}$$

dus $\sigma(z)$ is in $\mathrm{Wortels}(f)$. De conclusie is dat iedere σ in $\mathrm{Aut}(\mathbb{C})$ de elementen van $\mathrm{Wortels}(f)$ permuteert. De *Galoisgroep van f* is dan gedefinieerd als de verzameling $\mathrm{Gal}(f)$ van permutaties van $\mathrm{Wortels}(f)$ gegeven door elementen van $\mathrm{Aut}(\mathbb{C})$; omdat deze gesloten is onder samenstelling heet die verzameling dan een groep.

Deze constructie, die we te danken hebben aan Évariste Galois, rond 1830, lijkt te abstract om enig nut te hebben, maar het tegendeel is waar. Galois liet al zien dat $\mathrm{Gal}(f)$ bepaalt of de wortels van f te verkrijgen zijn door een eindig aantal keer de operaties $+$, $-$, $\cdot$, $/$ en r-de worteltrekken voor willekeurige gehele $r \geq 1$ toe te passen en te beginnen met 1, a_{n-1},...,a_0 (denk aan een rekenmachine met de a_i in een aantal registers, en de knoppen $+$, $-$, $\cdot$, $/$ en $x^{1/r}$ (radicalen!)). Hieruit volgt bijvoorbeeld dat vergelijkingen van graad 5 of hoger niet met radicalen oplosbaar zijn, materie die aan de universiteit in het 2e of 3e jaar onderwezen wordt. Er zijn algoritmen om Galoisgroepen uit te rekenen, en sinds kort schijnt er een praktisch algoritme te zijn voor polynomen van willekeurige graad, maar de details hierover moeten nog gepubliceerd worden (Claus Fieker en Jürgen Klüners). Het gratis downloadbare programma (met alle documentatie) PARI/GP (zie [PARI]) heeft het commando `polgalois` dat deze taak uitvoert voor polynomen van graad tot en met 11.

Laten we een paar voorbeelden van Galoisgroepen bekijken. Als alle wortels van f in $\mathbb{Q}$ liggen, dan zijn ze allemaal vast onder alle elementen van $\mathrm{Aut}(\mathbb{C})$, en is $\mathrm{Gal}(f)$ de triviale groep, met precies één element, de identiteit. Omgekeerd is ook waar: als $\mathrm{Gal}(f)$ triviaal is, dan is $\mathrm{Wortels}(f)$ bevat in $\mathbb{Q}$. Als f van graad 2 is en geen wortel in $\mathbb{Q}$ heeft, dan bestaat $\mathrm{Gal}(f)$ uit de identiteit en de verwisseling van de twee wortels. Deze verwisseling komt van de complexe conjugatie als de discriminant van f negatief is.



Figuur 3: $n = 5$

Belangrijke voorbeelden voor ons zijn de cyclotomische (cirkeldeling) polynomen. Laat $n > 1$, laat $f = x^n - 1$, en laat $z = \cos(2\pi/n) + i\sin(2\pi/n)$. Dan is de afbeelding:

$$\{0, 1, \ldots, n-1\} \to \mathrm{Wortels}(f), \quad a \mapsto z^a \tag{5.11}$$

bijectief, en noemen we die een *labelling* van de wortels. Hiernaast is een plaatje voor het geval $n = 5$. De algebraïsche relaties die er tussen de wortels van f zijn (het zijn allemaal machten van de wortel z) geven beperkingen op de permutaties gegeven door elementen σ in $\mathrm{Aut}(\mathbb{C})$. Voor zo'n σ geldt dat $\sigma(z^a) = \sigma(z)^a$, dus is de permutatie gegeven door σ bepaald door wat σ met z doet. Voor iedere σ in $\mathrm{Aut}(\mathbb{C})$ is er een unieke k in $\{0, \ldots, n-1\}$ met $\sigma(z) = z^k$. Omdat $z^n = 1$ en $z^d \neq 1$ voor alle d met $1 \leq d < n$, hebben we $\mathrm{ggd}(k, n) = 1$. Het volgt dan uit werk

