Vakantiecursus 1995
Kegelsneden en Kwadratische Vormen

SPREKERS

Prof. Dr. J. M. Aarts
Van Kinschotstraat 13, 2614 XJ Delft, 015-126448 wg. 015-785399/2697
e-mail: j.m.aarts@twi.tudelft.nl

Dr. A. G. van Asch
Beneden Molenweg 3D, 4112 NS Beusichem, 03453-1888, wg. 040-474280
e-mail: wsinaa@win.tue.nl

Prof. Dr. F. van der Blij
Ruysdaellaan 6, 3723 CC Bilthoven, 030-283168

Prof. Dr. A. W. Grootendorst
Aardbeistraat 11, 2564 TM Den Haag, 070-3232936

Dr. J. P. Hogendijk
Dassenlaan 17, 3734 HB Den Dolder, 030-287968 wg. 030-533697
e-mail: hogend@math.ruu.nl

Dr. Ir. C. W. A. M. van Overveld
Serlioweg 14, 5624 KA Eindhoven, 040-457953 wg.040-474416
e-mail: wsinkvo@win.tue.nl

Dr. P. Stevenhagen
p/a Fac. Wiskunde en Informatica Plantage Muidergracht 24, 1018 TV Amsterdam, wg. 020- 5255202, hs. 020- 6204588
e-mail: psh@fwi.uva.nl

Ten Geleide

Waar in de cursus 1994 een onderwerp aan de orde gesteld werd dat jong en in volle ontwikkeling is –de Computer Algebra– een onderwerp ook dat nog zijn entree op de scholen moet maken, staat thans, bij de cursus 1995, een thema centraal dat niet alleen een langdurige binding met het onderwijs heeft, maar dat ook gerekend kan worden tot de oudste gebieden van de wiskunde en dat zijn "roots" heeft in de Griekse Oudheid.

Traditioneel wordt de ontdekking van de ons bekende kegelsneden toegeschreven aan Menaechmus (ca. 350 v. Chr.). Typisch voor zijn methode was dat hij een kegel steeds sneed met een vlak dat evenwijdig verloopt aan een beschrijvende. De verschillende soorten kegelsneden ontstaan dan door de tophoek recht, stomp of scherp te nemen. Van zijn geschriften op dit gebied is niets bewaard gebleven, evenmin als van het werk van Aristaeus (eveneens midden vierde eeuw v. Chr.). Ook "onze" Euclides (ca. 300 v. Chr) zou, voortbouwend op Aristaeus, een werk "*κωνικά*" (konika) geschreven hebben dat verloren is gegaan.
Veel daarvan is echter opgenomen in het werk van Apollonius van Perga (circa 250 v. Chr.) die op zijn eigen, buitengewoon scherpzinnige wijze de kegelsneden invoerde en hun de bekende namen meegaf.

Met deze ontwikkelingen in de klassieke Oudheid, met name de constructie van kegelsneden bij Apollonius, zet de vacantiecursus 1995 in: zij vormen het onderwerp van de lezing van *Dr. J.P. Hogendijk.*

De volgende lezing is van *Prof. Grootendorst.* Deze verplaatst ons naar de 17e eeuw, waarin FERMAT (1601–1665), DESCARTES (1596–1650), WALLIS (1616–1703) en JOHAN DE WITT (1625–1672) de grondslagen legden voor wat wij thans analytische meetkunde noemen.
Deze lezing is gewijd aan het tweedelige werk van Johan de Witt, de "Elementa Curvarum Linearum" dat in 1661 voor het eerst in druk verscheen.
In het eerste deel daarvan zet deze zich af tegen de constructies van kegelsneden zoals Apollonius van Perga die gaf en stelt daar zijn eigen methode tegenover. Zijn bezwaar geldt namelijk de introductie van *vlakke* krommen via een *ruimtelijk* lichaam, i.c. een kegel. In het tweede deel introduceert hij de voorstelling van kegelsneden door middel van kwadratische vergelijkingen.

De volgende voordracht, die van *Prof. Aarts*, behandelt een generalisatie van de kegelsneden in de vorm van kwadrieken van dimensie twee en hoger.
Hierbij beperkt de spreker zich niet tot de Euclidische ruimte, maar gaat – na zorgvuldige voorbereiding– over op projectieve ruimten en komt dan op verrassende wijze tot de beantwoording van de vraag hoeveel rechten elk van een gegeven viertal kruisende rechten snijden.

De vierde voordracht, te geven door *Dr. van Asch*, is geheel gewijd aan kwadratische vormen en wel aan een aantal toepassingen daarvan: de zeer

practische methode van de "kleinste kwadraten", maar ook een meer abstracte toepassing als het introduceren van andere metrieken dan de "standaardmetriek", via kwadratische vormen en als derde toepassing de problemen van de metriek op de bol.

Reeds in de Griekse Oudheid was de vraag naar de mogelijkheid een kromme daadwerkelijk te construeren een zeer belangrijke en ook in de 17e eeuw stond dit probleem van de construeerbaarheid sterk in de belangstelling, met name bij Descartes.
Op practische en eigentijdse wijze gaat *Dr. Ir. van Overveld* hierop in. Aan de hand van veel illustraties met behulp van de computer toont hij aan hoe men d.m.v. zogenaamde discrete procedurele methoden (dit wordt zorgvuldig uitgelegd) in een eindig aantal stappen een redelijk beeld van een kromme kan ontwerpen, maar ook wijst hij erop dat zich hier complicaties kunnen voordoen.

In dit verband zij vermeld dat het gebruikelijke "uurtje zelfwerkzaamheid" dat zich in de laatste jaren een vaste plaats heeft verworven in de vacantiecursus, dit jaar vervangen is door drie kwartier computerdemonstratie waarin *Dr. Ir. Van Overveld* veel van wat de sprekers in hun voordrachten hebben behandeld, d.m.v. computer graphics in "geanimeerde" vorm zal vertonen.

Kwadratische vormen hebben sinds EULER (1707–1783), LAGRANGE (1736–1813) en GAUSS (1777–1855) hun plaats gevonden in de getallentheorie en het is dan ook niet verwonderlijk dat aan dit aspect van kwadratische vormen twee voordrachten zijn gewijd.

Dr. Stevenhagen zal o.a. spreken over de bijdragen die Gauss op dit gebied heeft geleverd. Zijn uitgangspunt is de volgende eenvoudige observatie van Diophantus van Alexandrië (ca. 250 A.D.) m.b.t. de splitsing van een natuurlijk getal in de som van twee kwadraten van gehele getallen:

$$65 = 7^2 + 4^2 = 8^2 + 1^2 \text{ omdat } 65 = 13 \times 5 = (3^2 + 2^2) \cdot (2^2 + 1^2).$$

Dit leidt tot de bestudering van kwadratische vormen, hun indeling in klassen van onderling equivalente vormen, de bijbehorende meetkundige voorstelling d.m.v. roosters in het complexe vlak en het rekenen daarmee. Het hoogtepunt is de bespreking van het aantal klassen $h(D)$ behorende bij de kwadratische vormen met gegeven discriminant D en in het bijzonder het geval waarin dit getal $h(D)$ de waarde 1 aanneemt.

Prof. van der Blij gaat eveneens uit van het splitsen van een natuurlijk getal in een som van twee kwadraten van gehele getallen. Deze situatie wordt daarna snel gegeneraliseerd. In de eerste plaats zal ook "modulo m" worden gerekend en ook zullen kwadratische vormen in meer dan twee variabelen worden bestudeerd. Deze weg zal leiden tot dieper liggende resultaten waarbij de belangrijke rol van de kwadratische vormen voor de getallentheorie nogmaals duidelijk wordt getoond. Veel aandacht wordt daarbij ook gegeven aan analytische methoden (o.a. theta-functies) en aan de invarianten van kwadratische

vormen onder bepaalde transformaties.
Een breed overzicht over de ontwikkelingen op dit gebied na 1945, besluit deze voordracht en tevens de vacantiecursus 1995.

Hiermee is dan in grote lijnen de inhoud van de cursus geschetst en men ziet dat dit klassieke onderwerp veel facetten vertoont en naar veel kanten kan worden uitgebouwd. Daarmee voldoet dit onderwerp naar de mening van de organisatoren goed aan de eisen die men stelt bij een vacantiecursus: het is gevarieerd en kan aanleiding geven tot verdere studie. Tevens bevat het ook voldoende materiaal dat voor geïnteresseerde leerlingen in de klas aan de orde gesteld kan worden en dat ongetwijfeld stimulerend zal werken en dat is – juist in deze tijd – van het grootste belang.

Traditioneel – maar het blijft ten volle gemeend– wordt ook dit "Ten Geleide" besloten met een oprecht woord van dank aan alle medewerk(st)ers van het CWI die er ook dit jaar weer in slaagden deze syllabus zo voortreffelijk uitgevoerd en op tijd uit te brengen. Dit jaar was het misschien een extra zware opgave door de vele illustraties die de tekst verduidelijken en die mede aan het boekje zo'n fraai uiterlijk geven.
De dank van de voorbereidingscommissie geldt echter niet alleen diegenen die deze syllabus vorm gaven, maar ook allen die de bekendmaking van de cursus, de registratie van de deelnemers en de goede verzorging tijdens de cursusdagen voor hun rekening namen. Speciaal onze dank aan hen die het mogelijk maakten dat nu ook in Eindhoven gemeenschappelijk geluncht kan worden. Zo'n gemeenschappelijke maaltijd geeft altijd een extra dimensie aan het geheel.
Van harte hopen de organisatoren dat deze cursus goed bezocht zal worden, zowel door "oude bekenden" als door nieuwkomers die eveneens van ganser harte welkom zijn en dat de cursisten twee genoeglijke dagen zullen beleven.

Daarna gaan we ons opmaken voor de vijftigste cursus die wij in 1996 hopen te organisern!

A.W. Grootendorst

Kegelsneden in de Griekse oudheid

J.P. Hogendijk

1. Inleiding

Kegelsnedes zijn omstreeks 350 voor Christus in de Griekse wiskunde ontdekt, en in de twee eeuwen daarna uitgebreid bestudeerd. De moderne termen *ellips*, *parabool*, *hyperbool* en *asymptoot* zijn ontleend aan een Grieks leerboek over kegelsneden, de *Conica* van Apollonius van Perga (ca. 200 v. Chr.).

In deze lezing proberen we een zeer globale indruk te geven van de theorie van de kegelsneden in de Griekse oudheid, met nadruk op zaken die verband houden met deze nu nog steeds gebruikte woorden. We zullen daarbij soms moderne notatie gebruiken, om de zaken voor ons eenvoudig voor te stellen, maar we moeten daarbij bedenken dat er toen niet zulke notatie was. De Grieken hadden geen reëel getalbegrip en ze gebruikten geen algebraïsche symbolen zoals x, y, $+$, $-$, $\times$, echter wel letters om punten en segmenten aan te duiden. Waar wij een "product $x \cdot y$" zouden gebruiken, zouden zij met een "rechthoek met zijden de segmenten A en B" werken, of met een uitdrukking als "de onder $BK\Gamma$", hetgeen betekent de rechthoek met zijden gelijk aan BK en $K\Gamma$.

We kunnen de Griekse wiskunde het beste vergelijken met een ijsberg, waarvan we alleen het gedeelte boven de waterspiegel kunnen zien. De geschiedenis van de kegelsneden moet hier en daar uit tekstfragmenten aan elkaar worden geplakt, en een aantal zaken zijn controversieel. Het is daarom belangrijk de bronnen aan te geven en feitenmateriaal van speculatie te scheiden.

2. Het begin van de kegelsneden. Het Delische altaar

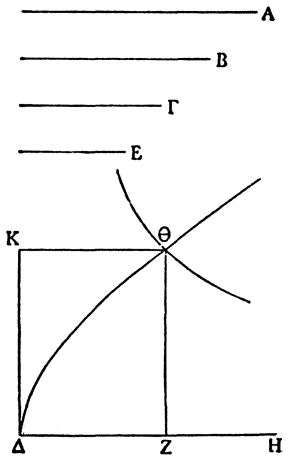
De oudste geschiedenis van de kegelsneden is verbonden met het zogenaamde Delische probleem. Volgens één van de versies van het verhaal heerste in Griekenland eens een epidemie. Het orakel van Delphi werd geraadpleegd en het bleek dat de goden de epidemie pas zouden laten ophouden als er een altaar zou worden geconstrueerd dat het dubbele was van het huidige altaar. Zo ontstond het probleem, een kubus te construeren met als inhoud twee maal de inhoud van een gegeven kubus.

Hippocrates van Chios (ca. 430 v. Chr.) ontdekte dat dit probleem zou kunnen worden opgelost, als men voor elk paar gegeven lijnstukken A en B twee middenproportionalen X en Y zou kunnen vinden, d.w.z. twee lijnstukken zodat $A : X = X : Y = Y : B$. Immers, als A de zijde is van de gegeven kubus, en $B = 2A$, dan is X de zijde van de te construeren kubus; in moderne notatie geldt $X = A\sqrt[3]{2}$. Hippocrates was helaas niet in staat X te vinden; wij kunnen tegenwoordig bewijzen dat dit met passer en lineaal onmogelijk is.

In het commentaar van Proclus (5e eeuw na Chr.) op de *Elementen* van Euclides wordt vermeld[1] dat de kegelsneden ontdekt zijn door Menaechmus. Dit is een meetkundige die omstreeks 350 voor Christus leefde en die connecties had met Plato. We weten dat Menaechmus zich met het Delische probleem

[1] Morrow, p. 91

heeft beziggehouden, en er een oplossing van heeft gegeven met kegelsneden. Die oplossing is bewaard door Eutocius (omstreeks 500 na Christus) in een commentaar op een boek van Archimedes over de bol en de cylinder (Figuur 1, zie de tekst en vertaling in THOMAS, vol. 1, pp. 278-283). De twee



Figuur 1:

gegeven segmenten zijn A en E. Menaechmus (of Eutocius) begint met de *analyse*, die we hier vrij zullen weergeven. In de analyse veronderstellen we dat de gevraagde segmenten er al zijn, en we proberen dan af te leiden hoe ze kunnen worden gevonden. Stel dus, dat de gevraagde segmenten B en Γ zijn, dat wil zeggen

$$A : B = B : \Gamma = \Gamma : E.$$

Teken twee lijnen $\Delta K = B$ en $\Delta Z = \Gamma$ loodrecht op elkaar, en maak rechthoek $\Delta Z \Theta K$. Nu is ten eerste $A : B = \Gamma : E$, dus rechthoek (A, E) = rechthoek (B, Γ) = rechthoek $\Delta Z \Theta K$. De oppervlakte van die rechthoek is dus bekend. Punt Θ ligt daarom op een hyperbool met asymptoten ΔK en ΔZ die door een gegeven punt gaat. (Wij denken daarbij aan een hyperbool met vergelijking $xy = c$ met $c = A \cdot E, x = \Delta Z, y = \Theta Z$). De twee loodrechte halve lijnen waarop ΔZ en ΔK zijn afgezet kunnen als gegeven worden beschouwd, en de hyperbool is dan bepaald.
Verder hebben we: $B^2 = A \cdot \Gamma$ (dwz het vierkant met zijde B heeft dezelfde oppervlakte als de rechthoek met zijden A en Γ). Dus geldt: $\Theta Z^2 = A \cdot \Delta Z$. Wij herkennen hierin de vergelijking van een parabool $y^2 = px$ met $p = A, x = \Delta Z, y = \Theta Z$. De tekst concludeert nu dat Θ op een parabool ligt door Δ. De as is ΔZ en de parameter is A; hier gebruiken we het woord parameter als

in de moderne vertalingen van de *Conica* van Apollonius. Menaechmus heeft waarschijnlijk een andere term voor p gebruikt, maar we weten niet welke, omdat de bewerker Eutocius de tekst volgens de normen van Apollonius gemoderniseerd heeft.
Dan volgt nu de *synthese*, dat is de echte constructie en het bewijs. Begin met twee loodrechte halve lijnen te kiezen die elkaar in Δ ontmoeten. Teken een hyperbool met asymptoten die beide lijnen, en zo dat elke rechthoek onder de hyperbool (als in Figuur 1) oppervlakte $A \cdot E$ heeft. Teken een parabool met top Δ, as één van beide lijnen en parameter A. Deze twee krommen snijden elkaar in een punt Θ. Laat uit Θ loodlijnen ΘZ en ΘK neer als in de figuur. ΔZ en ΔK zijn dan de gevraagde segmenten. Het bewijs gaat analoog aan de redenering in de analyse.
We concluderen dat Menaechmus de eigenschappen kende die overeenkomen met onze moderne vergelijkingen $y^2 = px$ van de parabool en $xy = c$ van de hyperbool. Het is zeer waarschijnlijk dat hij deze krommen als snede van een kegel en een vlak kende, en dat hij ook de ellips kende. Dat kunnen we opmaken uit een opmerking van de meetkundige Eratosthenes (ca. 250 v.Chr.), die zelf ook een constructie van het Delische probleem heeft bedacht met een toestel, en die daarop zeer trots was. Hij schreef een gedicht waarin stond dat zijn oplossing beter was dan vele andere, en waarin hij onder andere zei dat je de oplossing niet moest vinden "...door de kegel te snijden in de triaden van Menaechmus ...". De *triaden* worden meestal geïnterpreteerd als de drie kegelsneden die wij nu hyperbool, parabool en ellips noemen.[2] Hoe Menaechmus de eigenschappen van de kegelsneden uit de definitie van een kegel heeft afgeleid weten we niet, voor speculaties hierover zie bijvoorbeeld ZEUTHEN, pp. 455-469.
De parabool en de hyperbool van Menaechmus helpen ons niet bij het in de praktijk construeren van een altaar dat twee keer zo groot is als het oorspronkelijke. Zij laten wel zien, dat dit altaar bestaat in de ideale wiskundige wereld, tenminste als aannemelijk gemaakt kan worden dat kegelsneden daarin bestaan. De zuivere wiskunde van dit type heeft zich in de Griekse cultuur uitgebreid kunnen ontwikkelen en lang kunnen handhaven.

3. VAN MENAECHMUS NAAR APOLLONIUS

Over de volgende 150 jaar in de geschiedenis van de kegelsnedentheorie zijn we slecht ingelicht. We weten dat er leerboeken over kegelsneden geschreven zijn door de beroemde Euclides en Aristarchus, maar deze zijn verloren gegaan. Archimedes (ca. 287-212 v. Chr.) zegt af en toe wat over kegelsneden in zijn werken (die grotendeels bewaard zijn). Door dit materiaal te analyseren kan men er gedeeltelijk achter komen wat in zijn tijd bekend was. Zo blijkt dat

[2] Een afwijkende interpretatie is gegeven door W. KNORR in *The ancient tradition of geometric problems*, pp. 61-66. Deze stelt, dat Menaechmus de parabool en hyperbool niet als snede van een kegel met een vlak definiëerde, maar puntsgewijs, en dat de 'triaden' verwijzen naar de hyperbool en twee parabolen in twee oplossingen van het Delische probleem. Hiervan is alleen de eerste zeker van Menaechmus. KNORR moet aannemen dat Eratosthenes (en ook Proclus) de geschiedenis hebben vervalst door Menaechmus in verband te brengen met kegels. Daardoor is deze interpretatie weinig plausibel.

men in de tijd van Archimedes de kegelsnedes definieerde door een omwentelingskegel te nemen en die met een vlak loodrecht op een beschrijvende lijn te snijden. Afhankelijk van de tophoek waren er drie mogelijkheden, en men noemde de kegelsneden de *snede van een scherphoekige kegel* (onze ellips), *snede van een rechthoekige kegel* (onze parabool) en *snede van een stomphoekige kegel* (onze hyperbool).
Archimedes bewees veel stellingen over inhouden en zwaartepunten die met kegelsneden te maken hebben. Zo bepaalde hij de oppervlakte van een paraboolsegment en de inhoud van een figuur die ontstaat door een stuk van een parabool om zijn as te wentelen. Dit is een moderne uitdrukkingswijze, want voor Archimedes was een oppervlakte niet een reëel getal. 'Bepalen' betekende voor hem: bewijzen dat de figuur gelijk is (in oppervlakte of inhoud) aan een andere figuur, die uitgedrukt kon worden in een rechthoek of blok. Voorbeeld: We snijden van een parabool een segment af door een recht lijnstuk. Archimedes bewees dat dit segment gelijk is aan 4/3 maal de ingeschreven driehoek met basis het lijnstuk en top het snijpunt van de parabool met een lijn door het midden van deze basis evenwijdig aan de as van de parabool.[3]
Hieruit blijkt al, dat we de kennis over kegelsneden in de tijd van Archimedes niet moeten onderschatten. Er waren uiteraard ook dingen bekend die bij Archimedes niet werden aangehaald, bijvoorbeeld de brandpunt-richtlijneigenschap van de kegelsneden (dat deze bekend was blijkt uit informatie gegeven door de laat-Griekse schrijver Pappus).

4. Apollonius van Perga en zijn *Conica*.

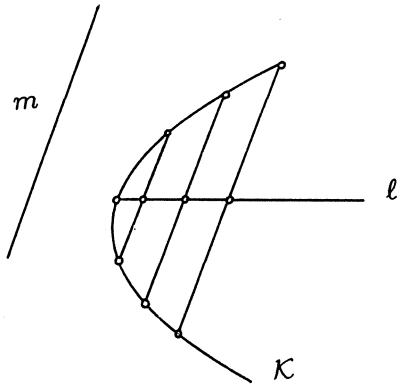
We zijn nu aangekomen bij de hoofdpersoon van dit verhaal, Apollonius van Perga. Zoals bij de meeste Griekse wiskundigen weten we weinig van zijn leven. Hij leefde omstreeks 200 voor Christus, had een zoon die ook Apollonius heette, en was behalve wiskundige ook astronoom. Zijn leerboek over kegelsneden is de *Conica*, die uit 8 "boeken" bestond, met "boek" bedoelen we hier een groot hoofdstuk. De boeken 1 tot 4 zijn in een Griekse versie bewaard, de boeken 5 tot 7 in een Arabische vertaling, en boek 8 is verloren gegaan. De eerste vier boeken zijn in 1566 in een Latijnse vertaling van Commandinus in Bologna gedrukt en ze werden zo in West-Europa bekend.
Naast de *Conica* schreef Apollonius nog een aantal kleinere werkjes over meetkundige constructies met passer en lineaal. Deze zijn op één na verloren gegaan, maar we weten er wel iets over uit beschrijvingen van de al eerder genoemde Pappus van Alexandrië. Zo weten we, dat de "cirkel van Apollonius" en het "raakprobleem van Apollonius" (een cirkel construeren die aan drie gegeven cirkels raakt) in deze verloren gegane werkjes behandeld werden.
De *Conica* zelf begint met een voorwoord, dat volgens kenners van de Griekse taal in een zeer goede stijl geschreven is. We zullen zometeen zien dat Apollonius behalve een goed wiskundige ook een taalkunstenaar was.
Apollonius begint zijn theorie van de kegelsneden als volgt. Hij bekijkt een willekeurige cirkel in een vlak en een punt niet in dat vlak. Alle halve rechten van dat punt naar de cirkel vormen samen een kegel (en de hele rechten twee

[3] Zie Dijksterhuis, *Archimedes*.

kegelnappen). Het punt heet de top van de kegel, de cirkel heet de basis van de kegel, en de lijn die de top met het middelpunt van de basis verbindt heet de as van de kegel. Hij gaat dan deze kegel met een vlak snijden en de snijfiguur bestuderen. Deze snijfiguur kan uit rechte lijnen bestaan of een cirkel of een kegelsnede zijn.
Deze definitie is in twee opzichten algemener dan die van de voorgangers van Apollonius. In de eerste plaats hadden de voorgangers alleen omwentelingskegels gebruikt, die ontstaan door een rechthoekige driehoek om een rechthoekszijde te wentelen. In zulke kegels is de as altijd loodrecht op het vlak van de basiscirkel. Bij Apollonius hoeft de as niet loodrecht op de basis te staan. Ten tweede snijdt Apollonius zijn kegels al in zijn definities met willekeurige vlakken, niet alleen met vlakken loodrecht op een beschrijvende lijn.
Apollonius behandelt eerst twee triviale gevallen: een vlak door de top snijdt de kegel in een driehoek (de basis wordt meegerekend), en een vlak evenwijdig aan de basis snijdt de kegel weer in een cirkel. Daarna stelt hij de vraag, of er nog andere vlakken zijn die de kegel ook in een cirkel snijden. Dit blijkt alleen zo te zijn, als de as niet loodrecht op de basis staat. Dan is er namelijk precies één vlak $\mathcal{V}$ door de as van de kegel loodrecht op het vlak van de basiscirkel $\mathcal{B}$. Er is ook precies één schaar onderling evenwijdige vlakken $\mathcal{W}$ die loodrecht staan op $\mathcal{V}$ en dezelfde hoek maken met de as als vlak $\mathcal{B}$ maar niet evenwijdig zijn aan $\mathcal{B}$. Apollonius noemt zulke vlakken $\mathcal{W}$ "tegengesteld". Hij bewijst ook dat alle andere vlakken dan de tot nu toe genoemde de kegel snijden in een kromme die geen cirkel is.
Apollonius onderzoekt zulke krommen op een manier die anders is dan een moderne aanpak met Cartesische coordinaten. Modern gezegd komt het erop neer, dat Apollonius ook met scheve coordinatenstelsels werkt, waarbij we opmerken dat moderne coördinaten (in de zin van reële getallen) bij Apollonius niet voorkomen. We vinden wel een paar begrippen die we als voorlopers van het huidige coordinaatsbegrip kunnen zien, namelijk diameter en ordinaat. Apollonius definieert deze voor een willekeurige kromme $\mathcal{K}$ (Figuur 2).

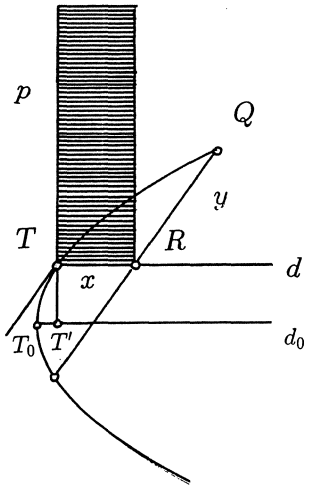
DEFINITIES:

1. een diameter van $\mathcal{K}$ is een rechte lijn ℓ met de volgende eigenschap: er is een andere rechte lijn m zodanig alle segmenten die evenwijdig zijn aan m en waarvan beide eindpunten op $\mathcal{K}$ liggen door ℓ middendoor worden gedeeld.
2. de helften van de segmenten die hierboven genoemd zijn heten *geordend* (Grieks: $\tau\varepsilon\tau\alpha\gamma\mu\acute{\varepsilon}\nu\omega\varsigma$) getrokken ten opzichte van de diameter ℓ. (De Latijnse vertaling gebruikt *ordinatim ducta*, vandaar dat we vaak het woord *ordinaat* voor zo'n helft van een segment zullen gebruiken. Voor de constante hoek tussen een diameter en de bijbehorende ordinaten gebruiken we het woord *ordinaatshoek*.)
3. een as is een diameter die loodrecht staat op de bijbehorende "geordend getrokken" lijnen (ordinaten).



Figuur 2:

Als $\mathcal{K}$ een cirkel is, dan is duidelijk dat elke rechte lijn door het middelpunt van de cirkel diameter is in de zin van Apollonius. De ordinaten staan loodrecht op de diameter.
Apollonius gebruikt nu de rest van Boek I, en een deel van de overige boeken van de *Conica* om uit te zoeken hoe de situatie wat dit betreft voor andere kegelsnedes is. Zijn bewijzen zijn zware kost, en we zullen er maar één voorbeeld van geven, dat geen recht doet aan de structuur en de moeilijkheid van het geheel. Maar eerst zullen we proberen het eindresultaat op een min of meer didaktische manier te presenteren. We beginnen met de parabool.
Als de kegel wordt gesneden met een vlak evenwijdig aan een lijn op de kegelmantel, onstaat een oneindige kromme, die Apollonius parabool noemt (op de naamgeving komen we zometeen terug). Elke parabool heeft één as, en elke lijn evenwijdig aan de as is een diameter van de parabool. In elk punt van de parabool kan precies één raaklijn getrokken worden. De ordinaten van een diameter zijn altijd evenwijdig aan de raaklijn in het punt waar de diameter de parabool snijdt. (Figuur 3) Stel nu dat d een diameter is die de parabool in T snijdt, kies een punt Q op de parabool, en laat QR de ordinaat door Q zijn die hoort bij de diameter d (die ordinaat is dus evenwijdig aan de raaklijn in T).
Dan is het vierkant van de ordinaat QR gelijk (d.w.z. gelijk in oppervlakte) aan de rechthoek waarvan één zijde gelijk is aan het stuk RT wat van de diameter afgesneden wordt (Latijn: abscissa), en de andere zijde gelijk aan een segment p wat alleen van de diameter d afhangt, maar niet van de keuze van punt Q (zie Figuur 3). Dit segment tekent Apollonius met eindpunt in T en loodrecht op d. De rechthoek komt dan ook echt in de figuur voor (het gearceerde gedeelte). Als we stellen $QR = y, RT = x$ dan ontstaat de vergelijking $y^2 = px$.
We komen nu op de naamgeving. In het Grieks zeggen we dat deze rechthoek langs p ligt, langsliggen is in het Grieks $\pi\alpha\rho\alpha\beta\alpha\lambda\lambda\epsilon\iota\nu$ (van $\pi\alpha\rho\alpha$, langs, en $\beta\alpha\lambda\lambda\epsilon\iota\nu$, werpen). Het segment p heet in het Grieks de *rechtopstaande zijde*, en



Figuur 3:

ook het *lijnstuk, waarlangs zij (de ordinaten) in vierkant gelijk zijn.* Hieruit is in de zeventiende eeuw het woord *parametrum* ontstaan: het lijnstuk, waarlangs gemeten wordt.[4] Het woord parameter heeft sindsdien een veel algemenere betekenis gekregen, en wordt ook op overdrachtelijke manier in gebieden buiten de wiskunde gebruikt.

Als d_0 de as is en d een willekeurige diameter, zijn de bijbehorende parameters p_0 en p niet gelijk. Stel d_0 snijdt de parabool in T_0 en laat T' de loodrechte projectie zijn van T op de as, dan geldt $p = p_0 + 4T_0T'$ (*Conica* VII:5).[5]

De hyperbool of de ellips kunnen ontstaan als het snijvlak niet door de top gaat en niet evenwijdig is aan een lijn op de kegelmantel. Als elke lijn in het kegeloppervlak door de top heen verlengd wordt, ontstaan twee nappen. Nu zijn er twee mogelijkheden:

a. het snijvlak snijdt maar één van de nappen, er ontstaat dan een ellips (tenzij het vlak evenwijdig aan de basis of tegengesteld aan de basis ligt, dan ontstaat een cirkel);

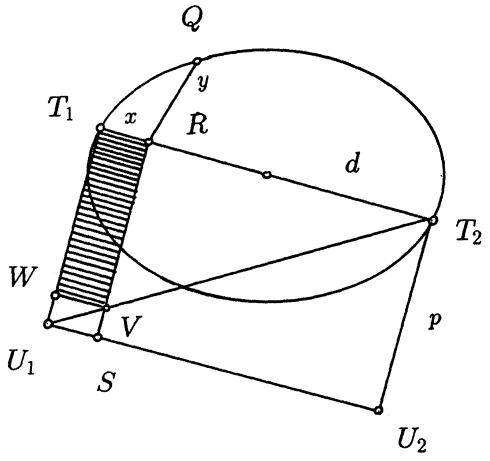
b. het snijvlak snijdt beide nappen, er ontstaan dan voor Apollonius twee hyperbolen. Een hyperbool bij Apollonius is één tak van wat wij een hyperbool noemen.

Bij elke ellips is er één speciaal punt, het middelpunt, met de eigenschap dat alle rechte lijnen door het middelpunt diameters zijn. In elk punt op de ellips kan één raaklijn aan de ellips getrokken worden, en de ordinaten van een diameter zijn weer evenwijdig aan de raaklijnen in de punten waar de diameter de ellips

[4]Nota bene: In het Nederlandse taalgebied wordt de term: parameter van een parabool meestal gebruikt voor de grootheid p in de vergelijking $y^2 = 2px$, in afwijking van wat in de huidige vertalingen van de *Conica* gebruikelijk is.

[5]Dit is de meest handzame manier waarop Apollonius dit verband aangeeft. In Boek I, prop. 49 geeft hij een minder doorzichtige manier.

snijdt (die twee raaklijnen zijn evenwijdig). Verder is het vierkant van elke ordinaat weer gelijk aan een rechthoek, die gearceerd is in Figuur 4, en die iets ingewikkelder is dan bij de parabool. Stel de diameter d snijdt de ellips in punten T_1 en T_2, kies een punt Q op de ellips dat niet met die twee punten samenvalt en trek de ordinaat QR door Q die bij d hoort (en die dus evenwijdig is aan de raaklijnen in T_1 en T_2). Er zijn nu segmenten T_1U_1=T_2U_2=p, die even lang zijn, loodrecht op d staan en onafhankelijk van Q zijn, zodat het volgende geldt. Trek de diagonaal T_2U_1, trek RS evenwijdig aan T_iU_i, die U_1U_2 in S snijdt en T_2U_1 in V. Trek VW evenwijdig aan d, het snijpunt met T_1U_1 is W. Nu geldt: QR^2 = rechthoek T_1RVW. Apollonius zei: het vierkant van



Figuur 4:

QR is gelijk aan een rechthoek, die langs $p = T_1U_1$ ligt, met breedte T_1R, en waaraan een (kleine) rechthoek VWU_1S "ontbreekt". Ontbreken is in het Grieks $\epsilon\lambda\lambda\epsilon\iota\pi\epsilon\iota\nu$, en Apollonius noemt deze kegelsnede daarom *ellips*. Het segment p heet weer o.a. parameter. Andere benamingen zijn *opstaande zijde* voor p en *dwarse zijde* voor $t = T_1T_2$.
Als we $y = QR, x = RT_1$ stellen, kunnen we $QR^2 = T_1RVW$ vertalen in de vergelijking $y^2 = px - \frac{p}{t}x^2$.
Als we een andere diameter $d' = T_1'T_2'$ kiezen en de notaties p', t' analoog aan p en t definiëren, dan geldt tussen p, t, p', t' het volgende verband:
$t^2 + pt = t'^2 + p't'$.[6]
Voor de hyperbool gelden soortgelijke redeneringen. Het vierkant van de ordinaat schiet in dit geval op analoge manier een kleine rechthoek VWU_1S over ten opzichte van de rechthoek T_1RSU_1 langs het segment p=T_1U_1, en overschieten

[6]Dit is een handzame vorm gebaseerd op Boek VII, prop. 12-13. In Boek I, prop. 50 geeft Apollonius het verband op een minder inzichtelijke manier.

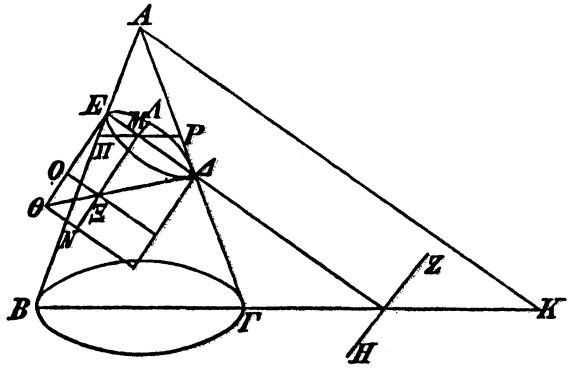
is in het Grieks $\hat{\upsilon}\pi\epsilon\rho\beta\alpha\lambda\lambda\epsilon\iota\nu$. Dit verklaart de naam hyperbool.[7]
We zouden kunnen zeggen: de ellips en de hyperbool ontlenen hun naam aan de $-$ of de $+$ in hun "vergelijkingen" $y^2 = px \pm \frac{p}{t}x^2$.
Aan het eind van Boek I (proposities 52-58) "vindt" Apollonius een kegelsnede met gegeven diameter, parameter en ordinaatshoek. Daarmee bedoelt hij: het vinden van een omwentelingskegel die het vlak in de gevraagde kegelsnede doorsnijdt. De top en een basiscirkel van die kegel worden door middel van een soort 3-dimensionale passer- en lineaalconstructie gevonden. De praktische betekenis van deze constructie is nihil; het gaat er waarschijnlijk alleen om dat een kegelsnede met gegeven diameter (of dwarse zijde), parameter en ordinaatshoek in filosofisch opzicht bekend kan worden verondersteld.
In Boek II bewijst Apollonius, dat er door het middelpunt van de hyperbool twee lijnen zijn met een speciale eigenschap, namelijk: zij hebben geen gemeenschappelijk punt met de hyperbool, maar de hyperbool heeft wel een snijpunt met elke rechte lijn door het middelpunt die in één van de hoeken tussen deze twee lijnen ligt. Samentreffen is in het Grieks $\sigma\upsilon\mu\pi\iota\pi\tau\epsilon\iota\nu$ en de twee speciale lijnen doen dit in geen enkel punt met de hyperbool, vandaar dat die twee speciale lijnen de asymptoten heten. Eigenlijk is deze term niet zo gelukkig, omdat er veel meer lijnen door het middelpunt zijn die de hyperbool niet ontmoeten. Apollonius bewijst vele stellingen over de hyperbool en zijn asymptoten, onder andere een stelling (*Conica* II:12) die zich laat vertalen in de moderne vergelijking $xy = c$ van een hyperbool ten opzichte van zijn asymptoten. We hebben deze stelling in de constructie van Menaechmus gezien. De *Conica* bevat nog veel meer stellingen over andere onderwerpen zoals pool en poollijn, brandpunten, normalen enzovoort, waar we nu niet op ingaan. We verwijzen de lezer daarvoor naar de boeken in de bibliografie aan het eind van dit artikel.

5. Een bewijs uit de *Conica*

We geven nu een voorbeeld van een stelling van Apollonius met bewijs, namelijk propositie 13 van Boek I (Apollonius ed. Heiberg vol. 1, p. 48-53; Ver Eecke p. 28-31). Het doel hiervan is nog een tipje van de sluier over zijn theorie op te lichten, en de lezer een gevoel te geven van de sfeer van het boek. In de stelling bewijst Apollonius de "vergelijking" $QR^2 = T_1RVW$ van de ellips, voor één speciale diameter. Dit is de "oorspronkelijke diameter", waarmee bedoeld wordt: de enige diameter waarvan Apollonius in de stereometrische figuur laat zien dat het een diameter is. De andere lijnen door het middelpunt van een ellips zijn ook diameters, maar het bewijs daarvan is veel ingewikkelder, en de kegel wordt daar niet opnieuw bij gebruikt.
Apollonius presenteert zijn stellingen volgens een vast ritueel, dat ook in de *Elementen* van Euclides wordt gebruikt. De stelling begint met de zogenaamde *protasis*, dat is een formulering in algemene termen (woorden tussen haakjes zijn door mij aan de tekst toegevoegd ter verduidelijking):

[7] Apollonius heeft in de benaming van de kegelsneden bestaande terminologie aangepast. In de tijd van Euclides en daarvoor bestonden al de parabolische, hyperbolische en elliptische aanpassingsproblemen. Zie hiervoor het in de bibliografie genoemde artikel van Grootendorst.

"Als een kegel gesneden wordt met een vlak door de as, en ook gesneden wordt door een ander vlak dat beide zijden van de driehoek door de as ontmoet, en dat niet evenwijdig aan de basis getrokken is en ook niet tegengesteld is, als verder het vlak van de basis van de kegel en het snijvlak elkaar ontmoeten in een rechte loodrecht op de basis van de driehoek door de as of (loodrecht op) het verlengde daarvan (d.w.z. van die basis), dan zal elke rechte die vanaf de kegelsnede getrokken is, evenwijdig aan de doorsnede van de vlakken, tot aan de diameter van de snede, in vierkant gelijk zijn aan een zekere oppervlakte die langs een zekere rechte ligt, (namelijk een rechte) waartoe de diameter van de kegelsnede een verhouding heeft die het vierkant van de rechte getrokken uit de top van de kegel, evenwijdig aan de diameter van de kegelsnede tot de basis van de driehoek, (heeft) tot de (rechthoek) die bevat wordt door de stukken die door hem (die rechte) tot de zijden van de driehoek worden afgesneden, met als breedte het stuk dat erdoor van de diameter wordt afgesneden tot de top van de kegelsnede, en die een figuur mist (*ἔλλειπον*) die gelijkvormig is en gelijkvormig ligt met de (rechthoek) bevat door de diameter en de rechte waarlangs zij (i.e. de ordinaten) in vierkant gelijk zijn (*παρ' ην δύνανται*). Laat zo'n kegelsnede een ellips genoemd worden."
Het is niet aan te nemen, dat de lezer na het lezen van deze volzin ook werkelijk begrijpt wat er aan de hand is. Het is dan ook niet duidelijk welk doel Apollonius nastreefde met zulke enorm gecompliceerde algemene formuleringen. Gelukkig volgt hierna de *ekthesis*, dat is opnieuw de stelling, maar nu met wat notatie en een figuur (Figuur 5, overgenomen uit Heiberg's editie). Die luidt hier aldus: "Laat er een kegel zijn met top punt A en basis cirkel $B\Gamma$.



Figuur 5:

Laat hij gesneden zijn met een vlak door de as die als snede driehoek $AB\Gamma$ maakt. Laat hij door een ander vlak gesneden zijn dat beide zijden van de driehoek door de as ontmoet en dat niet evenwijdig aan de basis van de kegel getrokken is en niet tegengesteld, en laat hij als snede in het oppervlak van de

kegel lijn ΔE maken. Laat de doorsnede van het snijvlak en het vlak van de basis van de kegel ZH zijn, die loodrecht is op $B\Gamma$. Laat de diameter van de (kegel)snede ΔE zijn, en laat in E loodrecht op $E\Delta$ (lijn) $E\Theta$ getrokken zijn, en door A evenwijdig aan $E\Delta$ (lijn) AK, en laat het zo gemaakt zijn dat het (vierkant) van AK tot de (rechthoek) onder $BK\Gamma$[8] is als ΔE tot $E\Theta$. Laat een of ander punt Λ op de snede genomen zijn, en laat door Λ evenwijdig aan ZH (lijn) ΛM getrokken zijn. Ik zeg dat ΛM in vierkant gelijk is aan een of ander oppervlak, dat langs $E\Theta$ ligt, als breedte (lijn) EM heeft, en een figuur mist die gelijkvormig is aan de (rechthoek) onder $\Delta E\Theta$."
Hierop volgt het bewijs:
"Want laat $\Delta\Theta$ getrokken zijn, en door M laat $M\Xi N$ evenwijdig aan ΘE getrokken zijn, en door Θ en Ξ laat ΘN en ΞO evenwijdig aan EM getrokken zijn, en laat door M evenwijdig aan $B\Gamma$ ΠMP getrokken zijn. Dan, omdat ΠP evenwijdig is aan $B\Gamma$, en ook ΛM evenwijdig is aan ZH, is het vlak door ΛM en ΠP evenwijdig aan het vlak door ZH en $B\Gamma$, dat is de basis van de kegel. Als dus door ΛM en ΠP een vlak wordt gelegd, zal de snede een cirkel met diameter ΠP zijn. En ΛM is een loodlijn daarop. Dus is de (rechthoek) onder ΠMP gelijk aan het (vierkant) van ΛM. Maar omdat het (vierkant) van AK staat tot de (rechthoek) onder $BK\Gamma$ als $E\Delta$ tot $E\Theta$, en de verhouding van het (vierkant) van AK tot de (rechthoek) onder $BK\Gamma$ samengesteld is uit (de verhouding die) AK heeft tot KB en AK staat tot $K\Gamma$, maar AK staat tot KB is gelijk aan EH staat tot HB, dat is EM tot $M\Pi$, en AK staat tot $K\Gamma$ is ΔH staat tot $H\Gamma$, dat is ΔM staat tot MP, daarom is de verhouding van ΔE staat tot $E\Theta$ samengesteld uit die van EM tot $M\Pi$ en die van ΔM tot MP. De verhouding, samengesteld uit die van EM tot $M\Pi$ en die van ΔM tot MP, is de verhouding van de (rechthoek) onder $EM\Delta$ tot de (rechthoek) onder ΠMP. Dus de (rechthoek) onder $EM\Delta$ staat tot de (rechthoek) onder ΠMP als ΔE staat tot $E\Theta$, dat is als ΔM staat tot $M\Xi$. Zoals ΔM staat tot $M\Xi$, als ME als gemeenschappelijke hoogte genomen wordt, zo staat de (rechthoek) onder ΔME tot de (rechthoek) onder ΞME. Dus de (rechthoek) onder ΔME staat tot de (rechthoek) onder ΠMP als de (rechthoek) onder ΔME staat tot de (rechthoek) onder ΞME. Dus de (rechthoek) onder ΠMP is gelijk aan de (rechthoek) onder ΞME. Er is aangetoond dat de (rechthoek) onder ΠMP gelijk is aan het (vierkant) van ΛM. Dus de (rechthoek) onder ΞME is gelijk aan het (vierkant) van ΛM. Dus ΛM is gelijk in vierkant aan MO, die langs ΘE ligt, breedte EM heeft, en die een figuur ON mist die gelijkvormig is aan de (rechthoek) onder $\Delta E\Theta$. Laat zo'n snede ellips genoemd worden, lijn $E\Theta$ (de lijn) waarlangs in vierkant gelijk zijn de (rechten) die naar ΔE geordend getrokken worden, en laat dezelfde rechte ($E\Theta$) ook opstaande (zijde) heten, $E\Delta$ dwarse (zijde)."
Dit bewijs moet gemakkelijk te volgen geweest zijn voor een lezer die ervaring had met het werken met rechthoeken en verhoudingen. De doorsnee lezer uit de Griekse oudheid had deze ervaring in ruime mate verkregen door het bestuderen

[8] In moderne termen BK maal $K\Gamma$. Apollonius gaat er niet van uit dat BK en $K\Gamma$ in de figuur een rechthoek maken, en dat is hier ook niet zo. Hij bedoelt een rechthoek waarvan de ene zijde gelijk is aan BK en de tweede zijde aan $K\Gamma$.

van de *Elementen* van Euclides. Het bewijs is gebaseerd op een eigenschap van de cirkel, $\Lambda M^2 = \Pi M \cdot MP$, die in de *Elementen* bewezen wordt. Modern komt dit neer op de vergelijking $y^2 = (r-x)(r+x)$ als we de oorsprong in het middelpunt van de cirkel nemen, de straal r stellen, en $\Lambda M = y, \Pi M = x$ kiezen. De rest van het bewijs berust op het gebruik van gelijkvormigheden. Hieruit komt de vergelijking van de ellips met $t = E\Delta, p = E\Theta = (\frac{BK \cdot K\Gamma}{AK^2}) \cdot t$. Uit het bewijs blijkt dat p en t onafhankelijk zijn van de ordinaat ΛM en alleen afhangen van de kegel en het snijvlak. De hoek tussen de diameter $E\Delta$ en de ordinaten ΛM is constant maar hoeft niet recht te zijn. Verder zit er nog een klein addertje onder het gras. Apollonius begint met het kiezen van een driehoek door de as van de kegel en hij snijdt de kegel daarna met een vlak dat de basis van de kegel snijdt in een lijn loodrecht op de basis van die driehoek. De lezer moet zelf inzien dat als een kegel met een vlak gesneden wordt dat als snede een ellips oplevert, het altijd mogelijk is een vlak door de as te kiezen dat de basis snijdt in een lijn loodrecht op de doorsnede van basis en snijvlak. Dit vlak door de as snijdt de kegel dan in de "driehoek door de as". Pas dan is duidelijk dat de stelling op dit snijvlak toegepast kan worden.
In het algemeen veronderstelt Apollonius nogal veel inzicht van de lezer. Dit blijkt ook uit het feit dat wanneer Apollonius een eerdere stelling uit de *Conica* noemt, hij dit bijna nooit vermeldt; kennelijk verwacht hij dat de lezer de verbanden zelf legt. Het bovenstaande bewijs is relatief gemakkelijk; de meeste bewijzen in de *Conica* zijn zeer veel gecompliceerder. In Boek V, over maximale en minimale afstanden van een gegeven punt in het vlak tot een gegeven kegelsnede, komen stellingen voor die zo ingewikkeld zijn dat het Griekse alfabet niet voldoende letters bevat om de punten in de figuren alle een naam te geven.
Een bijzonder aspect van de stijl van Apollonius is zijn gebruik van werkwoordsvormen (zie hiervoor HEATH, p. clxv). Hij zegt niet vaak "we trekken een lijn" maar bijna altijd "laat een lijn getrokken zijn". Dit taalgebruik past in de al eerder genoemde Griekse opvatting dat de wiskundige objecten zich bevinden in een eeuwige wiskundige wereld. Iets construeren is daarom in feite onmogelijk, omdat alles er al is. In deze wereld kunnen we alleen objecten aanwijzen.
Voor verdere voorbeelden van bewijzen van Apollonius verwijzen we de lezer naar de Franse vertaling van VER EECKE, of de Engelse parafrase van HEATH. Aanbevolen lectuur is bijvoorbeeld proposities 34 en 36 van Boek I, over de existentie en uniciteit van raaklijnen aan een ellips en een hyperbool in een gegeven punt. Apollonius bewijst dit op een correcte en fraaie manier, die verrassend anders is dan tegenwoordig gebruikte methoden.

6. WAAROM KEGELSNEDEN?

Er bestonden enkele toepassingen van kegelsneden in de theorie van brandspiegels. Het was in de tijd van Apollonius bekend dat een parabolische spiegel (die ontstaat door een parabool om zijn as te roteren) de zonnestralen evenwijdig aan de as naar het brandpunt terugkaatst. Er is echter geen aanwijzing dat zulke spiegels in de oudheid zijn gemaakt. Men kan daarom beter spreken

van een pseudo-toepassing, die in dezelfde categorie valt als de constructie van Menaechmus van een altaar met inhoud twee maal die van een gegeven altaar. Het heeft 1800 jaar geduurd voordat bekend werd dat de stellingen uit de *Conica* van Apollonius serieuze toepassingen hadden buiten de wiskunde. Dit werd duidelijk in 1604, toen Kepler ontdekte dat de planeten elliptische banen beschrijven met de zon in een brandpunt.
In de Griekse oudheid werden kegelsneden wel veel gebruikt om andere problemen uit de (zuivere) wiskunde op te lossen. Een eenvoudig voorbeeld is de trisectie van een gegeven hoek. Een veel ingewikkelder voorbeeld is het probleem van de *locus van drie of vier lijnen.* Gegeven: een verhouding α, vier rechte lijnen l_i, en vier andere rechte lijnen m_i, alle in hetzelfde vlak, zodat lijn m_i niet evenwijdig is aan lijn l_i voor $1 \leq i \leq 4$. Gevraagd alle punten P in het vlak die aan de volgende eigenschap voldoen: Van P trekken we een lijn evenwijdig aan m_i die l_i in Q_i snijdt. Er moet nu gelden

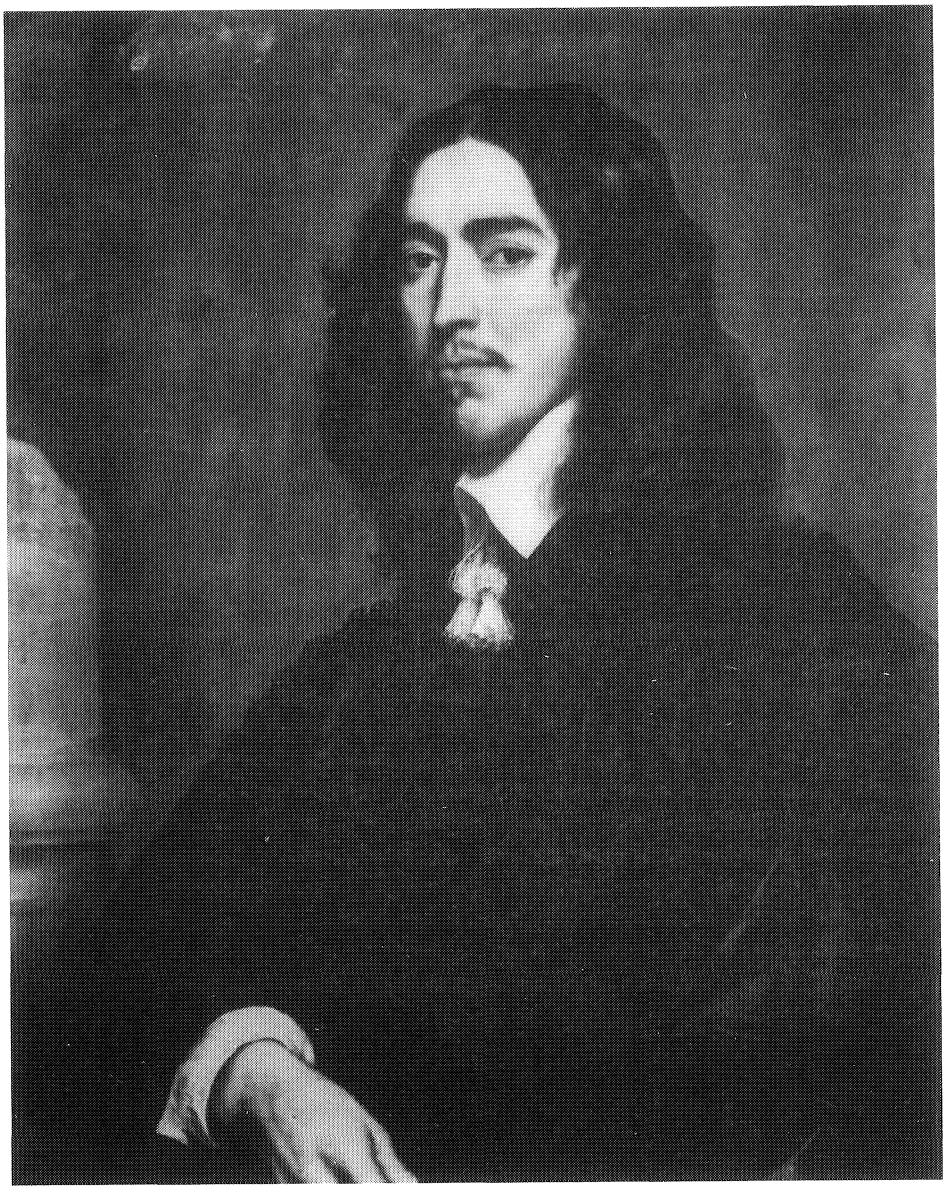
$$\frac{PQ_1 \cdot PQ_2}{PQ_3 \cdot PQ_4} = \alpha.$$

Het blijkt dat de verzameling punten P die aan deze eigenschappen voldoen uit twee kegelsnedes bestaat. Volgens Apollonius had Euclides dit probleem gedeeltelijk opgelost, en een deel van de motivatie van zijn eigen *Conica* is het geven van een volledige oplossing van dit probleem.[9]
Echter, voor Apollonius hebben zijn wiskundige stellingen een intrinsieke waarde, en toepassingen in andere wiskundige problemen of buiten de wiskunde waren voor hem niet essentieel. Zijn *Conica* is een monument van de zuivere wiskunde. Het is een wonder dat dit boek de late oudheid en de vroege middeleeuwen overleefd heeft. Het voortbestaan van de boeken V tot VII is te danken aan de Arabische cultuur, waar meer belangstelling voor zuivere wetenschap bestond dan in het middeleeuws Christelijke Europa.
Toen de Europese wiskundigen uit de zeventiende eeuw belangstelling kregen voor krommen, vonden zij in de *Conica* een enorme massa interessant materiaal. Zij hebben het arsenaal van krommen geweldig uitgebreid, en waren al spoedig in staat de kegelsneden op elegantere manier te behandelen, en daardoor had de *Conica* vanaf het eind van de zeventiende eeuw alleen nog historische waarde. De invloed van dit boek is desondanks enorm geweest, en de woorden ellips, parabool en hyperbool zullen ons daaraan blijven herinneren.

