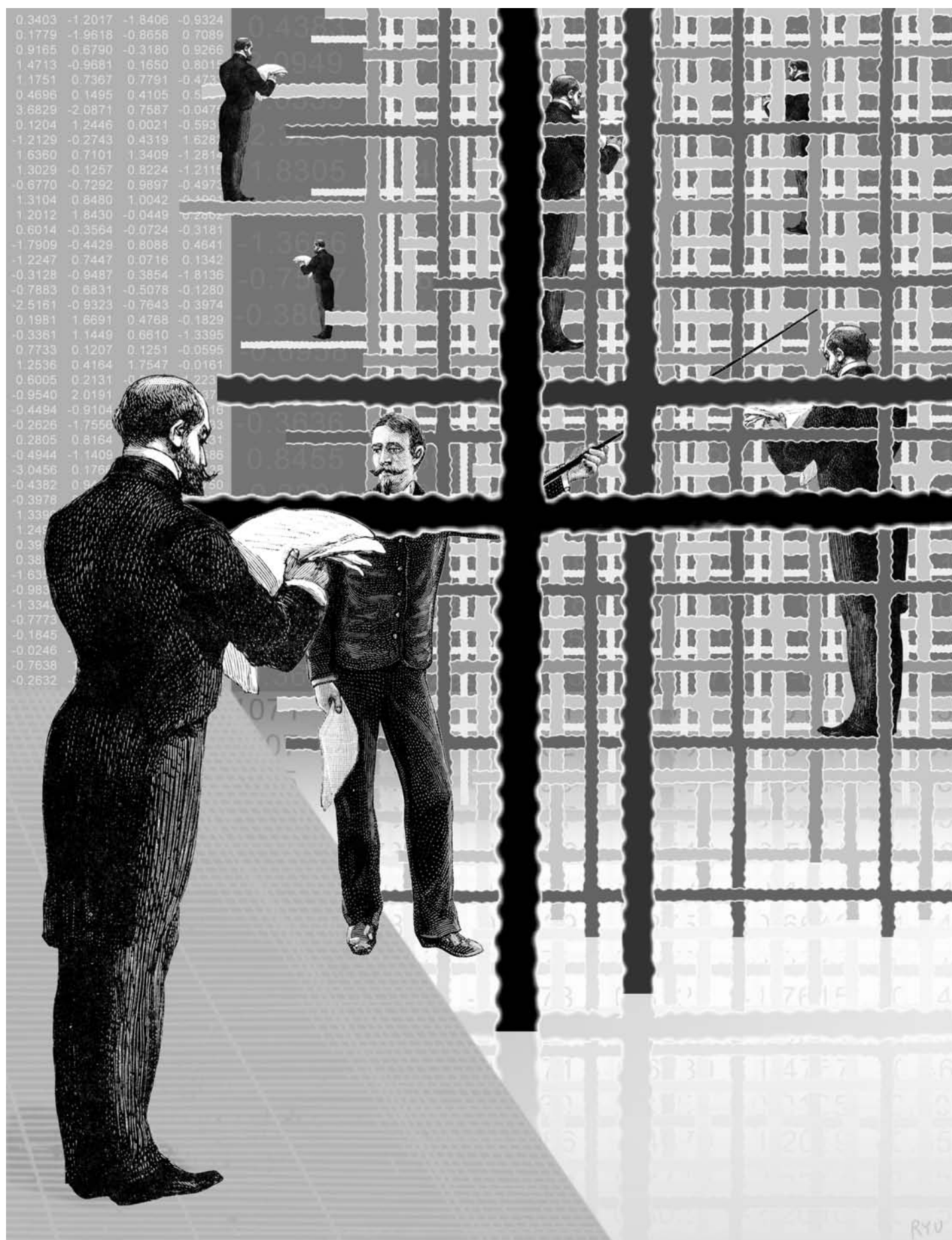




Henk van der Vorst

Mathematical Institute

Utrecht University, P.O. Box 80.010

3508 TA Utrecht, The Netherlands.

vorst@math.uu.nl

Afscheidsrede

Ik leg het nog één keer uit

Waarom wordt er nog research gedaan naar het oplossen van stelsels lineaire vergelijkingen? Toen Henk van der Vorst zijn onderzoeksloopbaan begon zag hij daar het belang ook niet echt van in. Pas toen hij, als werknemer van Reactor Centrum Petten, geconfronteerd werd met het praktisch en onmiddellijk belang van het oplossen van dergelijke stelsels, ging hij zich voor dit onderwerp interesseren. Het werd zijn levenswerk; zijn baanbrekend artikel over de zogenaamde Bi-CGSTAB-methode werd het meest geciteerde artikel ter wereld van de negentiger jaren. Nu is Henk van der Vorst de eerste die deze prestatie zal willen relativeren. Toch toont de impact van deze publicatie wel het maatschappelijk belang aan van zijn wiskundig onderzoek.

Bijna op de kop af veertig jaar geleden studeerde ik af aan deze universiteit als toegepast wiskundige. Wat doet een wiskundige voor de kost, of meer algemeen: wat kan je met wiskunde in de maatschappij? Dat vroeg ik me toen ook af. Wiskunde is er in vele varianten; ik had een flink aantal daarvan tijdens de studie voorbij zien komen. De weg van de fundamentele van de wiskunde naar de daadwerkelijke toepassing is een zeer lange. Ik zou mij in mijn loopbaan aan de zeer toepasbare kant gaan bevinden. In 1964 volgde ik het eerste college Numerieke Wiskunde van mijn latere promotor, professor van der Sluis, die toen net benoemd was in Utrecht. Hoe kon je met de computer wiskundige sommetjes uitrekenen? Dat leek me toen tamelijk voor de hand liggend, maar alras bleek dat dit verre van eenvoudig was. Als je bijvoorbeeld de wortels van een vierkantsvergelijking wilt uitrekenen via de bekende formules die je al op de middelbare school had geleerd, dan