van Gauss dat de Galoisgroep $\mathrm{Gal}(x^n-1)$ precies uit de permutaties bestaat die van de vorm zijn $a \mapsto ka \mod n$, met $0 \leq k < n$ en $\mathrm{ggd}(n,k) = 1$. Hierin staat $b \mod n$ voor de rest na deling van b door n. Bijvoorbeeld geeft de complexe conjugatie vermenigvuldiging met -1 (eigenlijk, met $n-1$, maar dat geeft dezelfde resten modulo n), want $\overline{z} = \overline{e^{2\pi i/n}} = e^{-2\pi i/n} = z^{-1}$. Conclusie: in termen van de labelling is $\mathrm{Gal}(f)$ gegeven door vermenigvuldigingen in het getal systeem $\mathbb{Z}/n\mathbb{Z}$, de getallen $\{0, 1, \ldots, n-1\}$ met optelling en vermenigvuldiging *modulo* n.

We komen nu bij een van onze twee hoofddoelen: het zo eenvoudig mogelijk beschrijven van wat 2-dimensionale Galoisrepresentaties zijn. Om die te definiëren is er nog één ingrediënt nodig: een geheel getal $n > 1$ zodat we het getalsysteem $\mathbb{Z}/n\mathbb{Z}$ hebben. In plaats van vermenigvuldigingen met elementen van $\mathbb{Z}/n\mathbb{Z}$ zoals in $\mathrm{Gal}(x^n-1)$ willen we nu vermenigvuldigingen met 2 bij 2 matrices.

Een 2-dimensionale Galoisrepresentatie mod n bestaat uit een polynoom

$$f = x^{n^2} + \cdots + a_1 x + a_0$$

van graad n^2, met a_i in $\mathbb{Q}$, en een labelling van $\mathrm{Wortels}(f)$ met vectoren $v = \binom{v_1}{v_2}$ met v_1 en v_2 in $\mathbb{Z}/n\mathbb{Z}$, zodat $\mathrm{Gal}(f)$ bestaat uit vermenigvuldigingen met matrices:

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} av_1 + bv_2 \\ cv_1 + dv_2 \end{pmatrix}.$$

Deze objecten spelen de hoofdrol in Andrew Wiles's bewijs van Fermat's laatste stelling (1993-1994). Sinds 40 jaar is er theorie over waar dit soort representaties vandaan komen: modulaire vormen, d.w.z., Langlands' programma. Een goede vraag is: kan men de Galoisrepresentaties waarvan het Langslands programma het bestaan garandeert ook efficiënt uitrekenen? Antwoord: het lijkt erop van wel. Een belangrijk verschil met het geval van eenheidswortels is dat bij het vermenigvuldigen van 2 bij 2 matrices de volgorde er toe doet (de vermenigvuldiging is niet commutatief) terwijl $\mathbb{Z}/n\mathbb{Z}$ wel commutatief is.

6 Maar wat hebben de $r_d(n)$ nu met Galoisrepresentaties te maken?

Hoe kan het nu dat Galoisrepresentaties zoals beschreven in de vorige sectie ook maar *iets* te maken hebben met de getallen $r_d(n)$? Inderdaad, dit is diepe materie, die ik zelf niet eerder dan aan het eind van mijn 5-jarige intensieve wiskundestudie ben begonnen te leren. Ik kan er hier dan ook geen uitleg over geven, behalve dan paar feiten die het gevraagde verband geven.

We bekijken het geval $d = 12$, het kleinste even getal waarin er geen elementaire formule is voor $r_d(n)$, vanwege de term a_n, de coëfficiënt van q^n in het product $q\prod_{m\geq 1}(1-q^{2m})^{12}$, zie (2.7). Een beroemd resultaat van Pierre Deligne, in 1968, bouwend op werk van Goro Shimura, Martin Eichler, Alexander Grothendieck en nog anderen, zegt het volgende.

Voor iedere $m > 1$ bestaat er een 2-dimensionale Galoisrepresentatie, dus een polynoom $f = x^{m^2} + \cdots + f_1 x + f_0$, met de coëfficiënten f_i in $\mathbb{Q}$, en een labelling van $\mathrm{Wortels}(f)$ met

vectoren $\binom{v_1}{v_2}$ met v_1 en v_2 in $\mathbb{Z}/n\mathbb{Z}$, zodat $\mathrm{Gal}(f)$ bestaat uit vermenigvuldigingen met matrices, en zodat voor ieder priemgetal $p \not| m$ het getal a_p modulo m snel berekend kan worden uit f.