[9] De generalisatie van dit probleem tot zes of meer lijnen is behandeld door René Descartes in zijn Géométrie (1637).

LITERATUUR

1. E.J. DIJKSTERHUIS, *Archimedes.* Kopenhagen 1956. (Engelse vertaling. Het hoofdstuk over de kwadratuur van de parabool is in de oorspronkelijke Nederlandse versie verschenen in het tijdschrift *Euclides* 17 (1940), pp. 31-40.)
2. A. GROOTENDORST, "Over de geometrische algebra van de Grieken en de oorsprong van de woorden parabool, ellips en hyperbool". In: A. GROOTENDORST, *Grepen uit de geschiedenis van de wiskunde.* Delft 1988, pp. 49-63.
3. T.L. HEATH, *Apollonius of Perga, treatise on conic sections.* Cambridge 1896.
4. J.L. HEIBERG (ED.), *Apollonii Pergaei quae Graece exstant cum commentariis antiquis,* Leipzig (Teubner) 1890. 2 vols.
5. J.L. HEIBERG (ED.), *Archimedes Opera Omnia.* 3 vols. Leipzig (Teubner) 1915.
6. W.R. KNORR, *The ancient tradition of geometric problems.* Boston 1986. Herdruk New York (Dover 1993).
7. G. MORROW, *Proclus. A commentary on the First Book of Euclid's Elements.* Princeton (Princeton University Press) 1970.
8. I. THOMAS, *Selections illustrating the history of Greek mathematics.* 2 vols. Cambridge - London, 1968 (Loeb Classical Library no. 335)
9. P. VER EECKE, *Les Coniques d'Apollonius de Perge.* Oeuvres traduites pour la première fois du Grec en Français, Brugge 1928.
10. B.L. VAN DER WAERDEN, *Ontwakende wetenschap,* Groningen 1950
11. H. ZEUTHEN, *Die Lehre von den Kegelschnitten im Altertum.* Kopenhagen 1886, herdruk Hildesheim 1966.



Johan de Witt (1625 – 1672)

Adriaen Hanneman (1601-1671) pinxit.
BRON: Museum Boymans - Van Beuningen, Rotterdam

De "Kegelsneden" bij Johan de Witt

A.W. Grootendorst

INLEIDING

1. In 1637 verscheen in Leiden -het Presidium Libertatis- bij de drukkerij van Jan Maire het eerste gedrukte werk van René Descartes, getiteld *"Discours de la Méthode pour bien conduire sa raison et chercher la vérité dans les sciences"*, gevolgd door drie appendices, Essais genoemd en getiteld: *"La Dioptrique"*, *"Les Météores"* en " La Géométrie".
De derde toevoeging -La Géométrie- betekende niet minder dan een revolutie in de beoefening van de wiskunde; hierin werd namelijk de (letter-) algebra toegepast in de meetkunde, waardoor de weg vrijgemaakt was voor datgene wat wij nu kennen als Analytische Meetkunde.

2. Tot degenen die direct al gefascineerd raakten door dit werk, behoorde onze landgenoot FRANS VAN SCHOOTEN DE JONGERE (1615–1660) die in 1646 zijn vader –FRANS VAN SCHOOTEN DE OUDERE (1581?–1645)– zou opvolgen als hoogleraar in de "Nederduitsche Mathematique en Wiskunde" aan de Academia Illustris, de aan de Leidse Universiteit verbonden ingenieursschool. Met "Nederduitsche Mathematique" wordt bedoeld wiskunde die van belang was voor landmeters en militaire ingenieurs; deze werd in de landstaal gedoceerd.
Van Schooten zag het belang van dit werk in, maar vond het moeilijk. Het resultaat van zijn grondige bestudering van de Géométrie, waarbij hij ook vele collegae betrok, was dat hij besloot dit werk, dat in het Frans geschreven was en daardoor voor velen minder toegankelijk, te vertalen in de Lingua Franca van de geleerde wereld uit die tijd: het Latijn.

3. De eerste editie van deze vertaling verscheen in 1649 –eveneens bij Jan Maire– onder de titel die wij hier beknopt weergeven:
"Geometria, a Renato Des Cartes Anno 1637 gallice edita, nunc autem cum Notis F. de Beaune..... in linguam Latinam conversa et commentariis illustrata opera atque studio Fr. à Schooten ....."
Inderdaad had Van Schooten zich uitgeput in een uitvoerig commentaar waarin hij op alle details inging en van grote belezenheid blijk gaf. Ook FLORIMOND DE BEAUNE (1601–1652) een vriend van Van Schooten, had –zoals de titel al aangeeft– zijn bijdragen geleverd.
Ook hierna bleef dit werk Van Schooten boeien en hij ontplooide zich in woord en geschrift tot een vurig profeet van de gedachten van Descartes en bracht zijn enthousiasme over op zijn studenten die van grote mathematische allure waren.

Tot hen behoorden o.a. CHRISTIAAN HUYGENS (1629–1695), HENDRICK VAN HEURAET (1634–1660?), JAN HUDDE (1628–1704) en degene wiens werk in deze voordracht centraal staat, JAN DE WITT (1625–1672).

4. Mede door het grote succes van zijn geannoteerde vertaling bereidde Frans van Schooten een tweede druk voor die in twee delen verscheen. Het eerste in 1659, het tweede kwam in 1661 uit, een jaar na zijn plotselinge dood in 1660. Het werd bezorgd door zijn broer PIETER VAN SCHOOTEN (1634–1679).
Deze tweede druk, uitgebracht door de firma Lodewijk Elzevier in Amsterdam, onderscheidde zich van de eerste, allereerst door vele verbeteringen, maar vooral ook door de toevoeging van bijdragen van leerlingen en collegae.
Het eerste deel bevatte –naast de inhoud van de eerste druk– twee brieven van Jan Hudde, burgemeester van Amsterdam. Een over het oplossen van vergelijkingen en een over een methode om (uiteraard zonder differentiatie) maxima en minima te bepalen met behulp van de befaamde "regel van Hudde" (zie ook litt. [9], pp. 81–106).
Tenslotte is er dan de brief van Van Heuraet over het bepalen van de lengte van krommen "Epistola de Transmutatione Curvarum Linearum in Rectas" (zie ook litt. [9], pp. 107–122).
Het tweede deel begint met een artikel van Van Schooten zelf, waarin de methode van Descartes in zijn algemeenheid uiteengezet wordt. Daarna volgen twee opstellen van De Beaune over algebraïsche vergelijkingen en dan –voorafgaande aan het slothoofdstuk van Van Schooten over kansrekening– komt de bijdrage van Jan de Witt getiteld "Elementa Curvarum Linearum" in twee delen, Liber Primus en Liber Secundus.
Voordat wij hierop nader ingaan nog eerst het verloop van de lotgevallen van het werk van Van Schooten. Na de liquidatie van het bedrijf van Elzevier in 1681 verscheen in 1683 een derde, ongewijzigde druk bij de firma Blaeu, eveneens in Amsterdam. In 1695 kwam de vierde en laatste editie, vermeerderd met opmerkingen van JACOB (I) BERNOULLI (1654–1705) uit bij de firma Knoch in Frankfurt.

5. Jan de Witt had zijn werk in essentie al in 1649 voltooid, maar eerst in 1656 de definitieve vorm ervan vastgesteld. In die tussenliggende tijd werd hij geroepen tot hoge ambten, die een zwaar beslag op hem legden. Zo werd hij in 1650 benoemd tot Pensionaris van Dordrecht, in 1653 tot Raadpensionaris van Holland.
Bij de voltooiing van zijn werk ondervond hij de medewerking van zijn leermeester die hem op een aantal onvolkomenheden had gewezen. Hiervoor bedankt Jan de Witt hem in de brief die, gedateerd 8 oktober 1658, aan het werk vooraf gaat.
In deze brief zet Jan de Witt zeer duidelijk zijn bedoelingen uiteen. Hij stoorde zich namelijk aan het feit dat de bekende kegelsneden: parabool, hyperbool en ellips die toch *vlakke* krommen, zijn via *ruimtelijke beschouwingen* gegenereerd worden als doorsnijding van een plat vlak en een ruimtelijk lichaam, een kegel. Wij laten hem hierover aan het woord:

"Toen ik echter de leerboeken van de overige kromme lijnen –voorzover deze door de Ouden zijn overgeleverd en door lateren zijn verklaard– nauwkeurig had bestudeerd, achtte ik het volslagen tegen de natuurlijke orde –die men in de wiskunde zoveel mogelijk in acht moet nemen– in te gaan dat men de oorsprong van deze krommen zoekt in ruimtelijke lichamen en hen vervolgens overbrengt naar het platte vlak".
Wat hem duidelijk voor ogen stond was een constructieve behandeling van de kegelsneden.
Ook Van Schooten hechtte daaraan; deze had in het vierde boek van zijn "Exercitationum mathematicarum libri quinque" ("Mathematische Oeffeningen"), getiteld "De organica conicarum sectionum in plano descriptione tractatus" etc. (Tuych-werckelijcke beschrijving der Kegel-sneden op een vlack) zich veel moeite gegeven voor een mechanische beschrijving van kegelsneden. Hiermee was mede een practisch belang gediend: kegelsneden waren van nut bij het vervaardigen van zonnewijzers, in de optica, zowel bij problemen van breking als van terugkaatsing (dioptrica en katoptrica) en ook in de leer van de perspectief. Op dit gebied is zijn invloed op Jan de Witt duidelijk. Zo vinden we de wijze waarop deze de ellips definieert, terug in het laatstgenoemde werk van Van Schooten.

6. Het werk, de Elementa, bestaat zoals gezegd uit twee boeken. Twee boeken van formeel zeer verschillende aard; het eerste louter verbaal, zonder enig wiskundig symbool of formule, het tweede uniek voor die tijd door de voorstelling van punten door getallenparen en van krommen door vergelijkingen.
De omvang van beide is bescheiden : iii + 85 pag. voor boek 1; 98 pag. voor boek 2, beide verduidelijkt door scherpe figuren, waarnaar helaas meestal niet expliciet verwezen wordt.
Een probleem bij eerste lezing is ook dat rechten steeds met twee letters zijn aangegeven bijv "recta AB", waarbij A en B dan nog geen specifieke punten representeren, maar daarna ineens de snijpunten voorstellen met later geïntroduceerde rechten of andere krommen. Ook betekent "recta" nu eens "lijnstuk", dan weer "lijn".
De denkwijze in het eerste boek is puur meetkundig. Een voorbeeld: de formule $AB^2 = DE.EF$ wordt als volgt omschreven: recta AB potest rectangulum DEF, hetgeen, letterlijk vertaald, zou betekenen: het lijnstuk AB vermag de rechthoek DEF.
De bedoeling is duidelijk: wanneer men het lijnstuk AB opvat als de zijde van een vierkant, dan is de oppervlakte daarvan dezelfde als die van de rechthoek met zijden DE en EF. Overigens is het interessant op te merken dat hier de oorsprong van de woorden macht, power, puissance, Potenz in hun meetkundige betekenis te zien is.

7. In deze voordracht zal de nadruk liggen op boek 1 en we zullen dan ook achtereenvolgens parabool, hyperbool en ellips bespreken op de wijze waarop en in de volgorde waarin, Jan de Witt dit doet in zijn "Elementa" met voornamelijk aandacht voor de manier waarop hij deze introduceert en identificeert

met de "klassieke" kegelsneden.
In de gedrukte tekst vindt men een nagenoeg volledige weergave van de inhoud van Liber Primus, weliswaar in zeer beknopte vorm, met vaak slechts een schetsmatige aanduiding van de bewijzen. Voor details wordt de lezer verwezen naar litt. [A].
Van Liber Secundus zullen in hoofdstuk 5 van deze tekst alleen de hoofdlijnen worden aangegeven.

8. Als algemene opmerking dit: De Witt legt geen expliciet verband tussen deze drie krommen. Slechts in hoofdstuk IV van zijn boek, getiteld "Alia Parabolam, Hyperbolam et Ellipsin in plano delineandi Methodus" ("Een andere methode om een parabool, een hyperbool en een ellips in het platte vlak te tekenen", hier samengevat in het hoofdstuk 4 "Parabool en Hyperbool revisited"), beziet hij parabool en hyperbool vanuit één standpunt, door deze op dezelfde wijze voort te brengen.
Ook zij er nog op gewezen dat de methode waarop De Witt de parabool en de hyperbool genereert, uitgelegd zou kunnen worden als projectieve voortbrenging met een punt op oneindig, maar dat is een gebruik van projectiviteit "avant la lettre" en zeker niet door hem als zodanig bedoeld.

JOHANNIS DE WITT

ELEMENTA

CVRVARVM LINEARVM.

Edita

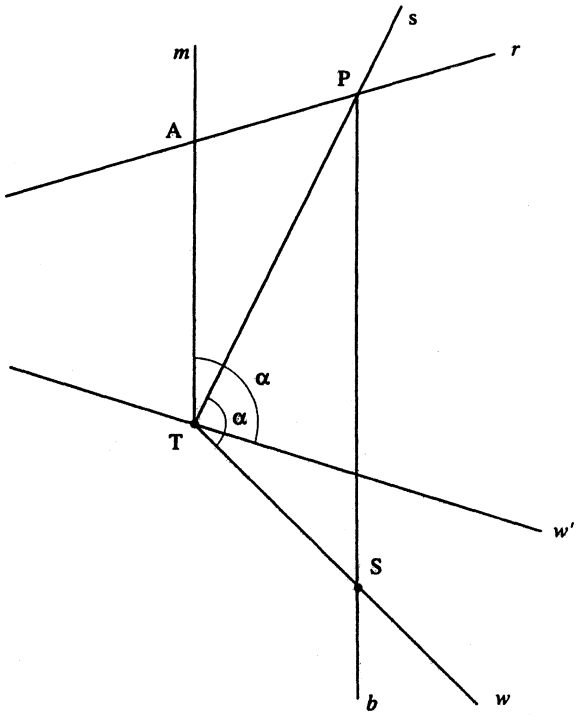
Operâ FRANCISCI à SCHOOTEN,
in Academia Lugduno-Batava Matheſeos
Profeſſoris.

AMSTELODAMI,
Ex Typographia BLAVIANA, MDCLXXXIII.
Sumptibus Societatis.

Titelpagina van Johan de Witt's *Elementa Curvarum Linearum*

1. DE PARABOOL

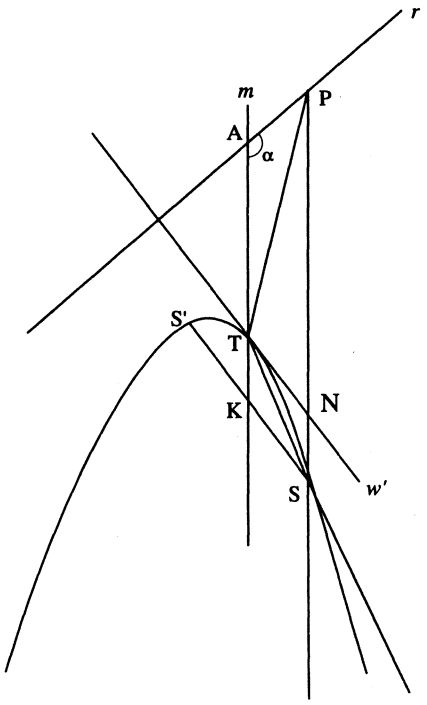
1.1. Voor het beschrijven van een parabool gaat Jan de Witt uit van een rechte r (de richtlijn), een punt T (de pool) en een rechte m door T die de richtlijn snijdt in het punt A; AT wordt het interval genoemd. De bewuste kromme ontstaat nu op de volgende wijze: een hoek α (de bewegende hoek of werkhoek) met benen s (het sleepbeen) en w (het werkbeen; de drager ervan heet werklijn) draait om de pool T als hoekpunt. Het draaiende been s snijdt de richtlijn daarbij in een variabel punt P. Door P wordt steeds een rechte b getrokken (de beschrijvende) evenwijdig aan m. Het gaat nu om de baan van het snijpunt S van de beschrijvende b met het werkbeen w van de hoek α (zie fig. 1.1).



FIGUUR 1.1

Deze baan zal blijken een parabool te zijn indien de hoek α gelijk is aan de hoek tussen m en r (aan dezelfde kant van m). Indien deze hoeken verschillend zijn, is de baan een hyperbool.

1.2. We bezien hier het eerste geval waarin dus $\angle PAT = \angle\alpha$ (zie fig. 1.2). In deze figuur is w' de werklijn "in de beginstand", d.w.z. de hoek ATN tussen m en w' is juist α. Eenvoudig is in te zien dat dan $AP = TN$ en $\angle ATP = \angle NTS$. Trekt men nog SK evenwijdig aan NT, dan geldt ook $SK = NT = PA$ en $\angle NTS = \angle KST$, dus $\angle ATP = \angle NTS = \angle KST$ en $\angle TAP = \angle ATN = \angle SKT$, zodat de driehoeken PAT en TKS gelijkvormig



FIGUUR 1.2

zijn, hetgeen inhoudt:

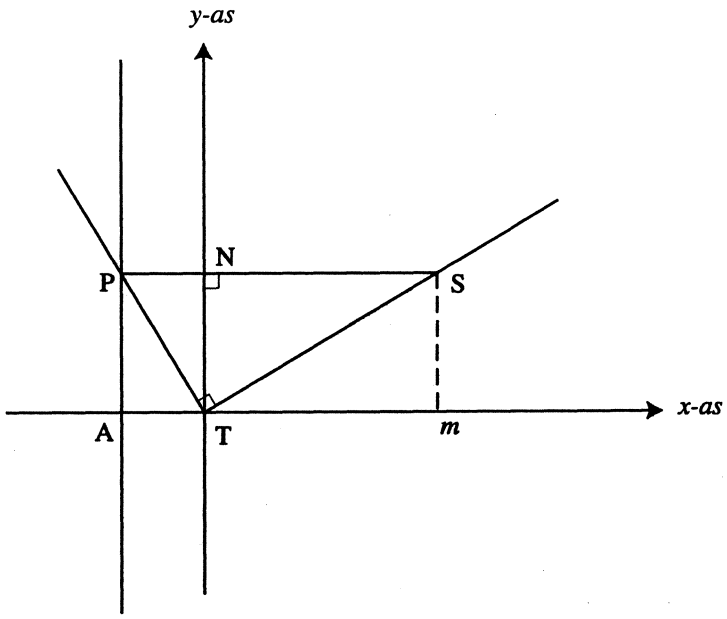
$$PA : TK = AT : KS,$$

maar $PA = KS$, dus

$$KS^2 = AT.TK.$$

Dit nu was voor Apollonius de kenmerkende eigenschap van een parabool met middellijn m en als top daarop het punt T (de pool) en met als *–bij deze middellijn bijbehorende–* parameter (latus rectum, rechte zijde [9]) het interval AT. Jan de Witt noemt dan ook de lijn m in het vervolg middellijn en –als deze loodrecht staat op w' –de as.
Kiest men T als oorsprong van een assenkruis met x– en y-as resp. m en w' en stelt men $AT = p$, dan is de vergelijking van deze parabool $y^2 = px$.
Een eenvoudig geval krijgt men indien de hoek α recht is en de richtlijn loodrecht staat op het interval (zie fig.1.3). In de rechthoekige driehoek PTS geldt $TN^2 = PN.NS$. Stelt men weer $AT = p$ en kiest men m en de drager van TN resp. als x- en y- as, dan wordt de parabool met vergelijking $y^2 = px$ beschreven. Terloops zij opgemerkt dat dit een eenvoudige manier geeft om met behulp van een tekendriehoek op gelinieerd papier punt voor punt een parabool te tekenen.

1.3. Van groot belang voor het vervolg is het begrip "geordend aangebracht" (ordinatim applicatus). Dit is niet alleen hier van belang maar ook bij de hy-

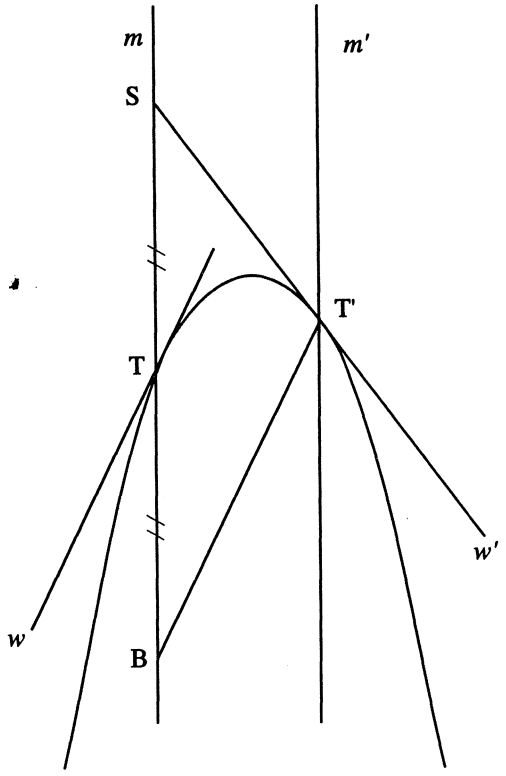


FIGUUR 1.3

perbool en de ellips. In het geval van de parabool heet een rechte geordend aangebracht op de middellijn m, indien deze rechte evenwijdig is met het werkbeen in de beginstand zoals bijvoorbeeld de rechte SKS' in fig. 1.2.
Uit deze wijze van voortbrengen van een parabool trekt Jan de Witt zijn eerste conclusies waarvan wij hier de belangrijkste laten volgen.

a. Een rechte die evenwijdig verloopt aan de middellijn snijdt de parabool in slechts één punt.
b. Een koorde die geordend is aangebracht op de middellijn wordt daardoor gehalveerd. Zo geldt immers in fig. 1.2 op $SKS'(//w')$ zowel SK^2 als $S'K^2 = AT.TK$, dus $SK = S'K$.
Het is niet moeilijk in te zien dat ook het omgekeerde waar is: een koorde die door de middellijn wordt gehalveerd, is daarop geordend aangebracht, d.w.z. verloopt evenwijdig met de werklijn in de beginstand.
c. De werklijn in de beginstand raakt aan de parabool. Jan de Witt bewijst dit door op te merken dat elk punt (m.u.v. de top T) en dus de gehele kromme , m.u.v. de top, krachtens de constructie "onder" de werklijn in beginstand ligt. Ook toont hij de eenduidigheid van deze raaklijn aan.
d. Tenslotte merkt hij op dat alle koorden die evenwijdig zijn aan de raaklijn in de top, gehalveerd worden door de middellijn en, omgekeerd, dat een lijn door de top die evenwijdig is aan een koorde die door de middellijn wordt gehalveerd, in de top aan de parabool raakt.

1.4. De tweede grote stelling in het hoofdstuk "parabool" houdt het volgende in: Indien men een willekeurig punt T' op de parabool aanneemt en daardoor een rechte m' evenwijdig aan de middellijn m trekt, dan is ook m' middellijn en T' de bijbehorende top van de gegeven parabool.
Hiermee wordt het volgende bedoeld: indien men met T' als top, met geschikt gekozen interval (= parameter) op m' en eveneens geschikt gekozen werkhoek volgens het bekende voorschrift een parabool construeert, dan blijkt deze samen te vallen met de gegeven parabool.



FIGUUR 1.4

Het bewijs is vrij gecompliceerd en moet hier achterwege blijven (zie hiervoor litt.[A]). De essentie is het volgende (zie fig. 1.4).
Laat de parabool beschreven zijn met middellijn m, top T, interval TA en w als werklijn in de beginstand; werkhoek en richtlijn zijn dan ook bekend.
De werklijn (in beginstand) behorende bij T' wordt nu als volgt bepaald: men brengt door T' een lijn $T'B$ geordend aan op m, d.w.z. evenwijdig aan w en

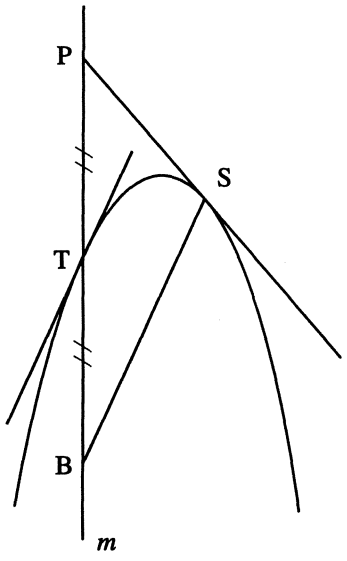
bepaalt vervolgens het punt S op m zodanig dat $TS = TB$. Voor de gezochte werklijn w' in de beginstand, behorende bij T', kiest men dan $T'S$.
Het bijbehorende interval (de parameter) p wordt gedefinieerd als de derde evenredige bij TS en $T'S$, dus $TS : T'S = T'S : p$.
Men kan dan bewijzen dat de parabool die beschreven wordt met deze gegevens als uitgangspunt, samenvalt met de oorspronkelijke parabool. Hiervan kan men dus ieder punt met bijbehorende m' evenwijdig aan m, opvatten als top, resp. middellijn; tevens kan men de aan m' toegevoegde richting construeren.
Een belangrijk gevolg is dat alles wat geldt voor de oorspronkelijke middellijn, nu ook geldt voor iedere middellijn. Zo is bijvoorbeeld de aan m' toegevoegde richting bekend (nl. $// ST'$).

1.5. Uit deze constructie trekt Jan de Witt enkele conclusies.

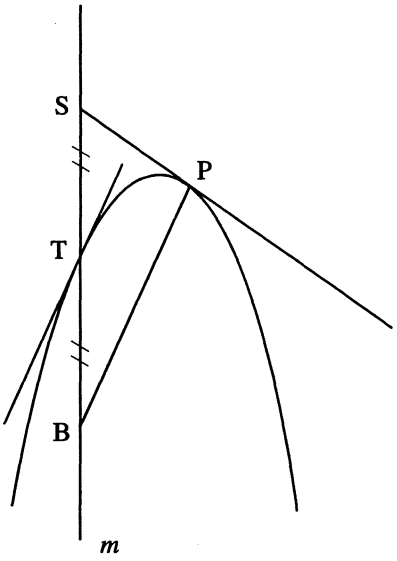
a. Op grond van 1.3.b kan men een willekeurige middellijn van een (als kromme) gegeven parabool vinden met de bijbehorende toegevoegde richting. Daartoe behoeft men slechts twee evenwijdige koorden te halveren en hun middens te verbinden. Hierbij zij opgemerkt dat Jan de Witt –sans scrupules– twee punten op de parabool kiest en deze verbindt om een koorde te vinden.
b. Uit de eenduidigheid van de raaklijn in een top van de parabool leidt hij af dat de rechte ST' in T' aan de parabool raakt indien $T'B$ geordend is aangebracht op m en tevens $BT = TS$.

1.6. Hierna stelt Jan de Witt de constructie van de raaklijn aan een parabool in of vanuit een punt P aan de orde. Hiertoe onderscheidt hij 4 gevallen.

i. P ligt op de parabool en wel op een middellijn waarvan de toegevoegde richting bekend is. Direct duidelijk is dat de lijn door P in deze richting raakt aan de parabool in P.
ii. P ligt op de parabool, maar niet op een bekende middellijn. We zagen in 1.5.a hoe men dan toch een middellijn m met de bijbehorende toegevoegde richting kan vinden. Laat deze de parabool snijden in T (fig. 1.5). We brengen vervolgens PB geordend aan op m en bepalen tenslotte het punt S op m zodanig dat $BT = TS$. Volgens 1.5.b raakt SP dan in P aan de parabool.
iii. P ligt buiten de parabool op het verlengde van een bekende middellijn m. (zie fig. 1.6). Volgens 1.4 kan men de aan m toegevoegde richting vinden. Laat m de parabool snijden in T. We bepalen dan het punt B op zodanig dat $PT = TB$ en brengen BS geordend aan op m. Volgens 1.5.b raakt PS dan in S aan de parabool.
iv. P ligt buiten de parabool, maar niet op het verlengde van een bekende middellijn. Via de bekende constructie is dan toch een middellijn, zeg m, met toegevoegde richting te vinden. Door P trekken we dan een rechte $m'//m$. Deze is dan eveneens middellijn en we zijn weer in de sub iii besproken situatie.



FIGUUR 1.5



FIGUUR 1.6

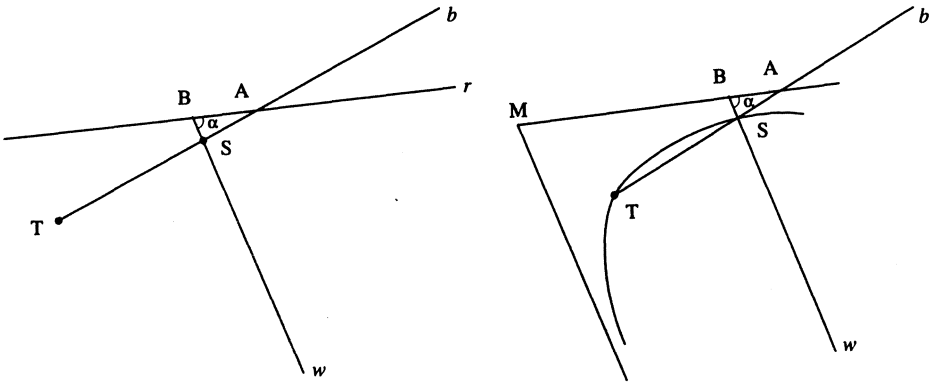
1.7. Als laatste probleem m.b.t. de parabool stelt Jan de Witt de vraag om van een parabool waarvan gegeven zijn: een middellijn, de top daarop, de parameter en de werkhoek, een andere middellijn te vinden met gegeven toegevoegde richting. Ook wordt de bijbehorende parameter gevraagd. Voor de uitwerking hiervan verwijzen we naar litt. [A].

1.8. Het zou te verwachten zijn dat Jan de Witt na deze behandeling van de parabool het geval zou gaan bespreken waarin de bewegende hoek (de werkhoek) verschilt van de hoek tussen het interval en de richtlijn. Hij zet echter omstandig uiteen dat hij dit niet in eerste instantie zal doen, maar daarop later zal terugkomen. Hij heeft namelijk voor de hyperbool een andere methode op het oog die in zekere zin duaal staat tegenover de methode die hij volgde bij de parabool.
Daarbij ging het immers om een draaiende hoek en een schuivende lijn. Voor de hyperbool zal hij een methode gebruiken die –zoals opgemerkt– in zekere zin duaal daartegenover staat: het zal daarbij namelijk gaan om een draaiende lijn en een schuivende hoek.
Jan de Witt stelt dit probleem direct na de introductie van de parabool inderdaad aan de orde, maar gaat eerst aan het einde van zijn boek hier nader op in. Wij zullen dit geval behandelen in hoofdstuk 4.

2. DE HYPERBOOL

2.1. Ook hier gaat Jan de Witt uit van een vaste rechte r (de richtlijn) en een vast punt T (de pool). Door T gaat een rechte b (de beschrijvende) die om T als pool ronddraait en de richtlijn r snijdt in een variabel punt A. Dit punt

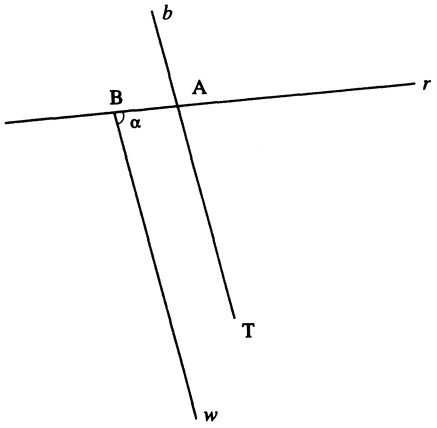
A is het eindpunt van het been BA met vaste lengte (het sleepbeen) van een hoek α van constante grootte (de bewegende hoek of de werkhoek). Het been BA ligt steeds op de richtlijn r en wordt –wanneer b om het punt T wentelt– voortgeschoven langs de richtlijn r.



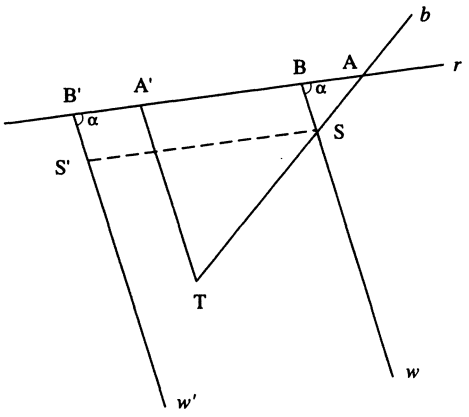
FIGUUR 2.1 en 2.1.a

Het gaat nu om de baan van het snijpunt S van de roterende beschrijvende b met het andere been w (het werkbeen) van de hoek α (zie fig. 2.1). Deze baan zal blijken een hyperbool te zijn, waarvan de beide takken "tegengestelde hyperbolen" genoemd worden. Jan de Witt spreekt steeds over een hyperbool als hij één van beide takken bedoelt.
Nu eerst enkele definities.

a. Wanneer de werklijn (d.i. de drager van het werkbeen) evenwijdig is aan de beschrijvende en deze elkaar dus niet snijden, dan zeggen we dat werklijn en beschrijvende in de beginstand zijn (fig. 2.2)
b. Het deel TA van de beschrijvende in de beginstand, dat ligt tussen de pool T en de richtlijn r, noemen we interval. Deze zelfde naam is ook gereserveerd voor de lengte van het sleepbeen AB (fig. 2.2).



FIGUUR 2.2



FIGUUR 2.3

2.2. Voor het bewijs dat de zo geconstrueerde kromme inderdaad een hyperbool is, bezien we fig. 2.3. Hierin is S een willekeurig punt op de kromme en zijn de beschrijvende TA' en de werklijn w' in de beginstand, hetgeen betekent: $TA'//w'$ en dus, omdat $A'B' = AB$, geldt: $AA' = BB'$.
We trekken nog SS' evenwijdig aan de richtlijn r. In $\Delta AA'T$ geldt dan:

$$AB : BS = AA' : A'T$$

en dus ook

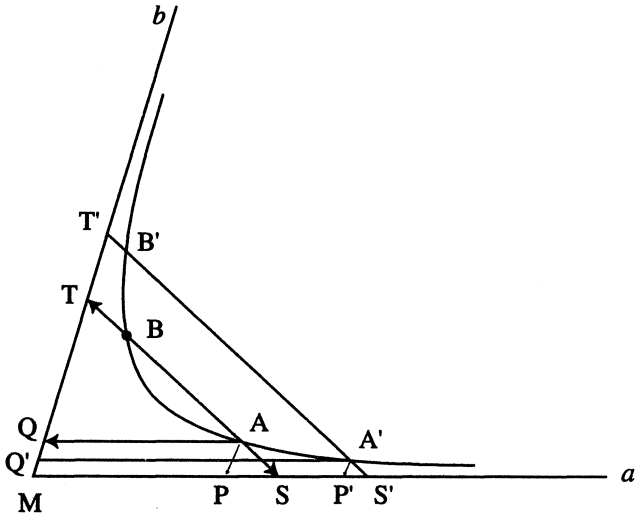
$$AB : BS = BB' : A'T$$

en dus

$$BS.SS' = BS.BB' = AB.A'T = \text{ constant.}$$

Jan de Witt merkt hierbij op dat dit nu juist de kenmerkende eigenschap is van de kromme die "de Ouden" een hyperbool noemden en waarvan de werklijn in de beginstand en de richtlijn de asymptoten zijn. Het snijpunt van beide lijnen noemt hij -voorbarig- het middelpunt.
Kiezen we deze laatste lijnen als resp. x- en y-as, dan luidt de vergelijking van deze hyperbool $xy = c$. Hierin is c het constante produkt van de beide intervallen, welk produkt de macht van de hyperbool wordt genoemd. Met zorg bewijst Jan de Witt ook dat de afstand tussen de kromme en elk van beide asymptoten kleiner gemaakt kan worden dan "elke willekeurige afstand".

2.3. Na deze eerste conclusie uit de definitie van een hyperbool leidt Jan de Witt enkele stellingen af die uitspraken doen over de afstanden van een willekeurig punt op de hyperbool tot de asymptoten, gemeten langs een koorde.



FIGUUR 2.4

We bezien hiervoor fig. 2.4. Hierin liggen A en B op een hyperbool waarvan a en b de asymptoten zijn.
Verder geldt

$$AQ//A'Q'//a;\ AP//A'P'//b;\ TAS//T'A'S'.$$

De bewering is nu

$$\begin{aligned} a. \qquad AT.AS &= A'T'.A'S' \\ b. \qquad AS.AT &= BS.BT \\ c. \qquad AS &= BT. \end{aligned}$$

Bewijs:

a. Uit de figuur lezen we af $AP : A'P' = AS : A'S'$ en $AQ : A'Q'$ $= AT : A'T'$.
Uit 2.2 volgt echter $AQ.AP = A'Q'.A'P'$, dus $AP : A'P' = A'Q' : AQ$.
Tezamen geeft dit

$$AS : A'S' = A'T' : AT, \text{ dus } AT.AS = A'T'.A'S'.$$

b. Dit bewijst men op analoge wijze.

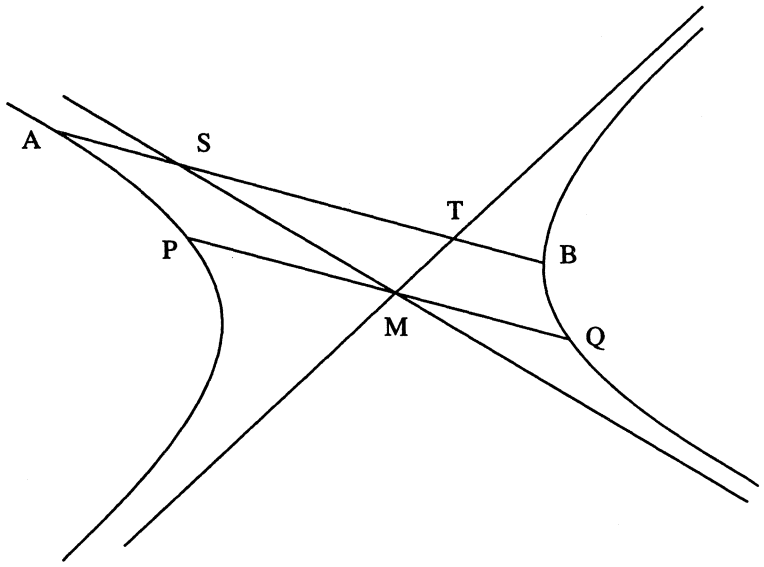
c. $AS.AT = AS(AB + BT)$ en $BS.BT = (AB + AS)BT$
maar

$$AS.AT = BS.BT, \text{ dus } AS.AB = AB.BT \text{ en } AS = BT.$$

Indien de punten A en B op "tegengestelde hyperbolen" liggen, dan gelden analoge stellingen. Let wel het gaat steeds om de afstanden van de betreffende punten tot elk van de asymptoten, gemeten langs de koorden.

2.4. Uit bovenstaande stellingen kunnen enkele conclusies getrokken worden waarvan de belangrijkste zijn:

a. Een koorde door het middelpunt die twee punten, gelegen op tegengestelde hyperbolen verbindt, wordt door het middelpunt gehalveerd. Voor het bewijs zie fig. 2.5 waarin $AB//PQ$.



FIGUUR 2.5

Hier geldt volgens 2.3.a : $PM^2 = AS.AT$ en $QM^2 = BS.BT$, volgens 2.3.b geldt echter ook $AS.AT = BS.BT$. Tezamen geeft dit $PM = QM$. Eerst nu is de naam middelpunt gerechtvaardigd. Overigens zij opgemerkt dat dit bewijs korter had gekund, immers formeel is direct te zien dat $PM^2 = QM^2$.

b. Een rechte die twee punten op een of op tegengestelde hyperbolen verbindt, heeft verder geen punt met de hyperbool gemeen. Volgens 2.3.b zou immers een eventueel derde snijpunt met een van de andere snijpunten moeten samenvallen.

c. Eenvoudig is te bewijzen dat een rechte door het middelpunt die een koorde halveert die twee punten op dezelfde of op tegengestelde hyperbolen verbindt, ook alle daarmee evenwijdige koorden halveert en dat deze onderling evenwijdige koorden de enige zijn die gehalveerd worden door deze rechte door het middelpunt.

d. Een rechte die de middens van twee evenwijdige koorden verbindt gaat door het middelpunt. Dit geeft dus een manier om het middelpunt te vinden.
Hierbij zij opgemerkt dat dit geldt voor evenwijdige koorden die twee punten op één hyperbool dan wel op tegengestelde hyperbolen verbinden.

2.5. Mede op grond van het sub 2.4 bewezene kunnen de volgende begrippen ingevoerd worden:

a. Een lijn door het middelpunt die twee punten op tegengestelde hyperbolen verbindt heet afgesneden middellijn, transversale middellijn of evt. middellijn zonder meer.

b. Een lijn door het middelpunt die verloopt in het gebied tussen twee tegengestelde hyperbolen en dus geen punt daarmee gemeen heeft, heet tweede middellijn.

c. Koorden die door een middellijn gehalveerd worden noemt men daarop geordend aangebracht.

d. Wanneer een tweede middellijn evenwijdig is met de koorden die op een transversale middellijn geordend zijn aangebracht, dan zegt men dat deze tweede middellijn toegevoegd is aan de betreffende transversale middellijn.

e. Middellijnen die onderling loodrecht zijn, noemt men assen.

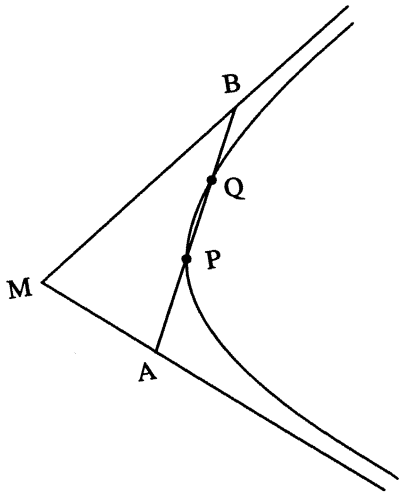
2.6

a. Uit de sub 2.4 genoemde stellingen is het duidelijk dat een lijnstuk AB (met A en B op de asymptoten) dat door een punt P op de hyperbool gehalveerd wordt, in P aan de hyperbool raakt. Jan de Witt bewijst dit door aan te tonen dat een eventueel tweede snijpunt Q samen zou vallen met P. Immers volgens 2.3.c geldt in fig. 2.6 : $PA = QB$; volgens het gegeven geldt ook $PA = PB$, dus $PB = QB$ en dus vallen P en Q samen. Formeel is dit echter nog geen bewijs dat AP een raaklijn is! Het omgekeerde van deze stelling is ook juist.

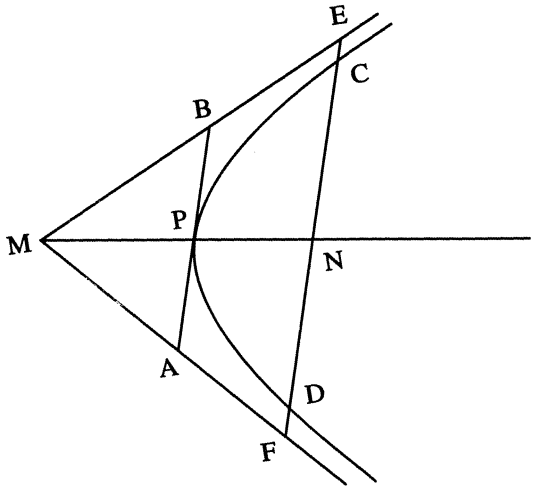
b. Uit het voorgaande volgt ook dit (zie fig. 2.7): indien een middellijn MP (met P op de hyperbool) een koorde CD halveert in N, dan zal de rechte door P evenwijdig aan CD, de hyperbool raken in P. Immers in de figuur geldt $EC = DF$, dus ook $EN = NF$ en dus $AP = PB$. Volgens het voorgaande raakt AB in P aan de hyperbool.

c. Tot slot bewijst Jan de Witt met behulp van het voorgaande dat de raaklijn in P eenduidig bepaald is.

2.7. Op dit punt in zijn betoog voert Jan de Witt enkele fundamentele begrippen in waarvan we er reeds één –de rechte zijde– ontmoetten bij de parabool,



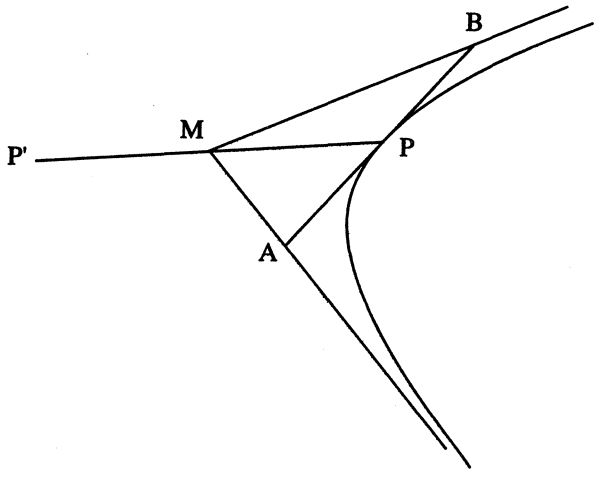
FIGUUR 2.6



FIGUUR 2.7

maar dan wel anders gedefinieerd. Analoga van de beide andere –de lengte van de middellijnen– zullen we later bij de ellips tegenkomen.

a. Wanneer P de top is op de transversale middellijn MP van een hyperbool en P' de top op de tegengestelde hyperbool, dan noemt men PP' (het dubbele dus van MP) de lengte van de transversale middellijn van de hyperbool en ook van die van de tegengestelde hyperbool (zie fig. 2.8).
b. We zagen al dat de aan $P'P$ toegevoegde tweede middellijn evenwijdig verloopt aan de raaklijn aan de hyperbool in P. De lengte van deze tweede middellijn wordt nu gedefinieerd als de lengte van het deel van de raaklijn in P dat tussen de asymptoten ligt.; in fig. 2.8 dus AB.
c. De parameter (latus rectum, dwarse zijde) p, behorende bij de gekozen transversale en tweede middellijn (in deze volgorde), wordt gedefinieerd als de derde evenredige bij PP' en AB, dus $PP' : AB = AB : p$.
Voor de achtergronden van dit begrip zij verwezen naar de voordracht van Dr. Hogendijk en naar litt. [A].

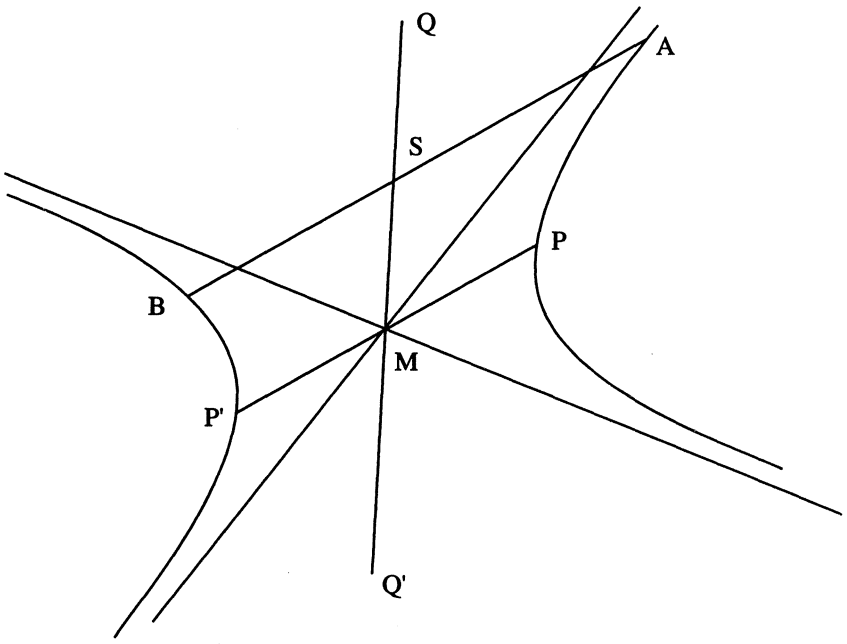


FIGUUR 2.8

2.8. Zonder bewijs vermelden we enkele stellingen m.b.t. het verband tussen een transversale middellijn en de bijbehorende tweede middellijn.

a. Een lijn door het (tweede) uiteinde van een transversale middellijn, evenwijdig met de raaklijn aan de hyperbool in de top op die middellijn, raakt aan de tegengestelde hyperbool.
b. Indien men een transversale middellijn kiest en op de daarbij behorende tweede middellijn een koorde geordend aanbrengt (d.w.z. dat deze koorde

door deze tweede middellijn gehalveerd wordt), dan is deze koorde evenwijdig aan de gekozen transversale middellijn. Zo is in fig. 2.9. QQ' geordend aangebracht op de transversale middellijn PP' en $AS = BS$. Er geldt dan: $AB//P'P$.



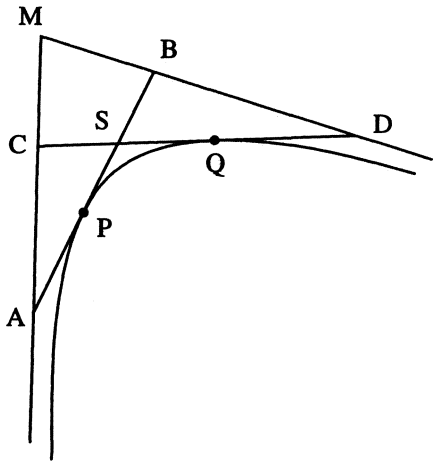
FIGUUR 2.9

Uitgaande van deze situatie kan men bewijzen dat het begrip "geconjugeerd met" oftewel "toegevoegd aan", een symmetrische relatie is. Jan de Witt gaat hier echter niet nader op in.

c. Het is mogelijk om bij gegeven toegevoegde middellijnen de assen van een hyperbool te construeren.

2.9. Bij de constructie van raaklijnen, waarmee Jan de Witt zijn hoofdstuk over de hyperbool zal besluiten, is een aantal eigenschappen van delen van raaklijnen van belang. Wij geven deze eigenschappen hier zonder bewijs (zie daarvoor litt.[A]) en verwijzen naar fig. 2.10.
In deze figuur raken AB en CD aan de hyperbool in resp. Pen Q en men kan dan bewijzen:

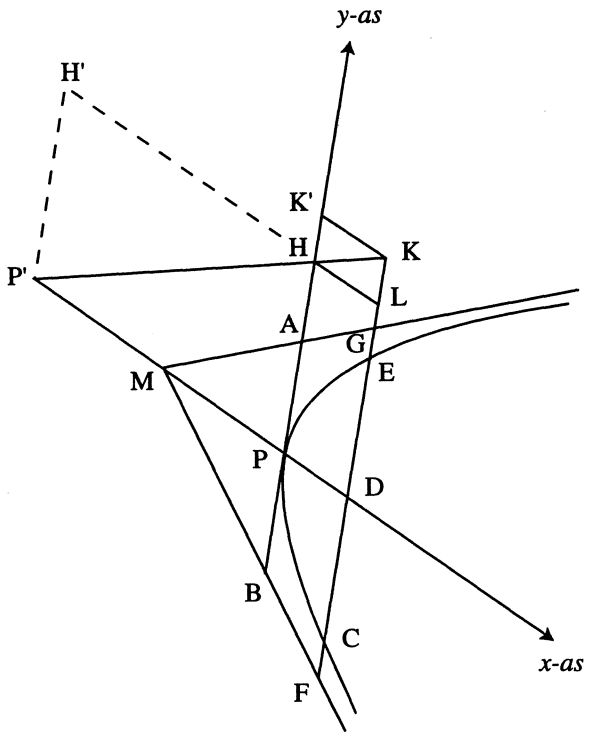
a.	$Opp.\Delta MAB$	$=$	$Opp.\Delta MDC.$
b.	$MA.MB$	$=$	$MD.MC$
c.	$AC : CM$	$=$	$DB : BM$
d.	$AS : SB$	$=$	$DS : SC$
e.	$AP : PS$	$=$	$DQ : QS.$



FIGUUR 2.10

2.10. Van zeer groot belang is de stelling die aantoont dat de kromme die Jan de Witt hier geïntroduceerd heeft juist het kenmerk heeft van de hyperbool zoals Apollonius die definieerde in het kader van de aanpassing van oppervlakten. Om dit verband in te zien hebben we de volgende hulpstelling nodig:
"Indien een willekeurige middellijn van een hyperbool getrokken is, dan is de verhouding tussen het vierkant op de tweede middellijn tot het vierkant op de transversale middellijn -oftewel de verhouding tussen de parameter en de transversale middellijn- gelijk aan de verhouding tussen het vierkant op een willekeurige geordend aangebrachte rechte tot de rechthoek die ingesloten wordt door die delen van de middellijn gelegen tussen de beide uiteinden van de transversale middellijn en de aangebrachte rechte".
Met opzet is hier de formulering van Jan de Witt in vertaling gekozen.
Voor een goed begrip bezien we fig. 2.11. Hierin is DE geordend aangebracht op de transversale middellijn $P'P$; op de drager van de topraaklijn AB is vanuit P de parameter PH uitgezet. In 2.7.c zagen we dat hiervoor geldt: $PH.P'P = AB^2$. Verder geldt $H'HL//P'MPenKK'//LH$.
De stelling zegt dan

$$AB^2 : P'P^2 = PH : P'P = DE^2 : P'D.PD$$



FIGUUR 2.11

Het bewijs verloopt aan de hand van deze figuur eenvoudig en wel als volgt:

$$GD^2 : AP^2 = MD^2 : MP^2$$

dus

$$(GD^2 - AP^2) : AP^2 = (MD^2 - MP^2) : MP^2$$

Ook geldt

$$AP^2 = CG.CF = CG.GE$$

dus

$$(GD^2 - CG.GE) : AP^2 = (MD^2 - MP^2) : MP^2$$

d.w.z.

$$[(DE + EG)^2 - (2DE + EG).GE] : AP^2 = (MD + MP)(MD - MP) : MP^2$$

oftewel

$$DE^2 : AP^2 = P'D.PD : MP^2$$

en dus

$$DE^2 : P'D.PD = AB^2 : P'P^2 \qquad (*)$$

Een analoge stelling geldt voor de ellips zoals we zullen zien in 3.2. Deze wordt echter vrijwel direct uit de definitie van de ellips afgeleid.
Als eerste toepassing van deze stelling leidt Jan de Witt een methode af om van een gegeven hyperbool de asymptoten te construeren. Wellicht ten overvloede zij opgemerkt dat onder "een gegeven hyperbool" verstaan moet worden een kromme die op papier gegeven is, behept met alle eigenschappen van een hyperbool en die naar hartelust gesneden mag worden met een rechte lijn.
Daartoe bepaalt men van de gegeven hyperbool (zie fig. 2.11) het middelpunt M (als snijpunt van twee middellijnen). Vervolgens kiest men een middellijn, zeg MD, die de hyperbool snijdt in P en brengt men door D een lijn CDE geordend aan op de middellijn MD. Dan verlengt men PM tot P' waarbij $PM = MP'$ en trekt door P een lijn evenwijdig met CE.
Tenslotte bepaalt men op deze raaklijn in P aan de hyperbool de punten A en B zodanig dat

$$DE^2 : P'D.PD = PA^2 : MP^2 = PB^2 : MP^2.$$

De rechten MA en MB zijn dan de asymptoten van de gegeven hyperbool.