komt daar in veel gevallen één nauwkeurig en één (soms veel) minder nauwkeurig bepaalde wortel uit. Via een eenvoudig trucje kun je beide wortels echter even nauwkeurig berekenen. Blijkbaar representeert de gegeven vergelijking de beide wortels even nauwkeurig, maar moet je er moeite voor doen om die informatie, rekenend met de computer in een eindige precisie, boven water te krijgen. En dat was dan nog maar een eenvoudige vierkantsvergelijking. Het zou al gauw nog aanzienlijk gecompliceerder worden omdat veel rekenwerk ook nog eens veel tijd vergt, zelfs op de allersnelste computers. Een fascinerend vak: snelheid en nauwkeurigheid speelden een cruciale rol. Een intellectuele variant van Formule 1 racen.

Lineaire stelsels in Petten

Tegen het eind van mijn doctoraalstudie volgde ik een seminarium over het oplossen van stelsels lineaire vergelijkingen aan de hand

van het toen pas verschenen boek van de Amerikaan Richard Varga. Die had het over de problemen bij het oplossen van grote stelsels. Niet alleen had ik toen geen flauw idee van wat groot was bij stelsels, ik had ook het idee dat de voorgestelde oplostecheniken een race tegen de electronica-techniek inhielden. Computers werden immers steeds sneller en krachtiger, zodat wat toen een probleem scheen, een paar jaar later mogelijk een trivialiteit zou zijn. Het reflecteerde evenwel de tijdgeest onder de toenmalige numeriek wiskundigen.

Niet veel later, zo begin jaren zeventig meenden velen onder hen dat de belangrijkste algoritmen al wel bedacht waren en dat de almaar toenemende computersnelheid de rest zou doen. In werkelijkheid leidden de krachtiger computers tot de vraag om oplossing voor steeds weer grotere problemen en de huidige zienswijze onder numerici is veel eerder dat we meestal met de handen in het haar zitten.