Het hoofdresultaat van [Ed-Co] en [Bruin] is dat de polynomen f waarvan Deligne het bestaan heeft aangetoond snel uitgerekend kunnen worden: in tijd polynomiaal in m.

Door a_p uit te rekenen modulo priemgetallen l tot een grens B krijgen we via de Chinese reststelling a_p modulo het product P van al deze priemgetallen. Dit product P is van de orde van grootte e^B. Een bovengrens voor $|a_p|$ is ook bekend: $|a_p| \leq 2p^{5/2}$ (alweer een beroemd resultaat van Deligne). Als P groter is dan $4p^{5/2}$, de lengte van het interval $[-2p^{5/2}, 2p^{5/2}]$, dan weten we a_p zelf, en niet alleen modulo P. Daaruit volgt dat het voldoende is om B van de orde van grootte $(5/2)\log p$ te nemen. Het eindresultaat is dan dat a_p uitgerekend kan worden in een tijd die hoogstens polynomiaal in $\log p$ is. Bekende formules laten zien hoe a_n uit te drukken is in de a_p waar p de priemdelers van n doorloopt. Dus, voor n gegeven met priemfactorisatie kan dan a_n uitgerekend worden in tijd polynomiaal in $\log n$.

Tot slot een opmerking over de terminologie van "elementaire formule voor $r_d(n)$". Voor alle even d zijn de getallen $r_d(n)$ te berekenen in tijd polynomiaal in $\log n$ (als de priemfactorisatie van n gegeven is), met behulp van 2-dimensionale Galoisrepresentaties. De elementaire gevallen zijn precies die waarin het al kan met behulp van 1-dimensionale Galoisrepresentaties. In dat geval zijn er ook expliciete en nuttige formules, zoals we hebben gezien.

7 Een voorbeeld van Johan Bosman

Het polynoom

$$\begin{aligned} f = x^{24} &- 2x^{23} + 115x^{22} + 23x^{21} + 1909x^{20} + 22218x^{19} + 9223x^{18} + 121141x^{17} \\ &+ 1837654x^{16} - 800032x^{15} + 9856374x^{14} + 52362168x^{13} - 32040725x^{12} \\ &+ 279370098x^{11} + 1464085056x^{10} + 1129229689x^9 + 3299556862x^8 \\ &+ 14586202192x^7 + 29414918270x^6 + 45332850431x^5 - 6437110763x^4 \\ &- 111429920358x^3 - 12449542097x^2 + 93960798341x - 31890957224 \end{aligned} \tag{7.1}$$

heeft Galoisgroep $\mathrm{PGL}_2(\mathbb{Z}/23\mathbb{Z})$, de groep van inverteerbare 2 bij 2 matrices met coëfficiënten in $\mathbb{Z}/23\mathbb{Z}$, op scalaire veelvouden na, voor een geschikte labelling van de wortels met de 24 lijnen door de oorsprong in $(\mathbb{Z}/23\mathbb{Z})^2$.

Referenties

[Bruin] Peter Bruin. *Modular curves, Arakelov theory, algorithmic applications.* Proefschrift, Leiden, September 2010. On-line beschikbaar op:
`http://hdl.handle.net/1887/15915`

[Ed-Co] Bas Edixhoven and Jean-Marc Couveignes, editors. *Computational aspects of modular forms and Galois representations.* With contributions by Johan Bosman, Jean-Marc

Couveignes, Bas Edixhoven, Robin de Jong, and Franz Merkl. Volume 176 of "Annals of Mathematics Studies", Princeton University Press, 2011. On-line beschikbaar op: http://www.math.univ-toulouse.fr/~couveig/book.htm

[Ha-Wr] G.H. Hardy and E.M. Wright. *An introduction to the theory of numbers.* Fifth edition. Clarendon Press, New York, 1979.

[PARI] C. Batut, K. Belabas, D. Bernardi, H. Cohen, and M. Olivier. *User's guide to PARI/GP (version 2.3.1).* http://pari.math.u-bordeaux.fr.

[Varma] Ila Varma. *Sums of Squares, Modular Forms, and Hecke Characters.* Master scriptie, Leiden, juni 2010. On-line beschikbaar op: http://www.math.leidenuniv.nl/nl/theses/196/