Zoals gezegd kan men met behulp van deze stelling de door Jan de Witt geconstrueerde kromme zien in het licht van de "aanpassingsproblemen" uit de Oudheid. Hiernaar verwijst hij dan ook met de volgende woorden: *"Indien men vanuit een punt op een hyperbool een lijnstuk geordend aanbrengt op een middellijn, dan geeft dit (lijnstuk, vert.) een oppervlakte, even groot als de oppervlakte grenzend aan de rechte zijde en met breedte het lijnstuk dat van de middellijn wordt afgesneden tussen de aangebrachte rechte en de top op die middellijn, vermeerderd met een figuur die gelijkvormig is met -en dezelfde stand heeft als- die welke ingesloten wordt door de dwarse zijde en de rechte zijde".*
In fig. 2.11 betekent dit dat de ruit met zijde DE en hoek PDE dezelfde oppervlakte heeft als het parallelogram $PDLH$ vermeerderd met het parallelogram $HLKK'$. Dit laatste parallelogram is immers gelijkvormig met het parallelogram $P'PHH'$ en "heeft dezelfde stand".
Het is duidelijk dat het gaat om een vierkant en rechthoeken dan wel een ruit en parallelogrammen, al naar dat de geordend aangebrachte rechte (i.c. DE) wel of niet loodrecht staat op de middellijn.
In formule komt dit neer op:

$$DE^2 = PD.HP + KL.KK'$$

Dit laatste nu is gemakkelijk af te leiden uit (*).
Immers uit

$$DE^2 : P'D.PD = AB^2 : P'P^2$$

volgt, omdat voor de rechte zijde PH geldt

$$AB^2 = P'P.PH$$

dat

$$DE^2 : P'P.PD = PH : P'P = DK : P'D = DK.PD : P'D.PD$$

Hieruit volgt

$$DE^2 = DK.PD = PD.(DL + LK).$$

Schrijft men dit als

$$DE^2 = PD.PH + KL.KK' \qquad (**)$$

dan is het gestelde duidelijk; het plusteken verklaart de naam "hyperbool" ($\upsilon\pi\varepsilon\rho\beta o\lambda\eta$, overschot).

Tot slot zetten we dit resultaat om in de notatie die wij in de analytische meetkunde hanteren. Daartoe kiezen we een assenkruis met oorsprong P en de dragers van PD en PA resp. als x-as en y- as. We geven E de coördinaten x en y en stellen de lengten van $P'P$ en AB voor door resp. 2a en 2b. Voor de parameter p geldt dan $p = 2b^2/a$.
Uit fig. 2.11 lezen we dan af:

$$LK : LH = PH : PP' \text{ oftewel } LK : x = p : 2a, \text{ zodat } LK = \frac{px}{2a}$$

waardoor (**) overgaat in

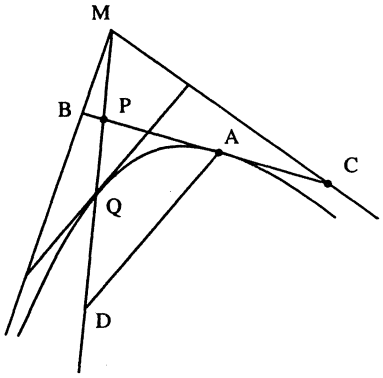
$$y^2 = px + \frac{px^2}{2a}$$

en dit is juist de topvergelijking van een hyperbool.

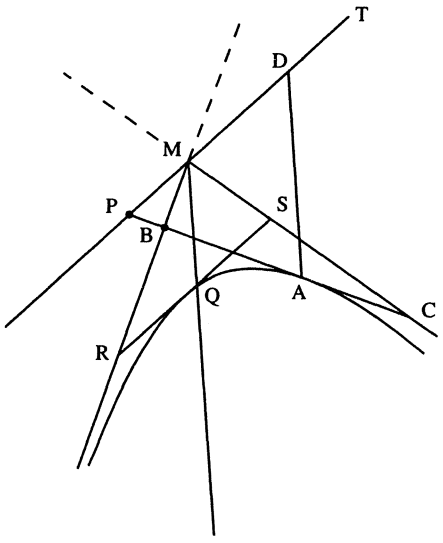
2.11. In de laatste paragraaf van het hoofdstuk dat gewijd is aan de hyperbool, behandelt Jan de Witt de constructie van een raaklijn in een punt op de kromme c.q. van een raaklijn vanuit een punt daarbuiten.
Daartoe leidt hij eerst twee stellingen af die gaan over de ligging van het snijpunt van een raaklijn met een transversale, c.q. tweede middellijn en wel de volgende:

a. Indien een raaklijn een transversale middellijn van een hyperbool snijdt en men vanuit het raakpunt een rechte geordend op deze middellijn aanbrengt, dan is het produkt van de delen van de middellijn die -gerekend vanaf het middelpunt- worden afgesneden door de raaklijn en de aangebrachte rechte, gelijk aan het kwadraat van deze halve transversale middellijn.
b. Indien een raaklijn een tweede middellijn van een hyperbool snijdt en men vanuit het raakpunt een rechte geordend op deze middellijn aanbrengt, dan is het produkt van de delen van de middellijn die -gerekend vanaf het middelpunt- worden afgesneden door de raaklijn en de aangebrachte rechte, gelijk aan het kwadraat van deze halve tweede middellijn.

In fig. 2.12.a is BC de raaklijn in een willekeurig punt A op een hyperbool; MQ is een willekeurige transversale middellijn die deze raaklijn snijdt in P.



FIGUUR 2.12a

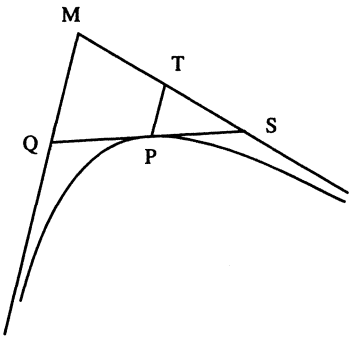


FIGUUR 2.12b

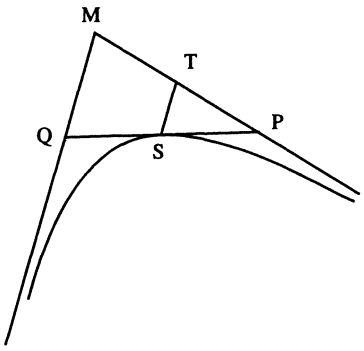
Door A is het lijnstuk AD geordend aangebracht op MQ, d.w.z. evenwijdig met de topraaklijn in Q.
De bewering is dan: $MQ^2 = MP.MD$.
In fig. 2.12.b is BC de raaklijn in een willekeurig punt A op een hyperbool; MT is een willekeurige tweede middellijn die deze raaklijn snijdt in P. Door A is het lijnstuk AD geordend aangebracht op MT, d.w.z. evenwijdig met de aan MT toegevoegde middellijn welke laatste de kromme snijdt in Q. Verder is RS de topraaklijn in Q. RS is dan per definitie de lengte van de tweede middellijn.
De bewering is dan: $QS^2 = MP.MD$.

Wij zullen deze stellingen hier niet bewijzen. Voor een bewijs wordt de lezer verwezen naar litt. [A].
Deze stellingen stellen Jan de Witt in staat een raaklijn aan een gegeven hyperbool te construeren in/vanuit een punt P. Daartoe onderscheidt hij vier gevallen.

i. Het punt P ligt op de hyperbool (fig. 2.13.a). Met de in 2.10 uiteengezette methode construeert men de beide asymptoten van de hyperbool. Vervolgens trekt men door P een rechte evenwijdig aan een daarvan die de andere snijdt in T en bepaalt op deze andere asymptoot een punt S



FIGUUR 2.13a



FIGUUR 2.13b

zodanig dat $TS = MT$.
Indien men nu SP snijdt met de eerste asymptoot in Q, dan is $PQ = PS$ en dus raakt QS in P aan de hyperbool op grond van 2.6.a.

ii. Het punt P ligt op een van de asymptoten (fig. 2.13.b). Men halveert dan PM in T, trekt door T een lijn evenwijdig aan de andere asymptoot. Indien deze lijn de kromme snijdt in S dan is het weer duidelijk dat PS in S aan de hyperbool raakt.

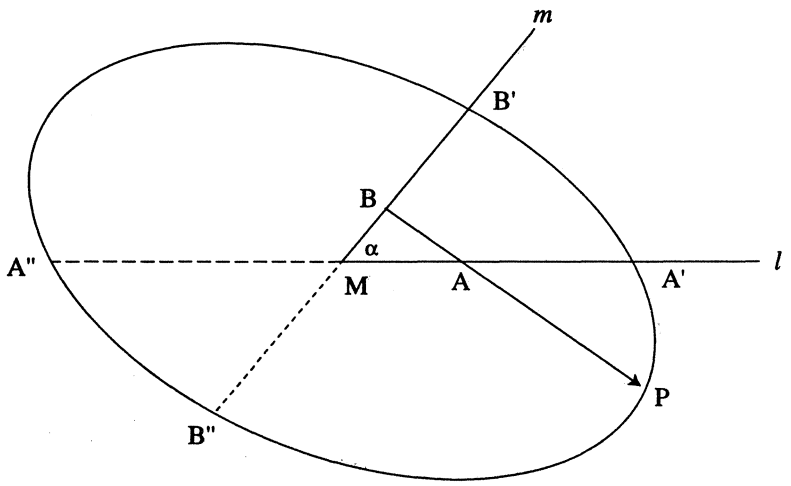
iii. Het punt P ligt binnen de hoek tussen de asymptoten waarbinnen ook de hyperbool ligt (fig. 2.12.a). Men trekt dan door P de middellijn MP die de kromme snijdt in Q. Vervolgens bepaaalt men op MQ het punt D zodanig dat $MP : MQ = MQ : MD$. Daarna brengt men door D een lijn geordend aan op de middellijn MQ, welke lijn de hyperbool snijdt in A(DA is dus evenwijdig aan de raaklijn in Q). Volgens 2.11.a en de eenduidigheid van de raaklijn in A zal dan PA in A aan de hyperbool raken.

iv. Het punt P ligt in een van de nevenhoeken van de hoek tussen de asymptoten waarbinnen de hyperbool ligt (fig. 2.12.b). Door P en het middelpunt M trekt men dan een (tweede) middellijn PT. Vervolgens trekt men de bijbehorende transversale middellijn die de hyperbool snijdt in Q. Deze transversale middellijn vindt men door een koorde, evenwijdig aan PT te halveren en het midden daarvan met M te verbinden. Ook trekt men de raaklijn RS in Q. Daarna bepaalt men op MT een punt D zodanig dat $PM : QS = QS : MD$ en trekt men een rechte evenwijdig aan de middellijn MQ, welke rechte de kromme snijdt in A. Volgens 2.11.b en de eenduidigheid van de raaklijn in A zal dan PA in A aan de hyperbool raken.

Het is duidelijk dat vanuit een punt in de hoek die de tegengestelde hyperbool

bevat, geen raaklijn aan de hyperbool mogelijk is. Over de andere raaklijn uit P spreekt Jan de Witt niet!

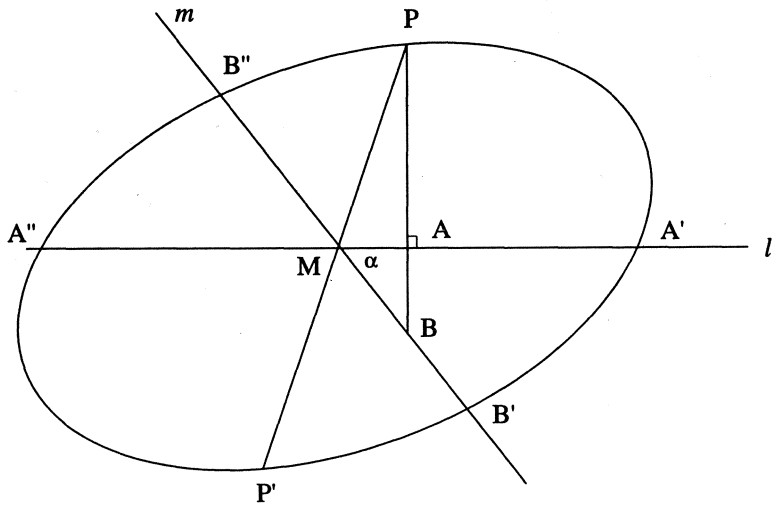
3. De Ellips



Figuur 3.1

3.1. De manier waarop Jan de Witt de ellips voortbrengt is thans nog bij velen bekend en heeft in vele vraagstukkenboeken een vaste plaats.
Uitgangspunt is hierbij een hoek α (de werkhoek) met hoekpunt M en met benen l en m (fig. 3.1). Een lijnstuk AB met vaste lengte (de beschrijvende) heeft een eindpunt (A) op l, het andere (B) op m. Op AB of het verlengde van AB dan wel van BA kiest men een punt P (het werkpunt). Het gaat nu om de baan van het punt P als AB zodanig beweegt dat A steeds blijft op l of het verlengde daarvan en B steeds op m of het verlengde daarvan.
De lijnstukken PA en PB heten de intervallen. Het punt M krijgt -voorbarigde naam "middelpunt".
Wanneer men langs l ter weerszijden van M het interval PB afpast, met eindpunten A' en A'', dan noemt men het lijnstuk $A'A''$ richtlijn. Analoog kan men het interval PA afpassen langs m met eindpunten B' en B''; ook $B'B''$ heet dan richtlijn (fig.3.1).
Men zegt dat de beschrijvende AB in de beginstand is indien deze loodrecht staat op de richtlijn $A'A''$. Het dubbele van MP in die stand noemt men

de secans, in fig. 3.2 is dit PP'. Uiteraard kan men analoge definities geven wanneer men uitgaat van de richtlijn $B'B''$.



FIGUUR 3.2

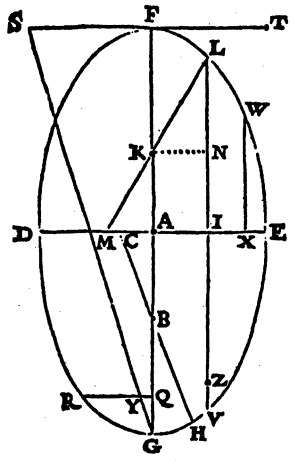
3.2. Uitgaande van deze constructie bewijst Jan de Witt voor de zo gegenereerde kromme een vijftal stellingen en leidt daaruit een aantal conclusies af, waaronder wel als zeer voornaam geldt dat deze kromme dezelfde is als de reeds uit de oudheid bekende ellips.
Allereerst de volgende fundamentele stelling:
Wanneer men door een willekeurig punt L op deze kromme een rechte LIV trekt, evenwijdig met de secans, welke rechte de richtlijn DE snijdt in I en de kromme ook nog in V, dan geldt

$$LI^2 : DI.IE = FG^2 : DE^2 \qquad \text{(fig.3.3)}$$

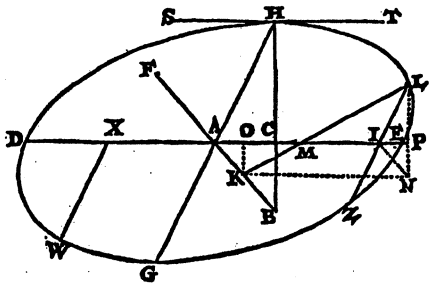
en

$$LI^2 : DI.IE = HG^2 : DE^2 \qquad \text{(fig.3.4)}$$

De figuren zijn ontleend aan het originele werk van Jan de Witt. De eerste figuur geldt voor het geval dat de werkhoek recht is; in de tweede figuur is de werkhoek niet recht. Dit tweede geval is tamelijk gecompliceerd (zie litt.[A]); we zullen ons dan ook beperken tot het eerste geval.
In fig.3.3 is KN evenwijdig aan DE getrokken. Er geldt zeker $KL = AE = AD$ (interval) en evenzo $LM = AF = AG$.



FIGUUR 3.3



FIGUUR 3.4

Omdat de werkhoek recht is (en dat maakt het geval zo eenvoudig) geldt:

$$LN^2 = KL^2 - KN^2 = AE^2 - AI^2 = (AE + AI)(AE - AI) = DI.IE.$$

Ook geldt

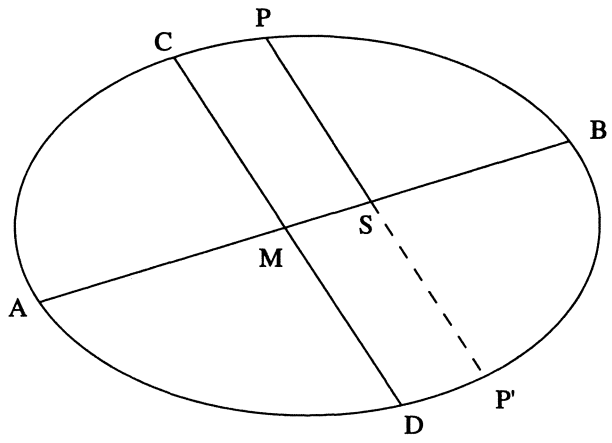
$$LI^2 : LN^2 = LM^2 : LK^2 = AF^2 : AE^2 = FG^2 : DE^2$$

en dus

$$LI^2 : DI.IE = FG^2 : DE^2.$$

Hieruit concludeert Jan de Witt "dat de voornoemde kromme dezelfde is als die welke door de Ouden ellips genoemd werd en dat de richtlijn en de secans dezelfde zijn die zij toegevoegde middellijnen noemden of, indien de werkhoek recht is, toegevoegde assen".
Immers reeds Menaechmus (ca. 350 v.Chr.) toonde aan dat de hierboven afgeleide eigenschap van gelijkheid van verhoudingen kenmerkend is voor de ellips. Deze eigenschap past geheel in de Euclidische traditie waarin verhoudingen zo centraal stonden. Ook ARCHIMEDES (287–212) zou zich hierop beroepen. In 3.5 zullen we zien dat de gegeven definitie ook in overeenstemming is met de wijze waarop Apollonius (ca. 200 v. Chr.) de ellips karakteriseerde via gelijkheid van oppervlakten.

3.3. Voordat Jan de Witt verder gaat introduceert hij enkele nieuwe begrippen waarvan we de analoga reeds bij de hyperbool ontmoetten.



FIGUUR 3.5

a. Allereerst voert hij de naam toegevoegde middellijnen in voor twee lijnen door het middelpunt, aan beide zijden begrensd door de kromme (koorden door het middelpunt dus) met de volgende eigenschap: trekt men door een punt op de ene een koorde evenwijdig aan de andere, dan geldt voor het kwadraat van de helft van deze aangebrachte koorde een eigenschap analoog aan de zojuist bewezen eigenschap. In fig. 3.5 betekent dit (indien $PS//CM$):

$$PS^2 : AS.SB = CD^2 : AB^2.$$

Hier heet AB de tranversale (dwarse) middellijn en CD de daaraan toegevoegde tweede middellijn. De symmetrie van de relatie "toegevoegd aan" zal worden aangetoond in 3.7.c.

b. Van zeer groot belang is ook hier het begrip "rechte zijde" (latus rectum, parameter) behorende bij een geordend paar toegevoegde middellijnen. Indien een transversale middellijn en de bijbehorende tweede middellijn resp. de lengte 2a en 2b hebben, dan is de parameter p die bij dit (geordende) paar behoort gedefinieerd als de derde evenredige bij 2a en 2b, dus $2a : 2b = 2b : p$ oftewel $p = 2b^2/a$.

c. Men zegt dat een rechte geordend is aangebracht op een middellijn indien deze rechte evenwijdig is met de bijbehorende tweede middellijn.

3.4. Uit het bovenstaande zijn enkele eenvoudige conclusies te trekken.

a. Indien men bij het bewijs in de redenering sub. 3.2. niet uitgaat van willekeurige middellijnen maar van assen, dan blijkt eenvoudig dat men daarin

de rol van de assen onderling kan verwisselen. In dit geval is dus een transversale as tevens tweede as en omgekeerd. In het geval van willekeurige middellijnen is de situatie moeilijker. Zie hiervoor 3.7.c.

b. Iedere koorde, evenwijdig aan een tweede middellijn wordt gehalveerd door de bijbehorende transversale middellijn. Zo geldt immers in fig. 3.5

zowel $PS^2 : AS.SB = CD^2 : AB^2$

als $P'S^2 : AS.SB = CD^2 : AB^2$,

dus $PS = P'S$.

c. Een rechte door het werkpunt in de beginstand en evenwijdig aan de richtlijn raakt in dit punt aan de ellips en heeft verder geen punt met de ellips gemeen. Jan de Witt bewijst dit door aan te tonen dat de kromme (m.u.v. het werkpunt in genoemde stand) geheel "onder de ellips ligt" (fig. 3.3 en 3.4).

d. Geordend aangebrachte rechten hebben hoogstens twee punten met de ellips gemeen. Dit volgt eenvoudig uit gevolg (b). Beschouw daartoe in fig. 3.4 het eventuele derde snijpunt Z van LI met de ellips; dit ligt vast door de eis $LI = IZ$ en valt dus samen met V.

Later zal blijken dat iedere lijn die de ellips snijdt geordend is aangebracht op een of andere middellijn en dat dus iedere lijn die de ellips snijdt hoogstens twee punten daarmee gemeen heeft.

3.5. Van zeer groot belang is de conclusie die, evenals bij de hyperbool, de verbinding legt met de wijze waarop Apollonius de ellips definieerde en die door Jan de Witt slechts gedeeltelijk wordt verduidelijkt. Ook hier laten we de tekst volgen in een directe vertaling van de Latijnse formulering van Jan de Witt

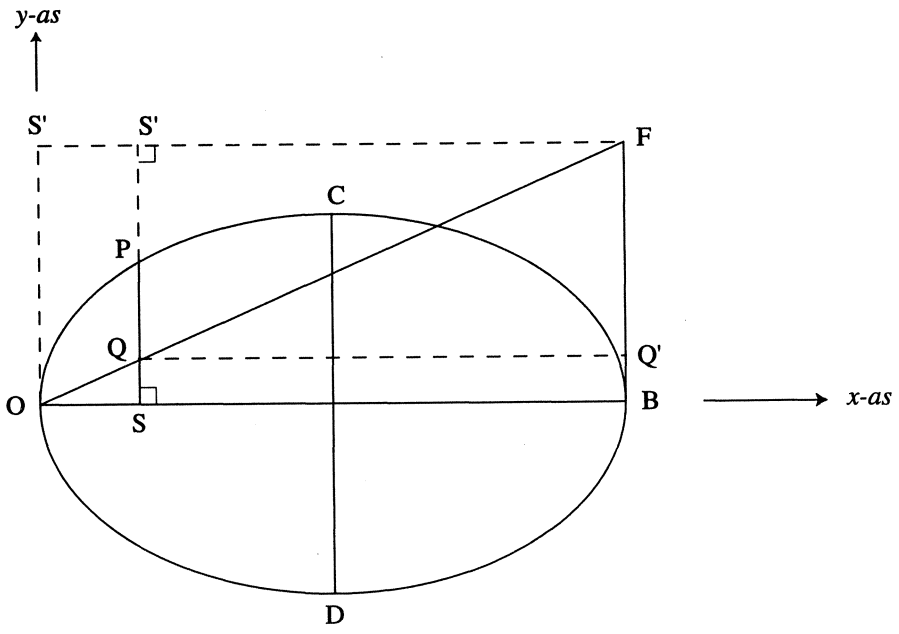
"*Indien men vanuit een punt op een ellips een lijnstuk geordend aanbrengt op een middellijn, dan geeft dit (lijnstuk, vert.) een oppervlakte, even groot als de oppervlakte grenzend aan de bijbehorende rechte zijde en met als breedte het lijnstuk dat van de middellijn wordt afgesneden tussen het aangebrachte lijnstuk en het uiteinde van de middellijn, verminderd met een figuur die gelijkvormig is met -en de zelfde stand heeft als- de figuur die ingesloten wordt door de dwarse en de rechte zijde*"

Ook hier is de formulering van Jan de Witt gebruikt om een indruk te geven van de stijl van zijn boek die, zoals hier, vaak cryptisch is.

Voor de betekenis kijken we naar fig. 3.6. Hierin is de situatie weergegeven waarin het om assen gaat; OB en CD zijn de assen van een ellips, BF staat loodrecht op OB en is de parameter (rechte zijde) behorende bij de transversale middellijn (dwarse zijde) OB en de tweede middellijn CD, d.w.z. $OB : CD = CD : BF$. Het punt P is willekeurig op de ellips gekozen; PQS staat loodrecht op OB.

De bewering is nu:

$$PS^2 = BF.SB - QS'.S'F.$$



FIGUUR 3.6

Hierin is PS^2 de oppervlakte van het vierkant met het aangebrachte lijnstuk als zijde; $BF.SB$ is de oppervlakte van de rechthoek met als zijden de rechte zijde BF en het bedoelde afgesneden stuk van de as, terwijl $QS'.S'F$ de oppervlakte is van de rechthoek $QS'FQ'$ die gelijkvormig is met -en dezelfde stand heeft als- de rechthoek $OS''FB$, ingesloten door de dwarse en de rechte zijde, resp. OB en BF.
De woorden "verminderd met" in de formulering weerspiegelen de oorsprong van de naam ellips ($\varepsilon\lambda\lambda\varepsilon\iota\psi\iota\sigma$ = tekort).
In het algemene geval gaat het niet om een vierkant en rechthoeken, maar om een ruit en parallelogrammen.
We zullen bewijzen:

$$PS^2 = SQ.SB$$

en dit is te schrijven als

$$PS^2 = (SS' - QS').SB = BF.SB - QS'.SB = BF.SB - QS'.S'F$$

Het bewijs dat $PS^2 = SQ.SB$ verloopt als volgt:
Aangezien BF de parameter is, geldt:

$$CD^2 = OB.BF$$

dus uit

$$CD^2 : OB^2 = PS^2 : OS.SB$$

volgt

$$CD^2 : OB^2 = OB.BF : OB^2 = BF : OB$$
$$= SQ : OS = SQ.SB : OS.SB$$

We zagen zo juist

$$CD^2 : OB^2 = PS^2 : OS.SB$$

dus

$$PS^2 = SQ.SB. \qquad (*)$$

We zagen hierboven dat dit leidt tot de meetkundige karakterisering die Jan de Witt formuleerde en dit is juist de wijze waarop ook Apollonius de ellips karakteriseerde.
Tot slot nemen we een rechthoekig assenkruis met O als oorsprong en als x- resp. y-as de drager van OB en de loodlijn in O daarop. Wanneer we dan stellen

$$BF = p, OB = 2a, CD = 2b, OS = x \text{ en } PS = y,$$

dan zien we eenvoudig:

$$SQ : OS = BF : OB,$$

dus

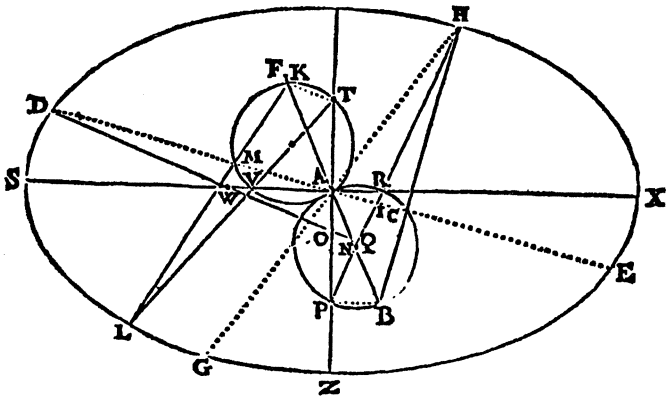
$$SQ = \frac{px}{2a} \text{ en } SB = 2a - x,$$

zodat (*) overgaat in

$$y^2 = px - \frac{px^2}{2a}$$

en dit is juist de topvergelijking van een ellips. Hier weerspiegelt het min-teken de oorsprong van de naam ellips.

3.6. We zagen al in 3.2 dat Jan de Witt een algemene definitie gaf van toegevoegde middellijnen, maar hij liet daarbij nog in het midden of er -naast de twee middellijnen die bij de constructie een fundamentele rol speelden- inderdaad nog andere paren toegevoegde middellijnen zijn.
Op deze vraag komen we nu terug. Jan de Witt leidt namelijk de volgende stelling af:
In een ellips kan elke middellijn optreden als transversale middellijn (=dwarse zijde) en heeft een daarbij passende tweede middellijn.
Het bewijs is tamelijk gecompliceerd en beslaat -inclusief de figuren- ruim vier pagina's. We zullen dit dan hier ook niet weergeven, maar volstaan met de leidende gedachte (zie ook fig. 3.7). Deze komt hierop neer: uitgaande van de assen van een ellips E construeert Jan de Witt bij een willekeurige middellijn van E op een speciale manier een andere middellijn. Met de hoek tussen deze en de gekozen middellijn en met geschikt gekozen intervallen als uitgangspunt,



FIGUUR 3.7

construeert hij vervolgens een nieuwe ellips E^*, waarvan de tweede middellijn op grond van de constructie -als secans- toegevoegd is aan de eerste. Vervolgens toont hij aan dat de ellipsen E en E^* samenvallen. De nieuwe middellijnen, waarvan de eerste geheel willekeurig was, vormen dan ook een paar toegevoegde middellijnen van de oorspronkelijke ellips E en daar ging het om.
Het belang van deze stelling is dat alle stellingen die voor de oorspronkelijke richtlijnen en secans golden, nu ook van toepassing zijn op ieder paar toegevoegde middellijnen.

3.7. Deze stelling leidt tot een aantal interessante conclusies

a. Als eerste toepassing leidt Jan de Witt af hoe men uitgaande van twee toegevoegde middellijnen twee onderling loodrechte middellijnen kan vinden. Hij doet dit door de hierboven aangeduide constructie in omgekeerde volgorde uit te voeren. Hiermee is tevens de existentie van de assen aangetoond.
b. Uit de genoemde constructie blijkt ook dat elke middellijn (d.w.z. een koorde door het middelpunt) gehalveerd wordt door het middelpunt. Dit rechtvaardigt eerst nu de naam middelpunt.
c. De gevolgde constructie leert ook dat, als de middellijn d' de tweede middellijn is bij de transversale middellijn d, ook d de tweede middellijn is bij d' als transversale middellijn. De relatie "toegevoegd aan" of "geconjugeerd met" is dus symmetrisch.

d. Op grond van 3.4.c kan men concluderen dat een lijn door het eindpunt van een middellijn, evenwijdig aan de bijbehorende transversale middellijn (d.w.z. geordend aangebracht), in dit punt aan de ellips raakt.

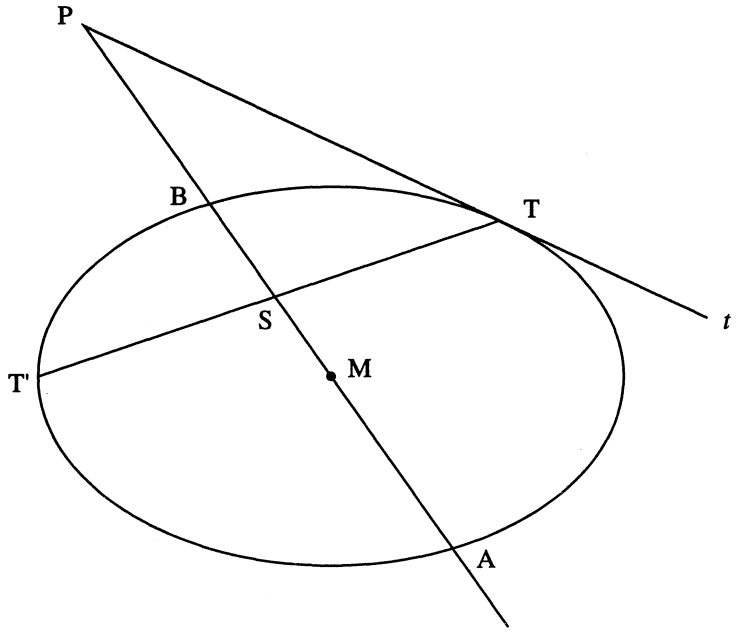
3.8. Aan het einde van het hoofdstuk over de ellips behandelt Jan de Witt enkele stellingen die van belang zijn voor constructies en wel

a. Een koorde die gehalveerd wordt door een middellijn gaat ofwel door het middelpunt of is geordend aangebracht op deze middellijn.
b. Het gevolg is dat een middellijn die een koorde halveert, ook alle daaraan evenwijdige koorden halveert en dat een rechte die twee evenwijdige koorden halveert noodzakelijkerwijze door het middelpunt gaat.
c. Dit laatste gevolg stelt ons in staat om van een gegeven ellips een middellijn te vinden: men behoeft slechts twee evenwijdige koorden te halveren en hun middens te verbinden. Deze middellijn geeft dan tevens de richting aan die aan de gekozen koorden is toegevoegd.
   Het middelpunt vindt men dan als het snijpunt van twee willekeurige middellijnen.
   Ook hier zij weer opgemerkt dat men een koorde verkrijgt door -sans scrupules- twee punten op de omtrek te kiezen en deze te verbinden.
d. Tenslotte kan men dan ook met behulp van een willekeurig paar toegevoegde middellijnen en de methode van 3.7.a de assen van een ellips construeren.

3.9 Als toepassing stelt Jan de Witt dan het volgende constructieprobleem:
In een willekeurige ellips bij een gegeven middellijn de daaraan toegevoegde middellijn te construeren.
Allereerst bepaalt men dan de assen (we zagen zoëven in 3.7.a hoe we dit moeten doen, uitgaande van een willekeurig paar toegevoegde middellijnen, die -zoals we zagen- gemakkelijk te vinden zijn). Maar dan zijn we in de uitgangssituatie van 3.6 waar Jan de Witt, uitgaande van de assen, bij een willekeurige middellijn de daaraan toegevoegde middellijn construeert.
Een belangrijke toepassing van deze constructie is de constructie van de raaklijn in een gegeven punt P op de ellips. Daartoe trekt men door P en het middelpunt een middellijn, bepaalt de daaraan toegevoegde middellijn en trekt door P een lijn evenwijdig aan deze toegevoegde middellijn. Dit is dan de gezochte raaklijn, waarvan Jan de Witt zeer zorgvuldig de eenduidigheid bewijst.

3.10. Het hoofdstuk over de ellips wordt afgesloten met de constructie van de raaklijnen vanuit een punt buiten de ellips. Hiertoe wordt eerst de volgende stelling afgeleid (fig. 3.8).
In deze figuur raakt de lijn t in T aan een ellips; AB is een willekeurige middellijn, die na verlenging deze raaklijn t snijdt in P. De lijn TT' is geordend aangebracht op AB en snijdt deze in S.
De bewering is nu:

a. $MB^2 = MS.MP$.
b. Het omgekeerde is ook juist: indien de lijn t de ellips snijdt in T en een middellijn AB snijdt in P zodanig dat $MB^2 = MS.MP$, dan raakt de lijn t in T aan de ellips. Hierbij is S gedefinieerd als hier boven.



FIGUUR 3.8

Ook hier is het bewijs vrij gecompliceerd (zie hiervoor litt. [A]; er worden de gevallen onderscheiden waarin AB al dan niet een as is.
Met behulp van deze stelling kan men dan een raaklijn aan een ellips construeren vanuit een punt daarbuiten.
Men gaat dan als volgt te werk : men trekt door P een middellijn, snijdt deze met de ellips; het (eerste) snijpunt zij B. Vervolgens bepaalt men dan het punt S op deze middellijn zodanig dat $MP : MB = MB : MS$ en brengt tenslotte door S een rechte geordend aan op PM. Volgens het bovenstaande is PT dan een raaklijn vanuit P aan de ellips. Over de tweede raaklijn spreekt Jan de Witt ook hier niet.

4. DE HYPERBOOL EN DE ELLIPS "REVISITED"
4.1. In het begin van hoofdstuk I van de "Elementa Curvarum Linearum" werd met een gegeven richtlijn, interval en bewegende hoek volgens een bepaald procédé een kromme beschreven die later een parabool bleek te zijn. In dit geval was echter de bewegende hoek in de beginstand gelijk aan de hoek die het interval maakt met de richtlijn (aan dezelfde kant van het interval). Een

voor de hand liggende vraag is dan "wat gebeurt er als deze hoeken verschillend zijn?"
Op deze vraag komen we nu terug.
In het slothoofdstuk van het eerste deel van zijn "Elementa" toont Jan de Witt aan dat er in dit geval een hyperbool beschreven wordt. Daartoe beschrijft hij op dezelfde wijze als bij de parabool -dus met draaiende werkhoek en schuivende beschrijvende- een kromme, maar nu met de werkhoek ongelijk aan de hoek tussen interval en richtlijn. Tijdens dit proces introduceert hij met behulp van de top, het interval en de richtlijn op vernuftige wijze twee rechten. Vervolgens neemt hij een willekeurig punt op de kromme en toont aan dat men de zojuist genoemde rechten kan opvatten als asymptoten van een hyperbool waarop het willeurig gekozen punt van de geconstrueerde kromme ligt.
Hiermee is bewezen dat de op deze wijze gegenereerde kromme inderdaad een hyperbool is. Het bewijs is gecompliceerd; we zullen het hier dan ook niet weergeven, maar verwijzen daarvoor naar litt. [A].

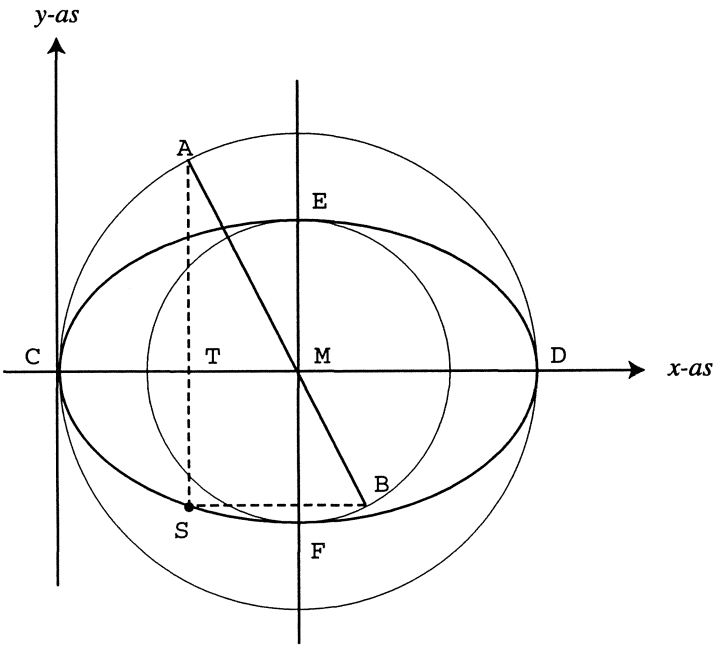
4.2. Het eerste deel van deze "Elementa" wordt besloten met enkele andere methoden om een ellips te beschrijven, waarvan de eerste reeds bekend was aan MYDORGE (1585–1647) en waarschijnlijk ook aan STEVIN (1548–1620) en ARCHIMEDES (287–212).

i. Het gaat hier om de bekende methode waarbij men uitgaat van twee concentrische cirkels. Met het gemeenschappelijke middelpunt M daarvan als centrum, draait een lijnstuk AMB rond (fig. 4.1). Men kiest nu twee vaste, onderling loodrechte lijnen en projecteert A en B steeds hierop. De kromme is nu de baan van het snijpunt P van deze projecterende lijnen. Het is niet moeilijk om langs synthetische weg aan te tonen dat $ST^2 : CT.TD = EF^2 : CD^2$. De baan is dus een ellips en-indien AM en MB gelijk zijn- een cirkel.
ii. Vervolgens beschouwt Jan de Witt het geval waarin de hoek AMB geen gestrekte hoek is (fig. 4.2) en construeert bij deze situatie een kromme op de volgende wijze: door A wordt een lijn l loodrecht op MA getrokken die de drager m van MB snijdt in S. Vervolgens laat hij de hoek AMB wentelen om M. We beschouwen een willekeurig stand $A'MB'$ van deze hoek. Hierbij wordt door A' een lijn l' getrokken evenwijdig aan l en door B' een lijn m', evenwijdig aan m. Het snijpunt zij S'. Het gaat nu om de baan van het punt S' als de hoek AMB wentelt om M.

Deze baan blijkt eveneens een ellips te zijn met middelpunt M. MS is hiervan een halve middellijn. De hieraan toegevoegde middellijn verloopt evenwijdig aan AS. De lengte hiervan is het dubbele van het lijnstuk MG, waarbij G zodanig is gekozen op het verlengde van AM dat de projectie ervan op m juist valt in het punt B.

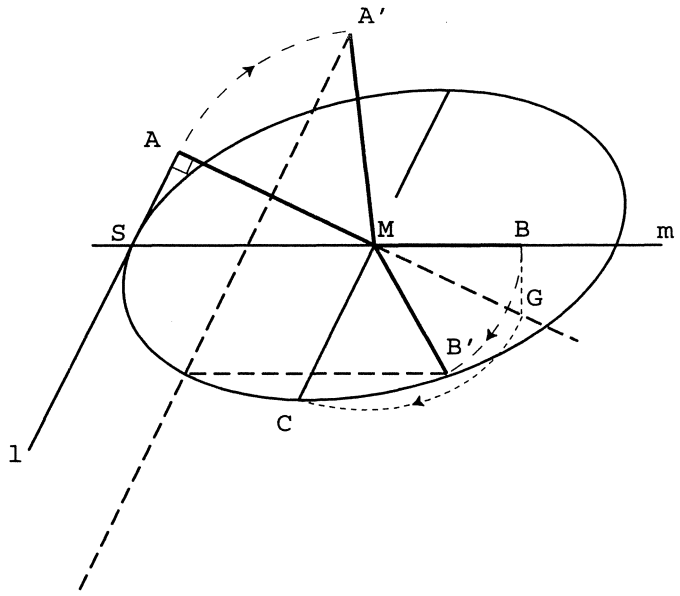
5. LIBER SECUNDUS

Tot slot nog een enkele opmerking over het tweede deel van de "Elementa",



FIGUUR 4.1

waaraan uiteraard een afzonderlijke voordracht gewijd zou kunnen worden, maar dan komen we buiten het bestek van deze Vacantiecursus. We zullen daarom volstaan met het aangeven van de hoofdlijnen ervan.
In dit deel worden de gedachten van Descartes -het toepassen van de letteralgebra in de meetkunde- consequent doorgevoerd, waardoor dit werk terecht het eerste leerboek van de Analytische Meetkunde genoemd mag worden.
Hier worden punten in het platte vlak gerepresenteerd door de ons bekende getallenparen (x, y) en krommen beschreven door vergelijkingen.
Jan de Witt gaat daarbij zeer systematisch te werk. Allereerst worden vergelijkingen opgesteld voor de rechte lijn in allerlei standen, al dan niet door de oorsprong, al dan niet evenwijdig aan de coördinaatassen etc.
Daarna worden vergelijkingen van de tweede graad onderzocht, eerst die waarbij als enige kwadratische term x^2 of y^2 voorkomt. In dit geval blijkt het om een parabool te gaan.
Vervolgens vergelijkingen waarin zowel x^2 als y^2 voorkomen, maar nog geen gemengde termen. Deze blijken dan een hyperbool of een ellips voor te stellen. Deze vondsten worden dan toegepast op de ons bekende problemen: het bepalen van de verzameling ("locus"- d.w.z. plaats- zegt Jan de Witt) van de



FIGUUR 4.2

punten die dezelfde afstand hebben tot een vaste rechte en een vast punt; het bepalen van de plaats van alle punten waarvoor de som resp. het verschil van de afstanden tot twee vaste punten constant is. Deze blijken dan resp. parabool, ellips of hyperbool te zijn, maar deze krommen worden nu niet *gedefinieerd* door deze eigenschappen. Ook worden eerst nu de brandpunten en ook de raaklijnen -op de ons bekende wijze als bissectrices van de hoek tussen de voerstralen- geïntroduceerd.
Tenslotte wordt de algemene vergelijking van de tweede graad aan een onderzoek onderworpen, weliswaar in een speciaal gekozen gedaante:

$$y^2 - \frac{2b}{a}yx + \frac{c}{a}x^2 - 2dy - 2ex + af = 0,$$

maar de klassificatie is er niet minder volledig door.
Hiermee sluit Jan de Witt zijn leerboek af. Een leerboek dat nooit aan een zelfstandige uitgave is toegekomen en heden ten dage een leven in de schaduw leidt, waarmee aan het mathematisch talent en de scherpzinnigheid van Jan de Witt niet de waardering is toegekomen die zij verdienen.

LITTERATUUR

A JOHAN DE WITT, *Elementa Curvarum Linearum*, Edita Opera Francisci à Schooten in Academia Lugduno - Batavo Matheseos Professoris. Amstelodami. Ex Typographia Blaviana, MDCLXXXIII. Dit werk is opgenomen in het tweede deel van Van Schootens "Geometria a Renato Descartes anno 1637 gallice edita, nunc autem ... in linguam latinam versa" (Amsterdam 1659–1661 en 1683). Van deze Elementa zal in het najaar van 1995 bij het Centrum voor Wiskunde en Informatica een uitgave verschijnen met tekst, vertaling en commentaar van de hand van A.W. Grootendorst.

1. C.B.BOYER, *History of Analytic Geometry.* New York, 1956.
2. J.L. COOLIDGE, *The Mathematics of Great Amateurs.* New York, 1963.
3. J.L. COOLIDGE, *A History of the Conic Sections and Quadratic Surfaces.* Oxford, 1945.
4. J.L. COOLIDGE, *A History of Geometric Methods.* New York, 1954.
5. RENÉ DESCARTES, *Oeuvres philosophiques.* Edition de F. Alquié, Tomes I, II, III. Paris, 1963, 1967, 1973.
6. *Dictionary of Scientific Biography* (XV Vol). New York, 1970–1978.
7. E. J. DIJKSTERHUIS, *Archimedes*, I & II. Groningen 1938.
8. P. VAN GEER, *Johan de Witt als wiskundige.* Nieuw Archief voor Wiskunde (2), XI, 1915, pp. 98–126.
9. A. W. GROOTENDORST, *Grepen uit de Geschiedenis van de Wiskunde.* Delft, 1988.
10. Sir TH. L. HEATH, *A History of Greek Mathematics*, Vol. I & II. Oxford, 1965.
11. Sir TH. L. HEATH, *Apollonius of Perga, Treatise on Conic Sections.* Cambridge, 1961.
12. Sir TH. L. HEATH, *The Works of Archimedes.* New York, 1953.
13. J.E. HOFMANN, *Frans van Schooten der Jüngere.* Wiesbaden, 1962.
14. N. JAPIKSE, *Johan de Witt.* Amsterdam, 1915.
15. M. KLINE, *Mathematical Thought from Ancient to Modern Times.* New York, 1972.
16. J.A. VAN MAANEN, *Facets of Seventeenth Century Mathematics in the Netherlands.* Utrecht 1987.
17. H.H. ROWEN, *Johan de Witt, Grand Pensionary of Holland, 1625–1672.* Princeton, New Yersey, 1978.
18. H.H. ROWEN, *John de Witt, Statesman of the True Freedom*, Cambridge, 1986.
19. J.F. SCOTT, *The Scientific Work of René Descartes*, New York, 1976.
20. D.E. SMITH, *The Geometry of Descartes*, translated from the French and Latin by DAVID EUGENE SMITH and MARCIA L. LATHAM. New York, 1954.
21. Vacantiecursus 1989, *Wiskunde in de Gouden Eeuw.* CWI Syllabus 25, Amsterdam 1989.

Kwadrieken van Dimensie Twee en Hoger

J.M. Aarts

SAMENVATTING
In de eerste paragraaf vertellen we iets over kwadrieken of kwadratische oppervlakken in $\mathbb{R}^3$. Dit zijn kwadrieken van dimensie twee. In het bijzonder bestuderen we het bestaan van rechte lijnen die geheel op een kwadriek liggen. In de volgende paragraaf gaan we dit bekijken in de drie-dimensionale ruimte $\mathbb{P}^3$. In de laatste paragraaf bestuderen we de verzameling van alle rechte lijnen in $\mathbb{P}^3$. De verzameling van rechte lijnen in $\mathbb{P}^3$ kan bijectief worden afgebeeld op een vier-dimensionale kwadriek in de vijfdimensionale ruimte $\mathbb{P}^5$. Er zijn interessante relaties tussen eigenschappen van deze vier-dimensionale kwadriek en eigenschappen van stelsels van lijnen.

1. RECHTE LIJNEN OP HET HALSVLAK

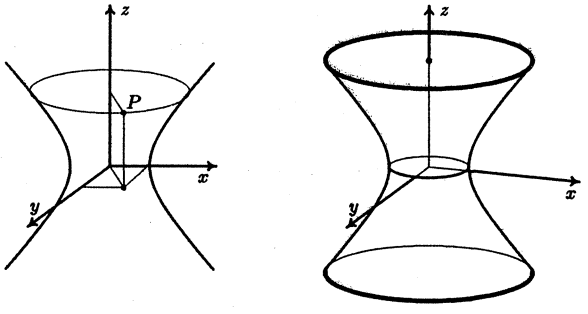
Op het halsvlak liggen twee stelsels van rechte lijnen. In deze paragraaf zullen we enkele eigenschappen van deze stelsels bespreken. Maar eerst, wat is een halsvlak?

1.1. Halsvlak en tweeblad

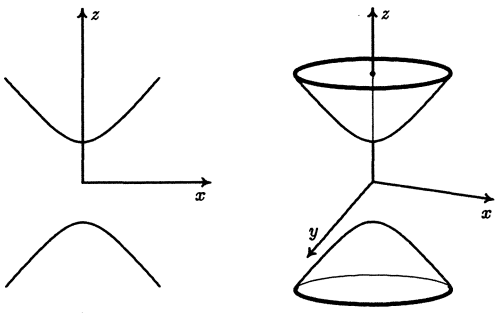
We beginnen met de beschrijving van het omwentelingshalsvlak dat ontstaat door wenteling van een hyperbool om een as. In het xz-vlak beschouwen we de hyperbool

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1; \quad y = 0. \tag{1}$$

Zie Figuur 1, links. De hyperbool ligt symmetrisch ten opzichte van de z-as. We wentelen deze hyperbool om de z-as. De figuur die zo ontstaat heet een *omwentelingshalsvlak* of een *éénbladige omwentelingshyperboloïde*. We leiden de vergelijking van het omwentelingshalsvlak af. De afstand van het punt $P(x, y, z)$ tot de z-as is $\sqrt{x^2 + y^2}$. Zie Figuur 1. Het punt $P(x, y, z)$ ligt op



Figuur 1: Het omwentelingshalsvlak voortgebracht door een hyperbool



Figuur 2: Het omwentelingstweeblad voortgebracht door een hyperbool

het omwentelingshalsvlak dan en slechts dan indien het punt $(\sqrt{x^2+y^2}, 0, z)$ op de hyperbool (1) ligt, oftewel indien

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = 1. \tag{2}$$

Dit is de vergelijking van het omwentelingshalsvlak. Door nu nog op de y-as de schaal te veranderen (y vervangen door $y\frac{a}{b}$) krijgen we de algemene vergelijking van het halsvlak.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1. \tag{3}$$

We merken nog op dat we een hyperbool op twee essentiëel verschillende wijzen in het zx-vlak kunnen plaatsen zó dat de hyperbool symmetrisch is ten opzichte van de z-as. De ene manier hebben we net gezien, de andere is geschetst in Figuur 2. De vergelijking van de hyperbool is in dit geval

$$\frac{z^2}{c^2} - \frac{x^2}{a^2} = 1; \quad y = 0. \tag{4}$$

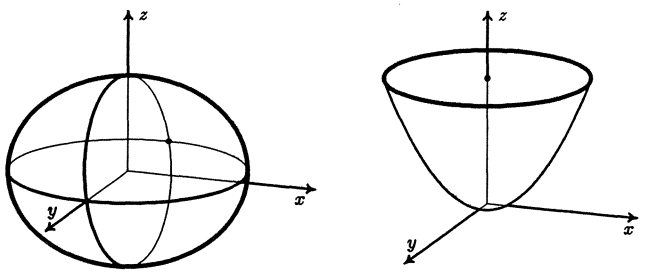
Deze hyperbool beschrijft bij wenteling om de z-as een *omwentelingstweeblad* of *tweebladige omwentelingshyperboloïde.* Door schaalverandering krijgen we de algemene vorm van het tweeblad. Op dezelfde wijze als boven vinden we voor het tweeblad de vergelijking

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \tag{5}$$

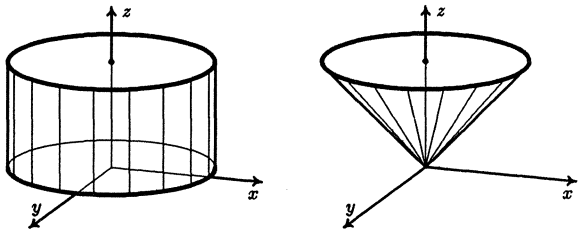
1.2. Kwadrieken

Het halsvlak en het tweeblad zijn voorbeelden van *kwadrieken* of *kwadratische oppervlakken* in $\mathbb{R}^3$; de vergelijking van zo'n oppervlak is van de tweede graad. Heel bekende voorbeelden van kwadrieken zijn de *ellipsoïde* en de *vaas* of *paraboloïde.* Deze zijn geschetst in Figuur 3.

We stellen nu een onderzoek in naar de aanwezigheid van rechte lijnen op kwadrieken. Omdat de ellipsoïde begrensd is, zien we direct in dat er geen



Figuur 3: De ellipsoïde (links) en de paraboloïde (rechts)



Figuur 4: De (elliptische) cilinder (links) en de kegel (rechts)

rechte lijnen zijn die geheel op de ellipsoïde liggen. In de volgende paragraaf zullen we vaststellen dat ook op de vaas en op het tweeblad geen rechte lijnen liggen. Verder hoeft het geen betoog dat de er bij de kwadrieken *cilinder* en *kegel* wel rechte lijnen zijn die op de kwadriek liggen. Kortheidshalve verwijzen we naar Figuur 4.

1.3. Rechte lijnen op het halsvlak

Nu is het halsvlak aan de beurt. De vergelijking (3) van het halsvlak kunnen we in een andere vorm schrijven:

$$\frac{x^2}{a^2} - \frac{z^2}{c^2} = 1 - \frac{y^2}{b^2}.$$

of

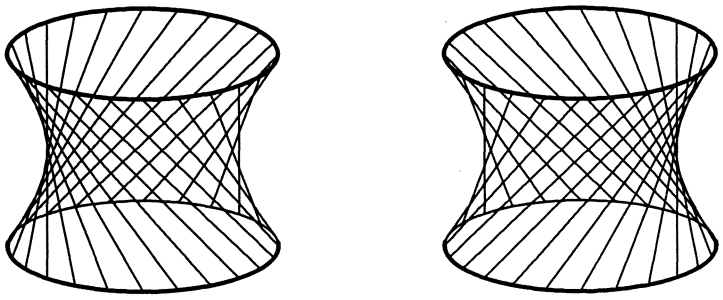
$$\left(\tfrac{x}{a} - \tfrac{z}{c}\right)\left(\tfrac{x}{a} + \tfrac{z}{c}\right) = \left(1 - \tfrac{y}{b}\right)\left(1 + \tfrac{y}{b}\right) \tag{6}$$

We beschouwen nu de volgende stelsels rechte lijnen:

$$\begin{cases} \lambda\left(\tfrac{x}{a} + \tfrac{z}{c}\right) &= \mu\left(1 + \tfrac{y}{b}\right), \\ \mu\left(\tfrac{x}{a} - \tfrac{z}{c}\right) &= \lambda\left(1 - \tfrac{y}{b}\right) \end{cases} ; \lambda \in \mathbb{R},\ \mu \in \mathbb{R},\ |\lambda| + |\mu| > 0, \tag{7}$$

en

$$\begin{cases} \rho\left(\tfrac{x}{a} + \tfrac{z}{c}\right) &= \sigma\left(1 - \tfrac{y}{b}\right), \\ \sigma\left(\tfrac{x}{a} - \tfrac{z}{c}\right) &= \rho\left(1 + \tfrac{y}{b}\right) \end{cases} ; \rho \in \mathbb{R},\ \sigma \in \mathbb{R},\ |\rho| + |\sigma| > 0. \tag{8}$$



Figuur 5: Twee stelsels lijnen op het halsvlak

Voor iedere keuze van λ en μ wordt door (7) een rechte lijn bepaald als doorsnijding van twee vlakken. Evenzo stelt (8) een rechte lijn voor bij iedere keuze van ρ en σ. Eerst komt de belangrijkste eigenschap van de stelsels aan de orde.

EIGENSCHAP 1.1 Iedere lijn uit het stelsel (7) en iedere lijn uit het stelsel (8) liggen geheel op het halsvlak (3).