De meesten onder u zullen op de middelbare school al wel kennis gemaakt hebben met het oplossen van eenvoudige stelsels vergelijkingen, bijvoorbeeld:

$$\begin{cases} 3x + 5y = 11, \\ x + 2y = 4. \end{cases}$$

Als je drie keer de tweede vergelijking van de eerste aftrekt, houd je over $y = 1$. Het kan ook wat ingewikkelder door eenderde keer de eerste vergelijking van de tweede af te trekken, maar dat levert natuurlijk hetzelfde op. De volgorde van de vergelijkingen voor het elimineren van de variabele x is dus van belang voor het rekenwerk. Dit zijn natuurlijk mooie getallen. Het blijkt alras, dat als je met gebroken getallen rekent en je neemt maar een eindig aantal decimalen mee, dat die volgorde ook van groot belang is voor de nauwkeurigheid van het berekende resultaat. Om die reden dacht men in de jaren veertig van de afgelopen eeuw nog dat men praktisch gesproken met een computer in het algemeen geen stelsels van meer dan twintig vergelijkingen met twintig variabelen op zou kunnen lossen, tenzij die stelsels een speciale structuur hadden.

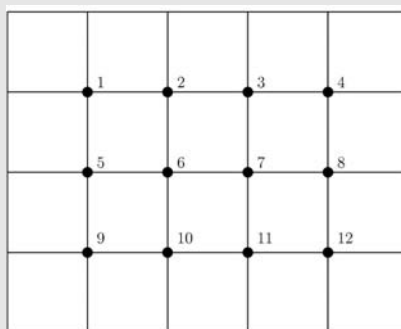
Het probleem van het oplossen van een stelsel vergelijkingen van het bovenstaande type is wiskundig gezien bijzonder eenvoudig en bevindt zich helemaal aan het eind van de wiskundeketen die van zeer fundamenteel naar zeer toegepast loopt, als je tenminste van zo'n keten kunt spreken. Het probleem is in wiskundige zin dan ook allang opgelost. Men weet heel precies wanneer zo'n stelsel een eenduidig bepaalde oplossing heeft en ook wanneer dat niet het geval is. Wat overblijft is het probleem om in concrete gevallen de oplossing uit te rekenen en dat is, omdat het basisrecept daarvoor bekend is, voor de mens alleen maar een vervelende klus. Tegenwoordig heb je daar computers voor en je kunt je dus afvragen waarom tal van wiskundigen over de hele wereld zich nog steeds buigen over dit probleem en ook waarom talloze wetenschappers, industrieel ontwerpers, en analisten ongeduldig staan te wachten op nieuwe inzichten. Er worden met grote regelmaat hele congressen aan dit type probleem gewijd. Dat wist ik ook nog niet toen ik pas afgestudeerd was en eigenlijk had ik ook nog geen goed idee waar grote stelsels optreden en hoe groot ze dan kunnen zijn. Om eerlijk te zijn: het probleem had ook niet mijn belangstelling en ik begon als jong onderzoeker met het numeriek integreren van functies, ofwel het met de computer bepalen van de oppervlakte onder een mogelijk ingewikkelde kromme. Daar was de aardigheid na dik twee jaar van af, waarna ik mijn promotiebaan in Utrecht opzegde en in 1969 ging werken bij het Reactor Centrum in Petten, het tegenwoordige Energie Centrum Nederland. Daar liep ik frontaal tegen grote stelsels op, ze deden met hun grote computer vrijwel niet anders. Die

stelsels ontstonden bij het numeriek modeleren van het gedrag van de kernreactor op hun terrein. Men wilde graag, voordat men stralingsexperimenten deed, weten hoe de reactor daar op zou reageren. Tenslotte stond die reactor vlakbij en dan wil je je best wel doen.

Warmtestroming in een plaat

De modellering die gebruikt wordt om het gedrag van een kernreactor te bestuderen is wat ingewikkeld en daarom wil ik u een eenvoudiger voorbeeld voorzetten. Stelt u zich een dunne rechthoekige plaat metaal voor. We kennen de temperatuur langs de zijden van de rechthoek en we willen weten hoe warm de plaat overal is. Dat zouden we natuurlijk gewoon kunnen meten, maar in werkelijkheid is zo'n plaat natuurlijk niet zeer dun en dan willen we de temperatuur in het inwendige kennen. Bovendien wil je in ontwerpprocessen van te voren weten wat je te wachten staat. We houden het voor het gemak dus maar even bij het vereenvoudigde voorbeeld. Het is slechts een vingeroefening en het wordt vanzelf erger. Overal de temperatuur berekenen is voorlopig wat veel, maar het zou al mooi zijn als we de temperatuur in een aantal punten zouden kunnen berekenen.