BEWIJS. Bekijk de lijn uit het stelsel (7) bij de parameterwaarden λ_0 en μ_0:

$$\begin{cases} \lambda_0\left(\frac{x}{a}+\frac{z}{c}\right) &= \mu_0\left(1+\frac{y}{b}\right), \\ \mu_0\left(\frac{x}{a}-\frac{z}{c}\right) &= \lambda_0\left(1-\frac{y}{b}\right), \end{cases} \tag{9}$$

Als $X(x,y,z)$ een punt op deze lijn is, dan voldoen x, y en z aan (9) en dus ook aan

$$\lambda_0\mu_0\left(\tfrac{x}{a}+\tfrac{z}{c}\right)\left(\tfrac{x}{a}-\tfrac{z}{c}\right) = \lambda_0\mu_0\left(1+\tfrac{y}{b}\right)\left(1-\tfrac{y}{b}\right). \tag{10}$$

Als $\lambda_0\mu_0 \neq 0$, dan kunnen we beide leden van (10) delen door $\lambda_0\mu_0$ en zien we dat X op het halsvlak (3) ligt. Is bijvoorbeeld $\lambda_0 = 0$ (maar dan is $\mu_0 \neq 0$), dan volgt uit (9) dat

$$1+\tfrac{y}{b} = 0 \text{ en } \tfrac{x}{a}-\tfrac{z}{c} = 0,$$

en dus dat x, y en z aan (6) voldoen. Ook nu ligt X op het halsvlak (3).

In Figuur 5 zijn de twee stelsels van lijnen die op het halsvlak liggen geschetst.

In de rest van deze paragraaf zullen we nog enkele eigenschappen van de stelsels van lijnen afleiden.

EIGENSCHAP 1.2 Door ieder punt van het halsvlak (3) gaat precies één lijn van het stelsel (7) en precies één lijn van het stelsel (8).

BEWIJS. Als $P(p_1,p_2,p_3)$ op het halsvlak (3) ligt, dan volgt met (6) dat

$$\left(\tfrac{p_1}{a}-\tfrac{p_3}{c}\right)\left(\tfrac{p_1}{a}+\tfrac{p_3}{c}\right) = \left(1-\tfrac{p_2}{b}\right)\left(1+\tfrac{p_2}{b}\right).$$

Neem eerst aan dat linker- en rechterlid van deze vergelijking niet gelijk aan nul zijn. Voor $\mu = 1$ en $\lambda = \left(\frac{p_1}{a} - \frac{p_3}{c}\right)/\left(1 - \frac{p_2}{b}\right)$ hebben we een lijn uit het stelsel (7) die door P gaat. Voor $\sigma = 1$ en $\rho = \left(\frac{p_1}{a} - \frac{p_3}{c}\right)/\left(1 + \frac{p_2}{b}\right)$ hebben we een lijn uit het stelsel (8) die door P gaat. Uit de eigenschap die we hierna afleiden volgt dat er ook niet meer dan een lijn uit elk stelsel door P gaat.
Is bijvoorbeeld $\left(1 - \frac{p_2}{b}\right) = 0$, dan is $\left(1 + \frac{p_2}{b}\right) \neq 0$. Voor $\lambda = 1$ en $\mu = \left(\frac{p_1}{a} + \frac{p_3}{c}\right)/\left(1 + \frac{p_2}{b}\right)$ hebben we een lijn uit het stelsel (7) die door P gaat. Met $\sigma = 1$ en $\rho = \left(\frac{p_1}{a} - \frac{p_3}{c}\right)/\left(1 + \frac{p_2}{b}\right)$ hebben we een lijn van het stelsel (8).

EIGENSCHAP 1.3 Twee verschillende lijnen uit hetzelfde stelsel kruisen elkaar.

BEWIJS. Neem twee verschillende lijnen uit het stelsel (7), zeg de lijn l_1 bij λ_1, μ_1 en de lijn l_2 bij λ_2, μ_2. Omdat de lijnen verschillend zijn, is $\lambda_1\mu_2 \neq \lambda_2\mu_1$. We laten eerst zien dat deze lijnen elkaar niet snijden. Het snijpunt $X(x_1, x_2, x_3)$ van de lijnen l_1 en l_2, zo het er mocht zijn, vinden we uit (Let op de volgorde van de vergelijkingen!)

$$\left.\begin{array}{llll} \lambda_1\frac{x}{a} & -\mu_1\frac{y}{b} & +\lambda_1\frac{z}{c} & = \mu_1 \\ \lambda_2\frac{x}{a} & -\mu_2\frac{y}{b} & +\lambda_2\frac{z}{c} & = \mu_2 \\ \mu_1\frac{x}{a} & +\lambda_1\frac{y}{b} & -\mu_1\frac{z}{c} & = \lambda_1 \\ \mu_2\frac{x}{a} & +\lambda_2\frac{y}{b} & -\mu_2\frac{z}{c} & = \lambda_2 \end{array}\right\} \qquad (11)$$

De gerande coëfficiëntenmatrix van dit stelsel is

$$A = \begin{pmatrix} \lambda_1 & -\mu_1 & \lambda_1 & \mu_1 \\ \lambda_2 & -\mu_2 & \lambda_2 & \mu_2 \\ \mu_1 & \lambda_1 & -\mu_1 & \lambda_1 \\ \mu_2 & \lambda_2 & -\mu_2 & \lambda_2 \end{pmatrix}.$$

Omdat $\det A = -4\left(\mu_1\lambda_2 - \mu_2\lambda_1\right)^2 \neq 0$, heeft A de rang 4. Het stelsel (11) heeft dus geen oplossing en de lijnen l_1 en l_2 snijden elkaar niet. De lijnen kunnen ook niet evenwijdig zijn. Om dat in te zien verplaatsen we de vlakken die l_1 en l_2 bepalen naar de oorsprong. De vergelijkingen van de verplaatste vlakken verkrijgen we door in (11) de rechterleden gelijk aan nul te stellen:

$$\left.\begin{array}{llll} \lambda_1\frac{x}{a} & -\mu_1\frac{y}{b} & +\lambda_1\frac{z}{c} & = 0 \\ \lambda_2\frac{x}{a} & -\mu_2\frac{y}{b} & +\lambda_2\frac{z}{c} & = 0 \\ \mu_1\frac{x}{a} & +\lambda_1\frac{y}{b} & -\mu_1\frac{z}{c} & = 0 \\ \mu_2\frac{x}{a} & +\lambda_2\frac{y}{b} & -\mu_2\frac{z}{c} & = 0. \end{array}\right\} \qquad (12)$$

Zouden nu de lijnen l_1 en l_2 evenwijdig zijn, dan zou het stelsel (12) een oplossing ongelijk aan de nuloplossing hebben. Maar, omdat A de rang 4 heeft, is de rang van de coëfficiëntenmatrix van (12) gelijk aan 3. Het stelsel (11) heeft dus alleen de nuloplossing, en de lijnen l_1 en l_2 kunnen dus niet evenwijdig zijn.

EIGENSCHAP 1.4 Twee lijnen uit verschillende stelsels liggen in één vlak.

BEWIJS. Beschouw een lijn l uit het stelsel (7) en een lijn m uit het stelsel (8):

$$l: \quad \begin{cases} \lambda\left(\frac{x}{a}+\frac{z}{c}\right) &= \mu\left(1+\frac{y}{b}\right) \\ \mu\left(\frac{x}{a}-\frac{z}{c}\right) &= \lambda\left(1-\frac{y}{b}\right) \end{cases} \tag{13}$$

$$m: \quad \begin{cases} \rho\left(\frac{x}{a}+\frac{z}{c}\right) &= \sigma\left(1-\frac{y}{b}\right) \\ \sigma\left(\frac{x}{a}-\frac{z}{c}\right) &= \rho\left(1+\frac{y}{b}\right) \end{cases} \tag{14}$$

Het vlak

$$\lambda\rho\left(\tfrac{x}{a}+\tfrac{z}{c}\right)+\mu\sigma\left(\tfrac{x}{a}-\tfrac{z}{c}\right)=\mu\rho\left(1+\tfrac{y}{b}\right)+\lambda\sigma\left(1-\tfrac{y}{b}\right)$$

is een lineaire combinatie zowel van de vlakken in (13) als van de vlakken in (14). Hieruit volgt dat zowel l als m tot het vlak behoren.

1.4. Toepassingen

De eigenschap dat er op het halsvlak twee stelsels rechte lijnen liggen wordt toegepast bij de constructie van koeltorens. Bij de constructie van koeltorens uit gewapend beton wordt het stalen vlechtwerk gevormd door de stelsels van rechte lijnen. Ook wordt de genoemde eigenschap gebruikt bij de constructie van tandwielen die een overbrenging tot stand moeten brengen tussen twee assen welke een scherpe hoek met elkaar maken.

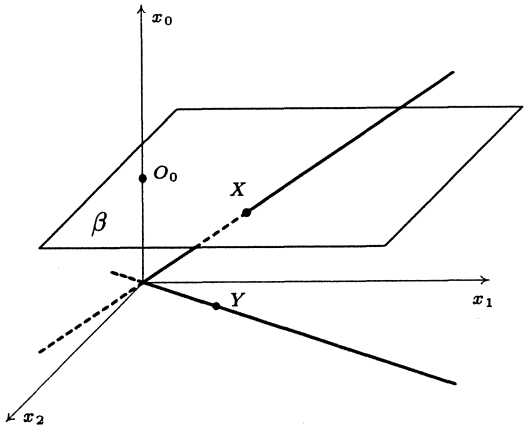
2. PROJECTIEVE KLASSIFICATIE VAN KWADRIEKEN

In de vorige paragraaf hebben we een aantal kwadrieken gezien. Zijn dat nu alle kwadrieken, of zijn er nog andere? Hoe weten we of de lijst van kwadrieken volledig is? We zullen deze problemen nu gaan aanpakken en een projectieve klassificatie van alle kwadrieken geven. Ter voorbereiding eerst iets over projectieve ruimten.

2.1. Het projectieve vlak $\mathbb{P}^2$

Ik vertel eerst iets over het projectieve vlak, omdat dat het eenvoudigst te doorzien is. Alle eigenschappen van het projectieve vlak zijn gemakkelijk te generaliseren naar hoger-dimensionale projectieve ruimten.

In een (gewoon) vlak α kan men een rechthoekig assenstelsel aanbrengen. Zoals bekend correspondeert dan met ieder coördinatenpaar (x_1, x_2), $x_1 \in \mathbb{R}$, $x_2 \in \mathbb{R}$ precies één punt van het vlak. Soms is er behoefte aan om ook de "richtingen" of "oneigenlijke punten" in de beschouwingen te betrekken. Door over te gaan op homogene coördinaten kan men zowel de eigenlijke als de oneigenlijke punten van α met drietallen van reële getallen aanduiden. Het punt $X(x_1, x_2)$ van het vlak α heeft *homogene coördinaten* (ξ_0, ξ_1, ξ_2), $\xi_0 \in \mathbb{R} \setminus \{0\}$ als geldt $\frac{\xi_1}{\xi_0} = x_1$ en $\frac{\xi_2}{\xi_0} = x_2$. Het punt X is bepaald door de verhouding van zijn homogene coördinaten: $X(\xi_0, \xi_1, \xi_2) = X\left(\frac{\xi_1}{\xi_0}, \frac{\xi_2}{\xi_0}\right)$. Een getallenvoorbeeld: het punt $(-3, 2)$ heeft homogene coördinaten $(1, -3, 2)$, maar ook $(2, -6, 4)$ of $(\frac{1}{3}, -1, \frac{2}{3})$. Bij de homogene coördinaten gaat het om de verhouding van de coördinaten. Met de homogene coördinaten $(\frac{1}{2}, -2, \frac{1}{4})$ correspondeert $(-4, \frac{1}{2})$ in gewone coördinaten.



Figuur 6: Het projectieve vlak

Aan de richtingsvector $u = (u_1, u_2)$ voegen wij toe het *oneigenlijke punt* $(0, u_1, u_2)$. Merk op dat er geen punt is met homogene coördinaten $(0, 0, 0)$. De punten van α tesamen met de oneigenlijke punten vormen het projectieve vlak. Een *punt van het projectieve vlak* $\mathbb{P}^2$ is een verhouding van drie reële getallen die niet alle drie gelijk zijn aan 0. We zeggen hier: een punt is een verhouding. Daarmee bedoelen we: als (x_0, x_1, x_2) een punt is van $\mathbb{P}^2$ dan is $(\lambda x_0, \lambda x_1, \lambda x_2)$, $\lambda \neq 0$, hetzelfde punt van $\mathbb{P}^2$.
Het volgende moge deze definitie rechtvaardigen. Een lijn l in het vlak α heeft de vergelijking

$$a_0 + a_1 x_1 + a_2 x_2 = 0. \tag{15}$$

De normaal van deze lijn is de vector (a_1, a_2) en de vector $(-a_2, a_1)$ is een richtingsvector van l. Als een punt $X(x_1, x_2)$ op de lijn (15) ligt dan voldoen de homogene coördinaten $(\lambda, \lambda x_1, \lambda x_2)$, $\lambda \neq 0$, van X aan de vergelijking

$$a_0 x_0 + a_1 x_1 + a_2 x_2 = 0. \tag{16}$$

De laatste vergelijking is de *vergelijking van een lijn* in het projectieve vlak. Het oneigenlijke punt $(0, -a_2, a_1)$, dat correspondeert met de richtingsvector van l, voldoet aan de vergelijking.
We kunnen ons op de volgende wijze een voorstelling maken van het projectieve vlak. De punten van het projectieve vlak zijn verhoudingen van drietallen van getallen die niet alle drie gelijk zijn aan 0. Bij een gegeven punt van het projectieve vlak liggen de bijbehorende drietallen op een rechte lijn in de ruimte die door dat punt en de oorsprong gaat; ieder punt op deze lijn dat verschillend is van de oorsprong bepaalt een drietal dat met het gegeven punt correspondeert. Dit is in de figuur tot uitdrukking gebracht. Ieder punt X met coördinaten $(1, x_1, x_2)$ wordt afgebeeld in het vlak β in de ruimte met

vergelijking $x_0 = 1$; alle drietallen die op de lijn door de oorsprong en $(1, x_1, x_2)$ liggen zijn, met uitzondering van de oorsprong, homogene coördinaten van het punt X. Ieder punt Y met coördinaten $(0, \cos\varphi, \sin\varphi)$ wordt afgebeeld in het x_1x_2-vlak; alle drietallen die op de lijn door de oorsprong en $(0, \cos\varphi, \sin\varphi)$ liggen zijn, met uitzondering van de oorsprong, homogene coördinaten van het punt Y. Het punt $(0, \cos\varphi, \sin\varphi)$ is het gemeenschappelijke punt van alle lijnen in het projectieve vlak van de vorm

$$a_0x_0 - \sin\varphi\, x_1 + \cos\varphi\, x_2.$$

De punten $O_0(1,0,0)$, $O_1(0,1,0)$ en $O_2(0,0,1)$ heten de *grondpunten* van het projectieve vlak. Het punt $E(1,1,1)$ heet het *eenheidspunt.*
We hebben gezien dat (16) de algemene vergelijking is van een lijn in het projectieve vlak. Hierbij zijn niet alle drie de a_i gelijk aan 0. We leiden een parametervoorstelling van de lijn (16) af. Laat $P(p_0, p_1, p_2)$ en $Q(q_0, q_1, q_2)$ twee verschillende vaste punten op deze lijn zijn en zij $X(x_0, x_1, x_2)$ een willekeurig punt van deze lijn. Dan geldt

$$\begin{array}{lllllll} a_0p_0 & + & a_1p_1 & + & a_2p_2 & = & 0 \\ a_0q_0 & + & a_1q_1 & + & a_2q_2 & = & 0 \\ a_0x_0 & + & a_1x_1 & + & a_2x_2 & = & 0. \end{array}$$

Omdat niet alle a_i gelijk aan 0 zijn is

$$\begin{vmatrix} p_0 & p_1 & p_2 \\ q_0 & q_1 & q_2 \\ x_0 & x_1 & q_2 \end{vmatrix} = 0. \tag{17}$$

Omdat P en Q verschillend zijn, moet de derde rij te schrijven zijn als lineaire combinatie van de eerste twee rijen:

$$(x_0, x_1, x_2) = \lambda(p_0, p_1, p_2) + \mu(q_0, q_1, q_2), \quad \text{kortweg,} \quad X = \lambda P + \mu Q. \tag{18}$$

Dit resultaat volgt ook uit de meetkundige voorstelling van het projectieve vlak. In de ruimte vallen het vlak met de vergelijking (16) en het vlak (de twee-dimensionale deelruimte) opgespannen door P en Q samen. Beide vlakken bepalen dan dezelfde lijn in het projectieve vlak.
Een belangrijk aspect van de projectieve meetkunde is het begrip *dualiteit.* De vergelijking (16) van de lijn in het projectieve vlak is geheel bepaald door het drietal (a_0, a_1, a_2). De drietallen (a_0, a_1, a_2) en $(\lambda a_0, \lambda a_1, \lambda a_2)$, $\lambda \neq 0$, bepalen hetzelfde vlak. We kunnen deze drietallen opvatten als homogene coördinaten van een punt in het projectieve vlak $\mathbb{P}^2$. Op deze wijze vormen de lijnen van het projectieve vlak zelf weer een projectief vlak, het *duale vlak.*

2.2. Projectieve transformaties $\mathbb{P}^2 \to \mathbb{P}^2$

Een lineair isomorfisme $S\colon \mathbb{R}^3 \to \mathbb{R}^3$ veroorzaakt een permutatie van de één-dimensionale deelruimten van $\mathbb{R}^3$. De permutatie bepaalt op zijn beurt een bijectieve afbeelding $\mathbb{P}^2 \to \mathbb{P}^2$. Zo'n transformatie heet een *projectieve transformatie* van $\mathbb{P}^2$. In de *projectieve meetkunde* bestudeert men eigenschappen die invariant zijn onder projectieve transformaties.

Een lineair isomorfisme van $\mathbb{R}^3$ is volledig bepaald indien van drie vectoren de beeldvectoren gegeven zijn. Bij projectieve transformaties hebben we iets meer speelruimte zoals blijkt uit de volgende stelling.

STELLING 2.1 Als $P(p_0,p_1,p_2)$, $Q(q_0,q_1,q_2)$, $R(r_0,r_1,r_2)$ en $F(f_0,f_1,f_2)$ vier punten van $\mathbb{P}^2$ zijn waarvan er geen drie op één lijn liggen, dan is er één en slechts één projectieve afbeelding S zó dat $S(O_0)=P$, $S(O_1)=Q$, $S(O_2)=R$ en $S(E)=F$.

BEWIJS. Hier en in het vervolg is het vaak handig om punten van $\mathbb{P}^2$ te schrijven als kolomvector. De beoogde transformatie S zal, zo hij bestaat, moeten voldoen aan de volgende voorwaarde: er bestaan λ, μ en ρ, ongelijk aan 0, zó dat

$$\begin{aligned} S(O_0) &= (\lambda p_0, \lambda p_1, \lambda p_2) \\ S(O_1) &= (\mu q_0, \mu q_1, \mu q_2) \\ S(O_2) &= (\rho r_0, \rho r_1, \rho r_3). \end{aligned}$$

Als we S opvatten als transformatie van vectoren in $\mathbb{R}^3$ dan is er precies één lineaire transformatie die dit doet. In matrixvorm is de transformatie $y=Sx$:

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \lambda p_0 & \mu q_0 & \rho r_0 \\ \lambda p_1 & \mu q_1 & \rho r_1 \\ \lambda p_2 & \mu q_2 & \rho r_2 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}.$$

Dat S bijectief is, hangt samen met het feit dat P, Q en R niet op één lijn liggen. Nu moet gelden

$$F = S(E) = \lambda \begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} + \mu \begin{pmatrix} q_0 \\ q_1 \\ q_2 \end{pmatrix} + \rho \begin{pmatrix} r_0 \\ r_1 \\ r_2 \end{pmatrix}$$

Er is precies één keuze van λ, μ en ρ waarvoor dit geldt. Omdat F niet collineair is met P en Q, noch met P en R, noch met Q en R zijn λ, μ en ρ alle drie ongelijk aan 0.

We krijgen een (nieuw) *coördinatenstelsel* in het projctieve vlak door drie grondpunten en een eenheidspunt aan te wijzen; van deze punten mogen er geen drie op één lijn liggen. Laat $P(p_0,p_1,p_2)$, $Q(q_0,q_1,q_2)$ en $R(r_0,r_1,r_2)$ de beoogde (nieuwe) grondpunten zijn. Hierbij horen homogene (oude) coördinaten $(\lambda p_0, \lambda p_1, \lambda p_2)$, $(\mu q_0, \mu q_1, \mu q_2)$ respectivelijk $(\rho r_1, \rho r_2, \rho r_3)$. Is nu $F(f_0,f_1,f_2)$ het beoogde (nieuwe) eenheidspunt dan willen we dat F de (nieuwe) coördinaten $(1,1,1)$ krijgt, m.a.w.

$$\begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} \lambda p_0 \\ \lambda p_1 \\ \lambda p_2 \end{pmatrix} + \begin{pmatrix} \mu q_0 \\ \mu q_1 \\ \mu q_2 \end{pmatrix} + \begin{pmatrix} \rho r_0 \\ \rho r_1 \\ \rho r_2 \end{pmatrix}$$

Hierdoor zijn λ, μ en ρ eenduidig bepaald. Omdat van de grondpunten en het eenheidspunt er geen drie op een lijn liggen, is λ, noch μ, noch ρ gelijk aan nul.

2.3. Kwadrieken in $\mathbb{P}^3$

Alle resultaten betreffende $\mathbb{P}^2$ kunnen op de voor de hand liggende wijze gegeneraliseerd worden tot $\mathbb{P}^n$, de n-dimensionale projectieve ruimte. In de projectieve ruimte $\mathbb{P}^3$ stelt de vergelijking

$$a_0x_0 + a_1x_1 + a_2x_2 + a_3x_3 = 0$$

een vlak voor. Dit vlak heeft de homogene coördinaten (a_0, a_1, a_2, a_3).
Een lijn in $\mathbb{P}^3$ wordt meestal gegeven door een parametervoorstelling. De parametervoorstelling van de lijn door de punten P en Q is (Vergelijk formule (18))

$$X = \lambda P + \mu Q.$$

De vergelijking van het halsvlak (3) wordt op homogene coördinaten

$$-x_0^2 + \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 0.$$

Algemeen wordt een kwadriek Γ in $\mathbb{P}^3$ gegeven door

$$F(X) = \sum_{i,k=0}^{3} a_{ik}x_ix_k = 0, \tag{19}$$

waarbij $a_{ik} = a_{ki}$ voor alle k en i. We zien dat $F(X)$ een homogene kwadratische vorm is. We onderzoeken de snijpunten van de kwadriek Γ met de rechte lijn l door de punten $P(p_0, p_1, p_2, p_3)$ en $Q(q_0, q_1, q_2, q_3)$. De parametervoorstelling van l is

$$x_i = \lambda p_i + \mu q_i, \quad i = 0, 1, 2, 3.$$

Het snijpunt van l met Γ vinden we met behulp van

$$F(\lambda P + \mu Q) = \sum_{i,k=0}^{3} a_{ik}(\lambda p_i + \mu q_i)(\lambda p_k + \mu q_k) = 0.$$

Uitwerken van deze vergelijking geeft

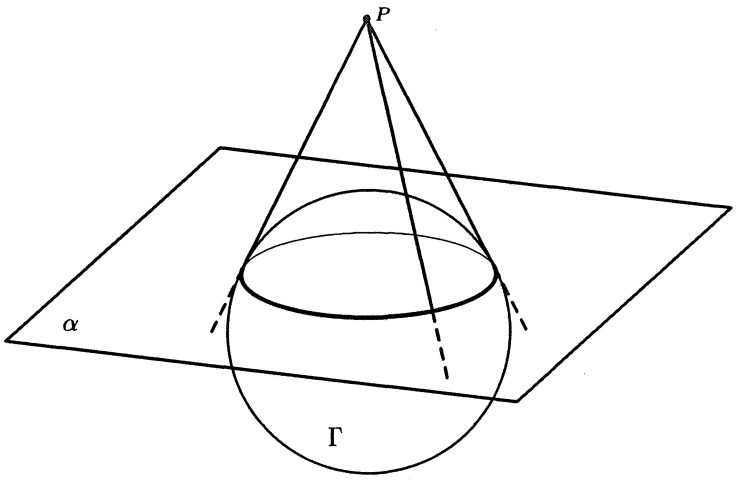
$$\lambda^2 \sum_{i,k=0}^{3} a_{ik}p_ip_k + 2\lambda\mu \sum_{i,k=0}^{3} a_{ik}p_iq_k + \mu^2 \sum_{i,k=0}^{3} a_{ik}q_iq_k = 0.$$

We voeren nu de volgende notatie in:

$$f(P, Q) = \sum_{i,k=0}^{3} a_{ik}p_iq_k.$$

Merk op dat $f(P, Q) = f(Q, P)$ voor alle P en Q. Verder is $f(P, P) = F(P)$ en dus $f(P, P) = 0$ dan en slechts dan als P op Γ ligt. De vergelijking gaat nu over in

$$\lambda^2 f(P, P) + 2\lambda\mu f(P, Q) + \mu^2 f(Q, Q) = 0 \tag{20}$$



Figuur 7: α is het poolvlak van P; ieder punt van α is poolverwant met P ten opzichte van de kwadriek Γ.

We zeggen dat $P(p_0, p_1, p_2, p_3)$ een *dubbelpunt* is van de kwadriek Γ (19) indien $\sum_{i=0}^{3} a_{ik}p_i = 0$ voor $k = 0, 1, 2, 3$. Neem eens aan dat P een dubbelpunt van Γ is. Dan geldt

$$f(P, X) = \sum_{i,k=0}^{3} a_{ik}p_i x_k = \sum_{k=0}^{3} \left(\sum_{i=0}^{3} a_{ik}p_i \right) x_k = 0.$$

voor alle $X(x_0, x_1, x_2, x_3)$. In het bijzonder geldt $P \in \Gamma$. Voor een willekeurig punt $Q(q_0, q_1, q_2, q_3)$ dat verschillend is van P vinden we met (20)

$$F(\lambda P + \mu Q) = \mu^2 f(Q, Q) = 0.$$

We zien dat $\mu = 0$ een dubbele wortel is van deze vergelijking en dat P dus een dubbel tellend snijpunt is van de lijn door P en Q. Dit rechtvaardigt de naam "dubbelpunt". Omgekeerd geldt dat als $\mu = 0$ een dubbele wortel is van de vergelijking (20) voor iedere Q, dan $f(P, Q) = 0$ voor iedere Q. Het punt P ligt dan op Γ en voldoet aan de dubbelpuntsvergelijking.

2.4. Poolverwantschap
We zeggen dat de punten P en Q *poolverwant* zijn ten opzichte van de kwadriek Γ (19) indien $f(P, Q) = 0$. Als P geen dubbelpunt is van Γ, dan stelt de vergelijking $f(P, X) = 0$ een vlak voor; dit vlak heet het *poolvlak* van P. We zullen de meetkundige betekenis van poolverwantschap uit de doeken doen. De eerste eigenschap is evident.

EIGENSCHAP 2.2. Neem aan dat de kwadriek Γ geen dubbelpunt heeft. Dan ligt P in het poolvlak van Q dan en slechts dan als Q in het poolvlak van P ligt.

EIGENSCHAP 2.3. Als $P \in \Gamma$ geen dubbelpunt van Γ is, dan is het poolvlak van P het raakvlak in P aan de kwadriek Γ. De vergelijking van het raakvlak in P luidt: $f(P,X) = 0$.

BEWIJS. Als Q in het poolvlak van P ligt, dan gaat de vergelijking (20) over in

$$F(\lambda P + \mu Q) = \mu^2 f(Q,Q) = 0.$$

We zien dat P een dubbeltellend snijpunt is van de lijn door P en Q met de kwadriek.

EIGENSCHAP 2.4. Het punt P ligt niet op de kwadriek Γ. Voor ieder punt $Q \in \Gamma$ geldt: de lijn door P en Q is raaklijn aan Γ dan en slechts dan als Q in het poolvlak van P ligt. De raaklijnen door P aan Γ vormen een kegel met P als top. De vergelijking van deze kegel is

$$\big(f(P,X)\big)^2 - f(P,P)f(X,X) = 0.$$

BEWIJS. Met $P \notin \Gamma$ en $Q \in \Gamma$ luidt de vergelijking (20)

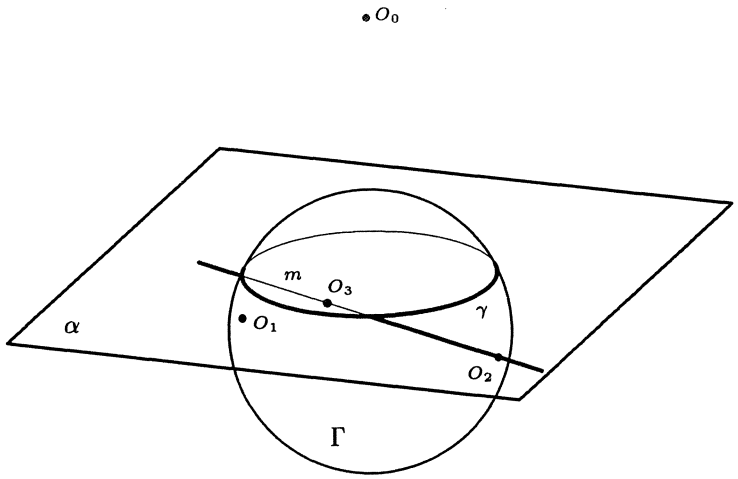
$$\lambda^2 f(P,P) + 2\lambda\mu f(P,Q) = 0.$$

Hierin is $f(P,P) \neq 0$. Dan is $\lambda = 0$ een dubbele wortel van de vergelijking (20) (en dus Q een dubbeltellend snijpunt van de lijn door P en Q met Γ) dan en slechts dan als $f(P,Q) = 0$. Dat bewijst het eerste deel van de eigenschap. Als Q een willekeurig punt is, dan is de lijn door P en Q een raaklijn aan de kwadriek Γ dan en slechts dan als de vergelijking (20) twee samenvallende wortels heeft, oftewel, als de discriminant gelijk is aan 0. Zo komen we aan de vergelijking van de kegel van raaklijnen.

2.5. Klassificatie van de kwadrieken

We gebruiken de theorie over de poolverwantschap om een overzicht te krijgen van de kwadrieken. Laat een kwadriek Γ door de vergelijking (19) gegeven zijn. We kiezen nieuwe grondpunten O_0, O_1, O_2 en O_3 zó dat O_i en O_j poolverwant zijn voor alle $i \neq j$. Dat dit mogelijk is, kan men als volgt inzien. Zie Figuur 8. We nemen hierbij aan dat de kwadriek Γ niet ontaard is. (De kwadriek Γ heet ontaard als $F(X)$ geschreven kan worden als product van twee lineaire factoren.) De behandeling van het geval dat Γ ontaard is, wordt aan de lezer overgelaten. Kies $O_0 \notin \Gamma$ willekeurig. Het poolvlak van O_0 snijdt Γ in een kegelsnede γ. We kiezen O_1 in het poolvlak van O_0, maar buiten Γ. De poolvlakken van O_0 en O_1 snijden elkaar volgens een rechte lijn die we m noemen. Als Γ een dubbelpunt D heeft dan ligt D op m; we kiezen $O_2 = D$ en O_3 op m, verschillend van D. Als Γ geen dubbelpunt heeft dan kiezen we O_2 op m willekeurig, doch niet op γ. De doorsnede van het poolvlak van O_2 met m is dan O_3. We leiden af hoe de vergelijking van Γ in het nieuwe coördinatenstelsel eruit ziet. Voor $i \neq j$ zijn O_i en O_j poolverwant en dus is $f(O_i,O_j) = a_{ij} = 0$. De vergelijking van Γ luidt

$$a_{00}x_0^2 + a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 = 0.$$

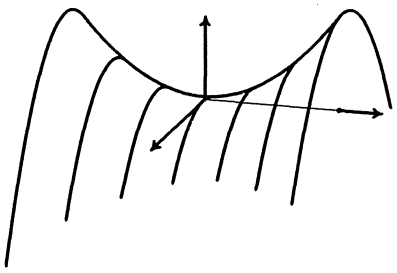


Figuur 8: De nieuwe grondpunten; α is het poolvlak van O_0.

Na schaalverandering, x_i vervangen door $\frac{x_i}{\sqrt{|a_{ii}|}}$, en eventuele permutatie van de grondpunten krijgen we volgende opsomming van projectief verschillende kwadrieken.

1. $x_0^2 + x_1^2 + x_2^2 + x_3^2 = 0$. Er is geen punt in $\mathbb{P}^3$ dat hieraan voldoet.
2. $x_0^2 + x_1^2 + x_2^2 - x_3^2 = 0$. Dit is een ellipsoïde.
3. $x_0^2 + x_1^2 - x_2^2 - x_3^2 = 0$. Dit is een halsvlak.
4. $x_0^2 + x_1^2 + x_2^2 = 0$. Alleen het punt O_3 voldoet.
5. $x_0^2 + x_1^2 - x_2^2 = 0$. Dit is een kegel met top (dubbelpunt) O_3.
6. $x_0^2 + x_1^2 = 0$. Alleen de lijn door O_2 en O_3 voldoet.
7. $x_0^2 - x_1^2 = 0$. Een ontaarde kwadriek: twee verschillende vlakken.
8. $x_0^2 = 0$. Een ontaarde kwadriek: twee samenvallende vakken.

Op de onder 2 genoemde kwadrieken liggen geen rechte lijnen. Hieronder vallen, naast de ellipsoïde, ook de tweebladige omwentelingshyperboloïde en de paraboloïde die in Paragraaf 1 genoemd zijn.
Op de onder 3 genoemde kwadrieken liggen stelsels van rechte lijnen. Hieronder valt ook het zadelvlak (zie Figuur 9). De Eigenschappen 1.1 t/m 1.4 gelden ook voor de stelsels lijnen op het zadelvlak; de genoemde eigenschappen zijn immers invariant onder projectieve transformaties.
Onder geval 5 vallen ook de cilinders.



Figuur 9: Het zadelvlak

OPMERKING 2.5 Het is duidelijk dat de kegelsneden in het projectieve vlak op dezelfde wijze geklassificeerd kunnen worden. Het blijkt dat alle niet ontaarde kegelsneden projectief equivalent zijn.

3. LIJNCOÖRDINATEN IN $\mathbb{P}^3$

We hebben gezien hoe aan een punt of vlak in de projectieve ruimte coördinaten worden toegevoegd. We beschrijven nu hoe aan een lijn homogene coördinaten kunnen worden toegevoegd. De lijncoördinaten zijn ingevoerd door PLÜCKER en worden naar hem genoemd. Aan iedere lijn worden zes homogene coördinaten toegevoegd; op deze manier komt er een correspondentie tot stand tussen de lijnen in de projectieve ruimte en punten uit de $\mathbb{P}^5$. De punten die hierbij optreden vormen een kwadriek van dimensie vier. Dit is de kwadriek van dimensie hoger dan twee waarvan sprake is in de titel van dit verhaal. We zullen verschillende malen gebruik maken van de volgende formule betreffende determinanten.

$$\begin{vmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \\ c_0 & c_1 & c_2 & c_3 \\ d_0 & d_1 & d_2 & d_3 \end{vmatrix} = \begin{vmatrix} a_0 & a_1 \\ b_0 & b_1 \end{vmatrix} \cdot \begin{vmatrix} c_2 & c_3 \\ d_2 & d_3 \end{vmatrix} - \begin{vmatrix} a_0 & a_2 \\ b_0 & b_2 \end{vmatrix} \cdot \begin{vmatrix} c_1 & c_3 \\ d_1 & d_3 \end{vmatrix}$$
$$+ \begin{vmatrix} a_0 & a_3 \\ b_0 & b_3 \end{vmatrix} \cdot \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \cdot \begin{vmatrix} c_0 & c_3 \\ d_0 & d_3 \end{vmatrix}$$
$$- \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \cdot \begin{vmatrix} c_0 & c_2 \\ d_0 & d_2 \end{vmatrix} + \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \cdot \begin{vmatrix} c_0 & c_1 \\ d_0 & d_1 \end{vmatrix}.$$

3.1. Plücker-coördinaten

De lijn l is gegeven door twee van zijn punten $A(a_0, a_1, a_2, a_2)$ en $B(b_0, b_1, b_2, b_3)$. We definiëren de getallen p_{ik} door

$$p_{ik} = \begin{vmatrix} a_i & a_k \\ b_i & b_k \end{vmatrix}, \qquad i, k = 0, 1, 2, 3.$$

Omdat A en B verschillend zijn, kunnen niet alle p_{ik} gelijk aan 0 zijn. We zien dat $p_{ii} = 0$ en $p_{ik} = -p_{ki}$ voor alle i en k. De informatie over de getallen p_{ik}

is dus vastgelegd door het zestal

$$P = (p_{01}, p_{02}, p_{03}, p_{23}, p_{31}, p_{12}). \tag{21}$$

Als we de coördinaten van A of B met λ vermenigvuldigen, dan worden die van p ook met λ vermenigvuldigd. Als we B vervangen door een van A verschillend punt C van l, zeg $C = \lambda A + \mu B$ met $\mu \neq 0$, dan worden de coördinaten van p met μ vermenigvuldigd. Hieruit zien we dat de verhouding van het zestal (21) onafhankelijk is van de keuze van de punten A en B van l. Het zestal (21) noemen we de *Plücker-coördinaten van de lijn l*. We kunnen de Plücker-coordinaten van de lijn nu opvatten als homogene coördinaten van een punt in $\mathbb{P}^5$. Door uitrekenen van de determinant

$$\begin{vmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \\ a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \end{vmatrix}$$

vinden we de volgende relatie tussen de p_{ik}

$$\Omega(P) = 2\big(p_{01}p_{23} + p_{02}p_{31} + p_{03}p_{12}\big) = 0. \tag{22}$$

Nu zullen we laten zien dat met ieder punt P dat voldoet aan $\Omega(P) = 0$ een lijn l correspondeert zó dat de coördinaten van P juist de Plücker-coördinaten van l zijn. Daartoe onderzoeken we eerst de meetkundige betekenis van de Plücker-coördinaten. Naar analogie van de formule (17) vinden we voor het vlak door de drie punten $A(a_0, a_1, a_2, a_3)$, $B(b_0, b_1, b_2, b_3)$ en $C(c_0, c_1, c_2, c_3)$ die niet op één lijn liggen de vergelijking

$$\begin{vmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \\ c_0 & c_1 & c_2 & c_3 \\ x_0 & x_1 & x_2 & x_3 \end{vmatrix} \tag{23}$$

Door nu voor C achtereenvolgens in te vullen O_0, O_1, O_2 en O_3 vinden de volgende vergelijkingen

$$\begin{array}{lccccccccc} \alpha_0: & & & p_{23}x_1 & + & p_{31}x_2 & + & p_{12}x_3 & = & 0 \\ \alpha_1: & p_{32}x_0 & & & + & p_{03}x_2 & + & p_{20}x_3 & = & 0 \\ \alpha_2: & p_{13}x_0 & + & p_{30}x_1 & & & + & p_{01}x_3 & = & 0 \\ \alpha_3: & p_{21}x_0 & + & p_{02}x_1 & + & p_{10}x_2 & & & = & 0 \end{array} \tag{24}$$

Als O_i op de lijn door A en B ligt dan zijn alle coëfficiënten van de de vergelijking α_i gelijk aan 0. We vinden de Plückercoördinaten van de lijn door A en B dus terug in de coördinaten van de vier vlakken door A, B en elk der grondpunten. Berekenen we de determinant van de coëfficientenmatrix van (24), die uiteraard 0 moet zijn, dan vinden we de waarde $\frac{1}{4}(\Omega(P))^2$. Dit is een bevestiging van (22). Nu geldt de volgende stelling.

STELLING 3.1 Een zestal $P = (p_{01}, p_{02}, p_{03}, p_{23}, p_{31}, p_{12})$ zijn Plücker-coördinaten van een lijn dan en slechts dan indien tenminste één p_{ik} ongelijk aan 0 is en $\Omega(P) = 0$.

BEWIJS. De "slechts dan"-kant is al bewezen. Neem aan dat $p_{23} \neq 0$ en dat $\Omega(P) = 0$. Vermenigvuldig de vergelijkingen van α_0, α_1 en α_2 met p_{03}, p_{13} respectievelijk p_{23} en tel de resultaten op. We vinden dat de drie vergelijkingen afhankelijk zijn. Op dezelfde wijze blijkt dat ieder drietal vergelijkingen in (24) afhankelijk is. Er volgt dat de rang van de coëfficiëntenmatrix in (24) gelijk is aan 2. Dan is de oplossing van dat stelsel een lijn; de Plücker-coördinaten van deze lijn komen overeen met het gegeven zestal.

OPMERKING 3.2 Uit het bewijs blijkt dus: de vergelijkingen (24) zijn de vergelijkingen van de lijn in $\mathbb{P}^3$ met de Plücker-coördinaten P.

STELLING 3.3 $P = (p_{01}, p_{02}, p_{03}, p_{23}, p_{31}, p_{12})$ zijn Plücker-coördinaten van de lijn l. De lijn m heeft Plücker-coördinaten $Q = (q_{01}, q_{02}, q_{03}, q_{23}, q_{31}, q_{12})$. Dan geldt dat l en m elkaar snijden dan en slechts dan als

$$\begin{aligned} \omega(P,Q) \quad = \quad & p_{01}q_{23} + p_{02}q_{31} + p_{03}q_{12} + \\ & + q_{01}p_{23} + q_{02}p_{31} + q_{03}p_{12} = 0. \end{aligned} \tag{25}$$

BEWIJS. Laat $A(a_0, a_1, a_2, a_3)$ en $B(b_0, b_1, b_2, b_3)$ verschillende punten van l zijn en laat $C(c_0, c_1, c_2, c_3)$ en $D(d_0, d_1, d_2, d_3)$ verschillende punten van m zijn. Nu snijden l en m elkaar dan en slechts dan indien A, B, C en D in een vlak liggen. Dit laatste geldt dan en slechts dan indien

$$\begin{vmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \\ c_0 & c_1 & c_2 & c_3 \\ d_0 & d_1 & d_2 & d_3 \end{vmatrix} = 0.$$

Ontwikkeling van de determinant geeft de vergelijking (25).

We definiëren de hyperkwadriek Ω door

$$\Omega(X) = 2x_0x_3 + 2x_1x_4 + 2x_2x_5 = 0. \tag{26}$$

Ω is een (vierdimensionaal) oppervlak in $\mathbb{P}^5$. Uit de Stelling ?? blijkt dat Ω precies de verzameling is van alle Plücker-coördinaten van lijnen in $\mathbb{P}^3$. De vergelijking (25) zegt niets anders dan dat P en Q poolverwant zijn ten opzichte van Ω.
Uit het voorgaande volgt de lijn l met Plücker-coördinaten P en de lijn m met Plückercoördinaten Q elkaar snijden dan en slechts dan als P en Q poolverwant zijn. Als P en Q poolverwant zijn, dan ligt de lijn door P en Q geheel op Ω. Immers,

$$\Omega(\lambda P + \mu Q) = \lambda^2\omega(P,P) + 2\lambda\mu\omega(P,Q) + \mu^2\omega(Q,Q) = 0.$$

Hieruit zien we dat Ω rechte lijnen bevat. Men kan bewijzen dat een rechte lijn op Ω correspondeert met een lijnenwaaier in $\mathbb{P}^3$.
Naast lijnen zijn er ook stelsels vlakken die op Ω liggen. Laat l, m en n lijnen zijn met Plücker-coördinaten P, Q respectievelijk R. Als de lijnen l, m en n elkaar twee aan twee snijden, dan ligt het vlak door P, Q en R geheel op Ω. Dit volgt uit de formule

$$\Omega(\lambda P + \mu Q + \nu R) = 0$$

welke men eenvoudig door uitrekenen kan controleren. Er zijn twee typen van vlakken die op Ω liggen. Eén type correspondeert met het geval dat l, m en n door één punt gaan, het andere type met het geval dat l, m en n in één vlak liggen.

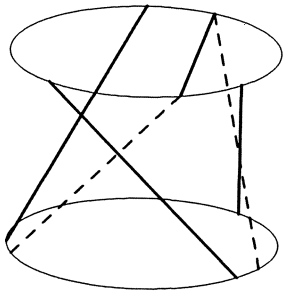
3.2. Transversalen van kruisende lijnen
We komen nu toe aan de behandeling van de vraag die in de aankondiging van deze voordracht werd gesteld:

hoeveel transversalen zijn er bij vier kruisende lijnen?

(Een transversaal is een lijn die de vier lijnen snijdt.) Die vraag kunnen we nu beantwoorden. Laat de vier lijnen Plücker-coördinaten P, Q, R en S hebben. Noem de Plücker-coördinaten van de gezochte transversaal X. Dan moet X voldoen aan de vergelijkingen

$$\omega(P,X) = 0, \quad \omega(Q,X) = 0, \quad \omega(R,X) = 0, \quad \omega(S,X) = 0.$$

Zijn deze vergelijkingen onafhankelijk (en dat zijn ze in het algemeen) dan is de oplossingsruimte een lijn v in $\mathbb{P}^5$. Niet ieder punt op v zal corresponderen met een lijn in $\mathbb{P}^3$. Dat doen alleen de snijpunten van v met de hyperkwadriek Ω. Als v dus niet in Ω ligt, dan zijn er 0, 1 of 2 transversalen. Als de vergelijkingen afhankelijk zijn, dan zijn er oneindig veel transversalen.
Ook al is hiermee de vraag enigzins beantwoord, dan weten we nog niet hoe we ons dit moeten voorstellen. Daarom gaan we eerst eens terug naar een vraagstuk dat vroeger populair was bij de MO-examens. We brengen in herinnering dat een *regelvlak* is opgebouwd uit rechte lijnen die drie gegeven ruimtekrommen (de richtkrommen) snijden. Vaak werd gevraagd om het regeloppervlak te bepalen met drie gegeven rechte lijnen als richtkrommen. Dat regeloppervlak bleek altijd een oppervlak van de tweede graad te zijn. Dat was geen toeval. Men kan bewijzen dat dit altijd het geval moet zijn. Uit het voorgaande weten we nu dat het gevraagde regeloppervlak dan een halsvlak of een zadelvlak moet zijn. We weten ook dat het regeloppervlak dan projectief equivalent is met het halsvlak dat in Paragraaf 1 bestudeerd is.
Bekijken we nu nog eens Figuur 5. En denken we nu nog eens na over de vraag hoeveel transversalen er zijn voor vier gegeven kruisende rechte lijnen. We bekijken eerst drie van de gegeven lijnen. De transversalen van deze drie lijnen vormen een regelvlak. Zonder beperking der algemeenheid mogen we aannemen dat dit regelvlak een halsvlak is. Uit de eigenschappen die we in



Figuur 10: De twee transversalen van vier kruisende rechten

Paragraaf 1 bestudeerd hebben weten we nu dat de transversalen één van de stelsels lijnen op het halsvlak vormen en dat de drie gegeven lijnen tot het andere stelsel moeten behoren. We hoeven nu alleen nog maar uit het stelsel van transversalen er twee(!) uit te zoeken die de vierde gegeven lijn snijden. Dit is in Figuur 10 uitgebeeld.
We merken nog op dat elk stelsel van lijnen op het halsvlak correspondeert met een op Ω gelegen kegelsnede. Kunt U dit verklaren?

Over het voorgaande en nog veel meer kan men lezen in de "oude boeken". We noemen slechts:
J.A. BARRAU, *Analytische Meetkunde*, delen I (Het Platte Vlak) en II (De Ruimte), Noordhoff, Groningen 1927

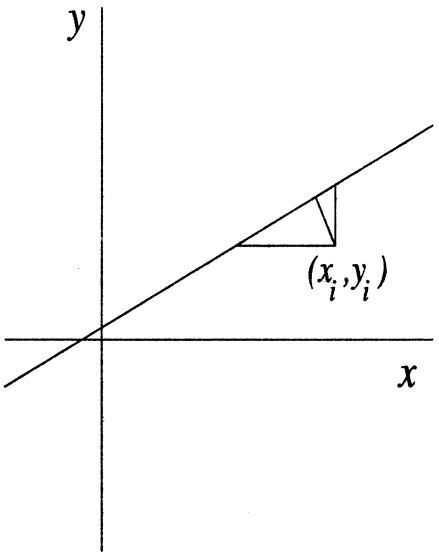
Kwadratische Vormen en Metriek

A.G. van Asch

1. Kleinste kwadraten

1.1. Door één punt in het vlak gaan oneindig veel rechte lijnen, door twee punten in het vlak gaat één rechte lijn, en door drie punten in het algemeen geen. In het geval van drie of meer punten in het vlak die min of meer op een rechte lijn liggen is men soms geïnteresseerd in die rechte lijn die het beste bij deze punten past. Stel dat er gegeven zijn punten (x_i, y_i), $i = 1, ..., m$. De "gewone" afstand van een punt (x_i, y_i) tot een rechte lijn met vergelijking $ax + by + c = 0$ wordt gegeven door

$$\left|\frac{ax_i + by_i + c}{\sqrt{a^2 + b^2}}\right|$$



Figuur 1:

Om nu een rechte lijn te bepalen die zo goed mogelijk bij deze punten past kunnen we bijvoorbeeld proberen a, b, c te vinden zó dat

$$\sum_{i=1}^{m} \left(\frac{ax_i + by_i + c}{\sqrt{a^2 + b^2}}\right)^2$$

minimaal is.
Maar we kunnen ook bijvoorbeeld naar de "horizontale" of naar de "verticale" afstand van zo'n punt tot de lijn kijken.
In het eerste geval zullen we a, b en c proberen te vinden zó dat

$$\sum_{i=1}^{m} \left(\frac{ax_i + by_i + c}{a}\right)^2$$

minimaal is, en in het tweede geval zullen we

$$\sum_{i=1}^{m} \left(\frac{ax_i + by_i + c}{b}\right)^2$$

proberen te minimaliseren.
In tal van practische situaties geniet dit laatste de voorkeur. Stel dat we een verband onderzoeken tussen grootheden x en y. We hebben een aantal waarden x_i en vinden daarbij meetgegevens (benaderingen) y_i. We zoeken (omdat we dat uit een model vermoeden) een lineaire functie $y = ax + b$ die dit verband zo goed mogelijk weergeeft. En dus proberen we de som van de "verticale" afstanden te minimaliseren. Dit betekent dat we waarden voor a en b zoeken zó dat

$$S = \sum_{i=1}^{m} \{y_i - (ax_i + b)\}^2$$

minimaal is.
Dit heet de methode van de kleinste kwadraten.
Bij gegeven punten (x_i, y_i), $i = 1, ..., m$, is S een functie van twee variabelen. Het minimum kan

dan bepaald worden via $\dfrac{\partial S}{\partial a} = 0$ en $\dfrac{\partial S}{\partial b} = 0$.
Dus

$$\sum_{i=1}^{m} x_i y_i = a \sum_{i=1}^{m} x_i^2 + b \sum_{i=1}^{m} x_i$$

$$\sum_{i=1}^{m} y_i \quad = a \sum_{i=1}^{m} x_i \; + bm \; .$$

Voeren we de notatie

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i \; , \qquad \bar{y} = \frac{1}{m} \sum_{i=1}^{m} y_i \; ,$$

in, dan volgt als oplossing

$$a = \frac{\sum_{i=1}^{m} (x_i - \bar{x})\, y_i}{\sum_{i=1}^{m} (x_i - \bar{x})^2} \; , \qquad b = \bar{y} - a\bar{x} \; .$$

In de volgende paragraaf bekijken we dit probleem vanuit een andere gezichtshoek. Tevens breiden we het probleem uit van $\mathbb{R}^2$ naar $\mathbb{R}^n$.

1.2. We hebben een stelsel van m lineaire vergelijkingen met n onbekenden:

$$\begin{array}{rcl} p_{11}z_1 + p_{12}z_2 + \ldots + p_{1n}z_n & = & q_1 \\ p_{21}z_1 + p_{22}z_2 + \ldots + p_{2n}z_n & = & q_2 \\ \vdots & & \vdots \\ p_{m1}z_1 + p_{m2}z_2 + \ldots + p_{mn}z_n & = & q_m \end{array}$$

We beschouwen $\begin{pmatrix} p_{11} \\ \vdots \\ p_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} p_{1n} \\ \vdots \\ p_{mn} \end{pmatrix}, \begin{pmatrix} q_1 \\ \vdots \\ q_m \end{pmatrix}$ als elementen van $\mathbb{R}^m$. Voor variabele $z_1, \ldots, z_n$ stellen de linkerleden van het stelsel elementen in een lineaire deelruimte van $\mathbb{R}^m$ voor. Als $\begin{pmatrix} q_1 \\ \vdots \\ q_m \end{pmatrix}$ in die deelruimte ligt dan is er dus een oplossing voor de z_i. Ligt $\begin{pmatrix} q_1 \\ \vdots \\ q_m \end{pmatrix}$ niet in de deelruimte dan zoeken we een punt in de deelruimte dat zo dicht mogelijk bij $\begin{pmatrix} q_1 \\ \vdots \\ q_m \end{pmatrix}$ ligt. Dat vinden we door vanuit $\begin{pmatrix} q_1 \\ \vdots \\ q_m \end{pmatrix}$ een loodlijn op de deelruimte te construeren en het voetpunt van deze loodlijn geeft de gezochte getallen $\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$.

We werken dit nu verder uit, en voeren de volgende notaties in.

$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mn} \end{pmatrix}$ is de coëfficiënten-matrix van het stelsel vergelijkingen; de kolommen van P noteren we kortweg als $\mathbf{p}_1, \ldots, \mathbf{p}_n$. De door $\mathbf{p}_1, \ldots, \mathbf{p}_n$ voortgebrachte lineaire deelruimte van $\mathbb{R}^m$ noteren we als $K(P)$: de kolommenruimte van P.

De rang van de matrix P wordt wel gedefiniëerd als de dimensie van $K(P)$; dit is dus het maximale aantal lineair onafhankelijke kolommen van P.

Voor

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{R}^n$$

kunnen we het matrix-product $P\mathbf{z}$ ook schrijven als

$$P\mathbf{z} = z_1\mathbf{p}_1 + \ldots + z_n\mathbf{p}_n \ .$$

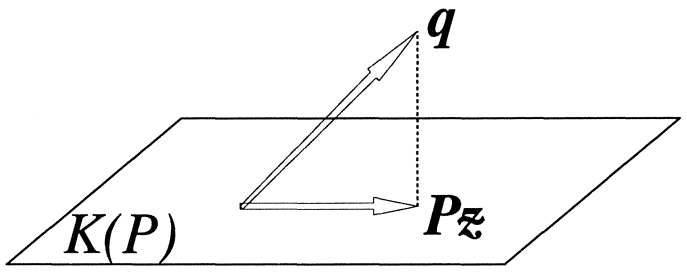
Het stelsel lineaire vergelijkingen laat zich herschrijven tot

$$P\mathbf{z} = \mathbf{q} \ ,$$

en er geldt

$$P\mathbf{z} = \mathbf{q} \text{ is oplosbaar } \Leftrightarrow \mathbf{q} \in K(P) \ .$$

Als $\mathbf{q} \notin K(P)$ dan heeft het stelsel $P\mathbf{z} = \mathbf{q}$ dus geen oplossingen. We kunnen in dat geval vragen naar een vector $\tilde{\mathbf{z}}$ zó dat $P\tilde{\mathbf{z}}$ zo dicht mogelijk in de buurt van $\mathbf{q}$ ligt. Nu is voor elke $\tilde{\mathbf{z}}$ de vector $P\tilde{\mathbf{z}}$ een element van $K(P)$. De afstand tussen $P\tilde{\mathbf{z}}$ en $\mathbf{q}$ moet minimaal zijn, dus de lengte van $P\tilde{\mathbf{z}} - \mathbf{q}$ moet minimaal zijn.



Figuur 2:

Dit betekent dat $P\tilde{\mathbf{z}} - \mathbf{q}$ loodrecht op $K(P)$ moet staan, met andere woorden we zoeken $\tilde{\mathbf{z}}$ zó dat $P\tilde{\mathbf{z}}$ precies de projectie van $\mathbf{q}$ op $K(P)$ is.
Voor een matrix P, resp. vector $\mathbf{p}$, noteren we met $P^\top$, resp. $\mathbf{p}^\top$ de getransponeerde (gespiegelde) van P, resp. $\mathbf{p}$. Dan is $\mathbf{p}^\top$ dus een rij-vector. Nu zoeken we $\tilde{\mathbf{z}}$ zó dat

$$(P\tilde{\mathbf{z}} - \mathbf{q}) \perp K(P) \ ,$$

dus

$$(P\tilde{\mathbf{z}} - \mathbf{q}) \perp \mathbf{p}_i \ , \qquad i = 1, ..., n \ .$$

Dit kunnen we herschrijven tot

$$\mathbf{p}_i^\top (P\tilde{\mathbf{z}} - \mathbf{q}) = 0 \ , \qquad i = 1, ..., n \ ,$$

ofwel

$$P^T (P\tilde{\mathbf{z}} - \mathbf{q}) = \mathbf{0} \ ,$$

dus

$$P^\top P\tilde{\mathbf{z}} = P^\top \mathbf{q} \ .$$

Deze vergelijking (of eigenlijk dit stelsel vergelijkingen) heet de bij het oorspronkelijke stelsel behorende normaalvergelijking.
We bekijken nu de situatie dat

$m > n$ (aantal vergelijkingen groter dan aantal onbekenden)

$\mathbf{q} \notin K(P)$.

We spreken dan van een overbepaald stelsel. De normaalvergelijking is altijd oplosbaar: òf er is precies één oplossing, òf er zijn oneindig veel oplossingen. De aard van de oplossingsverzameling hangt af van de rang van de matrix P. Er geldt nl.:

als rang$(P) = n$ (dus de kolommen van P zijn lineair onafhankelijk),

dan is de matrix $P^\top P$ inverteerbaar.

Dit betekent dus dat in het geval rang$(P) = n$ de normaalvergelijking precies één oplossing heeft, nl.

$$\tilde{\mathbf{z}} = (P^\top P)^{-1} P^\top \mathbf{q} \ .$$

De aldus verkregen vector $\tilde{\mathbf{z}}$ heet de kleinste kwadraten-oplossing van het stelsel vergelijkingen.
Het probleem uit 1.1 laat zich ook in deze vorm beschrijven. Het gaat daarbij om het overbepaalde stelsel lineaire vergelijkingen

$$\begin{aligned} ax_1 \ &+ \ b \ = \ y_1 \\ ax_2 \ &+ \ b \ = \ y_2 \\ &\vdots \\ ax_m \ &+ \ b \ = \ y_m \ . \end{aligned}$$

Gegeven zijn $x_1, ..., x_m$ en $y_1, ..., y_m$, gevraagd worden a en b. Nu moet dus de afstand van $\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ tot de door $\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ en $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ opgespannen deelruimte geminimaliseerd worden. Het kwadraat van de afstand van $\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ tot een vector $a \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} + b \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ in die deelruimte wordt gegeven door

$$\sum_{i=1}^{m} \{y_i - (ax_i + b)\}^2 \ ,$$

en we vinden dus precies de uitdrukking terug die we ook in 1.1 gebruikten.
Stellen we met behulp van de matrix $P = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}$ de normaalvergelijking op voor het stelsel vergelijkingen, dan vinden we

$$\begin{pmatrix} \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & m \end{pmatrix} \begin{pmatrix} \tilde{a} \\ \tilde{b} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} y_i \end{pmatrix} \ ,$$

en dat levert dus ook de vergelijkingen uit 1.1.

1.3. Het bepalen van de kleinste kwadraten-oplossing van een overbepaald stelsel kan vervelend rekenwerk met zich meebrengen. Speciaal als de coëfficiënten van de matrix P het resultaat zijn van metingen (en dus zeer waarschijnlijk geen mooie gehele getallen) kan het bepalen van $(P^\top P)^{-1}$ lastig zijn. Het ligt voor de hand voor dit rekenwerk een computer in te schakelen.
Computer-algebra programma's kunnen met matrices rekenen; tot op zekere hoogte kunnen resultaten zelfs exact worden bepaald. Worden de coëfficiënten van P en $\mathbf{q}$ als rationale getallen ingevoerd, dan wordt de kleinste kwadraten-oplossing ook in rationale getallen gegeven.
Het overbepaalde stelsel

$$\begin{pmatrix} 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 8 & 64 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 5 \\ 12 \end{pmatrix}$$

wordt b.v. met behulp van het programma DERIVE als volgt opgelost.
Noemen we de gevonden vector $\mathbf{z}$, dan kunnen we DERIVE ook $P\mathbf{z}$ laten uitrekenen:
Een programma dat specifiek gericht is op het rekenen met matrices is PCMATLAB. Met dit programma kan vrijwel uitsluitend numeriek gerekend worden. Bij een ingevoerde matrix P en vector $\mathbf{q}$ kan binnen dit programma de opdracht

$$P\backslash\mathbf{q}$$

gegeven worden. Dit levert op

òf een oplossing $\mathbf{z}$ van $P\mathbf{z} = \mathbf{q}$

òf de kleinste kwadraten-oplossing van dit stelsel vergelijkingen .

In het volgende voorbeeld is gebruik gemaakt van dit programma. De stand van het wereldrecord op de marathon (42.195 km) is in het onderstaande overzicht aangegeven voor zowel vrouwen (v) als mannen (m), steeds met een tussenperiode van 10 jaar. Ook zijn de gemiddelde snelheden, v_{gem}, vermeld.

jaar	v	v_{gem}(m/s)	m	v_{gem}(m/s)
1950	3.45.26	3.1195	2.25.39	4.8284
1960	3.27.45	3.3851	2.15.16	5.1990
1970	3.02.53	3.8453	2.08.33	5.4706
1980	2.25.41	4.8273	2.08.13	5.4849
1990	2.21.06	4.9841	2.06.50	5.5447

Gaan we in beide gevallen uit van een lineair verband tussen v en t (t in jaren, gemeten vanaf 1950), $v = at + b$, dan krijgen we de volgende twee overbepaalde stelsels:

1: $P := \begin{bmatrix} 1 & 2 & 4 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 8 & 64 \end{bmatrix}$

2: $q := [1, 3, 5, 12]`$

3: $(P` \cdot P)^{-1} \cdot P` \cdot q$

4: $\begin{bmatrix} \frac{31}{150} \\ \frac{1}{100} \\ \frac{11}{60} \end{bmatrix}$

TRANSFER PRINT FILE: Expressions Screen Window

Rekentijd: 0.0 seconden
Simp(3) Free:100% Derive Algebra

Figuur 3:

$$\begin{pmatrix} 0 & 1 \\ 10 & 1 \\ 20 & 1 \\ 30 & 1 \\ 40 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3.1195 \\ 3.3851 \\ 3.8453 \\ 4.8273 \\ 4.9841 \end{pmatrix} \quad \text{en}$$

$$\begin{pmatrix} 0 & 1 \\ 10 & 1 \\ 20 & 1 \\ 30 & 1 \\ 40 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 4.8284 \\ 5.1990 \\ 5.4706 \\ 5.4849 \\ 5.5447 \end{pmatrix} .$$

Met PCMATLAB berekenen we:

```
>> P =
```

6: P · z

7:

$$\begin{bmatrix} \frac{24}{25} \\ \frac{159}{50} \\ \frac{121}{25} \\ \frac{601}{50} \end{bmatrix}$$

TRANSFER PRINT FILE: Expressions Screen Window

Rekentijd: 0.1 seconden
Simp(6) Free:100% Derive Algebra

Figuur 4:

```
 0  1        >> q1 =              >> q2 =
10  1              3.1195               4.8284
20  1              3.3851               5.1990
30  1              3.8453               5.4706
40  1              4.8273               5.4849
                   4.9841               5.5447

>> P\q1

   ans =
         0.0517
         2.9980

>> P\q2
```

```
ans =
        0.0172
        4.9618
```

Voor vrouwen zou hier dus uit volgen

$$v = 0.0517t + 2.9980 ,$$

en voor mannen

$$v = 0.0172t + 4.9618 .$$

Voor $t = 56.9217$ zijn deze snelheden aan elkaar gelijk, dus rond het jaar 2007 zullen vrouwen en mannen de marathon even snel kunnen lopen!

2. Metriek in het vlak

2.1 De standaardmetriek in het vlak definiëert de afstand van twee punten met coördinaten (x_1, x_2) en (y_1, y_2) als $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. Duidelijk is dat we ook op andere manieren, niet met de wortel uit een kwadratische vorm, de afstand kunnen definiëren. Zo kunnen we b.v. de zogenaamde New-York metriek $|x_1 - y_1| + |x_2 - y_2|$ hanteren.
Beide metrieken zijn invariant onder translaties. Maar de Euclidische metriek, berustend op de stelling van Pythagoras, is ook invariant onder rotaties. Er is dus een verband tussen de kwadratische vorm en de groep van rotaties in het vlak. We zouden dus ook van rotaties, of van de groep van rotaties, de orthogonale groep, of zelfs van hoeken kunnen uitgaan, en van daaruit proberen een invariante afstand te definiëren. We zouden dan niet te afhankelijk van coördinaten in een lineaire ruimte willen werken. Het inproduct van twee vectoren bepaalt in wezen nu de orthogonale groep, het hoekbegrip en de metriek. In [1] werd een op Bourbaki gebaseerde introductie van het hoekbegrip gegeven. Hieronder volgen de hoofdlijnen.
Uitgangspunt is een 2-dimensionale vectorruimte E over $\mathbb{R}$, voorzien van een inproduct $\Phi(\mathbf{x}, \mathbf{y})$. Een bijectieve lineaire afbeelding $u : E \to E$ met de eigenschap $\Phi\left(u(\mathbf{x}), u(\mathbf{y})\right) = \Phi(\mathbf{x}, \mathbf{y})$ voor $\mathbf{x}, \mathbf{y} \in E$ heet een orthogonale afbeelding. Laat $\{\mathbf{e}_1, \mathbf{e}_2\}$ een orthonormale basis van E zijn, d.w.z. $\Phi(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij}$ $(= 1$ als $i = j$, 0 als $i \neq j)$. De coëfficiënten van de matrix van een orthogonale afbeelding u t.o.v. een orthonormale basis voldoen aan een aantal relaties. Als nl.

$$\begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$$

de matrix is, dan volgt uit

$$\Phi\left(u(\mathbf{e}_i), u(\mathbf{e}_j)\right) = \Phi(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij}$$

dat

$$\begin{cases} \alpha^2 + \beta^2 = 1 \\ \gamma^2 + \delta^2 = 1 \\ \alpha\gamma + \beta\delta = 0\,. \end{cases}$$

Hieruit volgt dat

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

en dus is $(\det(u))^2 = 1$, m.a.w. $\det(u) = \pm 1$.
De orthogonale afbeeldingen vormen een groep, de orthogonale groep; notatie O.
De verzameling

$$O^+ = \{u \in O \mid \det(u) = 1\}$$

is een ondergroep van O. De elementen van O^+ noemen we rotaties.
Het is nu gemakkelijk in te zien dat voor de matrix van een rotatie u naast de al eerder genoemde relaties ook nog geldt

$$\alpha = \delta \quad \text{en} \quad \beta = -\gamma\,.$$

Dat betekent dat elke $u \in O^+$ ten opzichte van een orthonormale basis een matrix-voorstelling

$$\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$$

heeft, waarbij $\alpha, \beta \in \mathbb{R}$, $\alpha^2 + \beta^2 = 1$.
De volgende propositie is niet moeilijk te bewijzen.

Propositie 1.
Als $\mathbf{x}, \mathbf{y} \in E\backslash\{\mathbf{0}\}$, en $\Phi(\mathbf{x},\mathbf{x}) = \Phi(\mathbf{y},\mathbf{y})$, dan is er precies één $u \in O^+$ met $u(\mathbf{x}) = \mathbf{y}$.

Het begrip halfrechte wordt als volgt ingevoerd. Laat $\mathbf{x} \in E$, $\mathbf{x} \neq \mathbf{0}$, dan heet

$$[\mathbf{x}] = \{\lambda\mathbf{x} \mid \lambda \in \mathbb{R},\ \lambda \geq 0\}$$

de door $\mathbf{x}$ voortgebrachte halfrechte.
Uit Propositie 1 volgt onmiddellijk

Propositie 2.
Als D_1 en D_2 halfrechten zijn, dan is er precies één $u \in O^+$ met $u(D_1) = D_2$.

Hoeken worden nu ingevoerd via geordende paren van halfrechten. We definiëren een relatie $\sim$ als volgt

$$(D_1, D_2) \sim (D_1', D_2') \Leftrightarrow \text{er is } u \in O^+ \text{ met } u(D_1) = D_1' \text{ en } u(D_2) = D_2'\,.$$

Deze relatie is een equivalentie-relatie. De equivalentie-klasse van (D_1, D_2) wordt als $\angle(D_1, D_2)$ genoteerd en dit wordt de hoek van de halfrechten D_1 en D_2 genoemd.
In de verzameling van alle hoeken kunnen we een optelling definiëren:

$$\angle(D_1, D_2) + \angle(D_3, D_4) = \angle(D_1, u(D_4)) \ ,$$

waarbij $u \in O^+$ zó dat $u(D_3) = D_2$.
Dit is een goed gedefiniëerde bewerking (d.w.z. onafhankelijk van de gekozen representant uit een equivalentie-klasse), en met deze bewerking wordt de verzameling H van alle hoeken een abelse groep. En op grond van de volgende propositie kunnen we hoeken identificeren met rotaties.

Propositie 3.
De afbeelding $h : H \to O^+$ gedefiniëerd door $h(\angle(D_1, D_2)) = u$, waarbij $u(D_1) = D_2$, is een isomorfisme.

In deze laatste propositie wordt dus precies het verband tussen hoeken en rotaties aangegeven. De introductie van beide is gebaseerd op het inproduct $\Phi(\mathbf{x}, \mathbf{y})$. Via dit inproduct kan ook een metriek gedefiniëerd worden door

$$\sqrt{\Phi(\mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y})} \ .$$

Uit de constructie is duidelijk dat deze afstands-functie invariant is onder elke rotatie.

2.2 In de vorige paragraaf lieten we zien dat de klassieke definitie van afstand met de wortel uit een kwadratische vorm de eigenschap heeft invariant onder translaties en rotaties te zijn. De New-York metriek is wel invariant onder translaties, maar niet onder rotaties. Het ligt voor de hand er bij de definitie van een afstands-functie in ieder geval naar te streven dat deze invariant onder translaties zal zijn. Dan is het in b.v. de n-dimensionale ruimte $\mathbb{R}^n$ voldoende een norm te definiëren, d.w.z. een functie $\mathbf{x} \to \|\mathbf{x}\|$ die aan vectoren reële getallen toevoegt zodanig dat

1. $\|\mathbf{x}\| \geq 0$ voor alle $\mathbf{x}$,
2. $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$,
3. $\|\lambda\mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$ voor alle $\mathbf{x}$, en alle $\lambda \in \mathbb{R}$,
4. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ voor alle $\mathbf{x}$, $\mathbf{y}$.

Als we vectoren $\mathbf{x}$ en $\mathbf{y}$ ook opvatten als punten X en Y in $\mathbb{R}^n$ dan kunnen we met behulp van zo'n norm de afstand tussen deze punten definiëren als $d(X, Y) = \|\mathbf{x} - \mathbf{y}\|$.
Deze metriek is invariant onder translaties. De tweede eis voor een norm impliceert dat de afstand tussen twee verschillende punten steeds positief is, en uit de vierde eis volgt de driehoeksongelijkheid:

$$d(X,Z) \leq d(X,Y) + d(Y,Z) \ .$$

(Voor de kenners een opmerking: In de speciale relativiteitstheorie voert men een "afstand" tussen gebeurtenissen in, wel met een kwadratische vorm, maar één die niet eigenschap 2 bezit. De afstand tussen twee gebeurtenissen (x,y,z,t) en (x',y',z',t') is in de speciale relativiteitstheorie immers

$$\sqrt{(x-x')^2+(y-y')^2+(z-z')^2-c^2(t-t')^2}$$

.)
We zullen als voorbeeld in de volgende paragraaf meetkundige eigenschappen onderzoeken als we een metriek in het vlak invoeren door middel van een norm die niet samenhangt met een kwadratische vorm.
In deze paragraaf bekijken we algemeen een klasse van normen op $\mathbb{R}^n$ gedefiniëerd door

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \ , \qquad p \geq 1 \ .$$

Voor $p=1$ is dit de New-York norm, voor $p=2$ de gewone Euclidische. We bewijzen nu dat door deze definitie inderdaad een norm gegeven wordt. De eigenschappen 1, 2 en 3 zijn onmiddellijk duidelijk. We bewijzen eigenschap 4 voor $p>1$. (Voor $p=1$ is eigenschap 4 duidelijk.)
Noem q het getal waarvoor geldt $\frac{1}{p}+\frac{1}{q}=1$. Door de functie $f(t)=\frac{t^p}{p}+\frac{t^{-q}}{q}$, $t>0$, te onderzoeken via $f'(t)$ vinden we dat $f(t) \geq f(1) = 1$ voor alle $t>0$. Door dan $t=\frac{a^{1/q}}{b^{1/p}}$, a en b positieve getallen, in te vullen volgt de ongelijkheid

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \ .$$

Laat nu

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \ , \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \ .$$

Dan geldt volgens bovenstaande ongelijkheid:

$$\frac{|x_i|}{\|\mathbf{x}\|_p} \cdot \frac{|x_i+y_i|^{p-1}}{\|\mathbf{x}+\mathbf{y}\|_p^{p/q}} \leq \frac{1}{p} \ \frac{|x_i|^p}{\|\mathbf{x}\|_p^p} + \frac{1}{q} \ \frac{|x_i+y_i|^{(p-1)q}}{\|\mathbf{x}+\mathbf{y}\|_p^p} \ ,$$

en

$$\frac{|y_i|}{\|\mathbf{y}\|_p} \cdot \frac{|x_i+y_i|^{p-1}}{\|\mathbf{x}+\mathbf{y}\|_p^{p/q}} \leq \frac{1}{p} \ \frac{|y_i|^p}{\|\mathbf{y}\|_p^p} + \frac{1}{q} \ \frac{|x_i+y_i|^{(p-1)q}}{\|\mathbf{x}+\mathbf{y}\|_p^p} \ ,$$

voor $i=1,...,n$.
Uit de eerste ongelijkheid volgt dan

$$\frac{1}{\|\mathbf{x}\|_p} \ \frac{1}{\|\mathbf{x}+\mathbf{y}\|_p^{p/q}} \sum_{i=1}^{n} |x_i| \, |x_i+y_1|^{p-1} \leq$$

$$\leq \frac{1}{p}\ \frac{1}{\|x\|_p^p}\ \sum_{i=1}^{n} |x_i|^p + \frac{1}{q}\ \frac{1}{\|\mathbf{x}+\mathbf{y}\|_p^p}\ \sum_{i=1}^{n} |x_i + y_i|^{(p-1)q} = 1\ ,$$

dus

$$\sum_{i=1}^{n} |x_i|\,|x_i + y_i|^{p-1} \leq \|\mathbf{x}\|_p\,\|\mathbf{x}+\mathbf{y}\|_p^{p/q}\ .$$

Op dezelfde manier volgt uit de tweede ongelijkheid

$$\sum_{i=1}^{n} |y_i|\,|x_i + y_i|^{p-1} \leq \|\mathbf{y}\|_p\,\|\mathbf{x}+\mathbf{y}\|_p^{p/q}\ .$$

En dus volgt

$$\|\mathbf{x}+\mathbf{y}\|_p^p = \sum_{i=1}^{n} |x_i + y_i|^p \leq \sum_{i=1}^{n} |x_i|\,|x_i + y_i|^{p-1} + |y_i|\,|x_i + y_i|^{p-1} \leq$$

$$\leq \|\mathbf{x}\|_p\,\|\mathbf{x}+\mathbf{y}\|_p^{p/q} + \|\mathbf{y}\|_p\,\|\mathbf{x}+\mathbf{y}\|^{p/q} = (\|\mathbf{x}\|_p + \|\mathbf{y}\|_p)\,\|\mathbf{x}+\mathbf{y}\|^{p/q}\ ,$$

en omdat $\frac{p}{q} = p - 1$ volgt hieruit eigenschap 4.
Voor elke $p \geq 1$ wordt er op deze manier dus een norm op $\mathbb{R}^n$ gedefiniëerd, en dus ook een metriek $d_p(X,Y) = \|\mathbf{x} - \mathbf{y}\|_p$. Alleen voor $p = 2$ hangt deze metriek samen met een inproduct; in $\mathbb{R}^2$ is dat precies de situatie zoals die in paragraaf 2.1 is beschreven.

2.3 We bekijken nu speciaal het geval $p = 4$, en we beperken ons tot $\mathbb{R}^2$. De afstands-functie wordt dus gegeven door

$$d_4(X,Y) = \sqrt[4]{(x_1 - y_1)^4 + (x_2 - y_2)^4}\ ,$$

waarbij X en Y de punten zijn met coördinaten (x_1, x_2) en (y_1, y_2). Deze metriek is invariant onder translaties. Er zijn slechts weinig lineaire afbeeldingen die deze metriek behouden.

Propositie 4.
De enige lineaire afbeeldingen die de metriek d_4 invariant laten zijn de lineaire afbeeldingen met de volgende matrix-voorstellingen:

$$\pm\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \pm\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$\pm\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \pm\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

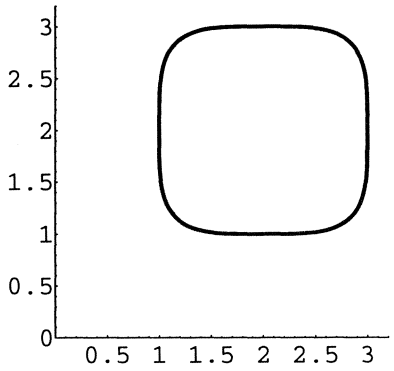
Een cirkel met straal r, en middelpunt M definiëren we op de voor de hand liggende manier als

$$\{X \mid d_4(X,M) = r\}\ .$$

Dit is dus de verzameling van de punten waarvan de coördinaten voldoen aan de vergelijking

$$(x_1 - m_1)^4 + (x_2 - m_2)^4 = r^4 \ .$$

B.v. de cirkel met middelpunt $(2, 2)$ en straal 1 ziet er als volgt uit:



Figuur 5:

Wanneer we een cirkel met middelpunt $(0, 0)$ beschouwen, dan kunnen we uit de vorm de in Propositie 4 genoemde lineaire afbeeldingen terugvinden: dit zijn de enige lineaire afbeeldingen die deze cirkel op zichzelf afbeelden.
Een in de meetkunde belangrijk begrip is "middelloodlijn" van een lijnstuk. Als A en B twee punten in het vlak zijn dan definiëren we de middelloodlijn van het lijnstuk AB als de verzameling van alle punten met gelijke afstanden tot A en B. De vorm van de middelloodlijn is afhankelijk van de ligging van het lijnstuk in het vlak. Elk lijnstuk kunnen we door een geschikte translatie overvoeren in een lijnstuk waarvan het midden precies in de oorsprong valt. Van een dergelijk lijnstuk, met eindpunten (a, b) en $(-a, -b)$, kunnen we gemakkelijk een vergelijking voor de middelloodlijn bepalen:

$$(x_1 - a)^4 + (x_2 - b)^4 = (x_1 + a)^4 + (x_2 + b)^4 \ ,$$

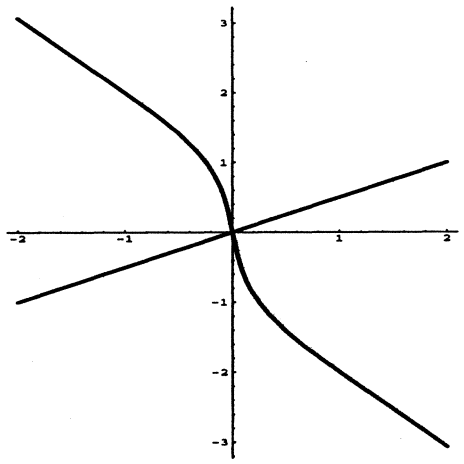
te vereenvoudigen tot

$$ax_1^3 + a^3x_1 + bx_2^3 + b^3x_2 = 0 \ .$$

Als nu $b = 0$ (lijnstuk langs de x_1-as), dan is de middelloodlijn de x_2-as; als $a = 0$ dan is de middelloodlijn de x_1-as. Er zijn nog enkele situaties waarin de middelloodlijn een rechte lijn is. In het algemeen is dat echter niet zo. Zie b.v.
Wel geldt:

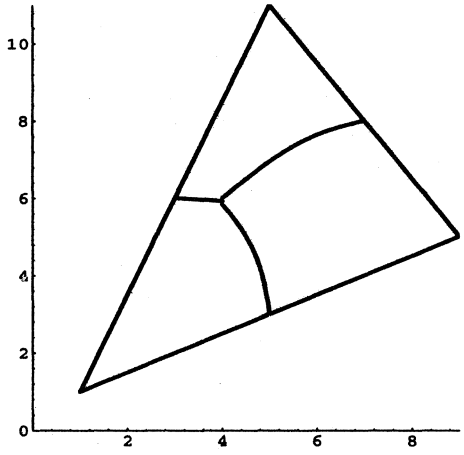
Propositie 5.
In een driehoek gaan de drie middelloodlijnen op de zijden door één punt.

Het "klassieke" bewijs blijft immers van kracht. Als de driehoek ABC is,



Figuur 6:

en de middelloodlijnen van AB en BC snijden elkaar in S, dan geldt dat $d_4(S,A) = d_4(S,B)$, en $d_4(S,B) = d_4(S,C)$, dus $d_4(S,A) = d_4(S,C)$. Maar dat betekent dat S ook op de middelloodlijn van AC ligt.



Figuur 7:

Een gevolg is dat elke driehoek een omgeschreven cirkel heeft. De straal van de omgeschreven cirkel hangt echter af van de ligging van de driehoek in het vlak.
Onder de afstand van een punt P tot een rechte lijn l verstaan we de minimale afstand van P tot een punt op l. Deze afstand is tevens gelijk aan de straal van een cirkel met P als middelpunt en zó dat deze cirkel aan l raakt. Als de lijn l gegeven wordt door de vergelijking

$$x_2 = ax_1 + b \ ,$$

dan is gemakkelijk na te gaan dat voor de afstand van de oorsprong tot l geldt

$$d_4(\mathrm{O}, l) = \frac{|b|}{(1 + a^{4/3})^{3/4}} \ .$$

En hieruit is onmiddellijk af te leiden dat voor willekeurige $X = (x_1, x_2)$ geldt

$$d_4(X, l) = \frac{|ax_1 - x_2 + b|}{(1 + a^{4/3})^{3/4}} \ .$$

Als l en m twee snijdende rechten zijn, dan verstaan we onder de bissectrice van de hoek, gevormd door l en m de verzameling van alle punten met gelijke afstanden tot l en m. Het is niet moeilijk een vergelijking voor deze bissectrice af te leiden. Stel dat l en m gegeven worden door resp. de vergelijkingen

$$x_2 = a_1x_1 + b_1 \quad \text{en} \quad x_2 = a_2x_1 + b_2 \ .$$

Dan bestaat de bissectrice dus uit alle punten (x_1, x_2) waarvoor geldt

$$\frac{|a_1x_1 - x_2 + b_1|}{(1 + a_1^{4/3})^{3/4}} = \frac{|a_2x_1 - x_2 + b_2|}{(1 + a_2^{4/3})^{3/4}} \ .$$

Door deze vergelijking worden twee rechte lijnen voorgesteld.
Analoog aan Propositie 5 geldt nu ook:

Propositie 6.
In een driehoek gaan de drie bissectrices door één punt.

Dit zijn uiteraard slechts enkele aspecten van de meetkunde die ontstaat als we een andere metriek dan de gewone Euclidische gebruiken.
We besluiten deze paragraaf met een aantal vragen :

1. Welke transformaties van het vlak zijn er die deze metriek invariant laten?
2. Kan met deze metriek een zinvolle hoekmaat gedefiniëerd worden?
3. Dezelfde vraag, nu voor de oppervlakte.

3. METRIEK OP EEN BOL

In de vorige paragraaf werden metrieken in het platte vlak behandeld. Wij leven echter op een bol, en in deze paragraaf besteden we kort aandacht aan het meten van afstanden tussen punten op een boloppervlak. We beschouwen daartoe een bol met straat R. Het boloppervlak kan opgevat worden als de verzameling van alle punten waarvan de coördinaten voldoen aan de vergelijking

$$x_1^2 + x_2^2 + x_3^2 = R^2 \ .$$

Voor een tweetal punten op de bol kunnen we de afstand definiëren als de gewone Euclidische afstand:

$$\sqrt{\sum_{i=1}^{3}(x_i - y_i)^2}\ .$$

Het onbevredigende daarbij is dat we dan de lengte van een lijnstsuk meten dat voor bewoners op het boloppervlak verborgen is. Het is wenselijk de afstand over het boloppervlak te meten. We voeren eerst een naam in:
onder een grote cirkel op het boloppervlak verstaan we een cirkel waarvan de straal (op de "gewone" manier gemeten) gelijk is aan R; dit is dus een cirkel op het boloppervlak met maximale omtrek, nl. $2\pi R$. Als X en Y twee punten op de bol zijn die niet diametraal tegenover elkaar liggen dan is er precies één grote cirkel waar X en Y op liggen. Door de punten X en Y wordt deze grote cirkel in twee cirkelbogen verdeeld. We definiëren nu als afstand $d(X,Y)$. de lengte van de kleinste van deze twee bogen. En voor de diametraal tegenover elkaar gelegen punten X en Y definiëren we $d(X,Y) = \pi R$.
Het is niet moeilijk deze afstand in de coördinaten van X en Y uit te drukken:

$d(X,Y)$ is het uniek bepaalde getal waarvoor geldt
$0 \leq d(X,Y) \leq \pi R$, en
$$\cos\left(\frac{d(X,Y)}{R}\right) = \frac{1}{R^2}\sum_{i=1}^{3} x_i y_i.$$

Per definitie is $d(X,Y) \geq 0$, en ook is gemakkelijk in te zien dat

$$d(X,Y) = 0 \quad \Leftrightarrow \quad X = Y.$$

We tonen nu aan dat deze afstands-functie ook aan de driehoeksongelijkheid voldoet. Laat gegeven zijn een drietal punten X, Y en Z op de bol. Noem A de matrix waarvan de kolommen gevormd worden door de coördinaten van resp. X, Y en Z, dus

$$A = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix}$$

Noem ter afkorting

$$\alpha = \frac{d(X,Y)}{R},\ \beta = \frac{d(X,Z)}{R} \text{ en } \gamma = \frac{d(Y,Z)}{R}.$$

Dan geldt dus $0 \leq \alpha, \beta, \gamma \leq \pi$.
Voor het matrix-produkt A^TA geldt dan:

$$A^TA = R^2 \begin{pmatrix} 1 & \cos\alpha & \cos\beta \\ \cos\alpha & 1 & \cos\gamma \\ \cos\beta & \cos\gamma & 1 \end{pmatrix}.$$

Bovendien: $\det(A^TA) = \det(A^T)\det(A) = (\det(A))^2 \geq 0$.

Dus $\begin{vmatrix} 1 & \cos\alpha & \cos\beta \\ \cos\alpha & 1 & \cos\gamma \\ \cos\beta & \cos\gamma & 1 \end{vmatrix} \geq 0.$

Uitwerken van deze determinant, en toepassen van enkele formules uit de goniometrie levert

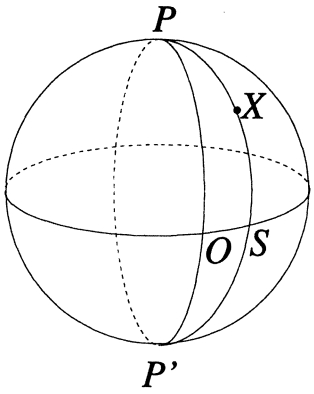
$4 \sin \frac{1}{2}(\alpha+\beta+\gamma) \sin \frac{1}{2}(\alpha+\beta-\gamma) \sin \frac{1}{2}(\alpha-\beta+\gamma) \sin \frac{1}{2}(-\alpha+\beta+\gamma) \geq 0.$

Nu geldt in ieder geval $0 \leq \alpha+\beta+\gamma \leq 3\pi$. Als $\alpha+\beta+\gamma > 2\pi$, dan is gemakkelijk in te zien dat $\frac{1}{2}(\alpha+\beta-\gamma)$, $\frac{1}{2}(\alpha-\beta+\gamma)$ en $\frac{1}{2}(-\alpha+\beta+\gamma)$ alle tussen 0 en π liggen. Maar dan geeft bovenstaand produkt van sinussen een negatief getal, en dat kan dus niet.
Dus geldt $0 \leq \alpha+\beta+\gamma \leq 2\pi$, en bijgevolg $\sin \frac{1}{2}(\alpha+\beta+\gamma) \geq 0$.
Nu kunnen de getallen $\alpha+\beta-\gamma, \alpha-\beta+\gamma$ en $-\alpha+\beta+\gamma$ niet alledrie negatief zijn, want dan zou ook de som negatief zijn. Ook is het niet mogelijk dat er twee negatief zijn, want door deze op te tellen zou dan volgen dat α, of β, of γ negatief is. Stel dat er één negatief is, b.v. $\alpha+\beta-\gamma < 0$. Omdat bovenstaand produkt van sinussen een niet-negatief getal moet opleveren moet in dat geval gelden $\alpha+\beta+\gamma = 2\pi$. Maar dan volgt dat $\gamma > \pi$, en dat kan ook niet.
Dus: $\alpha+\beta-\gamma \geq 0, \alpha-\beta+\gamma \geq 0$ en $-\alpha+\beta+\gamma \geq 0$. En hier staat in feite drie keer de driehoeksongelijkheid;

$$\text{b.v.} \qquad \alpha+\beta-\gamma \geq 0 \Leftrightarrow \frac{d(Y,Z)}{R} \leq \frac{d(X,Y)}{R} + \frac{d(Y,Z)}{R}$$

$$\Leftrightarrow d(Y,Z) \leq d(X,Y) + d(Y,Z).$$

De driehoeksongelijkheid had ook via een andere weg aangetoond kunnen worden. Er kan nl. bewezen worden dat de cirkelboog die we gebruikt hebben voor de definitie van $d(X,Y)$ de kortste weg over het boloppervlak tussen X en Y is, en daar is de driehoeksongelijkheid een gevolg van. Dit bewijs is echter ook niet eenvoudig. Een aardig, volledig meetkundig bewijs kan b.v. in [2,p.218 e.v.] gevonden worden.
De afstand wordt nu weliswaar gemeten over het boloppervlak, maar de uitdrukking voor de afstands-functie maakt gebruik van een coördinatenstelsel waarvan de oorsprong binnen het boloppervlak ligt. We brengen nu een coördinatenstelsel aan op het boloppervlak. We kiezen daarvoor een grote cirkel (met een richting) als x_1-as. Daarop wijzen we een vast punt O als oorsprong aan. Door de gekozen richting op de x_1-as kunnen we op de bol op voor de hand liggende manier (kurketrekkerregel bijvoorbeeld) een bovenste helft en een onderste helft onderscheiden. Noem P het punt op de bovenste helft zodanig dat de grote cirkel door P en O loodrecht op de x_1-as staat, en $d(P,O) = \frac{1}{2}\pi R$, en laat P' het diametraal tegenover P gelegen punt zijn (P en P' zijn dus a.h.w. de noord- en de zuidpool). Aan het punt X op de bol ($X \neq O, P, P'$) kennen we nu op de volgende manier coördinaten toe. Bepaal de grote cirkel door P en X. De cirkelboog PXP' snijdt de x_1-as in een punt S. De lengte van de cirkelboog OS, gemeten volgens de gekozen richting op de x_1-as, nemen we als x_1-coördinaat. En de lengte van SX, voorzien van een + teken als X op de bovenste helft ligt of een − teken als X op de onderste helft ligt nemen we als x_2-coördinaat.
Deze constructie werkt niet voor de punten O, P, P'; voor deze punten leggen we de coördinaten vast als resp. $(0,0), (0, \frac{1}{2}\pi R), (0, -\frac{1}{2}\pi R)$
De plaats van elk punt X op het boloppervlak wordt nu vastgelegd door twee coördianten (x_1, x_2) waarbij $0 \leq x_1 < 2\pi R$, $-\frac{1}{2}\pi R \leq x_2 \leq \frac{1}{2}\pi R$. In deze



Figuur 8:

coördinaten wordt de afstands-functie als volgt uitgedrukt:

$d(X,Y)$ is het uniek bepaalde getal waarvoor geldt
$0 \leq d(X,Y) \leq \pi R$, en

$$\cos\left(\frac{d(X,Y)}{R}\right) = \cos\left(\frac{x_1 - y_1}{R}\right)\cos\left(\frac{x_2}{R}\right)\cos\left(\frac{y_2}{R}\right) + \sin\left(\frac{x_2}{R}\right)\sin\left(\frac{y_2}{R}\right)$$

Deze uitdrukking lijkt dus in het geheel niet op een kwadratische vorm. Nu ziet het boloppervlak er lokaal min of meer plat uit. Neem twee punten op de bol die niet te ver van elkaar liggen:

$$X = (x_1, x_2), \quad Y = (x_1 + \triangle x_1, x_2 + \triangle x_2)$$

Dan geldt

$$\cos\left(\frac{d(X,Y)}{R}\right) = \cos\left(\frac{\triangle x_1}{R}\right)\cos\left(\frac{x_2}{R}\right)\cos\left(\frac{x_2 + \triangle x_2}{R}\right) + \sin\left(\frac{x_2}{R}\right)\sin\left(\frac{x_2 + \triangle x_2}{R}\right).$$

Door nu gebruik te maken van de tweede-orde benaderingen van sinus en cosinus vinden we na enig rekenwerk dat

$$d^2(X,Y) \approx \cos^2\left(\frac{x_2}{R}\right)(\triangle x_1)^2 + (\triangle x_2)^2.$$

Lokaal wordt de afstands-functie bij benadering dus wel door een kwadratische vorm gegeven.
In deze paragraaf is uitsluitend aandacht geschonken aan het vinden van een goede afstands-functie op de bol, en het bepalen van een formule hiervoor met

behulp van een coördinatenstelsel. We gaan verder niet in op de meetkunde die hier het gevolg van is. Zie daarvoor b.v. [3].

REFERENTIES

1. Asch, A.G. van, F. van der Blij, Hoeken en hun maat, CWI Syllabus 29, Amsterdam (1992).
2. Niven, I, Maxima and Minima without Calculus, The Mathematical Association of America (1981).
3. Schuh, F, Leerboek der Boldriehoeksmeting, G.B. van Goor Zonen's Uitgevers-Maatschappij, 's-Gravenhage (1940).

Roosters en Kwadratische Vormen

Peter Stevenhagen

SAMENVATTING
We bespreken de getaltheorie van definiete binaire kwadratische vormen aan de hand van de vermenigvuldiging van roosters in het complexe vlak. Tevens besteden we aandacht aan analytische klassengetalformules en het pas in de zestiger jaren opgeloste klassengetal-1-probleem.

1. INLEIDING

In de theorie van de kwadratische vormen spelen de vormen in twee variabelen, ook wel *binaire kwadratische vormen* genoemd, een bijzondere rol. In de eerste plaats zijn dit de 'oudste' en 'eenvoudigste' kwadratische vormen, maar wellicht belangrijker is het feit dat voor deze vormen door Gauss in de in 1801 verschenen *Disquisitiones Arithmetica* een arithmetische theorie ontwikkeld werd die diverse tot die tijd slechts gedeeltelijk begrepen numerieke eigenaardigheden van een theoretische verklaring voorzag. Deze theorie is de basis geworden voor wat men tegenwoordig *algebraïsche getaltheorie* noemt. Omdat de door Gauss en zijn opvolgers gelegde fundamenten door tijdgenoten als uiterst ingewikkeld ervaren werden heeft dit vakgebied tot ver in de twintigste eeuw een aureool van complexiteit en grote abstractie gehouden.
We beginnen ons verhaal met een opmerking die Diophantus van Alexandrië rond het jaar 250 maakte in zijn *Arithmetica*.

> Het getal 65 kan op natuurlijke wijze in twee kwadraten worden verdeeld, namelijk in $7^2 + 4^2$ en $8^2 + 1^2$; dit komt doordat 65 het produkt is van 13 en 5, die elk de som van twee kwadraten zijn. [Boek III, probleem 19]

Het schijnt dat Diophantus op de hoogte is van het feit dat produkten van sommen van twee kwadraten weer sommen van twee kwadraten zijn, en dat er een formule geldt als

$$(a^2 + b^2)(c^2 + d^2) = (ac \pm bd)^2 + (ad \mp bc)^2. \tag{1.1}$$

Immers, voor $65 = 13 \cdot 5 = (3^2+2^2)\cdot(2^2+1^2)$ geeft dit precies de opmerking van Diophantus. We zien dat we een getal n als een som van twee kwadraten kunnen schrijven indien we dit voor elk van de priemfactoren p van n kunnen. Als we dus willen weten welke gehele getallen sommen van twee kwadraten zijn, of, luxer gezegd, *gerepresenteerd worden* door de kwadratische vorm $X^2 + Y^2$, ligt het voor de hand te kijken welke priemgetallen p door $X^2 + Y^2$ gerepresenteerd worden. Deze vraag werd rond 1640 beantwoord door Fermat.

1.2. STELLING. *Voor een priemgetal p geldt*

$$p = x^2 + y^2 \text{ met } x, y \in \mathbb{Z} \Longleftrightarrow p = 2 \text{ of } p = 4k + 1.$$

Deze stelling is gemakkelijk 'empirisch' te verifiëren. Immers, omdat kwadraten altijd een viervoud of een viervoud+1 zijn is een priemgetal van de vorm $4k+3$ *nooit* een som van twee kwadraten; enig proberen laat zien dat dit voor priemen van de vorm $4k+1$ steeds *wél* het geval is. Een compleet bewijs is echter niet zo makkelijk te geven. Van Fermat is geen bewijs overgeleverd, en het eerste gepubliceerde bewijs voor deze stelling werd in 1749 door Euler in een brief aan Goldbach gegeven. Euler's bewijs is, geheel in de stijl van Fermat, gebaseerd op *descent-technieken*: uitgaande van een tegenvoorbeeld tegen de stelling wordt een nog kleiner tegenvoorbeeld geconstrueerd tot een tegenspraak volgt. De identiteit (1.1) speelt in dit bewijs een cruciale rol. Het bewijs laat zien dat ieder priemgetal p dat een *deler* is van een getal x^2+y^2 met $p \nmid xy$ zelf ook een som van twee kwadraten is, en men vindt een kwadraatvrije getal n van de vorm x^2+y^2 is dan en slechts dan als het geen priemfactoren van de vorm $4k+3$ bevat.
De vraagstelling laat zich generaliseren naar andere binaire kwadratische vormen, zoals X^2+nY^2 met $n \geq 1$. Omdat voor ieder geheel getal n de formule (1.1) vervangen kan worden door de soortgelijke formule

$$(a^2+nb^2)(c^2+nd^2) = (ac-bd)^2+n(ad+bc)^2 \tag{1.3}$$

ligt het weer voor de hand ligt eerst naar de representatie van priemgetallen te kijken. Analoga van stelling 1.2 voor de vormen X^2+2Y^2 en X^2+3Y^2 werden gevonden door Fermat, en de eerste bekende bewijzen komen ook hier van Euler. Het blijkt echter dat Eulers bewijzen voor hogere waarden van n op problemen stuiten. Reeds voor $n=5$ treedt er een merkwaardig probleem op. Door modulo 4 en 5, of in één keer modulo 20 te rekenen, bewijst men weer gemakkelijk de implicatie

$$p = x^2+5y^2 \text{ met } x,y \in \mathbb{Z} \implies p = 5 \text{ of } p \equiv 1,9 \bmod 20,$$

en numeriek overtuigt men zich er gemakkelijk van dat de omkering ook waar moet zijn. Euler's bewijs werkt echter hier niet, want anders dan in de eerdere gevallen kan het zo zijn dat een priemdeler $p|x^2+5y^2$ met $p \nmid xy$ zelf *niet* van de vorm x^2+5y^2 is. Een eenvoudig voorbeeld wordt gegeven door $21 = 4^2+5\cdot 1^2 = 3\cdot 7$. Deze problematische priemdelers te liggen in de restklassen 3 mod 20 en 7 mod 20, en Euler deed de verrassende ontdekking dat voor deze priemgetallen niet p zelf maar $2p$ van de vorm x^2+5y^2 is! Fermat had al eerder opgemerkt dat voor twee priemgetallen p en q die elk in de restklassen 3 mod 20 of 7 mod 20 liggen het produkt pq van de vorm x^2+5y^2 is. Fermat gaf toe dat hij geen bewijs had, en ook Euler slaagde er niet in zijn ontdekking van een sluitend bewijs te voorzien. Deze eer kwam toe aan Lagrange, die in 1773 de volgende identiteit vond:

$$(2p^2+2pq+3q^2)(2r^2+2rs+3s^2) = (2pr+qr+ps+3qs)^2+5(ps-qr)^2. \tag{1.4}$$

Lagrange liet zien dat de 'nieuwe' kwadratische vorm $2X^2+2XY+3Y^2$ naast $p=2$ alle priemgetallen in de restklassen 3 mod 20 en 7 mod 20 representeert. Hieruit volgt direct de opmerking van Fermat, en als we $p=1$ en $q=0$ nemen

krijgen we het vermoeden van Euler. Het is echter duidelijk dat deze 'vorm uit de hoge hoed' meer vragen oproept dan hij oplost.

2. Kwadratische vormen

Om de fenomenen uit de inleiding beter te begrijpen stellen we ons op een algemener standpunt, en vragen ons af welke priemgetallen gerepresenteerd worden door een kwadratische vorm $aX^2+bXY+cY^2$ met coëfficiënten $a, b, c \in \mathbb{Z}$. We geven zo'n vorm vaak kortweg aan met (a, b, c). We nemen hier en verder steeds aan dat de coëfficiënten a, b, c geen gemeenschappelijke factoren hebben; de vorm heet dan *primitief*.

Indien we alleen maar willen weten welke getallen een vorm representeert maakt het niet uit als we de vorm wijzigen door een transformatie

$$T : (X, Y) \mapsto (X + Y, Y) \qquad \text{of} \qquad S : (X, Y) \mapsto (Y, -X).$$

Men rekent gemakkelijk na dat S en T de coefficienten van een vorm op de volgende manier veranderen:

$$T : (a, b, c) \mapsto (a, b + 2a, c + b + a) \qquad \text{en} \qquad S : (a, b, c) \mapsto (c, -b, a).$$

Wat onder deze transformaties onveranderd blijft is de *discriminant* $D = b^2 - 4ac$ van een vorm. Deze belangrijke invariant is een viervoud voor vormen met even middencoefficient b en van de vorm $4k + 1$ voor vormen met oneven b. Discriminanten $D \equiv 2, 3 \bmod 4$ komen dus niet voor.

2.1. DEFINITIE *Twee kwadratische vormen heten equivalent als ze door herhaald toepassen van de transformaties S en T in elkaar kunnen worden overgevoerd.*

Omdat S en T de discriminant van een vorm niet veranderen hebben equivalente vormen dezelfde discriminant. We gaan nu, althans voor de zogenaamde *definiete* vormen die discriminant $D < 0$ hebben, kijken hoeveel equivalentieklassen van vormen er zijn bij vaste D. Hiertoe construeren we binnen elke equivalentieklasse van vormen een unieke *gereduceerde vorm*. Omdat er bij elke positief definiete vorm, zoals $F = X^2 + Y^2$, een bijbehorende negatief definiete vorm $-F$ is met dezelfde discriminant die dezelfde waarden met tegengesteld teken aanneemt, mogen we ons wel beperken tot het positief definiete geval. Dit betekent dat we vanaf nu alleen nog maar naar primitieve vormen (a, b, c) kijken met $a, c > 0$ en $b^2 - 4ac < 0$. Veel van wat volgt is in enigszins aangepaste vorm ook waar voor *indefiniete* vormen met discriminant $D > 0$, maar er treden enkele complicaties op die we willen vermijden.

2.2. DEFINITIE *Een positief definiete vorm (a, b, c) van discrimimant $D < 0$ heet gereduceerd als zijn coefficienten voldoen aan de ongelijkheden*

$$|b| \leq a \leq c,$$

waarbij in de grensgevallen $|b| = a$ en $a = c$ bovendien $b \geq 0$ geldt.

Door herhaald toepassen van S en T kan men een vorm gemakkelijk overvoeren in een equivalente gereduceerde vorm. We laten dit zien aan de hand van de vorm $(73, 54, 10)$ van discriminant $54^2 - 4 \cdot 73 \cdot 10 = -4$: $\stackrel{S}{\to}$

$$(73, 54, 10) \stackrel{S}{\to} (10, -54, 73) \stackrel{3T}{\to} (10, 6, 1) \stackrel{S}{\to} (1, -6, 10) \stackrel{3T}{\to} (1, 0, 1).$$

We zien dat de 'ingewikkelde' vorm $73X^2 + 54XY + 10Y^2$ exact dezelfde waarden representeert als de bekende vorm $X^2 + Y^2$.
Voor een gereduceerde vorm geldt $D = b^2 - 4ac \leq a^2 - 4a^2 = -3a^2$, en dus $a \leq \sqrt{-D/3}$. Omdat a positief en D negatief is zien we dat er bij *vaste* discriminant D maar eindig veel gereduceerde vormen bestaan: er zijn maar eindig veel mogelijkheden voor a, en bij elke a zijn er slechts eindig veel mogelijkheden voor b. Als a en b gekozen zijn ligt c vast door $c = (b^2 - D)/4a$. Dit bewijst het volgende.

2.3. STELLING. *Voor iedere discriminant D is het aantal equivalentieklassen van kwadratische vormen van discriminant D eindig.* □

Het aantal klassen in 2.3 heet het *klassengetal* $h(D)$ van de discriminant D.
Voor iedere discriminant D is er een unieke gereduceerde vorm met eerste coefficient $a = 1$, de *hoofdvorm* van discriminant D. Deze heeft coefficienten $(1, 0, -D/4)$ voor even D en $(1, 1, (1 - D)/4)$ voor oneven D.
Een gereduceerde vorm (a, b, c) van discriminant -4 heeft $a \leq \sqrt{4/3}$ en is dus gelijk aan de hoofdvorm $X^2 + Y^2$. Er volgt direct de gelijkheid $h(-4) = 1$. We laten zien hoe dit eenvoudige feit stelling 1.2 impliceert.

BEWIJS VAN STELLING 1.2. Zij p een priemgetal van de vorm $4k + 1$, en stel dat we een getal m kunnen vinden waarvoor $m^2 + 1$ deelbaar is door p, zeg $m^2 + 1 = rp$. We beschouwen dan de vorm $(p, 2m, r) = pX^2 + 2mXY + rY^2$ van discriminant $4m^2 - 4pr = -4$. Deze vorm representeert duidelijk p, en omdat we net zagen dat iedere vorm van discriminant -4 equivalent is met $X^2 + Y^2$ representeert de vorm $X^2 + Y^2$ ook p.
De opmerking dat voor ieder priemgetal $p \equiv 1 \bmod 4$ de congruentie $m^2 \equiv -1 \bmod p$ oplosbaar is is een speciaal geval van de kwadratische reciprociteitswet, en er zijn vele bewijzen van bekend. We laten dit als opgave aan de lezer
□

Doordat we in de definitie 2.2 van gereduceerdheid in de grensgevallen $b \geq 0$ eisen kan men zonder al te veel moeite bewijzen dat verschillende gereduceerde vormen niet onderling equivalent zijn, en dit geeft een eenvoudige methode om voor niet al te grote D het klassengetal te bepalen. Neemt men bijvoorbeeld $D = -47$, dan heeft iedere gereduceerde vorm coefficienten (a, b, c) met $|b| \leq a \leq \sqrt{47/3} < 4$. Er geldt nu $b \in \{\pm 1, \pm 3\}$ en we vinden $h(-47) = 5$ met representanten

$$(1, 1, 12) \qquad (2, \pm 1, 6) \qquad (3, \pm 1, 4).$$

Nog eenvoudiger is het geval $D = -20$ dat ten grondslag ligt aan de vermoedens van Fermat en Euler uit de vorige paragraaf. Hier vinden we $|b| \leq a \leq \sqrt{20/3} <$

3, en daarmee $h(-20) = 2$ met representanten

$$F_0 = X^2 + 5Y^2 \qquad \text{en} \qquad F_1 = 2X^2 + 2XY + 3Y^2.$$

Gelijkheid (1.3) en de miraculeuze identiteit (1.4) van Lagrange suggereren dat men een 'vermenigvuldiging' kan definiëren op de equivalentieklassen van kwadratische vormen met discriminant -20. Hiervoor geldt kennelijk $F_0 * F_0 = F_0$ en $F_1 * F_1 = F_0$. Moderne lezers zien al snel dat de verzameling $\{F_0, F_1\}$ zich kennelijk gedraagt als een *groep* van orde 2. Weet men eenmaal dat $F_0 * F_1 = F_1$ moet gelden, dan leidt gericht speurwerk tot een identiteit als

$$(p^2 + 5q^2)(2r^2 + 2rs + 3s^2) = 2X^2 + 2XY + 3Y^2$$

met $X = pr - qr - 3sq$ en $Y = ps + qs + 2qr$. Het is niet zo gemakkelijk om op deze manier een natuurlijke vermenigvuldiging te definiëren op bijvoorbeeld de 5 equivalentieklassen van vormen van discriminant -47. Het kan echter wel, en dergelijke 'compositieformules' werden aan het einde van de 18e eeuw gevonden door Legendre, vele decennia voor men abstracte begrippen als groepsaxioma's invoerde.
Legendre's definitie van equivalentie is iets verschillend van de onze, doordat hij de vormen (a, b, c) en $(a, -b, c)$, die immers dezelfde getallen representeren, als equivalent beschouwt. Dit blijkt tot gevolg te hebben dat de compositie van twee vormen niet meer eenduidig bepaald is, en er ontstaat geen groepsstructuur. Het is de grote verdienste van Gauss geweest dat hij er slaagde de 'juiste' definitie 2.1 van equivalentie te geven, en vervolgens op de verzameling $C(D)$ van equivalentieklassen van vormen van discriminant D een compositie te definiëren die $C(D)$ tot een abelse groep maakt. Gauss' formules zijn zeer gecompliceerd, en de verificatie van een 'standaardidentiteit' als $(F_1 * F_2) * F_3 = F_1 * (F_2 * F_3)$ leidt tot vele pagina's door weinigen begrepen rekenwerk. Veel latere wiskundigen gingen daarom op zoek naar 'natuurlijker' definities van de compositie. Wij geven in de volgende paragraaf een beschrijving gebaseerd op de vermenigvuldiging van roosters.

Opgaven.

1. Bereken de klassengetallen $h(-23)$, $h(-163)$ en $h(-164)$.
2. Zij p een priemgetal en x een geheel getal dat niet deelbaar is door p.
   a. Laat zien dat er een geheel getal y bestaat met $xy \equiv 1 \bmod p$, en dat $y \bmod p$ uniek bepaald is. We zeggen dat $y \bmod p$ de *multiplicatieve inverse* van x modulo p is.
   b. Bewijs: $x^2 \equiv 1 \bmod p \Longrightarrow x \equiv \pm 1 \bmod p$;
   c. Bewijs de *stelling van Wilson*: $(p-1)! \equiv -1 \bmod p$.[Hint: combineer ieder getal met zijn multiplicatieve inverse.]
   d. Neem nu aan dat $p \equiv 1 \bmod 4$ geldt. Bewijs: voor $u = (\frac{p-1}{2})!$ geldt $u^2 \equiv -1 \bmod p$.
3. Zij $p = 4k + 1$ een priemgetal. Laat zien dat het polynoom $X^{4k} - 1$ precies $4k$ verschillende nulpunten modulo p heeft. Concludeer dat de congruentie $x^2 \equiv -1 \bmod p$ precies $2k = \frac{p-1}{2}$ oplossingen heeft modulo p.