We leggen daartoe een denkbeeldig regelmatig rooster over de plaat aan. Dat rooster heeft 12 knooppunten en de lijntjes geven aan hoe de knooppunten aan elkaar gerelateerd zijn. De bolletjes geven de knooppunten aan waar we de temperatuur zouden willen kennen. Voor de temperatuur in een roosterpunt,



Numeriek model

- Fysische warmtevergelijking:
 $-T_{XX} - T_{YY} = 0$
- Benaderingsformule:
 $T_6 \approx \frac{T_5 + T_7 + T_{10} + T_2}{4}$
- Rekenmodel:
 $4T_6^h - T_5^h - T_7^h - T_{10}^h - T_2^h = 0$
- Twaalf vergelijkingen, twaalf onbekenden

in de figuur roosterpunt nummer 6, gebruiken we een eenvoudige benaderingsformule voor de fysische warmtevergelijking die het temperatuursverloop beschrijft. De temperatuur in elk roosterpunt geven we aan met het symbool T met als index het nummer van het roosterpunt. In het rekenmodel voegen we nog een index h toe om aan te duiden dat we niet de exacte temperatuur berekenen en dat de benaderde waarden van de roosterafstand h afhangen.



Complex rekenrooster

Dat leidt in dit geval tot 12 vergelijkingen met 12 onbekenden en die kunnen we gelukkig oplossen omdat de knooppunten die met randpunten verbonden zijn een bekend rechterlid opleveren. De verkregen benadering is grof, maar je kunt bewijzen dat de afwijking van het antwoord ten opzichte van de echte temperatuur afneemt met het kwadraat van de roosterafstand: een tien keer zo kleine roosterafstand h levert zo ongeveer een honderd keer zo kleine afwijking van het exacte antwoord op. Voor een plaat van 1 bij 1 meter hebben we ruwweg een rooster met roosterafstand 1 cm nodig om ongeveer 4 cijfers precisie in het resultaat te krijgen. Dat betekent dus $100 \times 100 = 10000$ vergelijkingen met even zoveel onbekenden. Driedimensionaal kom je zo voor een kubus van 1 kubieke meter aan een miljoen vergelijkingen met een miljoen onbekenden. Zou je dat zonder verder nadenken aan een computer aanbieden, om met de standaard eliminatieprocedure het antwoord te bepalen, dan zou dat ongeveer tien tot de achttiende rekenoperaties (optellingen, vermenigvuldigingen, ...) vergen. Een immense klus, waar de allersnelste computers op dit moment toch al gauw een week rekentijd voor nodig hebben.

Gauss

U zult zich nu kunnen voorstellen dat het voor de volgende situatie nog problematischer wordt: rond 1820 hield de Duitse wis-

kundige Gauss zich ook bezig met het praktisch oplossen van stelsels vergelijkingen. In zijn geval stelsels van vier vergelijkingen met vier onbekenden die optraden bij het nauwkeurig bepalen van afstanden en hoeken bij landmeting via driehoeken. De coëfficiënten in zijn stelsels waren geen mooie getallen en het rekenwerk met de hand werd al gauw erg vervelend: hoe nauwkeurig moest hij zijn tussenresultaten uitrekenen? Bovendien: hoe kwam je er achter dat je je niet vergist had? Hij bedacht daarom een methode waarmee hij snel bruikbare benaderingen kon uitrekenen. Een voorbeeld:

$$\begin{cases} 8x + y = 10, \\ x + 10y = 21. \end{cases}$$

In dit voorbeeld heb ik gehele getallen als coëfficiënten gekozen, maar daar was in het geval van Gauss geen sprake van. De oplossing van dit stelsel is $x = 1$ en $y = 2$, maar neem even aan dat we dat nog niet weten, dat de getallen in het voorbeeld niet zo mooi zijn en dat we het risico op vergissen willen beperken.

Gauss liet om te beginnen de y uit de eerste vergelijking weg:

$$\begin{cases} 8x = 10, \\ x + 10y = 21. \end{cases}$$

De oplossing is nu eenvoudig te bepalen: $x = 1.25$, $y = 1.975$. Niet zo heel best, maar we hebben de grootteorde redelijk te pakken. Dit werkt alleen maar als de coëfficiënten die we weg laten klein zijn ten opzichte van de overblijvende invloeden coëfficiënten en dat was bij Gauss het geval. Gauss paste nog een aantal trucs toe, zoals het geschikt veranderen van de volgorde van de vergelijkingen om de benadering nog iets te verbeteren. Vervolgens dacht hij na over het corrigeren van de benadering. Hij zag in dat de (nog onbekende) correcties voldeden aan dezelfde vergelijkingen, echter nu met een ander rechterlid. Daarom kan dezelfde truc, het weglaten van de coëfficiënt, rechtsboven weer toegepast worden om een benadering voor de correctie te verkrijgen. Met deze benaderde correctie kan de benaderde oplossing 1.25, 1.975 aangepast worden en dat leidt tot een betere benadering, om precies te zijn: 1.00313, 1.999687. Zo verder gaand komen we in een klein aantal stappen zeer dicht bij de oplossing $x = 1$, $y = 2$. Het aantrekkelijke voor Gauss was dat hij in elke rekenstap niet erg nauwkeurig hoef-

de te rekenen: hij mocht zich zelfs vergissen, want het proces corrigeert zichzelf in de volgende stappen. Hij schreef dan ook enthousiast aan zijn collega Gerling dat men dit proces half slapend kon doen of desnoods tijdens het rekenen aan iets anders kon denken. Dit was een eerste voorbeeld van een zogenaamd iteratief proces: elke rekenslag om een betere benadering te krijgen heet een iteratie.

Krylov

Na Gauss kwamen er nog aanpassingen op dit proces maar de basisgedachte bleef onaangetast. Jacobi bedacht een nog iets eenvoudiger proces, rond 1846, om stelsels van zeven vergelijkingen met zeven onbekenden te kunnen oplossen. Dat had hij nodig om de stabiliteit van de planetenbanen in ons zonnestelsel te kunnen bestuderen. Men dacht toen nog dat er maar zeven planeten waren en de gevonden afwijkingen waren er mede de oorzaak van dat men een achtste planeet op het spoor kon komen.

Het duurt dan ruim tachtig jaar voordat de volgende essentiële stap gemaakt kon worden. De Russische wiskundige Krylov kwam rond 1930 op het idee om in een gegeven stelsel een willekeurige benadering voor de oplossing in te vullen en dan het rechterlid uit te rekenen. Dat rechterlid vulde hij daarna ook weer in het stelsel in en kreeg zo het volgende rechterlid. De serie rechterleden die hij zo achtereenvolgens verkreeg vormde wat men later een Krylovruimte is gaan noemen [2].

Krylov bestudeerde stelsels van zes vergelijkingen met zes onbekenden, die optraden bij een ruwe modellering van stabiliteitsproblemen in de scheepsbouw. Hij genereerde zes achtereenvolgende rechterleden en probeerde deze zo te combineren, met geschikt gekozen coëfficiënten, dat ze samen nul opleverden. De zo gevonden coëfficiënten stelden hem in staat om de stabiliteitsfactoren (eigenwaarden) uit te rekenen. Voor zes vergelijkingen met zes onbekenden leverde dit proces nog redelijke stabiliteitsconstanten op. Voor stelsels met (veel) meer vergelijkingen liep het al gauw spaak, maar dat bleek pas (veel) later toen de computer het rekenen met grotere stelsels toeliet.

In 1952 – we zijn weer 20 jaar verder – paste de Hongaar Lanczos het proces van Krylov verder aan. Hij construeerde een betere basis voor de Krylovruimte, in plaats van de rechterleden die Krylov gebruikte, en kon zo nog weer grotere stelsels goedkoper oplossen [3].