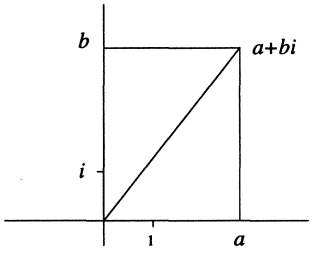
4. De door Fermat gevonden analoga van stelling 2.1 voor de vormen $X^2 + 2Y^2$ en $X^2 + 3Y^2$ zijn:

$$p = x^2 + 2y^2 \text{ met } x, y \in \mathbb{Z} \Longleftrightarrow p = 2 \text{ of } p \equiv 1, 3 \bmod 8;$$
$$p = x^2 + 3y^2 \text{ met } x, y \in \mathbb{Z} \Longleftrightarrow p = 3 \text{ of } p \equiv 1 \bmod 3.$$

Laat eerst zien dat het rechterlid betekent dat de respectievelijk de congruenties $x^2 \equiv -2 \bmod p$ en $x^2 \equiv -3 \bmod p$ een oplossing bezitten, en bewijs vervolgens middels een berekening van $h(-8)$ en $h(-3)$ de equivalentie met het linkerlid.

3. Vermenigvuldiging van roosters

Veel van de formules in het voorafgaande worden duidelijk wanneer men ze beschouwt als identiteiten tussen *complexe getallen*. Zoals bekend zijn dit getallen van de vorm $z = a + bi$ met $a, b \in \mathbb{R}$ reëel en $i^2 = -1$. We zeggen dat $a = \mathrm{Re}(z)$ het *reële deel* van z is en $b = \mathrm{Im}(z)$ het *imaginaire deel*. Neemt men a en b geheel, dan heet $a + bi$ een *geheel getal van Gauss*. Complexe getallen hebben een *norm* gegeven door $N(a+bi) = (a+bi)(a-bi) = a^2 + b^2$. Tekenen we de complexe getallen $a+bi$ op de bekende manier als roosterpunten (a, b) in het 'complexe vlak' $\mathbb{C} \approx \mathbb{R}^2$, dan is de norm wegens Pythagoras het kwadraat van de afstand $|a + bi|$ van (a, b) tot de oorsprong.



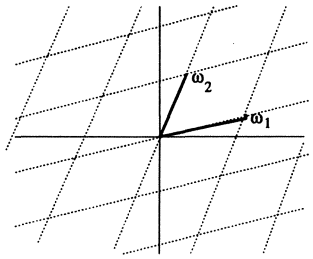
De normfunctie is duidelijk *multiplicatief*, hetgeen betekent dat voor $a, b, c, d \in \mathbb{R}$ de gelijkheid

$$N((a + bi) \cdot (c + di)) = N(a + bi) \cdot N(c + di)$$

geldt. Nu is echter $(a+bi)\cdot(c+di) = (ac-bd)+(ad+bc)i$, en deze identiteit is daarmee 'dezelfde' als de vertrouwde identiteit (1.1). Deze exercitie geeft enig vertrouwen in het nut van complexe getallen voor onze kwadratische vormen. We gaan nog verder gebruik maken van complexe getallen door te kijken naar roosters in het complexe vlak. Een *rooster* in $\mathbb{C}$ is een verzameling van de vorm

$$L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \{k_1\omega_1 + k_2\omega_2 : k_1, k_2 \in \mathbb{Z}\} \subset \mathbb{C}$$

voor twee elementen $\omega_1, \omega_2 \in \mathbb{C}$ die lineair onafhankelijk zijn over $\mathbb{R}$. Zoals onderstaand plaatje laat zien vormen de elementen van L inderdaad een roosterpatroon in $\mathbb{C}$.



We schrijven $L = [\omega_1, \omega_2]$ en zeggen dat $\{\omega_1, \omega_2\}$ een *basis* is voor het rooster L. Als $\lambda \in \mathbb{C}^*$ een complex getal is, dan geven we met λL het rooster $[\lambda\omega_1, \lambda_2\omega_2]$ aan. Merk op dat de basis $\{\lambda\omega_1, \lambda_2\omega_2\}$ ontstaat uit $\{\omega_1, \omega_2\}$ door een rotatie gevolgd door een vermenigvuldiging met $|\lambda|$. De roosters L en λL heten *gelijkvormig* of kortweg *equivalent*.
Door in het bovenstaande $\lambda = \omega_1^{-1}$ te kiezen zien we dat ieder rooster equivalent is met een rooster van de vorm $[1, \omega]$. Door eventueel ω door $-\omega$ te vervangen neemt men steeds ω met positief imaginair deel, d.w.z. in het 'complexe bovenhalfvlak'. Er zijn dan natuurlijke transformaties

$$T : [1, \omega] \mapsto [1, \omega - 1] \qquad \text{en} \qquad S : [1, \omega] \mapsto [1, -1/\omega]$$

die het rooster respectievelijk onveranderd laten en, met het oog op de identiteit $[1, -1/\omega] = 1/\omega \cdot [\omega, 1]$, in een equivalent rooster overvoeren. Men heeft in feite het volgende.

3.1. LEMMA. *De roosters $[1, \omega_1]$ en $[1, \omega_2]$ zijn dan en slechts dan equivalent als ze door herhaald toepassen van S en T in elkaar kunnen worden overgevoerd.*

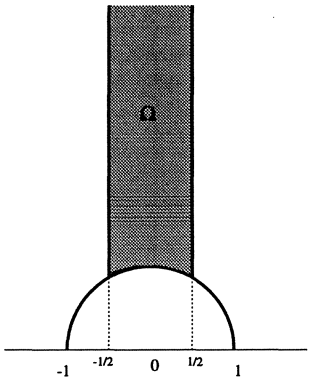
Het bewijs van 3.1, dat niet heel diep is, laten we hier weg. De oplettende lezer zal de analogie tussen 3.1 en definitie 2.1 niet ontgaan zijn. Er is ook een analogon van 2.2.

3.2. DEFINITIE. *Een basis $\{1, \omega\}$ van een rooster heet gereduceerd als ω een getal in het bovenhalfvlak is dat voldoet aan*

$$\mathrm{Re}(z) \leq 1/2 \qquad \textit{en} \qquad |z| \geq 1,$$

waarbij in de grensgevallen $\mathrm{Re}(z) = 1/2$ en $|z| = 1$ bovendien $\mathrm{Re}(z) \leq 0$ geldt.

Het gereduceerd zijn van de basis $\{1, \omega\}$ betekent dat ω in het gearceerde deel Ω van het complexe bovenhalfvlak ligt. Alleen het in het linkerhalfvlak $\{\mathrm{Re}(z) \leq 0\}$ gelegen deel van de rand van Ω behoort tot Ω. In een rooster met gereduceerde basis is de minimumafstand tussen de roosterpunten gelijk aan 1.



Men kan weer nagaan dat definitie 3.2 zo ingericht is dat er voor ieder rooster L een *uniek* element $\omega \in \Omega$ is waarvoor L en $[1,\omega]$ equivalent zijn. Met behulp van de transformaties S en T kan men iedere basis $\{1,\omega\}$ vervangen door een gereduceerde basis. Hierbij gebruikt men de gelijkheid $-1/(a+bi) = (-a+bi)/(a^2+b^2)$. Nemen we bijvoorbeeld $\omega = (-27+i)/146$, dan krijgen we

$$(-27+i)/146 \overset{S}{\to} (27+i)/10 \overset{3T}{\to} (-3+i)/10 \overset{S}{\to} 3+i \overset{3T}{\to} i.$$

Het rooster $[1,(-27+i)/146]$ is dus equivalent met met rooster $[1,i] = \mathbb{Z}[i]$ van de gehele getallen van Gauss. De lezer zal begrijpen dat de overeenkomst tussen de reductie van dit rooster en de reductie van de kwadratische vorm $73X^2 + 54XY + 10Y^2$ geen toeval is!
Alvorens deze analogie preciezer aan te geven gaan we proberen roosters te *vermenigvuldigen*, en kijken of het resultaat weer een rooster is.

3.3. DEFINITIE *Het produkt van twee roosters L en L' is gedefinieerd als*

$$L \cdot L' = \{\textstyle\sum_{i=1}^{n} \ell_i \ell_i' : \quad \ell_i \in L, \ell_i' \in L', n \in \mathbb{Z}_{\geq 1}\}.$$

Dit is de kleinste optelgroep in $\mathbb{C}$ die alle produkten $\ell\ell'$ met $\ell \in L$ en $\ell' \in L'$ bevat. In het algemeen zal $L \cdot L'$ geen rooster zijn. We zullen echter een voorwaarde aan onze roosters opleggen waaronder dit wel het geval is.

3.4. DEFINITIE *De multiplicatorring R_L van een rooster L is gedefinieerd door*

$$R_L = \{\lambda \in \mathbb{C} : \lambda L \subset L\}.$$

We zeggen dat R_L een *ring* is omdat met $\lambda_1, \lambda_2 \in R_L$ ook de som $\lambda_1 + \lambda_2$ en het produkt $\lambda_1 \cdot \lambda_2$ in R_L bevat zijn. Uit de definitie van een rooster is direct duidelijk dat R_L voor alle L de ring $\mathbb{Z}$ van gewone gehele getallen bevat, en zelfs dat $R_L \cap bfR = \mathbb{Z}$ geldt. Het is eveneens duidelijk dat L en een met L equivalent rooster λL dezelfde multiplicatorring hebben, zodat we voor de bepaling van de multiplicatorring R_L van L wel aan mogen nemen dat L van de

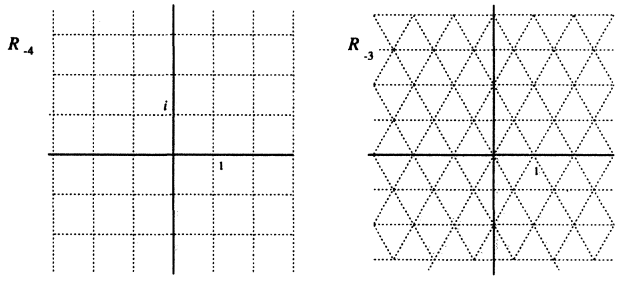
vorm $L = [1, \omega]$ is. In dat geval geldt voor $\lambda \in R_L$ de gelijkheid $\lambda \cdot 1 = \lambda \in L$, dus we vinden inclusies

$$\mathbb{Z} \subset R_L \subset L.$$

Er zijn nu twee mogelijkheden. In het geval dat R_L alleen reële getallen bevat hebben we $R_L = \mathbb{Z}$. Als er een (niet-reëel) complex getal is dat bevat is in R_L zeggen we dat L *complexe vermenigvuldiging* heeft. In dat geval volgt uit bovenstaande inclusies dat R_L een *deelrooster* van L is. Men kan $1 \in R_L$ als eerste basisvector nemen, zodat we $R_L = \mathbb{Z} + \mathbb{Z} \cdot \alpha$ krijgen. Nu is R_L de *ring* $\mathbb{Z}[\alpha]$, en er geldt $\alpha^2 \in R_L = [1, \alpha]$, dus α is een nulpunt van een kwadratische vergelijking van de vorm $\alpha^2 + m\alpha + n = 0$ voor $m, n \in \mathbb{Z}$. Geven we met $D = m^2 - 4n$ de discriminant van deze vergelijking aan, dan is D een negatief geheel getal dat een viervoud (voor m even) of een viervoud+1 (voor m oneven) is, en we hebben

$$R_L = \mathbb{Z}[\alpha] = \mathbb{Z}[\tfrac{m+\sqrt{D}}{2}] = \mathbb{Z}[\tfrac{D+\sqrt{D}}{2}] = \mathbb{Z} + \mathbb{Z} \cdot \tfrac{D+\sqrt{D}}{2}.$$

De ring R_L, die kennelijk alleen van de discriminant D afhangt, heet de *kwadratische orde van discriminant* D. Voor ieder negatief geheel getal $D \equiv 0, 1 \bmod 4$ hebben we zo'n ring, die we verder aan zullen geven met R_D. Merk op dat $R_{-4} = \mathbb{Z}[i]$ de ring van gehele getallen van Gauss is. De lezer kan nagaan dat de ring R_{-3} van *gehele getallen van Eisenstein* een rooster vormt dat het complexe vlak niet in vierkantjes maar in gelijkzijdige driehoekjes verdeelt. Algemener leiden de discriminanten $D \equiv 0 \bmod 4$ tot rechthoekige roosters, terwijl de discriminanten $D \equiv 1 \bmod 4$ tot driehoekige roosters aanleiding geven.



3.5. STELLING. *Laat L en L' roosters zijn die complexe vermenigvuldiging hebben met dezelfde kwadratische orde R_D. Dan is $L \cdot L'$ weer een rooster met multiplicatorring R_D.*

Het bewijs van deze stelling is niet heel moeilijk, en er geldt zelfs een soort omkering van de stelling. Door L en L' zo nodig door een equivalent rooster te vervangen herleid men tot het geval dat L en L' in R_D bevat zijn, en dan impliceert de geslotenheid van R_D onder vermenigvuldiging en optelling dat

$L \cdot L'$ bevat is in het rooster R_D. Een iets fijnere analyse laat zien dat de multiplicatorring van $L \cdot L'$, die duidelijk R_D bevat, in feite gelijk is aan R_D. We komen nu tot de hoofdstelling van deze paragraaf.

3.6. STELLING. *Het rooster* $[1,\omega]$ *heeft multiplicatorring* R_D *dan en slechts dan als er een primitieve kwadratische vorm* (a,b,c) *van discriminant* D *bestaat met* $\omega = (-b+\sqrt{D})/2a$. *De afbeelding*

$$(a,b,c) \longmapsto \left[1, \frac{-b+\sqrt{b^2-4ac}}{2a}\right]$$

beeldt gereduceerde vormen naar gereduceerde bases af en induceert een bijectie tussen de equivalentieklassen van de kwadratische vormen van discriminant D *en de equivalentieklassen van roosters met multiplicatorring* R_D.

BEWIJSSCHETS. Uit de inclusie $R_D = [1, \frac{D+\sqrt{D}}{2}] \subset L = [1,\omega]$ leidt men af dat er een positief getal a bestaat met $aL \subset R_D$. Neem a minimaal, dan geldt

$$\omega = \frac{-b+\sqrt{D}}{2a} \in \left[\frac{1}{a}, \frac{D+\sqrt{D}}{2a}\right]$$

voor een geheel getal b dat dezelfde pariteit heeft als D. Uit de relatie $\frac{D+\sqrt{D}}{2} \cdot \omega \in L$ krijgen we via een expliciete berekening

$$\frac{D+\sqrt{D}}{2} \cdot \omega = \frac{(1-b)D+(D-b)\sqrt{D}}{4a} = \frac{D-b^2}{4a} + \frac{D-b}{2}\omega + \in L.$$

Dit laat zien dat $c = (D-b^2)/4a$ een geheel getal is, en (a,b,c) een kwadratische vorm van discriminant D. Omgekeerd laat bovenstaande identiteit zien dat we voor een vorm (a,b,c) van discriminant D een rooster $[1, \frac{-b+\sqrt{D}}{2a}]$ krijgen met multiplicatorring R_D. We merken verder op dat de transformaties S en T op vormen en de transformaties S en T op roosters $[1,\omega]$ onder deze correspondentie van vormen en roosters overeenstemmen, zodat we een correspondentie tussen equivalentieklassen van vormen en roosters krijgen.
Het gereduceerd zijn van de basis $[1, \frac{-b+\sqrt{D}}{2a}]$ in de zin van 3.2 betekent dat er ongelijkheden

$$|\tfrac{-b}{2a}| \leq \tfrac{1}{2} \qquad \text{en} \qquad |\tfrac{-b+\sqrt{D}}{2a}| = |\tfrac{b^2-D}{4a^2}| = |\tfrac{c}{a}| \geq 1$$

gelden. Dit zijn precies de ongelijkheden uit definitie 2.2, en ook de ongelijkheden in de grensgevallen corresponderen. We krijgen hiermee een correspondentie van gereduceerde vormen en gereduceerde bases. □

Wegens 3.5 hebben we op de roosters met gegeven multiplicatorring R_D een natuurlijke vermenigvuldiging. Het is direct duidelijk dat deze vermenigvuldiging tot een vermenigvuldiging op de equivalentieklassen van roosters aanleiding geeft, daar $\lambda L \cdot L' = L \cdot \lambda L' = \lambda(L \cdot L')$ geldt. Stelling 3.6 laat zien hoe deze vermenigvuldiging zich vertaalt in een vermenigvuldiging van de elementen van $C(D)$, de equivalentieklassen van vormen van discriminant D. Dit

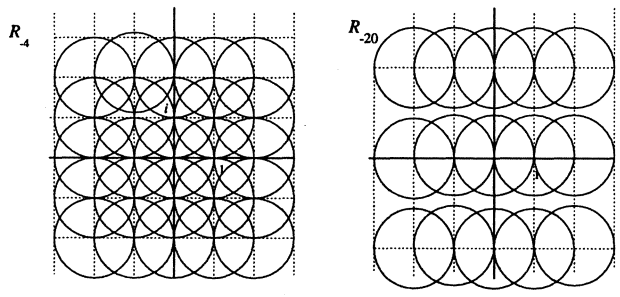
is de door Gauss gedefinieerde compositie. In termen van onze roosters is het niet moeilijk om in te zien dat $C(D)$ hiermee een eindige abelse groep wordt onder de vermenigvuldiging. De ingewikkelde compositieformules van Gauss voor vormen (a,b,c) en (a',b',c') van discriminant D kan men nu zelf als opgave proberen af te leiden door een basis te vinden van het corresponderende produktrooster opgespannen door de *vier* elementen

$$1, \qquad \tfrac{-b+\sqrt{D}}{2a}, \qquad \tfrac{-b'+\sqrt{D}}{2a'} \quad \text{en} \quad \tfrac{-bb'+D-(b+b')\sqrt{D}}{4aa'}.$$

Voor het produkt van de kwadratische vorm $2X^2+2XY+3Y^2$ met zichzelf krijgen we bijvoorbeeld het rooster L met voortbrengers

$$1, \qquad \tfrac{-1+\sqrt{-5}}{2} \quad \text{en} \quad \left(\tfrac{-1+\sqrt{-5}}{2}\right)^2 = \tfrac{-2+\sqrt{-5}}{2}.$$

Door $\{1/2, \sqrt{-5}/2\}$ als basis voor dit rooster te nemen zien we dat $L = \frac{1}{2} \cdot [1, \sqrt{-5}]$ correspondeert met de hoofdvorm X^2+5Y^2.
Men kan voor kleine discriminanten D nu ook de klassegetallen 'visualiseren'. Stel bijvoorbeeld dat $\{1,\omega\}$ een gereduceerde basis is van een rooster L dat complexe vermenigvuldiging toelaat met $R_{-4} = \mathbb{Z}[i]$. Dan is L een rooster dat het rooster $\mathbb{Z}[i]$ bevat, en omdat de basis voor L gereduceerd is bevat L geen roosterpunten op afstand < 1 van de oorsprong. Dit impliceert onmiddelijk dat L *gelijk* is aan $\mathbb{Z}[i]$, en dat $h(-4) = 1$ geldt. Immers, stel dat $\ell \in L$ en roosterpunt is dat buiten $\mathbb{Z}[i]$ ligt. Dan laat een plaatje direct zien dat er een punt $\alpha \in \mathbb{Z}[i]$ is waarvoor $|\ell - \alpha| < 1$ geldt, en omdat $\ell - \alpha$ in L bevat is geeft dit een tegenspraak.
Neemt men $D = -20$, dan is er behalve het 'hoofdrooster' $L = R_{-20} = \mathbb{Z}[\sqrt{-5}]$ een tweede rooster $L' \supset L$ met gereduceerde basis $\{1, \frac{1+\sqrt{-5}}{2}\}$. De 'extra punten' komen precies in de ruimte terecht die overblijft buiten de 'verboden schijfjes' met straal 1 om de roosterpunten van L.



Opgaven.

1. Interpreteer de identiteit 1.3 in termen van de norm van complexe getallen $a+b\sqrt{-n} = a+bi\sqrt{n}$.
2. Interpreteer de identiteit 1.4 in termen van de norm van complexe getallen $a+b\sqrt{-5}$.

3. Ga na dat de kwadratische orde van discriminant D gelijk is aan

$$R_D = \begin{cases} \mathbb{Z}[\sqrt{d}] & \text{if } D = 4d \equiv 0 \bmod 4; \\ \mathbb{Z}[\frac{1+\sqrt{D}}{2}] & \text{if } D \equiv 1 \bmod 4. \end{cases}$$

4. Laat zien dat $h(-12) = 2$ geldt, en teken R_{-12} en de gereduceerde bases van de corresponderende roosters in het vlak.
5. Bepaal de groepsstructuur van $C(-120)$ en $C(-68)$.
6. Stel dat het produkt $L \cdot L'$ van de roosters L en L' weer een rooster is. Bewijs dat L en L' roosters met complexe vermenigvuldiging zijn, en dat de discriminanten D en D' van de corresponderende kwadratische ordes een quotiënt hebben dat een *kwadraat* is in $\mathbb{Q}$.
7. Laat zien dat de *inverse* van een roosterklasse $[L] \in C(D)$ de klasse is van het complex toegevoegde rooster $[\overline{L}]$. Hoe vertaalt zich dit in termen van vormen?

4. KLASSENGETALFORMULES

De berekeningen van het klassengetal $h(D)$ die we in paragraaf 2 gaven verloopt via een rechtsstreeks aftellen van vormen en laat niet gemakkelijk enige regelmaat zien in de berekening van klassengetallen. De Duitse wiskundige Dirichlet liet echter in de veertiger jaren van de vorige eeuw zien dat men door gebruik te maken van complexe analyse formules voor $h(D)$ af kan leiden die de berekening van $h(D)$ tot een nog eenvoudiger telpartij reduceren. Het belangrijkste ingrediënt in zijn formules is het *kwadratische karakter* behorende bij de discriminant D. Omwille van de eenvoud zullen we in deze paragraaf veronderstellen dat D een *fundamentele* discriminant is. Dit betekent dat het getal D kwadraatvrij is, of, in het geval D even is, dat $D/4$ kwadraatvrij is. De bijbehorende ring R_D heet in dit geval een *maximale orde.*
Voor een willekeurig geheel getal $D \equiv 0, 1 \bmod 4$ definiëren we een bijbehorend *karakter* $\chi = \chi_D : \mathbb{Z}_{>0} \to 0, \pm 1$. De functie χ is *multiplicatief*, hetgeen betekent dat voor $m, n \in \mathbb{Z}_{>0}$ steeds $\chi(mn) = \chi(m)\chi(n)$ geldt. Er zal dus $\chi(1) = 1$ gelden, en χ wordt vastgelegd door zijn waarden op de priemgetallen. Deze zijn gedefinieerd als

$$\chi(p) = \begin{cases} 0 & \text{als } p \text{ een deler is van } D; \\ 1 & \text{als de congruentie } D \equiv x^2 \bmod p \text{ een oplossing} \\ & x \equiv 0 \bmod p \text{ heeft}; \\ -1 & \text{als de congruentie} D \equiv x^2 \bmod p \text{ geen oplossing heeft.} \end{cases}$$

Voor het priemgetal $p = 2$ is de definitie iets anders. Voor even D nemen we weer $\chi(2) = 0$, maar voor oneven D is $\chi(2) = 1$ als $D \equiv 1 \bmod 8$ en $\chi(2) = -1$ als $D \equiv 5 \bmod 8$. In plaats van $\chi(n)$ komt men ook vaak de notatie $\left(\frac{D}{n}\right)$ tegen: het *Jacobi-symbool*. Er is de beroemde door Euler gevonden en door Gauss bewezen *kwadratische reciprociteitswet*, die zegt dat het karakter χ een *periodieke* functie is: $\chi(n) = \chi(n + D)$. Voor $D = -4$ hebben we

$$\chi_{-4}(n) = \left(\frac{-4}{n}\right) = \begin{cases} 0 & \text{voor even } n; \\ 1 & \text{voor } n \equiv 1 \bmod 4; \\ -1 & \text{voor } n \equiv -1 \bmod 4; \end{cases}$$

In het geval dat n een priemgetal is kwamen we dit reeds tegen in het bewijs van 1.2 in paragraaf 2.
We geven nu zonder bewijs de *analytische klassengetalformule* van Dirichlet. Hierin geeft w_D het aantal roosterpunten van R_D aan dat zich op de eenheidscirkel in het complexe vlak bevindt. Er zijn twee bijzondere waarden $w_{-3} = 6$ en $w_{-4} = 4$, zoals men aan de hand van een plaatje direct nagaat. Voor *alle* andere waarden van D geldt $w_D = 2$.

4.1. STELLING. *Zij χ het karakter behorende bij de fundamentele discriminant D. Dan geldt*

$$h(D) = \frac{w_D\sqrt{|D|}}{2\pi}\sum_{n=1}^{\infty}\frac{\chi(n)}{n}.$$

Dit is een wonderlijke formule, die het gehele getal $h(D)$ beschrijft als het produkt van een eenvoudige oneindige som en een onwaarschijnlijke niet-rationale factor. Bij wijze van voorbeeld kunnen we $D = -4$ nemen. Wie de machtreeks voor de arctangens-functie in 1 kan evalueren vindt inderdaad

$$h(-4) = \frac{4\cdot 2}{2\pi}\left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \ldots\right) = \frac{4}{\pi}\cdot\frac{\pi}{4} = 1,$$

wie dat niet kan vindt uit $h(-4) = 1$ een fraaie oneindige somuitdrukking voor π.
Men kan zich natuurlijk afvragen of er geen eenvoudige gesloten formule bestaat voor $h(D)$ die geen oneindige sommen bevat. Deze formule bestaat inderdaad, en wel in twee iets verschillende in twee vormen.

4.2. STELLING. *Zij χ het karakter behorende bij de fundamentele discriminant $D < -4$. Dan geldt*

$$h(D) = \frac{1}{D}\sum_{n=1}^{|D|} n\chi(n) = \frac{1}{(2-\chi(2)}\sum_{n=1}^{[|D|/2]}\chi(n).$$

Deze formules hebben de eigenschap dat ze voor elke D de waarde $h(D)$ als een *eindige* som beschrijven. Nemen we bijvoorbeeld de discriminant $D = -11$, dan geeft de eerste vorm van de formule

$$h(-11) = \frac{1}{-11}(1 - 2 + 3 + 4 + 5 - 6 - 7 - 8 + 9 - 10) = \frac{1}{-11}(22 - 33) = 1,$$

terwijl de tweede vorm iets efficienter het resultaat

$$h(-11) = \frac{1}{2+1}(1 - 1 + 1 + 1 + 1) = 1$$

geeft. Merk op dat het uit geen van beide vormen van de formule duidelijk is dat $h(D)$ een *positief* getal is. Kennelijk is het zo dat χ vaker de waarde $+1$ aanneemt op de eerste helft van het rijtje gehele getallen $1, 2, 3, \ldots, |D|$ dan op de tweede helft. Zelfs in het eenvoudigste geval dat χ het kwadratische karakter bij een priemgetal $p = -D \equiv 3 \bmod 4$ is is er geen direct bewijs voor deze 'scheve verdeling' van de kwadraten en de niet-kwadraten modulo p.
Wat ook uit geen van de klassengetalformules duidelijk is, is het 'gemiddelde gedrag' van $h(D)$ voor $D \to -\infty$. Hiernaar zullen we in het laatste deel van dit artikel kijken.

5. Het klassengetal-1-probleem

Indien men tabellen bekijkt van klassengetallen van fundamentele discriminanten, dan merkt men op dat de klassengetallen zich weliswaar locaal nogal springerig gedragen, maar dat er een duidelijk tendens is voor $h(D)$ om te groeien met $|D|$. Neemt men de gemiddelde waarde van $h(D)$ voor de discriminanten in een wat langer interval $[N, N+s]$, dan ziet men dat deze zich gedraagt als een constante maal $\sqrt{N}$. Dit werd al geobserveerd door Gauss, die tevens in artikel 303 van de *Disquisitiones* voor een aantal kleine waarden van h een rij van discriminanten D geeft met klassengetal h. Gauss merkt op dat hij de klassengetallen van heel veel discriminanten uitgerekend heeft, waaronder alle negatieve waarden $D > -3000$, en dat het zeer waarschijnlijk is dat er geen grotere discriminanten bestaan met klassengetal h. Dit doet vermoeden dat er voor iedere vaste $h \geq 1$ de rij van discriminanten met klassengetal h eindig is:

> nullum dubium esse videtur, quin series adscriptae revera abruptae sint, et per analogiam conclusionem eandem ad quasvis alias qualificationes extendere licebit. (...) Demonstrationes autem *rigorosae* harum observationum perdifficilis esse videntur.
> [er schijnt geen twijfel aan te zijn dat de gegeven rijen daadwerkelijk afbreken, en naar analogie kan men deze conclusie tot willekeurige andere qualificaties (=h-waarden) uitbreiden.] Een *rigoureus* bewijs van deze observaties schijnt bijzonder moeilijk te zijn.]

Laten we eens kijken naar het speciale geval van discriminanten D met $h(D) = 1$. In dit geval ziet het lijstje van Gauss er als volgt uit.

5.1. Observatie *Er geldt* $h(D) = 1$ *voor de discriminanten*

$$D \in \{-3, -4, -7, -8, -11, -16, -19, -27, -28, -31, -43, -67, -163\}.$$

Alle andere $D > -10000$ *hebben* $h(D) > 1$.

Met name de grotere discriminanten in dit lijstje hebben vermakelijke eigenschappen. Men kan, uitgaande van de geslachtstheorie van Gauss, gemakkelijk laten zien dat een fundamentele discriminant $D < -8$ met $h(D) = 1$ van de vorm $D = -p$ is voor een priemgetal $p \equiv 3 \bmod 4$. De hoofdvorm van discriminant $-p$ is dan $X^2 + XY + \frac{p+1}{4}Y^2$, en we kijken nu naar het geassocieerde

polynoom

$$f_p(x) = x^2 + x + \tfrac{p+1}{4} = (x + \tfrac{1}{2})^2 + \tfrac{p}{4}.$$

Dit polynoom is symmetrisch om $x = -1/2$, en als we de waarden tabelleren voor $x = 0, 1, 2, 3, 4, \ldots$ doen we verrassende ontdekkingen. De eerste 10 waarden van $f_{43}(x) = x^2 + x + 11$ zijn achtereenvolgens

$$11,\ 13,\ 17,\ 23,\ 31,\ 41,\ 53,\ 67,\ 83,\ 101.$$

Door een wonderlijk toeval zijn dit allemaal *priemgetallen*. De hogere waarden in het lijstje van Gauss geven nog mooiere voorbeelden. Zoals Euler al opmerkte geeft het polynoom $f_{-163}(x) = x^2 + x + 41$ geeft voor $x = 0, 1, 2, \ldots, 39$ een onwaarschijnlijk lange rij van priemgetallen: 41, 43, 47, 53, 61, 71, 83, 97, 113, 131, 151, 173, 197, 223, 251, 281, 313, 347, 383, 421, 461, 503, 547, 593, 641, 691, 743, 797, 853, 911, 971, 1033,1097, 1163, 1231, 1301, 1373, 1447, 1523, 1601. Pas $f(40) = 1681 = 41^2$ is een niet-priemgetal. We laten het aan de lezer over om de 'priemproducerende' eigenschappen van de andere polynomen f_{-p} te onderzoeken.
Een ander curieus fenomeen is het gedrag van de complexe e-macht $e^{2\pi iz}$ op de waarden in het complexe bovenhalfvlak die samenhangen met de roosters R_D voor de D uit het lijstje van Gauss. In deze gevallen vinden we voor $z = \sqrt{D}$ we getallen die verbazend dicht bij gehele getallen liggen, en wie gewend is aan rekenmachines zou denken dat een waarde als

$$e^{\pi\sqrt{163}} = 262537412640768743.999999999999$$

op een geheel getal duidt. De verklaring van dit fenomeen wordt gegeven door de theorie van *modulaire functies*, die laat zien dat $e^{2\pi iz}$ voor waarden van z met groot imaginair deel lijkt op de zogenaamde j-functie, waarvan de waarden uiterst interessante arithmetische eigenschappen blijken te hebben. Een exacte formulering valt ruim buiten het kader van dit artikel.
We dienen ons natuurlijk afvragen of het lijstje in 5.1 van discriminanten met klassengetal 1 daadwerkelijk compleet is. Dit heet het *klassengetal-1-probleem*. In 1908, ruim een eeuw na het verschijnen van de *Disquisitiones*, merkt Weber in zijn *Lehrbuch der Algebra* hierover het volgende op:

> Daß nicht mehr Diskriminante dieser Art existieren, kann bis jetzt nur daraus geschlossen werden, daß, soweit man die Berechnung der Klassenzahlen fortsgesetzt hat, weitere Zahlen dieser Art nicht gefunden sind.

Het bewijs van de compleetheid heeft lang op zich laten wachten. Een bewijs uit 1954 van Heegner was onvoldoende duidelijk om de wiskundige gemeenschap te overtuigen. Pas in 1968 verschenen er twee onafhankelijke bewijzen, van Stark en van Baker. Toen men daarop het bewijs van Heegner nog eens onder de loep nam bleek dit in essentie correct te zijn.

Inmiddels is voor een aantal kleine waarden van h, waaronder alle oneven waarden tot 21, de volledigheid van de lijstjes van Gauss en latere wiskundigen bewezen. Hierbij is gebruik gemaakt van een groot aantal verschillende technieken. De theorie van elliptische krommen, die vorig jaar tot een bewijs van Fermat's laatste stelling door Wiles heeft geleid, speelt hierin wederom een prominente rol.

Opgaven.

1. Laat zien dat een niet-constant polynoom $f \in \mathbb{Z}[X]$ oneindig veel waarden aanneemt op $\mathbb{Z}$ die niet priem zijn.

2. Definieer g door $g(x) = \sqrt[3]{744 - e^{\pi\sqrt{x}}}$. Benader de waarde $g(x)$ voor $x = 11, 19, 43, 67, 163$.

Voor verdere informatie over de in dit artikel behandelde onderwerpen is het boek van D. A. Cox getiteld *Primes of the form* $x^2 + ny^2$ (Wiley-Interscience, 1989) een uitstekende referentie.

Gedeelde vreugde is dubbele vreugde, of krommen, voortgebracht door recursieve procedures

Kees van Overveld

1. Inleiding: meetkunde voor pixels

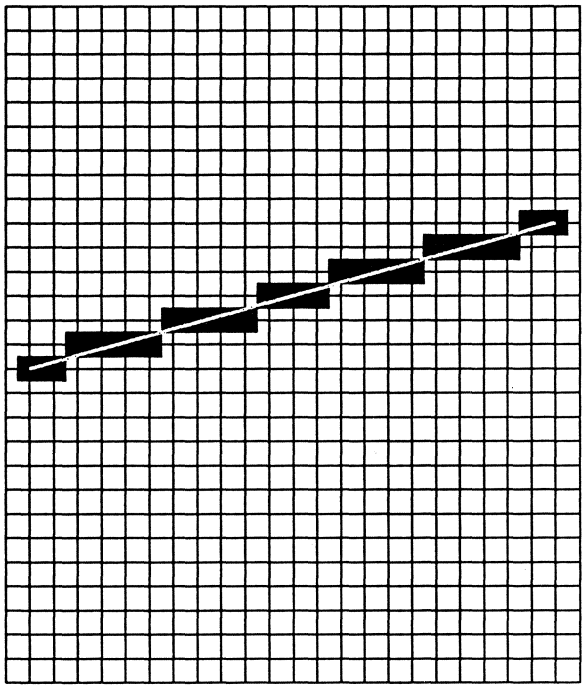
De meetkundige begrippen 'punt' , 'cirkel', 'kegelsnede', en dergelijke, worden doorgaans geassocieerd met objecten die gedefinieerd zijn in de continue ruimte, $\mathbb{R}^2$ of $\mathbb{R}^3$. Onder andere door de komst van de computer is er echter ook grote belangstelling ontstaan voor meetkundige objecten die in een discrete ruimte bestaan, $\mathbb{Z}^2$ of $\mathbb{Z}^3$. Deze belangstelling is te verklaren door de werking van een computer beeldscherm. Het beeld op een computerbeeldscherm is opgebouwd uit een grote collectie punten (*pixels* genaamd, een afkorting voor het Engelse *p*icture *el*ements). Deze punten zin gerangschikt in een rechthoekig rooster en de afmetingen van dat rooster noemen we de *resolutie* van het beeld. Resoluties die veel voorkomen zijn bijvoorbeeld 640×400, 800×600, 1024×720, enzovoort. Hoe groter de resolutie is, hoe meer details we kunnen onderscheiden in beelden die worden gerepresenteerd door de pixels verschillende helderheden of verschillende kleuren te geven.

1.1. Aliasing: een onvermijdelijke kwaal van rasterbeelden

Om afbeeldingen van continue meetkundige objecten op een beeldscherm te maken, moeten dus pixelverzamelingen geconstrueerd worden en van een kleur voorzien worden die lijken op de voor te stellen meetkundige objecten. Bijvoorbeeld een punt met heeltallige coordinaten wordt voorgesteld door een pixel op de plaats die door die coordinaten in het rooster wordt aangewezen. Om een punt met coordinaten die niet heeltallig zijn, zeg (12.3, 45.6) in het pixel-rooster te kunnen weergeven moeten we gaan afronden. In dit geval wordt dat (12,46). Maar dat betekent dus dat er verschillende punten zijn die op dezelfde pixel terecht komen: (12.1,45.9) komt op hetzelfde pixel terecht. Dit is de reden dat de weergave van een continu object op een rooster altijd met informatieverlies gepaard gaat: we kunnen geen onderscheid maken tussen punten die dichter bij elkaar liggen dan de afstand tussen twee naburige pixels. Dit is een fundamenteel verschijnsel dat uitgebreid bestudeerd is in de literatuur over signaalbewerking. De diepere oorzaak van het verschijnsel heeft te maken met het feit dat we een continu signaal (het beeld) gaan *samplen* op een periodiek rooster, waardoor er een bovengrens ontstaat aan de frequenties van de Fourier-componenten die in het beeld kunnen voorkomen. Als er componenten in het beeld voorkomen met frequenties die te hoog zijn (overeenkomend met details die een kleinere afmeting hebben dan de afstand tussen twee naburige pixels) dan veroorzaken die componenten hinderlijke vervormingen in het laag-frequente deel van het beeld.

1.2. Onscherpe beelden, grijswaarden en anti-aliasing

We zien in Figuur 1 een representatie van een rechte lijn onder een flauwe hoek



Figuur 1: Representatie van een rechte lijn op een rooster van pixels.

met de horizontale richting die op een rooster met pixels weergegeven wordt. De signaaltheorie levert ons overigens wel een methode om van die hakkeltjes af te komen, en de remedie ligt eigenlijk ook wel voor de hand. Immers: het euvel wordt veroorzaakt door de hoogfrequent componenten uit het beeld (de scherpe zwart-wit overgangen). Als we dus, alvorens we het beeld gaan vertalen naar de pixelrepresentatie, eerst ervoor zorgen dat er geen scherpre zwart-wit overgangen meer in het beeld voorkomen dan bereiken we daarmee dat die laagfrequente verstoringen ook niet meer optreden. De scherpe zwart-wit overgangen kunnen we uit het beeld verwijderen door het beeld te *filteren* ofwel door na te bootsen dat we er een onscherpe foto van maken (we weten dat in een onscherpe foto de harde zwart-wit overgangen ook wazig worden). Dit proces heet dan ook *anti-aliasing.* Om anti-aliasing te kunnen doen, moeten we echter kunnen beschikken over verschillende kleuren (of grijswaarden) per pixel.

1.3. Lijnstukken: enkele makkelijke gevallen

Omdat we ons in dit betoog zouden bezig houden met uitsluitend zwarte en witte pixels, moeten we hier de filtering-methoden voor anti-aliasing verder laten rusten, en we zullen bekijken in hoeverre we toch meetkundige objecten op discrete roosters kunnen afbeelden, waarbij we de jaggies maar voor lief zullen nemen.

Na de losse punten zijn lijnstukken de eerstvolgende in aanmerking komende objecten om af te beelden. Er zijn drie typen lijnstukken die heel gemakkelijk op een rooster afgebeeld kunnen worden. De horizontale en vertikale lijnstukken kunnen we immers weergeven door louter naast elkaar of bovenelkaar liggende pixels zwart te maken. Hierbij treden nergens hakkeltjes op: we hebben hier kennelijk een geval waarin we van aliasing toevallig geen last hebben. Het derde type lijnstukken zijn de diagonale lijnstukken. Hiervoor moeten we een collectie pixels zwart maken waarbij elke pixel de rechter (of linker) bovenbuur is van de vorige. Dat betekent dat we kunnen zeggen dat er tussen elk tweetal opvolgende pixels een hakkeltje optreedt. Maar dat betekent dat bij niet al te lage resoluties de hakkeltjes zo dicht bij elkaar liggen dat ze met het blote oog nauwelijks te onderscheiden zijn, en dan hebben we er dus ook geen last van.
Iets minder triviaal zijn de lijnstukken onder andere hellingshoeken. We moeten hier een *algoritme* gaan introduceren om vast te stellen, voor een gegeven tweetal punten a en b waarbij we voorlopig voor het gemak veronderstellen dat a en b heeltallige coordinaten hebben, welke pixels zwart gemaakt moeten worden om de beste benadering voor het lijnstuk ab te verkrijgen.

1.4. Algoritmen: de receptenboeken van de wiskunde

Een algoritme is een recept dat bestaat uit een eindig aantal basishandelingen die op een eenduidige manier gedefinierd zijn en die 'gedachtenloos' uitgevoerd kunnen worden. Het idee van algoritmen is al heel oud: de Euclidische meetkunde is opgebouwd rond een groot aantal algoritmen, namelijk de constructies met passer en lineaal om tot allerlei bewijzen te komen in de vlakke meetkunde. De basishandelingen zijn beperkt in aantal en nauwkeurig omschreven, zoals 'zet een punt in het vlak', 'leg een lineaal langs twee punten', 'trek een lijn langs een lineaal', 'neem de afstand tussen twee punten als de afstand van de benen van een passer', 'trek met een passer een cirkel om een gegeven punt', enzovoort. We merken op dat, behalve de basishandelingen, er bij een algoritme ook altijd sprake is van de *objecten* waarop die basishandelingen uitgevoerd moeten worden: in het geval van de Euclidische meetkunde zijn dat de punten en de lijnen die tijdens de uitvoering van een algoritme gecreeerd worden. We zullen die objecten aanduiden met het woord *variabelen.*
Ook al sinds de oude Grieken zijn er voorbeelden van algoritmen waarin de optredende variabelen geen meetkundige objecten zijn, maar getallen: bijvoorbeeld het algoritme om de grootste gemene deler van twee gehele getallen p en q te bepalen:
'Zolang p en q ongelijk zijn, vervang de grootste van de twee door het verschil van de grootste en de kleinste. Als p en q gelijk aan elkaar zijn, hebben ze de waarde van de grootste gemene deler van de (oorspronkelijke) p en q.
Tegenwoordig schrijven we dat op een wat zakelijkere manier op, op een wijze die ontleend is aan de talen waarin computers geprogrammeerd worden. Dat hoeft geen verbazing te wekken: computers zijn feitelijk niets anders dan programmeerbare rekenmachines waarbij de afzonderlijke stappen uit een algoritme de rekenbewerkingen zijn die in de *processor* (het hart van de computer) uitgevoerd worden; de optredende variabelen verwijzen naar stukjes uit het

geheugen van de computer. Het bovenstaande algoritme zouden we kunnen opschrijven als volgt:

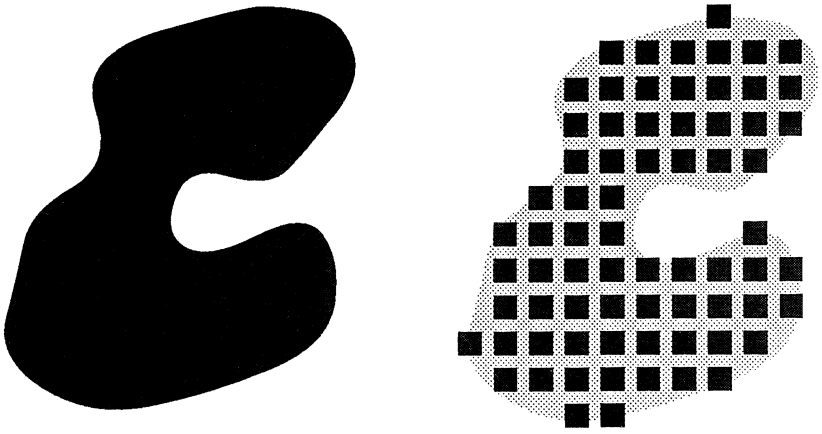
```
int p,q;                        /* hier definieren
                                 * we twee geheeltallige
                                 * variabelen */

while (p!=q)                    /* zolang p en q ongelijk */
   if(p>q) p=p-q ;              /* als p de grootste is, vervang p
                                 * door het verschil */
   else    q=q-p ;              /* anders vervang q
                                 * door het verschil */

print(p);                       /* als we klaar zijn, zijn p en q
                                 * gelijk aan elkaar geworden
                                 * en dan drukken we
                                 * een van de twee af.*/
```

1.5. Algoritmen voor pixelverzamelingen

In het navolgende gaan we bezien hoe een algoritme gemaakt zou kunnen worden om de pixels voort te brengen die we zwart moeten maken om een zo goed mogelijke benadering te verkrijgen van een lijnstuk. Voordat we dat kunnen doen moeten we echter nog afspreken wat we eigenlijk verstaan onder 'zo goed mogelijk'. Het zal blijken dat we op verschillende manieren te werk kunnen gaan, afhankelijk van wat voor definitie van 'een lijnstuk' we uitgaan. Als we de parametervoorstelling van een lijnstuk als uitgangspunt nemen kunnen we het meest gebruikte algoritme vinden: een algoritme dat in 1965 door Bresenham gepubliceerd is. Voor de rest van ons verhaal is het echter aardiger om een wat exotischere definitie van een lijnstuk te nemen, en wel een definitie die zelf al de vorm van een algoritme heeft, en wel een *recursief* algoritme (we zullen verderop uitleggen wat 'recursief' betekent). Zo komen we op het Jonkers-Corthout algoritme dat pas ca.1985 'ontdekt' werd. Het aardige van dit Jonkers-Corthout algoritme is dat, ondanks zijn ogenschijnlijke eenvoud, de voortgebrachte pixelcollecties in feite veel ingewikkelder zijn dan de pixelcollecties van het Bresenham algoritme, en zij hebben zelfs trekjes van fractals - hetgeen een voorbode is van wat ons verderop nog te wachten staat. *1.6. Krommen*

Vervolgens stappen we over van rechten op krommen. Ook hier is het ons uiteindelijk begonnen om algoritmen, uit te voeren door een computer, voor het voortbrengen van pixelverzamelingen, maar we maken eerst even een tussenstop bij een ander resultaat dat al in de oudheid bekend was: het maken van afrondingen van een hoek tussen twee rechten door een limiet-proces van zaagsneden. Wij zullen zien dat de afrondingen die op die manier voortgebracht worden, kwadratische krommen zijn waarmee de link met het thema van deze cursus gelegd is. We blijven echter niet al te lang stilstaan bij het ge-

val van de kwadratische afrondingen, want een eenvoudige generalisatie van het zaagsneden-algoritme levert krommen van een hogere graad die op een prettige manier aan een aantal practische randvoorwaarden voldoen. Deze krommen zijn sedert ca. 30 jaar in gebruik in de auto-industrie en recenter ook in andere takken van het industrieel ontwerp, We praten hier over Bezier-krommen: een belangrijke telg uit de familie der zogenaamde *spline*-krommen.



Figuur 2: Een continu meetkundig object en zijn close discretisatie.

2. Algoritmen voor pixelverzamelingen.

De bedoeling bij het vertalen van een continu meetkundig object, zoals een lijnstuk, naar een collectie pixels, is dat de gelijkenis tussen het meetkundige object en de collectie pixels zo groot mogelijk is. Het afleiden van een algoritme is echter een wiskundige bezigheid, en dat betekent dat de specificatie van een algoritme (de omschrijving van wat dat algoritme gaat uitrekenen of voortbrengen) ook op een wiskundige manier gegeven moet zijn. We moeten dus een wiskundige definitie geven van 'zo groot mogelijke gelijkenis'. We hadden al eerder gezien dat voor een punt zo'n definitie makkelijk te geven is: we zien zonder moeite in dat het afronden van de coordinaten van het punt op geheeltallige pixel-coordinaten, de beste benadering levert. In het algemeen is echter zo'n operationale wijze van doen niet geschikt. We moeten dan een relatie introduceren, tussen een continue verzameling V die een deel is van $\mathbb{R}^2$ en een discrete verzameling pixels P, deel van $\mathbb{Z}^2$, die uitspreekt dat ze voldoen aan de volgende twee eisen (zie Figuur 2):

- voor elk punt p uit P is er een punt v uit V zodanig dat p en v 'dicht bij elkaar' liggen;
- voor elk punt v uit V is er een punt p uit P zodanig dat p en v 'dicht bij elkaar' liggen.

We merken op dat inderdaad beide delen van deze definitie nodig zijn: als we alleen het tweede deel zouden eisen, is voor elke V, de oplossing $P = \mathbb{Z}^2$ afdoende. Omgekeerd is de lege verzameling als keuze voor P voldoende als we alleen de eerste eis zouden stellen.
We merken op dat we nog moeten afspreken wat we verstaan onder het 'dicht bij elkaar liggen' van twee punten, maar het is duidelijk dat dit een stuk minder lastig is dan het 'dicht bij elkaar liggen' van twee willekeurige verzamelingen. Laten we hiervoor vooralsnog de (verderop te definieren) relatie $D(p, v)$ tussen de punten p en v introduceren, dan kunnen we onze eis van 'zo goed mogelijk op elkaar lijken' opschrijven als:

$$(\forall p \in P : \exists v \in V : D(p, v)) \tag{1}$$

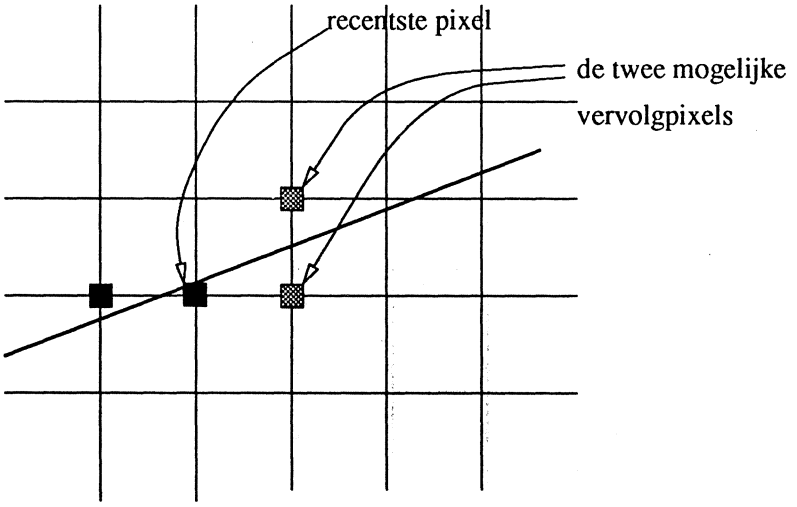
en

$$(\forall v \in V : \exists p \in P : D(p, v)). \tag{2}$$

Bovenstaande relatie tussen P en V zullen we aanduiden als *closeness* ('nabijheid'): we zeggen dat P een close discretisatie van V is.
De relatie D moeten we toch wel zorgvuldig kiezen. Als we een heel strikte definitie voor $D(p, v)$ kiezen, bijvoorbeeld, 'p=v', dan zal het niet mogelijk zijn om voor een willekeurige V een close discretisatie te vinden. Omgekeerd, als we een te soepele definitie van D hanteren zal een close discretisatie niet 'de best mogelijke' gelijkenis tussen P en V garanderen. In het algemeen kunnen we $D(p, v)$ opvatten als $v \in B_p$ waar B_p een goed gekozen omgeving van p is. We kunnen makkelijk inzien dat dan en slechts dan als de B_p voor alle pixels p juist de gehele $\mathbb{Z}^2$ overdekken, voor elke V een close discretisatie bestaat. Andersom, als de B_p allen disjunct zijn, is een close discretisatie ook uniek, want voor elke v is er dan precies een pixel p waar deze 'bij hoort'. Omdat we waarschijnlijk ook willen dat alle B_p dezelfde vorm hebben, zijn er maar weinig keuzen voor B_p te maken. In de practijk kiezen we de B_p meestal als vierkanten met zijde 1, gecentreerd om p, waar de linker- en de bovenrand bij het vierkant horen, en de rechter- en de onderrand niet.
Nu we hebben vastgelegd wat we verstaan onder 'de best gelijkende pixelverzameling' P bij een gegeven meetkundig object V, namelijk de close discretisatie, kunnen we stelselmatig voor allerlei typen meetkundige objecten V algoritmen gaan maken om die P's voort te brengen. Voor het geval dat V een 'los' punt is, is ons eerdere 'afrondingsalgoritme' inderdaad de juiste keuze (afhankelijk van de wijze waarop we 0.5 afronden), zoals zich eenvoudig laat verifieren.
Voor een lijnstuk zijn er verschillende manieren waarop we te werk kunnen gaan. We kunnen een lijn voorstellen door een algebraische vergelijking, $px + qy = r$, maar het blijkt dat deze voorstelling geen intuïtieve manier is om punten te genereren die aan die vergelijking voldoen. Ook is het moeilijk om aan te geven dat het slechts een lijnstuk, in plaats van een oneindig lange rechte betreft. (Deze vorm is natuurlijk wel uiterst geschikt om te verifieren of een punt (x, y) op de lijn ligt, en ook het snijpunt van twee lijnen kunnen we er handig mee uitrekenen.)

Wat dat betreft is een parametervoorstelling, $V = \{a + \lambda(b - a)\}$ waarbij λ tussen 0 en 1 varieert, en a en b twee punten in het vlak zijn, veel aanschouwelijker. Ten eerste is onmiddellijk duidelijk op welke manier de eindpunten van het lijnstuk, a en b in de parametervoorstelling voorkomen. Ten tweede geeft de parametervoorstelling ook al een hint om een algoritme te maken om een discretisatie op te leveren: als λ in voldoend kleine discrete stapjes λ_i varieert ontstaat een collectie punten $a + \lambda_i(b - a)$. Voor elk van die punten passen we ons 'afrondingsalgoritme' toe, en zo ontstaat een close discretisatie van een lijnstuk.
Tot voor 1965 werden op plotters (rasterbeeldschermen bestonden toen nog niet) op deze manier lijnen getrokken. De computers uit die tijd waren echter nog niet erg krachtig, en het 'algoritme' heeft nogal wat vervelende eigenschappen die ervoor zorgen dat er veel onnodig rekenwerk wordt gedaan - wat dus een zwaar beslag legde op de schaarse rekencapaciteiten uit die tijd. Met name is het zo dat

- voor elke waarde van λ_i moeten twee floating-point vermenigvuldigingen uitgevoerd worden plus nog een tweetal optellingen (merk op dat de vector $b - a$ slechts eenmaal berekend wordt).
- het is niet a-priori duidelijk hoeveel waarden van λ_i er precies nodig zijn; te weinig waarden geeft gaten in de collectie pixels P, en te veel waarden is een verspilling van rekencapaciteit.



Figuur 3: Bij Bresenham's algoritme wordt de keuze gedaan voor het beste van twee mogelijke pixels om de lijn voort te zetten.

In 1965 deed Jack Bresenham van IBM de volgende constatering:
Als we veronderstellen dat de lijn een helling heeft die ten hoogste 45 graden is, en we hebben een aantal pixels al gevonden, dan zijn er voor het eerstvolgende

pixel slechts twee keuzen mogelijk: de rechter buur en de rechtsboven buur (zie Figuur 3). Voor beide pixels kan ik de afstand tot de te discretiseren lijn uitrekenen, en ik kies dat pixel met de kortste afstand. Als die afstanden vanaf 'scratch' uitgerekend moeten worden, is dat een dure en onaantrekkelijke berekening. Het is echter heel goedkoop mogelijk om, steeds als ik naar het volgende pixel overstap, met een simpele berekening (niet meer dan twee of drie geheeltallige optellingen) de waarde van die afstanden bij te houden. Met behulp van deze zogenaamde incrementele berekening leidde hij het volgende algoritme af:

```
int x,y,d,d1,d2;        /* we veronderstellen een lijn
                         * van (0,0) naar (X,Y) waar
                         * 0<Y<X */
 x=0;
 y=0;
 d=2*Y-X;
 d1=2*Y;
 d2=2*(Y-X);
 while(x<X)             /* zolang we nog niet klaar zijn */
 {
     pixel(x,y);        /* maak een pixel zwart op plaats x,y */
     x=x+1;             /* het voldende pixel staat in ieder
                         * geval rechts van de
                         * huidige plaats */
     if(d>0)            /* we moeten naar rechtsboven */
     {
        d=d+d2;
        y=y+1;
     } else             /* we moeten naar rechts */
     d=d+d1;
  }
```

Het is op het eerste gezicht geheel niet duidelijk dat dit algoritme inderdaad een collectie pixels oplevert die een close discretisatis is van het lijnstuk $(0,0)$ naar (X,Y). Een zorgvuldige analyse (die buiten het bestek van deze syllabus valt) laat echter zien dat zulks inderdaad het geval is. Indien het lijnstuk niet vanaf het punt (0,0) loopt, maar vanaf een willekeurig ander punt is het natuurlijk triviaal om de pixels over de juiste vector te verschuiven. Ook de gevallen waar het lijnstuk een helling heeft die niet tussen 0 en 45 graden ligt laten zich eenvoudig uitprogrammeren.
Wij laten echter een verdere analyse van het Bresenham algoritme hier rusten, en concentreren ons liever op een andere wijze van het voortbrengen van een close discretisatie van een lijnstuk die straks bruikbaar blijkt om te generaliseren naar krommen.
Wij kunnen een lijnstuk als volgt definieren: 'het lijnstuk van a naar b is de vereniging van de lijnstukken van a naar $\frac{a+b}{2}$ en van $\frac{a+b}{2}$ naar b'. Deze definitie ziet er wat wonderlijk uit; het is een voorbeeld van een *recursieve* definitie.

Een recursieve definitie is een definitie waarbij het te definieren begrip (het *definiens*) ook in de inhoud van de definitie (het *definiendum*) terugkomt, zoals bijvoorbeeld 'een paard is een paard met vier benen'. Soms is dit toegestaan, en zelfs heel handig, zoals bijvoorbeeld bij het voortbrengen van de natuurlijke getallen:

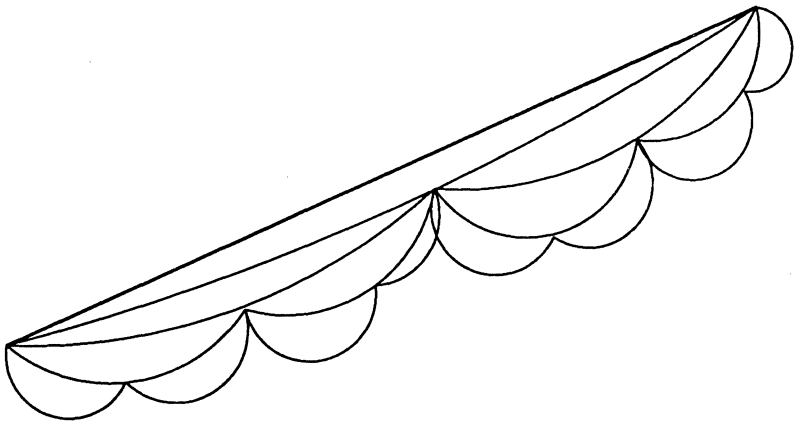
- nul is een natuurlijk getal;
- als p een natuurlijk getal is, is de opvolger van p ook een natuurlijk getal.

We merken echter op dat dit alleen maar een goede recursieve definitie is omdat we 'van de grond' kunnen komen: in de meeste gevallen is de definitie inderdaad echt recursief (het tweede deel van de definitie), maar in een bijzonder geval is er een uitweg uit de vicieuze cirkel door een gewone, niet-recursieve defintie. Onze 'definitie' van een lijnstuk hierboven is wat dat betreft niet compleet omdat die 'nooduitgang' ontbreekt. Intuitief zien we steeds kortere lijnstukjes ontstaan, maar het zijn allemaal nog steeds echte lijnstukjes omdat in het continue geval het opdeel-proces niet eindigt. We bedenken ons echter dat we, in het geval van pixels, we een natuurlijke ondergrens aan de lengte van een lijnstuk hebben: een lijnstuk korter dan een pixel is zinloos, dus we kunnen de definitie herformuleren:
'als a en b verder van elkaar liggen dan 1, is het lijnstuk van a naar b de vereniging van de lijnstukken van a naar $\frac{a+b}{2}$ en van $\frac{a+b}{2}$ naar b; als a en b niet verder van elkaar liggen dan 1 is het lijnstuk het pixel op de afgeronde plaats van $\frac{a+b}{2}$.'
Hiermee hebben we een geldige definitie van een lijnstuk gemaakt, en we kunnen dit ook nog op een heel aantrekkelijke manier in een computertaal opschrijven ook:

```
lijnstuk(punt a,b)
{ punt m;
  if('a en b voldoende ver van elkaar')
  {
    m=gemiddelde van a en b;
    lijnstuk(a,m);
    lijnstuk(m,b);
  } else
  {
    pixel(afgeronde plaats van a);
    pixel(afgeronde plaats van b);
  }
}
```

Dit is een voorbeeld van een recursief algoritme. Net zoals in een recursieve definitie het definiens in het definiendum voorkomt, komt bij een functie in een recursief algoritme de aanroep van die functie voor tijdens het uitvoeren van die functie zelf. In dit voorbeeld is dat de functie 'lijnstuk' die het stuk van de algoritme bevat dat nodig is om een lijnstuk voort te brengen. En

net zoals bij een recursieve definitie moeten we 'van de grond komen' door een uitzonderingsgeval: hier het geval dat a en b dicht bij elkaar liggen. De correcte werking van dit algoritme (als je eenmaal gewend bent aan recursie) is veel makkelijker in te zien dan de correcte werking van Bresenham's algoritme, omdat dit zogenaamde Jonkers-Corthout algoritme een rechtstreekse vertaling is van de wiskundige definitie van een lijnstuk met behulp van recursie.
Overigens is het niet helemaal eerlijk om deze versie van het J.-C.-algoritme direct met Bresenham's algoritme te vergelijken. Immers: als de coordinaten van de punten a en b reele getallen zijn, ontstaan door het steeds opnieuw delen door 2 breuken (zelfs als in het begin a en b toevallig op roosterposities liggen). En dat betekent dat binnen de aanroep *pixel* een afronding moet plaatsvinden. Als daarentegen de punten a en b geheeltallige coordinaten hebben, gebeurt er iets heel wonderlijks. De eerste maal vindt afronding plaats om het middelpunt van het lijnstuk ab te vinden. Dat betekent dat het pixel dat dit middelpunt voorstelt, niet meer dan 0.5 van de correcte waarde afligt. We gaan echter verder met dit (foutieve) punt om de twee helften elk weer op te delen, en we maken opnieuw een afrondfout, die bijtelt bij de afrondfout die we al hadden.



Figuur 4: Schematische weergave van de cumulative afrondfouten bij het oorspronkelijke Jonkers-Corthout algoritme.

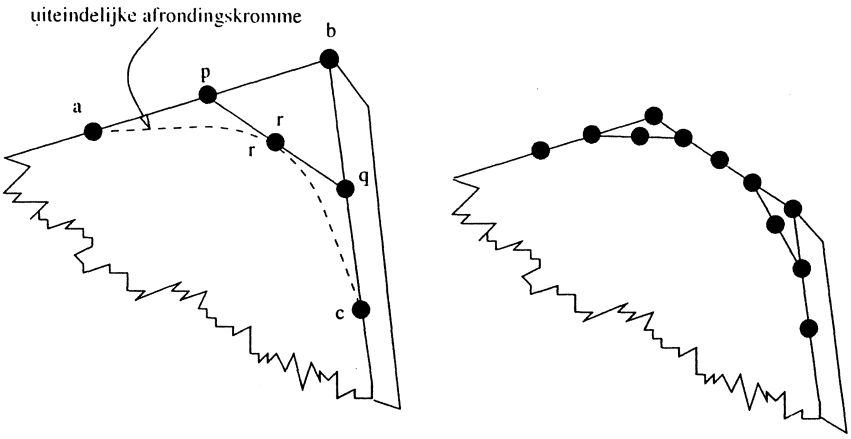
Dus de punten op $\frac{1}{4}$ en $\frac{3}{4}$ kunnen al lijden (als we pech hebben) aan een grotere afrondfout. De punten op $\frac{1}{8}$, $\frac{3}{8}$, $\frac{5}{8}$ en $\frac{7}{8}$ kunnen nog weer een grotere afrondfout hebben, enzovoort. De afrondfouten zijn dus heel grillig over de lijn verdeeld, en sterk overdreven ziet de collectie pixels er uit als in Figuur 4. Dit lijkt dus helemaal niet op een keurige rechte lijn tussen a en b, en we zouden kunnen betogen dat het resulterende object wel iets weg heeft van een fractal: een object dat op alle lengteschalen structuur bezit, hoever we ook inzoomen. We merken dus op dat het met recursieve algoritmen mogelijk is om met weinig moeite fractal-achtige objecten te genereren, zelfs als we misschien de bedoeling hadden om een glad object te maken (in dit geval een recht lijnstuk).

Het is overigens wel mogelijk om het J.-C.-algoritme een beetje aan te passen zodanig dat er een keurige close discretisatie van een rechte uitkomt. Om precies te zijn: dezelfde rechte als door Bresenham's algoritme wordt voortgebracht (dit is ook niet zo verwonderlijk als we ons herinneren dat we de definitie van closeness zo gemaakt hadden dat een close discretisatie uniek is: het doet er dan immers niet meer toe op welke wijze we die close discretisatie voortbrengen. We zullen dat niet helemaal uitwerken hier; de basisgedachte komt erop neer dat bij elk punt een apart getal wordt bijgehouden dat aangeeft hoever de echte lijn van dat punt afligt, ongeveer opdezelfde manier als de d, $d1$ en $d2$ dat in Bresenham's algoritme doen. En afhankelijk van de waarde van die afwijkingsgetallen wordt beslist of naar boven ofwel naar beneden afgerond wordt. Dus ongeveer als volgt:

```
lijnstuk(roosterpunt a,b; int da,db)
{ roosterpunt m;
  int dm;
  if('a en b voldoende ver van elkaar')
  {
    m=gemiddelde van a en b waarbij voor de afronding
    met da en db rekening gehouden wordt;
    dm=..... zodanig dat dm een maat is voor de afwijking bij m;
    lijnstuk(a,m,da,dm);
    lijnstuk(m,b,dm,db);
  } else {
    pixel(a);
    pixel(b);
  }
}
```

3. Recursieve algoritmen voor krommen

In de vorige paragraaf kregen we, door ondoordachte afronding, in het oorspronkelijke J.-C.-algoritme al pixelverzamelingen die krommen voorstellen. De vraag is natuurlijk of we dat niet een beetje systematischer voor elkaar kunnen krijgen. Dit is inderdaad mogelijk, en wel op een manier die al heel lang, sinds de tijd van de oude Grieken bekend is. Beschouw drie punten in het vlak, a, b en c. Stel voor dat a b c een hoek aan een plank voorstelt (zie onderstaande figuur). De vraag is nu hoe de timmerman die hoek kan afronden. Hij kan als volgt te werk gaan: ten eerste zoekt hij het punt $p = \frac{a+b}{2}$ midden tussen a en b, en het punt $q = \frac{b+c}{2}$, midden tussen b en c. Langs het lijnstuk pq zaagt hij de driehoek pbq eraf. Nu is de hoek bij b vervangen door twee stompere hoeken: apq en pqc. Dit is echter nog niet glad genoeg: hij wil het proces nog een keer herhalen. Hij zoekt daartoe het punt $r = \frac{p+q}{2}$, midden tussen p en q. Vervolgens merkt hij op dat de drie punten a, p, r met z'n drieen net zo behandeld kunnen worden als de oorspronkelijke a, b, c. En zo ook voor r, q, c. Dus hij zoekt opnieuw de gemiddelden op, verbindt ze met twee lijnstukjes en zaagt de top-driehoeken eraf. Nu is de oorspronkelijke hoek a b c al vervangen door vier stompe hoeken (zie Figuur 5), en in de practijk is dit al zo glad dat



Figuur 5: Twee fasen van het zaagsneden-algoritme.

hij de rest wel met een vijl of met een stuk schuurpapier afkan. Wiskundig echter kunnen we dit proces willekeurig lang voortzeten, en de reeks overblijvende lijnstukjes zal steeds meer op een gladde kromme gaan lijken (we hebben dat nu uiteraard nog niet bewezen; straks zal blijken dat het resultaat inderdaad een gladde kromme is).
We merken op dat we hier een kromme op een wat merkwaardige manier gedefinieerd hebben: niet als vergeljiking, $f(x,y) = 0$, of als parametervoorstelling, $x = x(t)$, $y = y(t)$ en zelfs niet eens met een meetkundige constructie, zoals een kegelsnede. We zullen deze manier van het definieren van een kromme een *procedurele* definitie noemen, omdat we alleen maar een procedure ofwel een algoritme gegeven hebben waarmee die kromme voortgebracht wordt. In dit geval is het zelfs een recursief algoritme, en dat kunnen we wat beter zien als we het zaagsnede algoritme in computertaal gaan opschrijven:

```
zaagsnede(punt a,b,c)
{
punt p,q,r;
  if('a en b en c ver van elkaar')
  {
    p=gemiddelde van a en b;
    q=gemiddelde van b en c;
    r=gemiddelde van p en q;
    zaagsnede(a,p,r);
    zaagsnede(r,q,b);
  } else
  {
    verbind a en b met een recht lijntje;
    verbind b en c met een recht lijntje;
  }
}
```

Het meest essentiele van dit algoritme is dat het werkt doordat de gehele kromme gedefinieerd door a, b en c wordt opgevat als de vereniging van twee helften van die kromme, ieder gedefinieerd door a, p, r respectievelijk r, p, b. Dit lijkt heel erg op de definitie van het lijnstuk zoals we die tegenkwamen bij het J.-C.-algoritme, waar een lijnstuk werd opgevat als de vereniging van de twee helften van dat lijnstuk, ieder gedefinieerd door a, m, respectievelijk m, b. We merken ook nog op dat voor de kromme, de punten p, q en r werden verkregen door middelingen uitgaande van a, b, en c, terwijl voor de rechte het punt m het resultaat was van een middeling van a en b.
We kunnen dus met recht ervan spreken dat de op deze manier op een recursieve procedurele gedefinieerde kromme een *generalisatie* is van het recursief gedefinieerde lijnstuk.
Ofschoon we nog niet veel van deze kromme weten, kunnen we alvast een paar eigenschappen vaststellen.

- de kromme heeft a en c als eindpunten.
- als de kromme differentieerbaar (liever gezegd: als een goedgekozne *parameterisatie* van de kromme, $f = f(t)$, differentieerbaar is), raakt hij aan ab en aan bc.
- omdat de kromme uit de twee halve krommen bestaat die in punt r bij elkaar komen terwijl pr en rq in elkaars verlengde liggen, geldt ook weer dat als de kromme differentieerbaar is, is pq een raaklijn in het punt r.

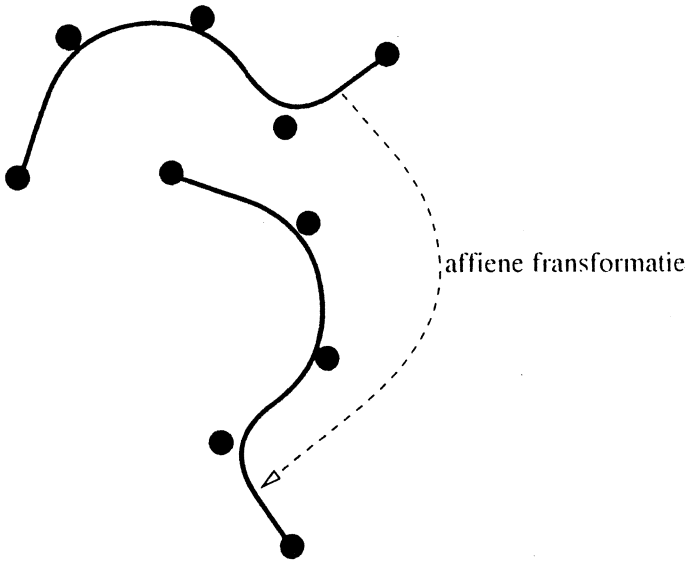
We gaan nu eerst laten zien dat de geheimzinnige kromme die we op deze manier ingevoerd hebben een eenvoudige parameterkromme is. We voeren hiertoe een klasse parameterkrommen in, naar de ontdekker ervan *Bezier krommen* genoemd, die we schrijven als

$$f(t) = \Sigma_{i=0}^{i=n} p_i B_i^n(t) \tag{3}$$

waarin p_i een serie punten in het platte vlak, of in de ruimte zijn, en B_i^n de zogenaamde *Bernstein* polynomen van de ne graad, $B_i^n(t) = C_i^n t^i (1-t)^{n-i}$. Hierin is $C_i^n = \frac{n!}{i!(n-i)!}$, ofwel de binomiaalcoefficient n boven i. De Bernstein polynomen hebben een groot aantal interessante eigenschappen. Voor ons zijn op dit moment de belangrijkste:

- voor vaste n sommeren $B_i^n(t)$ tot 1, ongeacht de waarde van t. Daardoor is de kromme $f(t)$ die we verkrijgen door alle punten p_i een vaste vector δ op te schuiven gelijk aan $f(t) + \delta$, de opgeschoven kromme.
- ze zijn allemaal positief. Daardoor blijft de kromme in zijn geheel binnen de convexe omhulling van de p_i.
- $B_0^n(0) = 1$, verder zijn alle $B_i^n(0)$ gelijk aan 0. Analoog is $B_n^n(1) = 1$, en verder zijn alle $B_i^n(1)$ gelijk aan 0.

De punten p_i *controleren* de vorm van de kromme, en daarom noemen we deze punten ook wel de *controlepunten.* Omdat $f(t)$ een lineaire (en zelfs convexe) combinatie is van de controlepunten voor elke t geldt dat als we een willekeurige affiene transformatie A op de controlepunten toepassen, het effect gelijk is aan



Figuur 6: Bij een affiene transformatie van de controlepunten wordt de voortgebrachte kromme op dezelfde manier getransformeeerd.

het toepassen van A op $f(t)$. Met andere woorden, de kromme zit 'vast' aan de controlepunten (zie figuur 6). Definitie 3 ziet er, in zijn meest algemene vorm, misschien wat onaantrekkelijk uit. Voor willekeurige n is het een polynomische kromme in t die door $n+1$ controlepunten bepaald wordt. Laten we eerst eens kijken waartoe deze reduceert voor enkele lage waarden van n. Voor $n = 0$ omvat de sommatie slechts een term, $i = 0$, en er staat

$$f(t) = p_0. \tag{4}$$

Voor $n = 1$ komt er een eerste graads kromme:

$$f(t) = p_0(1-t) + p_1 t \tag{5}$$

ofwel een rechte lijn door p_0 en p_1. Voor t tussen 0 en 1 is het een lijnstuk met kennelijk $f(0) = p_0$ als beginpunt en $f(1) = p_1$ als eindpunt. In het algemeen geldt overigens dat 3 voor $t = 0$ door het punt p_0 gaat en voor $t = 1$ door het punt p_n, zoals we eenvoudig inzien. Door 3 te differentieren naar t zien we met enig rekenwerk dat $f'(0) = n(p_1 - p_0)$ en $f'(1) = n(p_n - p_{n-1})$. Voor het geval $n = 1$ klopt dit inderdaad keurig: de raaklijn in $t = 0$ is gelijk aan de raaklijn in $t = 1$ is gelijk aan de rechte $p_1 - p_0$ zelf.
Voor het geval $n = 2$ krijgen we een kwadratisch polynoom. Uitgeschreven is dit

$$f(t) = p_0(1-t)^2 + 2p_1 t(1-t) + p_2 t^2 \tag{6}$$

en deze kromme heeft als raakvector in $t = 0$ de vector $p_1 - p_0$ en in $t = 1$ de vector $p_2 - p_1$.
De eigenschap dat een rechte de vereniging is van twee halfrechten kunnen we met onze parametervoorstelling makkelijk uitdrukken. Bijvoorbeeld voor de linker halfrechte moeten we de vervangingen doen:

$$p_0 \rightarrow p_0 \tag{7}$$
$$p_1 \rightarrow \frac{p_0 + p_1}{2} \tag{8}$$
$$t \rightarrow 2t \tag{9}$$

Als we deze nieuwe waarden invullen in 3 voor $n = 1$ krijgen we, niet onverwacht,

$$f(t) = p_0(1 - t) + p_1 t \tag{10}$$

waarbij nu voor t tussen 0 en 1 de linker helft van het lijnstuk beschreven wordt. Voor de kwadratische versie kunnen we precies zo'n vervanging doen:

$$p_0 \rightarrow p_0 \tag{11}$$
$$p_1 \rightarrow \frac{p_0 + p_1}{2} \tag{12}$$
$$p_2 \rightarrow \frac{p_0 + 2p_1 + p_2}{4} \tag{13}$$
$$t \rightarrow 2t \tag{14}$$

en opnieuw komt er (na enig rekenen) weer precies de oorspronkelijke uitdrukking uit

$$f(t) = p_0(1 - t)^2 + 2p_1 t(1 - t) + p_2 t^2. \tag{15}$$

die nu de linkerhelft (d.w.z. het stuk van de kromme 'aan de kant van' p_0) beschrijft.
Er laat zich eenvoudig verifieren dat we ook de vervanging

$$p_0 \rightarrow \frac{p_0 + 2p_1 + p_2}{4} \tag{16}$$
$$p_1 \rightarrow \frac{p_1 + p_2}{2} \tag{17}$$
$$p_2 \rightarrow p_2 \tag{18}$$
$$t \rightarrow 2t - 1 \tag{19}$$

hadden kunnen doen waarmee we, voor de oorspronkelijke t tussen 0.5 en 1, de rechter helft van de kromme vinden.

We hebben hiermee een heel belangrijke eigenschap voor de parameterkrommen van de vorm 3 gevonden. Voor willekeurige n kan deze ook geformuleerd worden: als we de vervanging van de controlepunten p_i en de parameter t uitvoeren volgens

$$p_i \rightarrow \frac{\Sigma_{j=0}^{j=i} C^i{}_j p_j}{2^i} \tag{20}$$

$$t \rightarrow 2t \tag{21}$$

is de resulterende kromme, met de oorspronkelijke t tussen 0 en 0.5, gelijk aan de linkerhelft van de oorspronkelijke kromme. En een analoge algemene uitdrukking geldt voor de rechterhelft.
We hebben dus een gemakkelijke manier om dergelijk krommen op een recursieve manier te construeren door steeds helften en helften van helften, etcetera, te maken.
Als we nu terugdenken aan de zaagsnede constructie, zien we dat we hier precies een vervanging volgens bovenstaand recept voor het kwadratisch geval uitvoeren. Met andere woorden, we construeren een kwadratische parameterkromme met controle punten a, b, en c. En inderdaad is het zo dat deze kromme differentieerbaar is, met, zoals boven aangetoond, $b-a$ en $c-b$ als raak-richtingen in de eindpunten.

4. Rationale splines

Dit is zover een aardig resultaat, maar omdat deze syllabus over kegelsneden handelt is het interessant om te kijken of de op deze manier verkregen kromme ook inderdaad een kegelsnede is. We zien eenvoudig in dat dit in het algmeen zeker niet het geval zal zijn. We kunnen een parabool construeren, door te zetten $x = t$ en $y = pt^2 + qt + r$ ofwel door de x-coordinaten van de controlepunten zo te kiezen dat de constante term en de kwadratische term juist 0 zijn, maar dit is geen algemene aanpak, en bovendien krijgen we zo geen ellips of hyperbool. Om wel kegelsneden te krijgen moeten we onze toevlucht nemen tot rationale functies in plaats van polynomische functies. Voor controlepunten p_i, $i = 0, 1, 2$ was de uitdrukking 3:

$$f(t) = \Sigma_i p_i B_i^2(t) \tag{22}$$

waarin $B_i^2(t)$ de kwadratische polynomen $B_i^2(t) = C_i^2 t^i (1-t)^{2-i}$ zijn. Laat nu, behalve de controlepunten p_i ook de serie (scalaire) coefficienten s_i, $i = 0, 1, 2$ gegeven zijn. Dan kunnen we de *rationale* kromme

$$r(t) = \frac{\Sigma_i s_i p_i B_i^2(t)}{\Sigma_i s_i B_i^2(t)} \tag{23}$$

maken. Als alle gewichten gelijk zijn, volgt uit de normeringseigenschap van de polynomen $\Sigma_i B_i^2 = 1$ voor willekeurige t dat $r(t) = f(t)$. Als we de gewichten verschillend kiezen, kunnen we echter invloed uitoefenen op de 'belangrijkheid' van de verschillende controlepunten. We kunnen aantonen dat met de juiste

keuze van de gewichten s_i inderdaad alle kegelsneden gemaakt kunnen worden. Als we de gewichten $s_0 = s_2 = 1$ kiezen en $s_1 = \frac{r}{1-r}$ dan kan aangetoond worden dat voor $r = 0.5$ een parabool ontstaat; $r > 0.5$ een hyperbool; $r < 0.5$ een (deel van) een ellips en voor $r = 0$ een rechte. Het bewijs hiervan valt buiten het bereik van deze syllabus, maar kan gevonden worden in de meeste handboeken over (rationale) splines.

5. CUBISCHE SPLINES

De controlepunten p_0, p_1, p_2 van een kwadratische Bezier kromme definieren nogal wat relevante meetkundige eigenschappen van die kromme: twee punten waar de kromme doorheen gaat (p_0 en p_2), en de raakrichtingen in die twee punten ($p_1 - p_0$ en $p_2 - p_1$). Dat maakt op zichzelf de kromme een belangrijk gereedschap in handen van industriele ontwerpers en gebruikers van computer tekenpaketten. Het is echter jammer dat de twee raakrichtingen in de uiteinden niet onafhankelijk van elkaar te definieren zijn. Als we een raakvector wijzigen moeten we p_1 opschuiven en dan wijzigt dus in het algemeen ook de andere raakvector. Dat is ook wel begrijpelijk omdat een 2e graads kromme slechts 3 vrije parameters heeft. En we kunnen dus niet onafhankelijk van elkaar zowel de eindpunten als de raakvectoren in de eindpunten voorschrijven.
Dit is wel mogelijk als we een graad hoger gaan zitten: de cubische Bezier kromme heeft vier controlepunten, p_0 tot en met p_3 en daarvan bepalen p_0 en p_3 de eindpunten (als we, zoals gewoonlijk, t tussen 0 en 1 laten varieren) en $p_1 - p_0$ bepaalt de raakvector in p_0 en $p_3 - p_2$ bepaalt de raakvector in p_3. Door dat extra controlepunt hebben we een veel vrijere controle over de vorm van de kromme. Toch, zoals alle Bezierkrommen, heeft ook de cubische Bezierkromme precies zo'n zelfde opdeel-algoritme als zijn neven van lagere graad:

$$p_0 \rightarrow p_0 \tag{24}$$

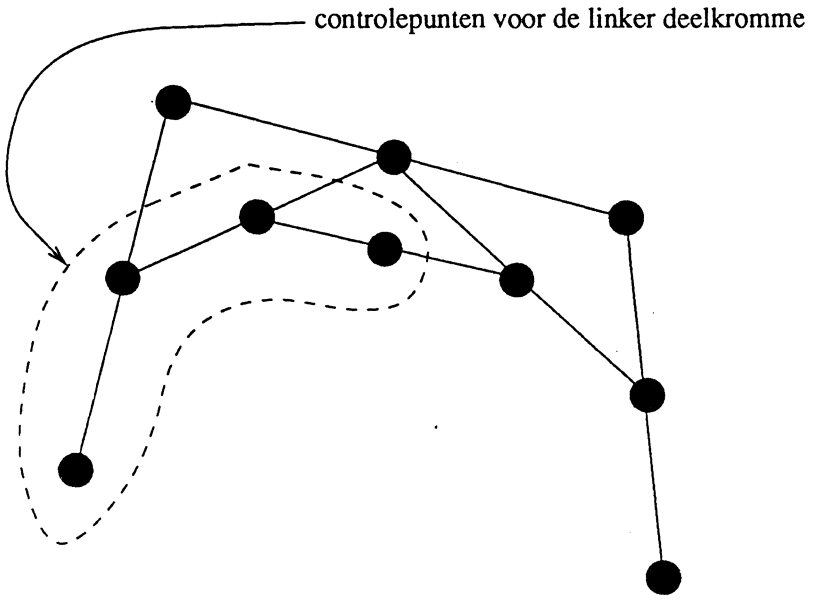
$$p_1 \rightarrow \frac{p_0 + p_1}{2} \tag{25}$$

$$p_2 \rightarrow \frac{p_0 + 2p_1 + p_2}{4} \tag{26}$$

$$p_3 \rightarrow \frac{p_0 + 3p_1 + 3p_2 + p_3}{8} \tag{27}$$

$$t \rightarrow 2t \tag{28}$$

en analoog voor de rechterhelft. Meetkundig komt dit overeen met de volgende opdeelconstructie (zie Figuur 7). Cubische krommen zijn tegenwoordig een van de meest gebruikte zgn. spline krommen, door hun vrijwel ideale compromis van lage graad (hogere graads polynomen zijn duur om uit te rekenen en geven soms aanleiding tot numerieke problemen - het opdeel-algoritme is weliswaar stabiel, maar het is niet altijd de snelste manier om een kromme te tekenen) en toch een intuitief stel randvoorwaarden. Omdat rationale krommen nog het bijkomende voordeel hebben dat ze zowel kegelsneden kunnen representeren als ook gewone polynomiale (Bezier-) splines, worden echter in de meest moderne teken- en ontwerppaketten vaak rationale krommen gebruikt.



Figuur 7: Controlepunten van de gehele kromme en controlepunten van de linker deelkromme voor een cubische Bezier kromme.

6. Generalisatie

Als we de controlepunten van de cubische Bezier kromme voorstellen als een vector in een abstracte 4-dimensionale ruimte, $P = (p_0, p_1, p_2, p_3)^T$, kunnen we het opdeelalgoritme schematisch opschrijven als $P_{left} = MP$ en $P_{right} = NP$ met

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \\ 1/8 & 3/8 & 3/8 & 1/8 \end{pmatrix}$$

$$N = \begin{pmatrix} 1/8 & 3/8 & 3/8 & 1/8 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In dit hoofdstuk zullen we eens nagaan wat er gebeurt als we voor de matrices M en N andere keuzen doen. Het ligt voor de hand dat we dan een veel uitgebreidere familie krommen kunnen voortbrengen, want we hebben maar liefst 32 parameters waar we aan kunnen draaien. Toch zullen de meeste keuzen voor de coefficienten in M en N weinig interessants opleveren; bijvoorbeeld omdat de herhaalde toepassing van M en N op P niet convergeert, omdat er

gaten in de resulterende kromme ontstaan, en dergelijke. In het navolgende zullen we een aantal voor de hand liggende eisen opstellen om de vrijheid in de 32 parameters geleidelijk terug te brengen waardoor we een interessante subklasse van krommen overhouden.

6.1. Eindpunten
Als we willen dat de eindpunten p_0 en p_3 voor elke keuze van P geinterpoleerd worden door de kromme, moet

$$M_{0*} = (1,0,0,0) \tag{29}$$

$$N_{3*} = (0,0,0,1) \tag{30}$$

Hier betekent het sterretje een lopende index, dus we hebben respectievelijk de bovenste rij van M en de onderste rij van N voorgeschreven. Het aantal vrij te kiezen parameters reduceert van 32 naar 24.

6.2. Raaklijnen in de eindpunten
Wil de resulterende kromme ooit differentieerbaar kunnen zijn, dan moet het punt dat in de plaats komt van p_1 op de rechte p_0p_1 liggen, en het punt dat in de plaats komt van p_2 op de rechte p_2p_3. Voor de matrices betekent dit:

$$M_{1*} = (M_{10}, M_{11}, 0, 0) \tag{31}$$

$$N_{2*} = (0, 0, N_{22}, N_{23}) \tag{32}$$

Het aantal vrij te kiezen parameters reduceert van 24 naar 20.

6.3. Continuiteit
Wil de resulterende kromme ooit continu kunnen zijn, dan moet het punt dat in de plaats komt van p_3 van de linker deelkromme gelijk zijn aan het punt dat in de plaats komt van p_0 van de rechter deelkromme. Voor de matrices M en N betekent dit:

$$M_{3*} = N_{0*} \tag{33}$$

Het aantal vrij te kiezen parameters reduceert van 20 naar 16.

6.4. Affiene invariantie
Als de rijsom van elke rij uit M en N gelijk is aan 1 zijn de resulterende deelkrommen affien invariant; dat betekent dat, op dezelfde wijze als voor de kromme in zijn geheel, affiene transformaties op de controlepunten een affiene transformatie van de kromme tot gevolg heeft. Deze eis luidt:

$$\Sigma_{i=0}^{3} M_{ij} = 1, i = 0, \cdots, 3 \tag{34}$$

$$\Sigma_{i=0}^{3} N_{ij} = 1, i = 0, \cdots, 3 \tag{35}$$

Omdat de eerste rij van M en de laatste rij van N hier al aan voldoen, en de eerste rij van N al gelijk is aan de laatste rij van M, reduceert het aantal vrije parameters verder tot 11.

6.5. Reflectie invariantie
De cubische Bezier krommen hebben de eigenschap dat de vorm van de kromme niet verandert als de controlepunten in de omgekeerde volgorde genummerd worden. Deze invariantie kunnen we uitdrukken door te eisen:

$$M_{ij} = N_{3-i,3-j}, i = 0, \cdots, 3; j = 0, \cdots, 3. \tag{36}$$

Daardoor is de matrix N compleet gedefineerd door matrix M. Omdat N nog 4 vrije parameters over had, reduceert het aantal vrije parameters tot 7.

6.6. Raaklijn in het middelpunt
Wil de resulterende kromme ooit differentieerbaar zijn, dan moeten de nieuwe punten p_2 en p_3 van de linker deelkromme met de nieuwe punten p_0 en p_1 van de rechter deelkromme colineair zijn. Voor een nader te bepalen parameter χ, ongelijk 0, geeft dit:

$$(M_{21} + M_{22} + M_{23} - M_{31} - \frac{1}{2}) = \chi(M_{23} + M_{31} - \frac{1}{2}) \tag{37}$$

$$(M_{31} - M_{21}) = \chi(M_{22} - M_{32}) \tag{38}$$

$$(M_{32} - M_{22}) = \chi(M_{21} + M_{31}) \tag{39}$$

$$(\frac{1}{2} - M_{31} - M_{23}) = \chi(\frac{1}{2} - M_{21} - M_{22} - M_{23} + M_{31}) \tag{40}$$

Uit vgl. 38 volgt

$$M_{21} = M_{31} + \chi(M_{31} - M_{22}), \tag{41}$$

en als we 41 gebruiken in 37, 39, en 40 volgt

$$(M_{23} - M_{22})(1 - \chi) = \frac{1}{2}(1 - \chi) \tag{42}$$

dus $\chi = 1$ of $M_{23} = M_{22} + \frac{1}{2}$;

$$M_{31}(1 - \chi^2) = M_{22}(1 - \chi^2) \tag{43}$$

dus $|\chi| = 1$ of $M_{31} = M_{22}$;

$$M_{31}(\chi^2 - 1) + M_{23}(\chi - 1) - M_2 2\chi(\chi - 1) - \frac{1}{2}(\chi - 1) = 0 \tag{44}$$

dus $\chi = 1$ of $M_{31}(\chi + 1) + M_{23} - M_{22}\chi - \frac{1}{2} = 0$. Uiteindelijk volgt, voor het geval $|\chi|$ ongelijk is aan 1:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 - M_{11} & M_{11} & 0 & 0 \\ \frac{1}{2} - 3M_{22} & M_{22} & M_{22} & M_{22} + \frac{1}{2} \\ \frac{1}{2} - M_{22} & M_{22} & M_{22} & \frac{1}{2} - M_{22} \end{pmatrix}$$

en dat geeft nog slechts 3 vrije parameters: χ, M_{11} en M_{22}. Maar met vgl. 44, $M_{22} = 0$, ontstaat het ontaarde geval waar de nieuwe p_2 en de nieuwe p_3 beiden gelijk zijn aan $\frac{p_0+p_3}{2}$. Dit geval zullen we verder niet beschouwen.

We merken nog wel terloops op dat met $\chi = -1$ in elk punt de raaklijn van richting omkeert. Inderdaad: $\chi = -1$ betekent dat overal de raaklijn in het middelpunt tegengesteld is aan p_0p_1, hoe klein $|p_0p_1|$ ook is. Het is duidelijk dat de kromme in dat geval nergens differentieerbaar is. We zien dus dat de familie van krommen die op deze manier voortgebracht wordt zowel polynomiale krommen bevat als fractal-achtige krommen. In het vervolg zullen we uitsluitend het geval $\chi = 1$ beschouwen. Dat betekent dat niet alleen de raaklijn in het middelpunt vanaf de linker deelkromme dezelfde is als vanaf de rechter deelkromme, maar zelfs de 'snelheid' waarmee het middelpunt van links en van rechts benaderd worden zijn gelijk. We hebben dan 4 vrije parameters over. Als we verder A voor M_{11}, B voor M_{31}, C voor M_{23} en D voor M_{22} schrijven, krijgen we de matrix M:

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1-A & A & 0 & 0 \\ 1-2B-C & 2B-D & D & C \\ \frac{1}{2}-B & B & B & \frac{1}{2}-B \end{pmatrix}$$

Deze matrix levert voor $A+\frac{1}{2}$, $B = \frac{3}{8}$, $C = 0, D = \frac{1}{4}$ weer precies het recursieve opdeelschema voor de cubische Bezier kromme op.
Ofschoon de bovenstaande 4 parameters een overzichtelijk kleine groep vormen, is hun meetkundige interpretatie niet onmiddellijk duidelijk. Laten we nog wat gevallenonderscheid doen op de richting van de nieuwe p_2p_3.

6.6.1. Parallel aan p_2p_1
Dit levert op, voor een of andere reele α:

$$\begin{aligned} &(\frac{1}{2}-B)p_0 + Bp_1 + Bp_2 + (\frac{1}{2}-B)p_3 \\ -(1-2B-C)p_0 &- (2B-D)p_1 - Dp_2 - Cp_3 \\ &= \alpha(p_2 - p_1). \end{aligned}$$

Wil dit kunnen gelden voor willekeurige $p_0, \cdots, p_3$ dan moet $C + B = \frac{1}{2}$, maar aan deze conditie wordt zelfs niet voldaan door de cubische Bezier kromme.

6.6.2. Parallel aan p_3p_0
Dit levert op, voor een of andere reele β:

$$\begin{aligned} &(\frac{1}{2}-B)p_0 + Bp_1 + Bp_2 + (\frac{1}{2}-B)p_3 \\ -(1-2B-C)p_0 &- (2B-D)p_1 - Dp_2 - Cp_3 \\ &= \beta(p_3 - p_0). \end{aligned}$$

Wil dit kunnen gelden voor willekeurige $p_0, \cdots, p_3$ dan moet $B = A$, maar ook aan deze conditie wordt zelfs niet voldaan door de cubische Bezier kromme.

6.6.3. In het vlak opgespannen door p_2p_1 en p_3p_0
Dit levert op, voor een of andere reele α en β:

$$(\frac{1}{2}-B)p_0 + Bp_1 + Bp_2 + (\frac{1}{2}-B)p_3$$

$$-(1-2B-C)p_0-(2B-D)p_1-Dp_2-Cp_3 = \alpha(p_2-p_1)+\beta(p_3-p_0).$$

Wil dit kunnen gelden voor willekeurige $p_0,\cdots,p_3$ dan moet

$$B-D=\alpha \tag{45}$$

$$\frac{1}{2}-B-C=\beta. \tag{46}$$

Dit legt geen restrictie op aan krommen van deze familie; voor de cubische Bezier kromme geldt $\alpha=\beta=\frac{1}{8}$. Een parameterisatie in termen van α en β is echter intuitief duidelijker omdat dan de twee bijzondere gevallen uit de twee voorgaande paragrafen ontstaan door een van de twee parameters, α of β gelijk aan nul te maken. Dus wij suggereren nu een parameterset

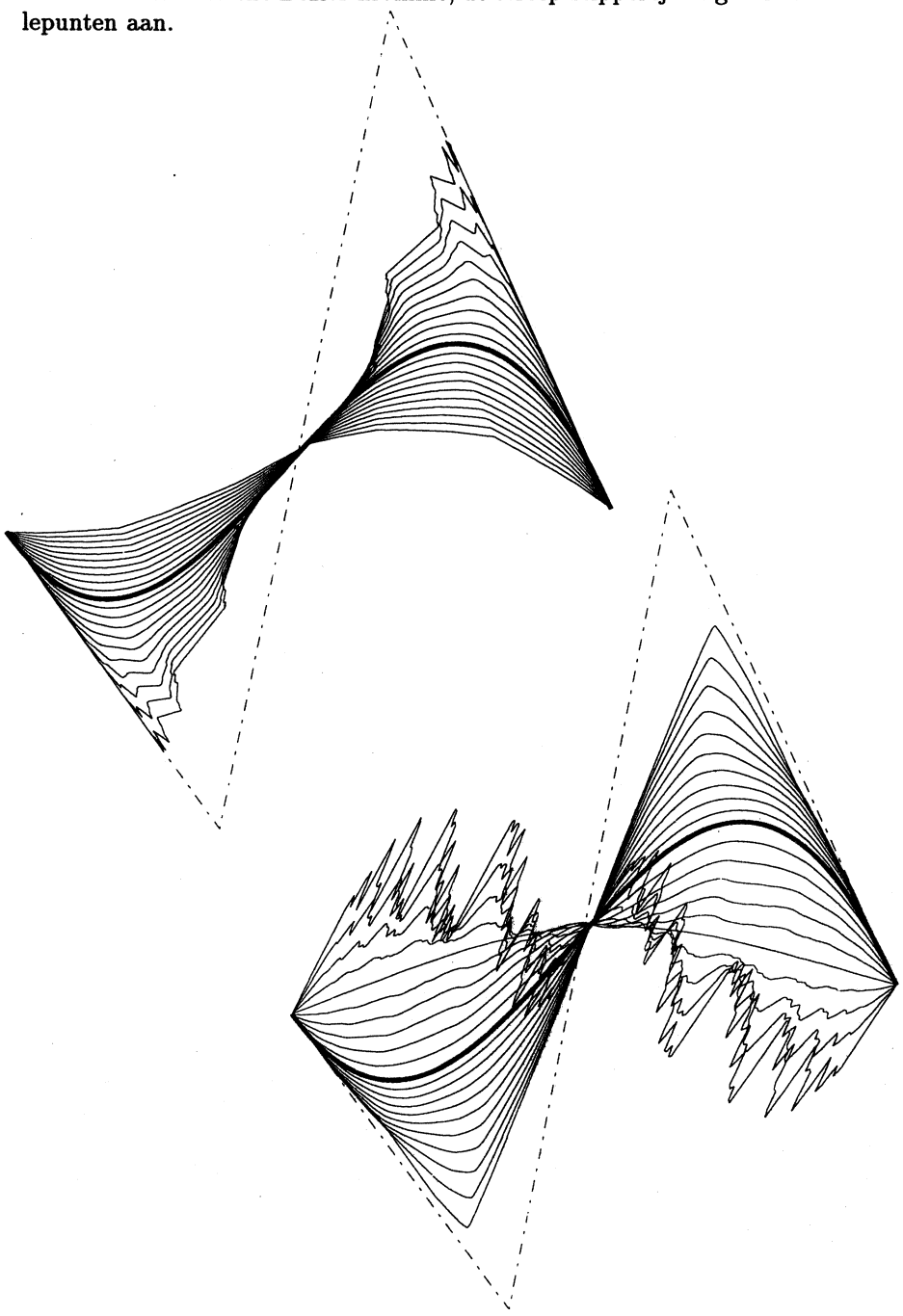
$$(A,B,\alpha,\beta) \tag{47}$$

waarbij $C=\frac{1}{2}-B-\beta$ en $D=B-\alpha$. De interpretatie van de nieuwe parameterset is als volgt. De parameter A bestuurt het gedrag van de kromme in de buurt van de eindpunten. Kleine waarden van A zorgen ervoor dat de kromme 'snel' de richting van de raaklijnen, p_0p_1 en p_2p_3, verlaat, terwijl grote A zorgt dat de raaklijnen 'lang' gevolgd worden. De parameter B bestuurt het gedrag van de kromme in de buurt van het middelpunt. Dit middelpunt ligt ergens op de lijn tussen het gemiddelde van p_1 en p_2 en het gemiddelde van p_0 en p_3. Kleine waardes zorgen ervoor dat het middelpunt dicht bij het gemiddelde van p_0p_3 ligt, en grote waardes juist meer bij het gemiddelde van p_1p_2. De overige twee parameters besturen de orientatie van de nieuwe zijde p_2p_3 met betrekking tot de zijden p_0p_3 en p_1p_2. De bijgaande figuren (8, 9, 10, 11, 12, 13, 14) laten de invloed van de parameters in kwalitatieve zin zien.

7. Afsluiting

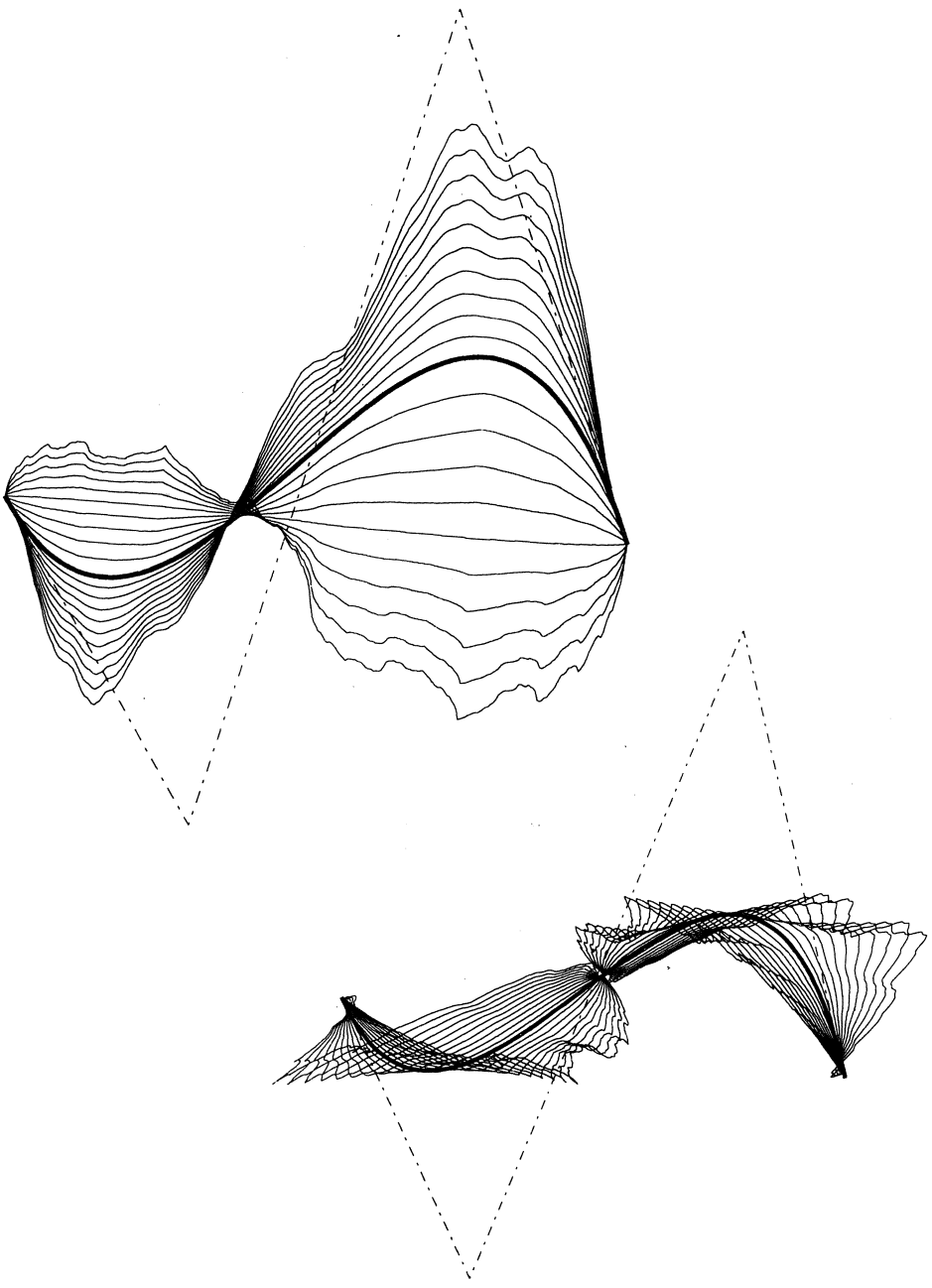
De wereld van gladde krommen en de wereld van chaotische, fractal-achtige krommen lijken heel ver van elkaar af te liggen. Toch, door krommen op een procedurele (algoritmische) manier te definieren kunnen we een familie van krommen introduceren die exemplaren van beide soorten bevat; sterker nog, door op continue wijze een reeel-waardige parameter te varieren kunnen we we krommen continue laten verlopen tussen 'glad' en 'chaotisch'.
Op zichzelf zijn chaotische meetkundige objecten al heel lang bekend, en sinds het werk van Mandelbrot in de 70-er jaren, ook heel populair door de fascinerend mooie computerafbeeldingen. Door chaotische meetkudndige figuren te beschrijven met controlepunten, zoals we dat ook met splines doen, wordt er echter een dimensie (sic!) aan fractals toegevoegd: hun globale vorm wordt namelijk bestuurbaar of controleerbaar. Door eerlijk te delen wordt de chaos getemd.

Figuur 8: Afhankelijkheid van parameter A. A loopt van 0 naar 1. De dikke kromme is de cubische Bezier-kromme; de streep-stippellijnen geven de controlepunten aan.



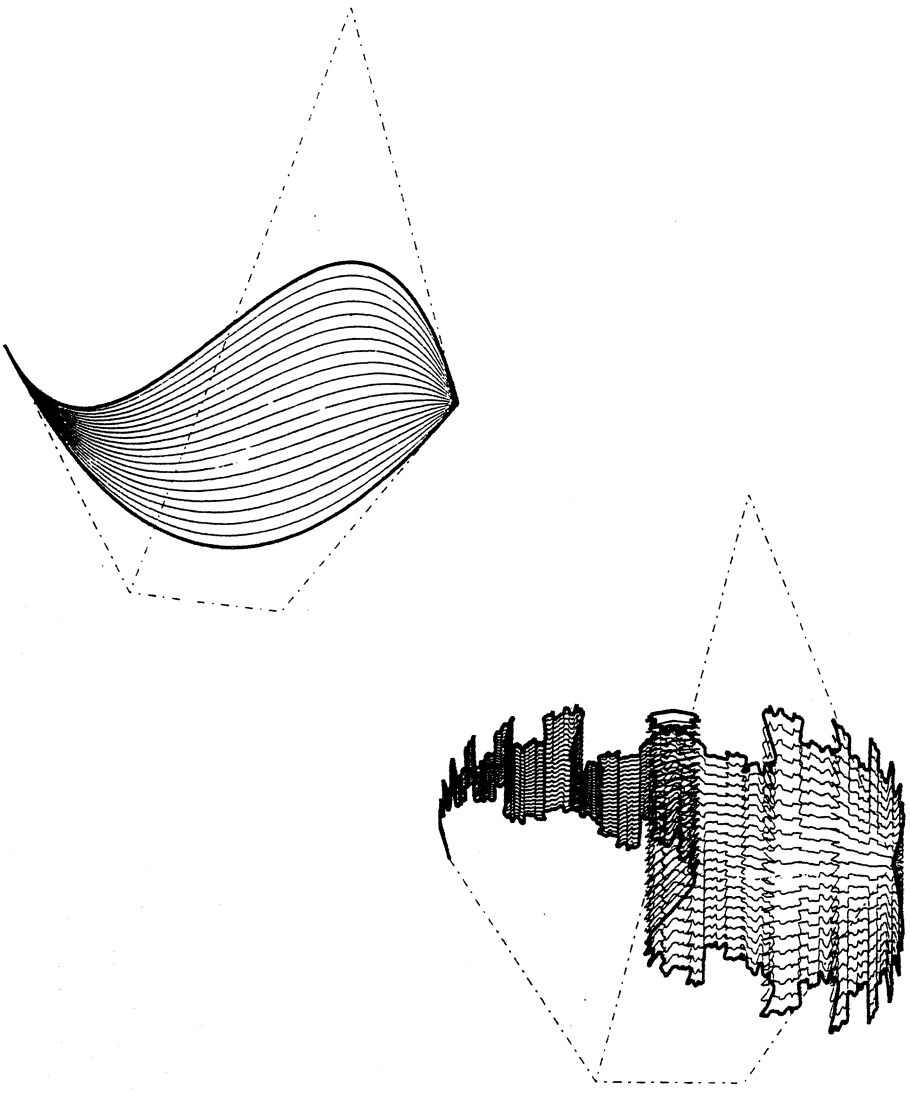
Figuur 9: Afhankelijkheid van parameter B. B loopt van -0.4 naar 0.8.

Figuur 10: Afhankelijkheid van parameter α. α loopt van -0.5 naar 0.5.

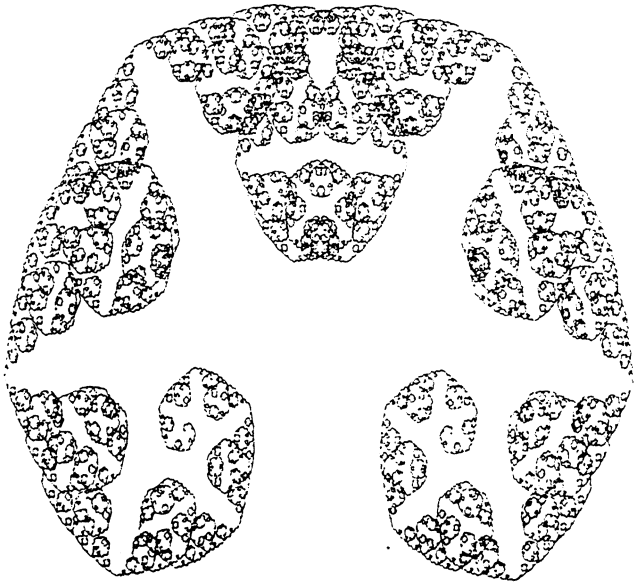


Figuur 11: Afhankelijkheid van parameter β. β loopt van -0.5 naar 0.7.

Figuur 12: Afhankelijkheid van de controlepunten in het geval van een gladde Bezier-kromme; de streep-stippellijnen geven de controlepunten aan in de twee uiterste situaties.



Figuur 13: Afhankelijkheid van de controlepunten in het geval van een fractal-achtige kromme.



Figuur 14: Voorbeeld van een meer 'regelmatige' fractal.

Sommen van Kwadraten

F. van der Blij

0. INTRODUCTIE

Het onderstaande is geen wetenschappelijke verhandeling, maar een vertelling. Stellingen worden zonder bewijs vermeld, misschien wordt soms wel eens een enkele voorwaarde vergeten of een bizonder geval niet als uitzondering vermeld. Deze bijdrage kan op verschillende manieren gelezen worden. Wellicht is het verstandig eerst alleen § 7 te lezen. Maar men kan natuurlijk ook gewoon bij § 1 beginnen en dan verder lezen. Men kan na § 1 direct § 7 lezen en daarna desgewenst nog de overige §§. Men kan ook direct met § 7 beginnen en daarna § 1 òf § 1 tot en met § 7 lezen.

In de vele leerboeken en monografieën over ons onderwerp is alles met bewijzen en strenge definities te vinden. Maar de lectuur van deze boeken vraagt wel enige voorzichtigheid omdat er nogal wat terminologische verwarring is. Woorden als gehele kwadratische vorm, determinant van een kwadratische vorm, discriminant van een kwadratische vorm, gewicht van een modulaire vorm zijn niet bij alle schrijvers op dezelfde manier gedefinieerd. Ik noem hier even twee boeken die ik bijzonder aanbeveel:

J.W.S. CASSELS: Rational Quadratic Forms. London 1978.
(met een literatuurlijst van 12 pagina's)

JEAN-PIERRE SERRE: Cours d'arithmétique. Paris 1970.
= A course in Arithmetic. Berlin 1973.

waarmee ik echter niets ten nadele van andere boeken over dit onderwerp wil zeggen. Bij het bladeren in de verschillende boeken moet men dus wel op de daar geldende definities letten.

1. INLEIDING

Een kwadratische vorm in n variabelen definiëren we door

$$Q(\boldsymbol{x}) = \sum a_{ij} x_i x_j, \quad a_{ij} = a_{ji}, \quad 1 \leq i \leq n;\ 1 \leq j \leq n.$$

Soms zullen we matrixnotatie gebruiken, de getransponeerde van een matrix of vector geven aan door een bovenindex ^{T}. Aan de kwadratische vorm Q voegen we de matrix

$$\boldsymbol{A} = (a_{ij})$$

toe, dan geldt dus

$$Q(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}.$$

Laat $\boldsymbol{C}$ een inverteerbare $\boldsymbol{n}$ bij $\boldsymbol{n}$ matrix zijn. Door

$$\boldsymbol{x} = \boldsymbol{C}\boldsymbol{y}$$

wordt een transformatie van de kwadratische vorm gedefinieerd door

$$P(\boldsymbol{y}) = \boldsymbol{y}^T \boldsymbol{C}^T \boldsymbol{A} \boldsymbol{C} \boldsymbol{y}.$$

Kwadratische vormen die door een inverteerbare transformatie in elkaar overgevoerd kunnen worden noemen we equivalent. Soms stellen we nog de extra eis dat de determinant van de matrix $\boldsymbol{C}$ gelijk moet zijn aan 1.
Beschouwen we kwadratische vormen over de rationale of over de reële getallen dan is iedere kwadratische vorm te transformeren in een diagonaal-vorm dat wil zeggen in een kwadratische vorm

$$\Sigma a_i x_i^2 \qquad 1 \leq i \leq n$$

In het algemeen kan dit voor ieder lichaam waarin $1 + 1 \neq 0$.
Bij kwadratische vormen met gehele rationale coëfficiënten kan dit niet altijd met een gehele transformatie gebeuren, een voorbeeld is de vorm

$$x^2 + xy + y^2.$$

Er is al lange tijd discussie hoe een kwadratische vorm met gehele rationale coëfficiënten te definiëren. We kunnen in bovenstaande formule eisen dat alle coëfficiënten a_{ij} geheel zijn, we kunnen ook eisen dat alle coëfficiënten a_{ii} en $(a_{ij} + a_{ji}) = 2a_{ij}$ met $i \neq j$ geheel zijn. Aan de kwadratische vorm voegen we een determinant of discriminant toe. Bij rationale of reële vormen worden deze grootheden bij transformaties met een kwadraat vermenigvuldigd. Bij een kwadratische vorm

$$Q(\boldsymbol{x}) = ax^2 + bxy + cy^2 \text{ met gehele } a, b \text{ en } c$$

zouden we als determinant / discriminant kunnen definiëren:

$ac - \frac{1}{4}b^2$ of $4ac - b^2$ of (denk aan de vierkantsvergelijking) $b^2 - 4ac$.

We definiëren voor vormen met een even aantal variabelen de discriminant door

$$disc(Q) = (-1)^{\frac{1}{2}n} det(\frac{\partial^2}{\partial x_i \partial x_j} Q(x)),$$

hetgeen voor de binaire vorm juist $b^2 - 4ac$ levert.
Door deze definitie heeft de vorm xy discriminant 1 en $x^2 + y^2$ de discriminant -4. We willen ons in deze vertelling speciaal bezighouden met de vraag naar de gehele oplossingen van de vergelijking

$$Q(\mathrm{x}) = N.$$

Enkele vertrouwde voorbeelden zijn

$$x^2 + y^2 - z^2 = 0$$

(driehoeken van Pythagoras) en

$$x^2 - 2y^2 = \pm 1$$

(rationale benaderingen van wortel 2) en

$$x^2 + y^2 = N$$

(het aantal roosterpunten op een cirkel met straal wortel N).
Wij zullen ons in het vervolg als regel bezighouden met positief definiete kwadratische vormen in n variabelen, dat wil zeggen vormen die door een reële transformatie overgevoerd kunnen worden in de som van n kwadraten. We zullen ons als regel beperken tot vormen met gehele coëfficiënten.

2. Binaire kwadratische vormen

We bestuderen nu positief definiete kwadratische vormen

$$ax^2 + bxy + cy^2$$

met gehele getallen a, b en c en discriminant $D = b^2 - 4ac$. We merken op $D < 0$. De equivalentieklassen onder gehele transformaties met determinant 1 noemen we kortweg klassen. De reductietheorie van kwadratische vormen geeft methoden om in iedere klasse een representant te kiezen. Bij gegeven discriminant D is er steeds een klasse te vinden met $-a < b \leq a \leq c$, terwijl $b \geq 0$ als $a = c$. We merken nog op dat $ax^2 + bxy + cy^2$ en $ax^2 - bxy + cy^2$ in elkaar te transformeren zijn met een gehele transformatie met determinant -1. De aantallen gehele oplossingen van

$$ax^2 + bxy + cy^2 = N \text{ en van } ax^2 - bxy + cy^2 = N$$

zijn evengroot.
Enkele voorbeelden:

Voor $D = -3$ is er slechts één klasse, namelijk die waartoe $x^2 + xy + y^2$ behoort.
Voor $D = -4$ is er slechts één klasse, namelijk die waartoe $x^2 + y^2$ behoort.
Voor $D = -23$ zijn er drie klassen, namelijk die waartoe $x^2 + xy + 6y^2, 2x^2 + xy + 3y^2$ en $2x^2 - xy + 3y^2$ behoren.

De oplossingen van $x^2 + y^2 = N$ zijn eenvoudig te vinden door gebruik te maken van de getaltheorie van de gehele complexe getallen van Gauss.
Het rationale priemgetal 2 is te ontbinden als $-i(1 + i)^2$, de rationale priemgetallen congruent 3 modulo 4 zijn ook priemgetallen van Gauss, de rationale priemgetallen congruent 1 modulo 4 zijn het product van twee toegevoegd complexe priemgetallen van Gauss: $p = (a + bi)(a - bi)$.
In de ring van de gehele complexe getallen van Gauss gelden de regels van de bekende rekenkunde, er is een eenduidige ontbinding in priemfactoren, enzovoorts. Daarmee is het aantal oplossingen te bepalen uit de factorisatie van N

in complexe priemfactoren. Met deze rekenkundige eigenschappen hangt ook het bestaan van de identiteit

$$(a^2 + b^2)(u^2 + v^2) = (au - bv)^2 + (av + bu)^2$$

samen.
We vinden voor het aantal oplossingen van $x^2 + y^2 = N$:

$$r(N) = 4\Sigma \left(\frac{-4}{d}\right) \text{ de som over alle positieve delers } d \text{ van } N.$$

Hierin is $(\frac{-4}{d})$ het kwadratische restsymbool; in dit speciale geval te definiëren als 0 voor even waarden van d en als $(-1)^{\frac{1}{2}(d-1)}$ voor oneven waarden van d.
Voor een willekeurige gehele kwadratische vorm $ax^2 + bxy + cy^2$ kunnen we ook in het lichaam dat uit de rationale getallen ontstaat door adjunctie van wortel D een ring van gehelen definiëren, maar als regel is er geen eenduidige ontbinding in priemfactoren, de definitie van de grootste gemene deler vraagt extra aandacht. Door de invoering van idealen (een soort remplaçant voor de grootste gemene deler) is wel een theorie op te bouwen, maar het aantal oplossingen van de betreffende vergelijking is niet steeds te vinden. Dit hangt samen met het feit dat de boven vermelde identiteit over de sommen van kwadraten nu een essentieel andere vorm krijgt.
In § 7 zullen we een voorbeeld uitvoeriger bespreken.

3. KWADRATISCHE VORMEN MET VIER OF MEER VARIABELEN

Bij de som van vier kwadraten zou men de kwaternionen van Hamilton kunnen proberen te gebruiken en bij willekeurige kwaternaire vormen met kwadratische discriminant een generalisatie van deze kwaternionen. In het eerste geval gebruikt men de identiteit

$$(a^2 + b^2 + c^2 + d^2)(u^2 + v^2 + w^2 + z^2) =$$

$$(au - bv - cw - dz)^2 + (av + bu + cz - dw)^2 +$$

$$(aw - bz + cu + dv)^2 + (az + bw - cv + du)^2.$$

In het algemene geval is de theorie vrij gecompliceerd.
Er bestaat een identiteit voor de sommen van acht kwadraten en daarmee verbonden is de theorie van de octaven (een niet-commutatief, niet-associatief getalsysteem). Daarmee zou in speciale gevallen iets over aantal oplossingen van kwadratische vergelijkingen in acht geheeltallige onbekenden gezegd kunnen worden.
Voor het analoge geval van zes variabelen is echter zo'n theorie niet voorhanden.
In het verleden zijn formules voor het aantal manieren waarop een getal als de som van vier kwadraten van gehele getallen te schrijven is op geheel andere manieren afgeleid.

We schetsen een analytisch bewijs voor de formule voor het aantal representaties van een getal als de som van vier kwadraten in de trant van een bewijs van Jacobi (1804–1851).
We gaan de analyse gebruiken en definiëren

$$\theta(z) = \Sigma q^{n^2},\ -\infty < n < \infty,\ q = e^{\pi i z},\ |q| < 1.$$

Laat $r_4(N)$ het aantal gehele oplossingen van

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = N$$

voorstellen. Dan geldt

$$\theta^4(z) = \Sigma r_4(N) q^N, \quad 0 \leq N < \infty.$$

We voeren nu nog in de functie

$$f(z) = 1 + 8\Sigma \frac{nq^n}{1-q^n} - 8\Sigma \frac{4nq^{4n}}{1-q^{4n}}, \quad 1 \leq n < \infty.$$

Jacobi bewees nu

$$\theta^4(z) = f(z).$$

We proberen enkele lijnen te schetsen van een bewijs van deze formule. (dit bewijs is volledig te vinden in Chapter X van K. CHANDRASEKHARAN: *Elliptic Functions*, Berlin 1985). We introduceren een functie

$$\eta(z) = q^{1/12}\Pi(1-q^{2n}), \quad 1 \leq n < \infty,$$

Hiervoor geldt

$$\frac{d}{dz}(\log \eta(z)) = \frac{\pi i}{12} - 2\pi i \Sigma \frac{nq^{2n}}{1-q^{2n}}.$$

Wat simpel manipuleren geeft

$$f(z) = \frac{8i}{\pi} \frac{d}{dz} \left[\log \eta(\frac{z}{2}) - \log \eta(2z)\right].$$

Voor de êta functie bestaan klassieke transformatieregels

$$\eta\left(\frac{-1}{z}\right) = \sqrt{\frac{z}{i}}\eta(z), \quad \eta(z+1) = e^{\frac{\pi i}{12}}\eta(z).$$

Daaruit volgt $f(-1/z) = -z^2 f(z)$.

De functie

$$K(z) = \frac{\theta^4(z) - f(z)}{\eta^4(z)}$$

voldoet aan

$$K(z+2) = hK(z) \text{ met } h^3 = 1,$$

$$K(-1/z) = K(z).$$

Dit gedrag voert tot de invariantie van de absolute waarde van de functie K onder de groep van gebroken lineaire transformaties

$$z \to \frac{\alpha z + \beta}{\gamma z + \delta},\ \alpha, \beta, \gamma, \delta \in \mathbb{Z},\ \alpha\delta - \beta\gamma = 1$$

die door de transformaties $z \to z + 2$ en $z \to -1/z$ wordt voortgebracht.
Wanneer we punten in de bovenste helft van het complexe vlak die door transformaties uit deze groep in elkaar overgaan identificeren verkrijgen we een complexe variëteit waarop de functie K gedefinieerd is. Uit de structuur van deze variëteit en het analytische gedrag van deze functie kan nu worden afgeleid dat hij constant 0 is. (een vergelijkbare stelling zegt dat een overal op de complexe bol holomorfe functie constant is)
Dus geldt $\theta^4(z) = f(z)$.
De coëfficiënten van $\theta^4(z)$ als machtreeks in q zijn de aantallen $r_4(N)$, de coëfficiënten van $f(z)$ als machtreeks in q zijn te vinden door de breuken

$$\frac{nq^n}{1-q^n} \text{ en } \frac{4nq^{4n}}{1-q^{4n}}$$

in machtreeksen te ontwikkelen.
We vinden zo de formules

$r_4(N) = 8\Sigma d$ gesommeerd over alle delers van N als N oneven is

$r_4(N) = 24\Sigma d$ gesommeerd over alle oneven delers van N als N even is.

De gevallen van zes, acht, tien, twaalf kwadraten zijn met analoge middelen te onderzoeken, alleen zes en acht kwadraten geven nog enigzins eenvoudige formules.

4. De analytische methoden van Hardy en Littlewood

We gaan opnieuw uit van de theta functie en bestuderen deze nu in het bijzonder in de omgeving van een punt $z = 2a/b$.
We schrijven

$$\varepsilon = exp(2\pi i a/b) \text{ met gehele } a \text{ en natuurlijke } b.$$

en

$$z = -\frac{i}{\pi}\log r + 2\frac{a}{b}$$

zodat

$$e^{\pi i z} = r\varepsilon$$

en

$$\begin{aligned}
\theta(z) &= 1 + 2\Sigma r^{n^2}\varepsilon^{n^2}, && 1 \le n < \infty,\ 0 < r < 1,\\
&= 1 + 2\Sigma r^{(kb+j)^2}\varepsilon^{(kb+j)^2}, && 0 < j \le b,\ 0 \le k < \infty,\\
&= 1 + 2\Sigma_j\, \varepsilon^{j^2}\Sigma_k\, r^{(kb+j)^2}, && 0 < j \le b,\ 0 \le k < \infty.
\end{aligned}$$

We schrijven voor de som van Gauss $\Sigma_j \varepsilon^{j^2}$ voortaan $G(a/b)$.
We gaan het gedrag van deze functie voor r klimmend naar 1 bekijken. Voor $r = 1$ divergeert de reeks. We schatten de som van de reeks met de overeenkomstige integraal:

$$\sum_k r^{(kb+j)^2} \approx \int_0^\infty r^{(xb+j)^2} dx \approx \frac{1}{2} b^{-1} \pi^{\frac{1}{2}} (\log 1/r)^{-\frac{1}{2}}.$$

Kombineren we alle opmerkingen dan lijkt het zinnig om $\theta^{2m}(z)$ in de buurt van het punt $2a/b$ te vergelijken met

$$\pi^m b^{-2m} G^{2m}(a/b)(log(1/r))^{-m}.$$

We vervangen nu nog , en als we er wat langer over nadenken is dat niet eens zo gek, (gezien het gedrag van deze functies in de buurt van $r = 1$) de factor

$$(\log(1/r))^{-m}$$

door

$$(\Gamma(m))^{-1} \Sigma n^{m-1} r^n, \quad 1 \leq n < \infty, \quad \Gamma(m) = (m-1)!$$

We strekken nu een som uit over alle breuken a/b en verkrijgen dan een functie, die in veel punten op de reële as eenzelfde (singulier) gedrag vertoont als de macht van de theta-functie.
Alles tezamen definiëren we de functie (op passende wijze sommerend over a, b en n)

$$\Theta_{2m}(z) = 1 + \Sigma \pi^m G^{2m}(a/b) b^{-2m} (\Gamma(m))^{-1} \Sigma n^{m-1} e^{\pi i (z - 2a/b)n}$$

Deze functie lijkt op $\theta^{2m}(z)$. We citeren nu even Hardy zelf. (wel een beetje vrij) (G.H. HARDY, *Ramanujan, Twelve lectures on subjects suggested by his life and work.* Cambridge 1940, Chapter IX: The representation of numbers as sums of squares)

... we may expect that $\Theta_{2m}(z)$ will mimic $\theta^{2m}(z)$ near all rational points $2a/b$. To say this is to say that $\Theta_{2m}(z)$ mimics $\theta^{2m}(z)$ very comprehensively, so comprehensively that there should be a very close relation between the coëfficiënts of the two functions. If this be so, then the way will be open at any rate to an approximate determination of $r_{2m}(N)$ ...

We moeten nu dus de coëfficiënten van $\Theta_{2m}(z)$ als machtreeks in q bepalen.
We beperken ons om extra complicaties in het vervolg te vermijden tot het geval dat m een viervoud is. Met enig klassiek rekenwerk vinden we hiervoor

$$\rho_{2m}(N) = \frac{\pi^m}{\Gamma(m)(1 - 2^{-m})\zeta(m)} \sigma^*_{m-1}(N)$$

Hierin is $\zeta(s)$ de zeta functie van Riemann, gedefinieerd door

$$\zeta(s) = \Sigma n^{-s} = \Pi(1 - p^{-s})^{-1}, \ 1 \leq n < \infty, \ p \text{ doorloopt de priemgetallen.}$$

Verder is

$$\sigma_m^*(N) = \sum_{d/N} d^m \qquad \text{als N oneven is ,}$$

$$\sigma_m^*(N) = \sum_{d/N, d \text{ even}} d^m - \sum_{d/N, d \text{ oneven}} d^m \qquad \text{als N even is.}$$

Nu leert een klassieke formule uit de analyse dat

$$\sum_{n=1}^{\infty} n^{m-1} e^{\pi i n w} = \frac{\Gamma(m)}{\pi^m} \sum_{k=-\infty}^{\infty} \frac{1}{(2k-w)^m}.$$

Gebruiken we deze in de formule voor $\Theta_{2m}(z)$ dan vinden we:

$$\Theta_{2m}(z) = 1 + \sum_{a,b,k} \frac{G^{2m}(\frac{a}{b})}{b^m} \frac{1}{[2(kb+a)-bz]^m}.$$

De som over a, b en k heeft de vorm van een Eisensteinreeks. De eenvoudigste Eisensteinreeks wordt gedefinieerd als een som over alle gehele b en c, niet beide nul,

$$E_{2g}(z) = \Sigma \frac{1}{(c+bz)^{2g}}.$$

De Eisensteinreeks E_{2g} is invariant onder de transformatie z naar $z+1$ en wordt vermenigvuldigd met z^{2g} onder de transformatie z naar $-1/z$. Hieruit is af te leiden dat de functie $\Theta^{2m}(z)$ zich als functie van z onder een ondergroep van de groep van alle gebroken lineaire transformaties gedraagt als de functie $\theta^{2m}(z)$. Op een manier analoog aan hetgeen we aan het einde van de vorige § schetsten kan nu bewezen worden dat voor $m = 4$ de twee functies gelijk zijn. Daarmee is dan de formule voor het aantal manieren waarop N als som van 8 kwadraten geschreven kan worden, gevonden. Voor grotere waarden van m zijn de functies niet meer gelijk en we vinden slechts benaderingen voor het aantal manieren waarop N als som van kwadraten te schrijven is. Het geval $m = 12$ is door Ramanujan (1887–1920) al onderzocht en we vermelden voor oneven waarden van N het resultaat:

$$r_{24}(N) = \frac{16}{691} \Sigma d^{11} + \frac{33152}{691} \tau(N) \qquad \text{som over alle delers van } N$$

waarin $\tau(N)$ wordt gedefinieerd door

$$\Sigma \tau(k) x^k = x \Pi (1-x^n)^{24}, \quad 1 \leq k < \infty, \quad 1 \leq n < \infty.$$

Het vreemde getal 691 in deze formule komt van de waarde van de zetafunctie in $s = 12$. De expliciete waarden van de zetafunctie voor even natuurlijke waarden van s wordt gegeven door:

$$\zeta(2m) = \frac{(2\pi)^{2m}}{2(2m)!} |B_{2m}|,$$

waarin B_{2m} de Bernoulli-getallen voorstellen, eenvoudig te berekenen uit de symbolische formule (exponenten als onder-indices op te vatten):

$$(B+1)^n = B^n,$$

en dus

$$B_1 = -\frac{1}{2}, B_3 = B_5 = B_7 = B_9 = B_{11} = B_{13} = \ldots = 0,$$

$$B_2 = 1/6,\ B_4 = -1/30,\ B_6 = 1/42,\ B_8 = -1/30,\ B_{10} = 5/66,$$

$$B_{12} = -691/2730,\ B_{14} = 7/6, \ldots$$

en hiermede is het optreden van het vreemde getal 691 in de formule voor $r_{24}(N)$ verklaard.

De methode van Hardy en Littlewood berust op het principe uit de complexe functietheorie dat de coëfficiënten van een machtreeks berekend kunnen worden door een geschikte contourintegraal. Voor de theta functie is de eenheidscirkel de natuurlijke grens. De te kiezen contour moet dus binnen deze cirkel liggen. Hardy en Littlewood construeren nu een samenstel van cirkelbogen, gebruik makend van geschikte met breuken samenhangende deelpunten op de eenheidscirkel als contour. Met behulp van schattingen voor de integralen vinden zij als schatting voor de coëfficiënten van machten van theta functies in principe dezelfde benadering als we boven met de heuristische methode vonden. Omdat de bewuste punten op de eenheidscirkel singuliere punten voor de integrand zijn, heeft deze methode ook wel de naam gekregen van de methode met de singuliere reeks. Deze methode is ook voor veel andere problemen uit de getaltheorie bruikbaar, verschillende toepassingen zijn te vinden in R.C. VAUGHAN, *The Hardy - Littlewood Method,* Cambridge 1981.

5. ALGEBRAÏSCH-ARITHMETISCHE THEORIE

Er is ook nog een heel andere theorie over de voorstelling van getallen door kwadratische vormen.
Als een gehele oplossing van de vergelijking

$$Q(x) = N$$

bestaat bestaan er natuurlijk voor ieder m oplossingen van de congruentie

$$Q(x) \equiv N(\bmod m).$$

Als oplossingen modulo iedere m bestaan zou er dan ook een gehele rationale oplossing bestaan? Het is duidelijk dat dit niet waar is; voor iedere waarde van m is er een niet-triviale oplossing van de congruentie

$$x^2 + y^2 + z^2 + w^2 \equiv 0 \pmod m$$

en het is duidelijk dat er geen niet-triviale gehele rationale oplossing kan bestaan. We zouden als extra eis kunnen stellen dat er ook een reële oplossing moet bestaan. Zou de stelling dan gelden?

Analoge vragen kunnen gesteld worden over de equivalentie van kwadratische vormen door transformaties met gehele coëfficiënten, rekenend modulo m en reëel rekenend. Bij deze vragen komt de vraag naar voren of klassen van onder gehele transformaties equivalente kwadratische vormen bepaalde invarianten bezitten, met behulp waarvan de equivalentie onderzocht kan worden. Klassieke onderzoekingen, we noemen die van Minkowski (1864–1909) geven zulke invarianten, die met congruenties modulo machten van priemgetallen samenhangen.
We merken even op dat de twee binaire kwadratische vormen $x^2 + 55y^2$ en $5u^2 + 11v^2$ rationaal in elkaar getransformeerd kunnen worden door

$$x = 5/4u - 11/4v \quad y = 1/4u + 1/4v$$

en het is duidelijk dat er geen gehele transformatie kan zijn die deze twee binaire vormen in elkaar overvoert.
Van belang is de stelling (Hasse 1923) dat als voor ieder gehele m de congruentie

$$Q(x) \equiv N(\bmod m)$$

oplosbaar is en er bovendien een reële oplossing van de vergelijking

$$Q(x) = N$$

bestaat er ook een rationale oplossing van deze vergelijking bestaat.
We kunnen de voorwaarde dat voor iedere m een oplossing van de congruentie bestaat verzwakken. Voor priemgetallen die niet in N en niet in de discriminant van Q opgaan eisen we de oplosbaarheid modulo dit priemgetal. Voor de eindig veel andere priemgetallen eisen we de oplosbaarheid modulo een geschikt gekozen macht van het priemgetal. Daaruit volgt vanzelf de oplosbaarheid modulo iedere m.
Deze voorwaarden modulo m of beter gezegd modulo machten van priemgetallen kunnen we sinds de invoering van de p-adische getallen door Hensel ook in termen van oplossingen in p-adische getallen formuleren. (de p-adische getallen zijn een completering van de rationale getallen met behulp van een speciale absolute waarde. Deze getallen vormen een lichaam, waarin zowel analyse als rekenkunde bedreven kan worden. Van belang is zich te realiseren dat alle lichamen van p-adische getallen en het reële lichaam de rationale getallen omvatten.) Kortgezegd volgt uit reële en voor ieder priemgetal p-adische representatie de rationale representatie. Maar dan behoeft er nog geen gehele representatie te bestaan. Een fundamenteel resultaat van SIEGEL (1935) geeft helderheid en kwantitatieve resultaten. We zullen kwadratische vormen verwant noemen als ze reëel en voor alle p ook p-adisch equivalent zijn. Dit is een equivalentierelatie, de equivalentieklassen noemen geslachten. Een geslacht omvat één of meer klassen die door de equivalentie met gehele transformaties met determinant 1 gedefinieerd zijn . Enerzijds beschouwt Siegel nu een gewogen gemiddelde van de representatie-aantallen van alle klassen in een vast geslacht van kwadratische vormen. We geven dit aan met $\mathcal{A}(Q, N)$.

Anderzijds is uit de aantallen oplossingen van de betreffende congruenties modulo opklimmende machten van p een genormaliseerd aantal p-adische oplossingen te construeren, we noemen dit aantal $a_p(Q,N)$ en tenslotte kan nog via integraalbeschouwingen een soort genormaliseerd aantal reële oplossingen $a_\infty(Q,N)$ gedefinieerd worden. De stelling van Siegel luidt nu (voor vormen met drie of meer variabelen)

$$\mathcal{A}(Q,N) = a_\infty(Q,N)\Pi a_p(Q,N)$$

Het rechterlid blijkt overeen te stemmen met de in de vorige § gedefinieerde benadering voor het aantal representaties (de singuliere reeks) in het geval van sommen van kwadraten. Omvat een geslacht slechts één klasse dan vinden we het exacte aantal oplossingen van de betreffende vergelijking. Dit is bijvoorbeeld het geval bij de representatie als sommen van 4, 6 en 8 kwadraten, bij 10 kwadraten zijn er meer klassen in het betreffende geslacht en vinden we dus slechts benaderingen voor of anders gezegd gemiddelden van deze aantallen. In een studie met als titel *Analytische Arithmetik der positiven quadratischen Formen* (1940) gaat HECKE in op eigenschappen van de resttermen bij de benaderingen van het aantal representaties. Fundamenteel is de verzameling van functies met een bepaald gedrag onder een ondergroep van de groep van de gebroken lineaire transformaties. We kwamen al de thetareeksen als voorbeelden tegen en evenzo de Eisensteinreeksen E. Nu blijkt een basis te bestaan voor de verzameling van deze functies, die bestaat uit Eisensteinreeksen en zogenaamde spitsvormen, dat zijn modulaire vormen met een voorgeschreven gedrag in enkele bijzondere punten in het complexe bovenhalfvlak. De coëfficiënten van de Eisensteinreeksen en van de spitsvormen hebben een zwak multiplicatief karakter. Meer precies gezegd als deze coëfficiënten $c(N)$ genoemd worden geldt

$$\sum_{N=1}^{\infty} \frac{c(N)}{N^s} = \prod_p \left(1 - \frac{c(p)}{p^s} + \frac{\varepsilon_p p^k}{p^{2s}}\right)^{-1} \quad \text{waarin } |\varepsilon_p| = 1.$$

We noemen dit product over alle priemgetallen een Euler-product, het is een generalisatie van het product voor de Riemann zeta functie.
Hecke voert in deze theorie de later naar hem genoemde T_n operatoren in, een werktuig dat we tot de functionaalanalyse zouden kunnen rekenen.

6. DE ONTWIKKELING NA 1945

Na de tweede wereldoorlog zijn er verschillende nieuwe methoden in de behandeling van (gehele) kwadratische vormen naar voren gekomen. Allereerst wordt de karakterisering van de kwadratische vorm door zijn coëfficiënten of door de matrix vervangen door een meer intrinsieke behandeling. Een kwadratische vorm is een functie op een lineaire ruimte die voldoet aan

$$Q(a+b) - Q(a) - Q(b) = B(a,b),$$

waarin B een bilineaire vorm in a en b is. (als $1+1 \neq 0$ geldt $B(a,a) = 2Q(a)$). De matrix behorende bij de kwadratische vorm hangt af van de keuze van de

basis in de lineaire ruimte. Duidelijk is dat nu de orthogonale groep, in dit geval de groep van de transformaties die de kwadratische vorm invariant laten, een belangrijke rol gaat spelen. Een standaard werk van M. EICHLER heet dan ook *Quadratische Formen und orthogonale Gruppen* (1952). Meer algebraïsche structuren gaan nu een rol spelen, naast geslachten treffen we nu spinor-geslachten aan. Dit gezichtspunt voert tot een nieuwe opzet van de Minkowski - Hasse - Siegel behandeling van de kwadratische vormen. De lichamen van de reële en p-adische getallen worden samengevoegd tot een grote topologische constructie, de adeles. Bij deze constructie is een Haarmaat te formuleren, een soort metriek die compatibel is met groepsstructuren. De stellingen van Siegel kunnen nu als analytische identiteiten in deze theorie geformuleerd worden. (A. WEIL, 1961) In deze vorm kan de theorie betrekkelijk eenvoudig gegeneraliseerd worden van gehele rationale getallen naar gehele algebraïsche getallen.

7. VOORBEELDEN

Als voorbeelden van de theorie behandelen we twee speciale gevallen, namelijk binaire positief definiete gehele kwadratische vormen met discriminant – 23 en positief definiete gehele kwadratische vormen met discriminant 1. Van de laatsten moet het aantal variabelen noodzakelijk een achtvoud zijn. Positief definiete, gehele kwadratische vormen met acht variabelen en discriminant 1 spelen mede een rol in de theorie van de exceptionele Liegroepen. Een voorbeeld is

$$Q_8(\boldsymbol{x}) = x_1^2 + x_2^2 + \ldots + x_8^2 - x_1x_3$$

$$-x_2x_4 - x_3x_4 - x_4x_5 - x_5x_6 - x_6x_7 - x_7x_8.$$

Er is maar één klasse van positief definiete kwadratische vormen in acht variabelen met discriminant 1. De discriminant van kwadratische vormen, waarvan het aantal variabelen een achtvoud is wordt gedefinieerd als de determinant

$$\left|\frac{\partial^2 Q(\boldsymbol{x})}{\partial x_i \partial x_j}\right|$$

Het aantal gehele oplossingen van de vergelijking

$$Q(\boldsymbol{x}) = N$$

geven we aan met $r(Q, N)$. Nu is

$$r(Q_8, N) = 240\Sigma d^3 \text{ gesommeerd over alle positieve delers van } N.$$

Vormen we een kwadratische vorm met discriminant 1 in 16 variabelen door

$$Q_{16}(\boldsymbol{x}, \boldsymbol{y}) = Q_8(\boldsymbol{x}) + Q_8(\boldsymbol{y})$$

dan geldt

$$r(Q_{16}, N) = 480\Sigma d^7 \text{ weer gesommeerd over alle positieve delers van } N.$$

Vormen we op dezelfde manier een kwadratische vorm in 24 variabelen

$$Q_{24}(x,y,z) = Q_8(x) + Q_8(y) + Q_8(z)$$

dan geldt

$$r(Q_{24},N) = \frac{65520}{691}\Sigma d^{11} + \frac{432000}{691}\tau(N),$$

waarin $\tau(N)$ gedefinieerd is door

$$\Sigma\tau(N)q^N = q\Pi(1-q^n)^{24}, \quad 1 \leq N < \infty, \quad 1 \leq n < \infty.$$

Er is een andere kwadratische vorm in 24 variabelen met discriminant 1 waarvoor geldt

$$r(F_{24},N) = \frac{65520}{691}\Sigma d^{11} + \frac{697344}{691}\tau(N).$$

De getallen $\tau(N)$ van Ramanujan hebben een aantal merkwaardige eigenschappen. Allereerst zijn ze van aanzienlijk kleinere orde dan de term die de som is van de elfde machten van de delers. De laatste is van de orde $N^{11+\varepsilon}$. De getallen van Ramanujan zijn van de orde $N^{11/2+\varepsilon}$. Verder voldoen ze aan bepaalde multiplicative relaties die we formuleren als

$$\sum_{N=1}^{\infty} \frac{\tau(N)}{N^s} = \prod_p \left(1 - \frac{\tau(p)}{p^s} + \frac{p^{11}}{p^{2s}}\right)^{-1}.$$

Met andere woorden

$$\tau(MN) = \tau(M)\tau(N) \text{ als de grootste gemene deler van } M \text{ en} N \text{ gelijk 1 is}$$

$$\tau(p^{k+2}) = \tau(p^{k+1})\tau(p) - p^{11}\tau(p^k), \; p \text{ is een priemgetal.}$$

terwijl bovendien

$$|\tau(p)| < 2p^{11/2}.$$

Het tweede voorbeeld gaat over positief definiete kwadratische vormen

$$Q(x,y) = ax^2 + bxy + cy^2$$

met discriminant $D = b^2 - 4ac$.
We zullen ons speciaal bezig houden met het geval waarin de discriminant gelijk is aan $D = -q$ met een priemgetal q dat een 24-voud min 1 is, dus bijvoorbeeld $D = -23$, $D = -47$, $D = -71$ enzovoorts.

We hebben een speciaal geval van het kwadratisch restsymbool nodig, we definiëren

$\left(\frac{a}{q}\right) = 1$ als er een gehele x bestaat met $x^2 \equiv a(\bmod q)$, a en q onderling ondeelbaar,

$\left(\frac{a}{q}\right) = 0$ als q een deler van a is,

$\left(\frac{a}{q}\right) = -1$ in de overige gevallen.

We zoeken nu kwadratische vormen met gehele a, b en c en met discriminant - 23. Vormen die door een gehele transformatie met determinant 1 in elkaar over te voeren zijn noemen we equivalent. Er zijn drie equivalentieklassen van kwadratische vormen met discriminant - 23, als representanten van deze klassen kunnen we kiezen

$$G_1(x,y) = x^2 + xy + 6y^2,$$
$$G_2(x,y) = 2x^2 + xy + 3y^2,$$
$$G_3(x,y) = 2x^2 - xy + 3y^2.$$

Als $a(G,N)$ het aantal gehele oplossingen van de vergelijking

$$G(x,y) = N$$

voorstelt kan men bewijzen dat

$$a(G_2,N) = a(G_3,N),$$

$$a(G_1,N) + a(G_2,N) + a(G_3,N) = 6\Sigma\left(\frac{-23}{d}\right) \quad \text{gesommeerd over alle delers van } N$$

$$a(G_1,N) - a(G_2,N) = 2t(N),$$

waarbij

$$\Sigma t(n)x^n = x\Pi(1-x^n)(1-x^{23n}), \quad 1 \leq n < \infty.$$

Bovendien geldt

$$\sum_{n=1}^{\infty} \frac{t(n)}{n^s} = \Pi\left(1 - \frac{t(p)}{p^s} + (\frac{-23}{p})\frac{1}{p^{2s}}\right)^{-1} \quad \text{(product over alle priemgetallen)},$$

zodat

$t(mn) = t(m)t(n)$ als m en n onderling ondeelbaar zijn,

$t(p^{k+2}) = t(p)t(p^{k+1}) - (\frac{-23}{p})t(p^k)$, p is een priemgetal.

en

$$|t(p)| \leq 2.$$

Voor kwadratische vormen met discriminant $-q$ waarbij q een priemgetal congruent 23 modulo 24 is kunnen enkele van deze eigenschappen gegeneraliseerd worden. Wel kunnen er meer dan drie klassen van kwadratische vormen optreden.
Dat we speciaal 24- vouden plus 23 noemen komt omdat hier een klassieke functie uit de theorie van de modulaire functies een rol speelt. De êta functie gedefinieerd door

$$\eta(x) = x^{1/24}\Pi(1-x^n), \quad 1 \leq n < \infty$$

kwamen we ook in het eerste voorbeeld tegen. Producten van êta functies die machtreeksen in x zijn spelen hier een speciale rol.
Ook vormen in meer variabelen kunnen soms in verband gebracht worden met producten van het boven geschetste type.
Zo kan men representatie aantallen van vormen in vier variabelen, waarvoor het getal 11 een bizondere rol speelt in verband brengen met de coëfficiënten van

$$x\Pi(1-x^n)^2(1-x^{11n})^2 \quad 1 \leq n < \infty.$$

Bewijzen van de in de voorbeelden behandelde stellingen kan men onder andere vinden in

F. VAN DER BLIJ: Binary quadratic forms of discriminant - 23. Proc. Kon. Ned. Ak. Amsterdam, A **55** , (1952) , 498-503.
B. SCHOENEBERG: Bemerkungen über einige klassen von Modulformen. Proc Kon. Ned. Ak. Amsterdam, A 70, (1967) , 177 - 182.
F. VAN DER BLIJ: Even quadratic forms with determinant unity. *The Quarterly Journal of Mathematics,* Oxford; second series 5, (1954) , 297 - 300.

BIBLIOGRAFIE
Voor een uitvoerige lijst van publicaties over de getaltheorie van kwadratische vormen verwijzen we naar de bibliografie van het in § 0 genoemde boek van Cassels. Wanneer men de bibliografiën van de daar genoemde werken weer raadpleegt enzovoorts zal men een collectie van honderden verwijzingen vinden.
Publicaties vóór 1919 zijn te vinden in het overzichtswerk van L.E. DICKSON: *History of the Theory of Numbers*, volume 2 (1920) hoofdstukken VI tot XX en volume 3 (1923), hoofdstukken I tot XI.
Elementaire bewijzen over sommen van kwadraten en analoge problemen zijn te vinden in het standaardwerk:

G.H. HARDY and E.M. WRIGHT: *An Introduction in the Theory of Numbers.*

Enkele fundamentele artikelen zijn:

C. G. J. JACOBI: *Journal für Mathematik 80,* (1875) zie ook Werke.
H. MINKOWSKI: *Mémoire sur la théorie des formes quadratiques à coefficients entiers,* 1884.
H. HASSE: Ueber die Darstelbarkeit von Zahlen durch quadratische Formen

im Körper der rationalen Zahlen; *J. Reine und angew. Math.* **152** (1923) 129–148.
C.L. SIEGEL: Ueber die analytische Theorie der quadratische Formen; *Ann. of Math.* **36** (1935), 527 - 606.
A. WEIL: *Adeles and algebraic groups* (1961).
A. WEIL: Sur la formule de Siegel dans la théorie des groupes classiques; *Acta Mathematica,* **113** (1965) 1 - 87.

De achtergronden van theta functies, identiteiten van theta functies die getaltheoretische resultaten geven, modulaire functies, functies invariant onder ondergroepen van de groep van gebroken lineaire transformaties enzovoorts zijn in vele studies behandeld. We noemen er enkelen:

B. SCHOENEBERG: *Elliptic Modular Functions,* Berlin, 1974.
ANDRÉ WEIL: *Elliptic Functions according to Eisenstein and Kronecker,* Berlin, 1976.
R.A.RANKIN: *Modular Forms and Functions,* Cambridge, 1977.

De methode van Hardy en Littlewood is onder andere behandeld in:
L.E. DICKSON: *Studies in the Theory of Numbers,* Chicago,1930.
M.I. KNOPP: *Modular Functions in Analytic Number Theory,* Chicago, 1970.

Moderne leerboeken over de getaltheorie van kwadratische vormen zijn in een willekeurige selectie, behalve de reeds genoemden:

B.W. JONES: *The Arithmetic Theory of quadratic Forms,* Carus Math. Monographs, 1950.
N. BOURBAKI: *Eléments de Mathématique,* I - II - Chapitre IX, Formes sesquilinéaires et Formes quadratiques, Paris, 1959.
G.L. WATSON: *Integral quadratic Forms,* Cambridge, 1960.
O.T.O'MEARA: *Introduction to quadratic Forms,* Berlin, 1963.
B.L. VAN DER WAERDEN UND H. GROSS: *Studien zur Theorie der quadratischen Formen,* Basel, 1968.
J. MILNOR - D. HUSEMOLLER: *Symmetric Bilinear Forms,* Berlin, 1973.
H. PETERSON: *Modulfunktionen und quadratische Formen,* Berlin, 1982.
Y. KITAOKA: *Arithmetic of Quadratic Forms,* Cambridge, 1993.

MC SYLLABI

1.1 F. Göbel, J. van de Lune. *Leergang besliskunde, deel 1: wiskundige basiskennis.* 1965.
1.2 J. Hemelrijk, J. Kriens. *Leergang besliskunde, deel 2: kansberekening.* 1965.
1.3 J. Hemelrijk, J. kriens. *Leergang besliskunde, deel 3: statistiek* 1966
1.4 G. de Leve, W. Molenaar. *Leergang besliskunde, deel 4: Markovketens en wachttijden.* 1966.
1.5 J. Kriens, G. de Leve. *Leergang besliskunde, deel 5: inleiding tot de mathematische besliskunde.* 1966.
1.6a B. Dorhout, J. Kriens. *Leergang besliskunde, deel 6a: wiskundige programmering 1.* 1968.
1.6b B. Dorhout, J. kriens, J.Th. van Lieshout. *Leergang besliskunde, deel 6b: wiskundige programmering* 2. 1977. 1.7a G. de Leve. *Leergang besliskunde, deel 7a: dynamische programmering 1.* 1968.
1.7b G. de Leve, H.C. Tijms. *Leergang besliskunde, deel 7b: dynamische programmering 2.* 1970.
1.7c G. de Leve, H.C. Tijms. *Leergang besliskunde, deel 7c: dynamische programmering 3.* 1971.
1.8 J. Kriens, F. Göbel, W. Molenaar. *Leergang besliskunde, deel 8: minimaxmethode, netwerkplanning, simulatie.* 1968.
2.1 G.J.R. Förch, P.J. van der Houwen, R.P. van de Riet. *Colloquium stabiliteit van differentieschema's, deel 1.* 1967.
2.2. L. Dekker, T.J. dekker, P.J. van der Houwen, M.N. Spijker. *Colloquium stabiliteit van differentieschema's, deel 2.* 1968.
3.1. H.A. Lauwerier. *Randwaardeproblemen, deel 1.* 1967
3.2. H.A. Lauwerier. *Randwaardeproblemen, deel 2.* 1968
3.3. H.A. Lauwerier. *Randwaardeproblemen, deel 3.* 1968
4 H.A. Lauwerier. *Representaties van groepen.* 1968
5 J.H. van Lint, J.J. Seidel, P.C. Baayen. *Colloquium discrete wiskunde.* 1968.
6. K.K. Koksma. Cursus ALGOL 60. 1969.
7.1 *Colloquium moderne rekenmachine, deel 1.* 1969.
7.2 *Colloquium moderne rekenmachine, deel 2.* 1969.
8 H. Bavinck, J. Grasman. *Relaxatietrillingen.* 1969.
9.1 T.M.T. Coolen, G.J.R. Förch, E.M. de Jager, H.G.J. Pijls. *Colloquium elliptische differentiaalvergelijkingen, deel 1.* 1970
9.2 W.P. van den Brink, T.M.T. Coolen, B. Dijkhuis, P.P.N. de Groen, P.J. van der Houwen, E.M. de Jager, N.M. Temme, R.J. de Vogelaere. *Colloquium elliptische differentiaalvergelijkingen, deel 2.* 1970
10 J. Fabius, W.R. van Zwet. *Grondbegrippen van de waarschijnlijkheidsrekening.* 1970
11 H. Bart, M.A. Kaashoek, H.G.J. Pijls, W.J. de Schipper, J. de Vries. *Colloquium halfalgebra's en positieve operatoren.* 1971.
12 T.J. Dekker. *Numerieke algebra.* 1971
13 F.E.J. Kruseman Aretz. *Programmeren voor rekenautomaten; de MC ALGOL 60 vertaler voor de EL X8.* 1971
14 H. Bavinck, W. Gautschi, G.M. Willems. *Colloquium approximatietheorie.* 1971
15.1 T.J. Dekker, P.W. Hemker, P.J. van der Houwen *Colloquium stijve differentiaalvergelijkingen, deel 1.* 1972.
15.2 P.A. Beentjes, K. Dekker, P.W. Hemker, S.P.N. van Kampen, G.M. Willems. *Colloquium stijve differentiaalvergelijkingen, deel 2.* 1973.
15.3 P.A. Beentjes. K. Dekker, P.W. Hemker, M. van Veldhuizen *Colloquium stijve differentiaalvergelijkingen, deel 3.* 1975.
16.1 L. Geurts. *Cursus programmeren, deel 1: de elementen van het programmeren.* 1973
16.2 L. Geurts. *Cursus programmeren, deel 2: de programmeertaal ALGOL 60.* 1973
17.1 P.S. Stobbe. *Lineaire algebra, deel 1.* 1973.
17.2 P.S. Stobbe. *Lineaire algebra, deel 2.* 1973.
17.3 N.M. Temme. *Lineaire algebra, deel 3.* 1976.
18 F. van der Blij, H. Freudenthal, J.J. de Songh, J.J. Seidel, A. van Wijngaarden. *Een kwart eeuw wiskunde 1946-1971, syllabus van de vakantiecursus 1971.* 1973.
19 A. Hordijk, R. Potharst, J.Th. Runnenburg. *Optimaal stoppen van Markovketens.* 1973.
20 T.M.T. Coolen, P.W. Hemker, P.J. van der Houwen, E. Slagt. *ALGOL 60 procedures voor begin- en randwaardeproblemen.* 1976.
21 J.W. de Bakker (red.). *Colloquium programmacorrectheid.* 1975.
22 R. Helmers, J. Oosterhoff, F.H. Ruymgaart, M.C.A. van Zuylen. *Asymptotische methoden in de toetsingstheorie; toepassingen van naburigheid.* 1976.
23.1 J.W. de Roever (red.). *Colloquium onderwerpen uit de biomathematica, deel 1.* 1976.
23.2 J.W. de Roever (red.). *Colloquium onderwerpen uit de biomathematica, deel 2.* 1977.
24.1 P.J. van der Houwen. *Numerieke integratie van differentiaalvergelijkingen, deel 1: eenstapsmethoden.* 1975.
25 *Colloquium structuur van programmeertalen.* 1976.
26.1 N.M. Temme (ed.). *Nonlinear analysis, volume 1.* 1976.
26.2 N.M. Temme (ed.). *Nonlinear analysis, volume 2.* 1976.
27 M. Bakker, P.W. Hemker, P.J. van der Houwen, S.J. Polak, M. van Veldhuizen. *Colloquium discretiseringsmethoden.* 1976.
28 O. Diekmann, N.M. Temme (eds.). *Nonlinear diffusion problems.* 1976.
29.1 J.C. Bus (red.). *Colloquium numerieke programmatuur, deel 1A, deel 1B.* 1976.
29.2 H.J.J. te Riele (red.). *Colloquium numerieke programmatuur, deel 2.* 1977.
30 J. Heering, P. Klint (red.). *Colloquium programmeeromgevingen.* 1983.
31 J.H. van Lint (red.). *Inleiding in de coderingstheorie.* 1976.
32 L. Geurts (red.). *Colloquium bedrijfssystemen.* 1976.
33 P.J. van der Houwen. *Berekening van waterstanden in zeeën en rivieren.* 1977.
34 J. Hemelrijk. *Oriënterende cursus mathematische statistiek.* 1977.
35P.J.W. ten Hagen (red.). *Colloquium computer graphics.* 1978.
36 J.M. Aarts, J. de Vries. *Colloquium topologische dynamische systemen.* 1977.
37 J.C. van Vliet (red.). *Colloquium capita datastructuren.* 1978.
38.1 T.H. Koornwinder (ed.). *Representations of locally compact groups with applications, part 1.* 1979.
38.2 T.H. Koornwinder (ed.). *Representations of locally compact groups with applications, part II.* 1979.
39 O.J. Vrieze, G.L. Wanrooy. *Colloquium stochastische spelen.* 1978.
40 J. van Tiel. *Convexe analyse.* 1979.
41 H.J.J. te Riele (ed.). *Colloquium numerical treatment of integral equations.* 1979.
42 J.C. van Vliet (red.). *Colloquium capita implementatie van programmeertalen.* 1980.
43 A.M. Cohen, H.A. Wilbrink. *Eindige groepen (een inleidende cursus).* 1980.
44 J.G. Verwer (ed.). *Colloquium numerical solution of partial differential equations.* 1980.
45 P. Klint (red.). *Colloquium hogere programmeertalen en computerarchitectuur.* 1980.
46.1 P.M.G. Apers (red.). *Colloquium databankorganisatie, deel 1.* 1981.
46.2 P.M.G. Apers (red.). *Colloquium databankorganisatie, deel 2.* 1981.
47.1 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60: general information and indices.* 1981.
47.2 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 1: elementary procedures; vol. 2: algebraic evaluations.* 1981.
47.3 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 3A: linear algebra, part 1.* 1981.
47.4 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 3B: linear algebra, part II.* 1981.
47.5 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 4: analytical evaluations; vol. 5A: analytical problems, part 1.* 1981.
47.6 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 5B: analytical problems, part II.* 1981.
47.7 P.W. Hemker (ed.). *NUMAL, numerical procedures in ALGOL 60, vol. 6: special functions and constants; vol. 7: interpolation and approximation.* 1981.
48.1 P.M.B. Vitányi, J. van Leeuwen, P. van Emde Boas (red.). *Colloquium complexiteit en algoritmen, deel 1.* 1982.
48.2 P.M.B. Vitányi, J. van Leeuwen, P. van Emde Boas (red.). *Colloquium complexiteit en algoritmen, deel 2.* 1982.
49 T.H. Koornwinder (ed.). *The structure of real semisimple Lie groups.* 1982.
50 H. Nijmeijer. *Inleiding systeemtheorie.* 1982.
51 P.J. Hoogendoorn (red.). *Cursus cryptografie.* 1983

CWI SYLLABI

1 Vacantiecursus 1984: *Hewet - plus wiskunde.* 1984.

2 E.M. de Jager, H.G.J. Pijls (eds.). *Proceedings Seminar 1981-1982. Mathematical structures in field theories.* 1984.

3 W.C.M. Kallenberg, et al. *Testing statistical hypotheses: worked solutions.* 1984.

4 J.G. Verwer (ed.). *Colloquium topics in applied numerical analysis, volume 1.* 1984.

5 J.G. Verwer (ed.). *Colloquium topics in applied numerical analysis, volume 2.* 1984.

6 P.J.M. Bongaarts, J.N. Buur, E.A. de Kerf, R. Martini, H.G.J. Pijls, J.W. de Roever. *Proceedings Seminar 1982-1983. Mathematical structures in field theories.* 1985.

7 Vacantiecurus 1985: *Variatierekening.* 1985.

8 G.M. Tuynman. *Proceedings Seminar 1983-1985. Mathematical structures in field theories, Vol.1 Geometric quantization.* 1985.

9 J. van Leeuwen, J.K. Lenstra (eds.). *Parallel computers and computations.* 1985.

10 Vacantiecursus 1986: *Matrices.* 1986.

11 P.W.H. Lemmens. *Discrete wiskunde: tellen, grafen, spelen en codes.* 1986.

12 J. van de Lune. *An introduction to Tauberian theory: from Tauber to Wiener.* 1986.

13 G.M. Tuynman, M.J. Bergvelt, A.P.E. ten Kroode. *Proceedings Seminar 1983-1985. Mathematical structures in field theories, Vol.2.* 1987.

14 Vacantiecursus 1987: *De personal computer en de wiskunde op school.* 1987.

15 Vacantiecursus 1983: *Complexe getallen.* 1987.

16 P.J.M. Bongaarts, E.A. de Kerf, P.H.M. Kersten. *Proceedings Seminar 1984-1986. Mathematical structures in field theories, Vol.1.* 1988.

17 F. den Hollander, H. Maassen (eds.). *Mark Kac seminar on probability and physics. Syllabus 1985-1987.* 1988.

18 Vacantiecursus 1988. *Differentierekening.* 1988.

19 R. de Bruin, C.G. van der Laan, J.R. Luyten, H.F. Vogt. *Publiceren met LATEX.* 1988.

20 R. van der Horst, R.D. Gill (eds.). *STATAL: statistical procedures in Algol 60, part 1.* 1988.

21 R. van der Horst, R.D. Gill (eds.). *STATAL: statistical procedures in Algol 60, part 2.* 1988.

22 R. van der Horst, R.D. Gill (eds.). *STATAL: statistical procedures in Algol 60, part 3.* 1988.

23 J. van Mill, G.Y. Nieuwland (eds.). *Proceedings van het symposium wiskunde en de computer.* 1989.

24 P.W.H. Lemmens (red.). *Bewijzen in de wiskunde.* 1989.

25 Vacantiecursus 1989: *Wiskunde in de Gouden Eeuw.* 1989.

26 G.G.A. Bäuerle et al. *Proceedings Seminar 1986-1987. Mathematical structures in field theories.* 1990.

27 Vacantiecursus 1990: *Getallentheorie en haar toepassingen.* 1990.

28 Vacantiecursus 1991: *Meetkundige structuren.* 1991.

29 A.G. van Asch, F. van der Blij. *Hoeken en hun Maat.* 1992.

30 M.J. Bergvelt, A.P.E. ten Kroode. *Proceedings seminar 1986-1987. Lectures on Kac-Moody algebras.* 1992.

31 Vacantiecursus 1992: *Systeemtheorie.* 1992.

32 F. den Hollander, H. Maassen (eds.). *Mark Kac seminar on probability and physics. Syllabus 1987-1992.* 1992.

33 P.W.H. Lemmens (ed.). *Meetkunde van kunst tot kunde, vroeger en nu.* 1993.

34 J.H. Kruizinga. *Toegepaste wiskunde op een PC.* 1992.

35 Vacantiecursus 1993: *Het reële getal.* 1993.

36 Vacantiecursus 1994: *Computeralgebra.* 1994.

37 G. Alberts. *Wiskunde en praktijk in historisch perspectief. Syllabus.* 1994.

38 G. Alberts, J. Schut (eds.). *Wiskunde en praktijk in historisch perspectief. Reader.* 1994.

39 E.A. de Kerf, H.G.J. Pijls (eds.). *Proceedings Seminar 1989-1990. Mathematical structures in field theory.* 1994.

40 Vakantiecursus 1995: *Kegelsneden en kwadratische vormen.* 1995.