Een aardige bijkomstigheid was dat je voor een stelsel van n vergelijkingen met n on-

bekenden hooguit n rekenslagen (iteraties) hoefde uit te voeren om aan de gewenste oplossing van het stelsel te komen, als je tenminste exact rekende. Bij Gauss was dat aantal iteraties in principe nog ongelimiteerd. Voor beide aanpakken gold dat je in veel gevallen ook met minder dan n stappen al dicht genoeg bij de oplossing kon zijn.

De strijd tussen wiskundige en computer

Nu hebben computers grote moeite met exact rekenen. Daar kwam men bij de Eidgenössische Technische Hochschule (ETH) in Zürich al snel achter. Men bestudeerde daar het gedrag van een aan één kant ingeklemde balk en dat leidde tot het voor die tijd al fantastische aantal van 64 vergelijkingen met 64 onbekenden. Bij het rekenen in ongeveer 14 decimalen constateerde men dat de Lanczos-methode na 64 iteraties nog ver van de oplossing verwijderd bleef; men had wel 200 iteraties nodig om een nauwkeurige oplossing te verkrijgen. Dat was verdacht en de in 1959 gepubliceerde resultaten [4] wierpen dan ook een smet op de aanvankelijk zo veelbelovende Lanczosmethode. De rekenende onderzoekswereld wierp zich met grote energie op andere varianten van de oorspronkelijke Gaussmethode, waarbij men uit opeenvolgende Gaussstappen een betere oplossing trachtte te extrapoleren door nauwkeurig naar de trend in die benaderingen te kijken: de zogenaamde Successieve Overrelaxatie Methoden. Deze methoden, die tot bloei kwamen rond 1970, mede door het werk van de Amerikanen Young en Varga, werden redelijk populair in de industrie. Ze moesten wedijveren met de klassieke eliminatiemethode, die door het werk van de Engelsman Wilkinson rond 1960 ook goed bruikbaar voor computers gemaakt was en die door de steeds sneller wordende computers ook attractief werd voor steeds grotere stelsels. Voor de ingeklemde balk kon die methode al redelijk wedijveren met de Lanczos methode voor 64 vergelijkingen in 1959.

Echter, doordat computers sneller werden konden ze grotere problemen aan en werd de vraag naar deze grotere problemen ook groter. Ik noem dat gekscherend wel eens de frustratie van de numeriek wiskundige: nauwelijks heb je een methode bedacht om een rekenprobleem van zekere omvang nauwkeurig en efficiënt op te lossen of de toepassende wereld wil vrijwel onmiddellijk nog weer grotere problemen oplossen. Nagenieten van het succes is er dus niet bij, je wordt constant uitgedaagd door nieuwe en moeilijker problemen. Dat is ook de grote aantrekkingskracht

van dit vak. Je zou dat kunnen vergelijken met topsport. Het geldt overigens voor vrijwel alle wetenschapsbeoefening.

Rekenboer in Utrecht

Na mijn tijd bij het Reactorcentrum in Petten ging ik in 1971 werken bij het computercentrum van de Utrechtse universiteit, als wiskundig adviseur. Tal van wetenschappers, voornamelijk chemici en fysici, moesten geholpen worden met hun rekenwerk. Dat was niet altijd eenvoudig want die onderzoekers wisten meestal zelf vrij goed van wanten. Alleen als ze volkomen vast zaten kwamen ze langs voor advies en dat betekende dat je vrij goed op de hoogte moest blijven van een uiteenlopend spectrum aan rekenmethoden en de achterliggende wiskunde. Het rekencentrum was gevestigd in hetzelfde gebouw als het wiskunde-instituut, maar de intellectuele afstand tussen beide instellingen werd als groot ervaren. Wij voelden ons rekenboeren. Of, zoals mijn latere promotor, de in 2004 overleden professor van der Sluis ooit eens zei: wij hielden ons bezig met intellectueel zaklopen. De wiskunde gaf voor tal van problemen elegante en compact geformuleerde oplossingen, maar als je die daadwerkelijk in getalvorm wilde uitrekenen kreeg je te maken met de beperkingen van de computer: de eindige rekensnelheid en de eindige rekenprecisie.

In die tijd had ik contact met de bij Shell werkende numericus Koos Meijerink en samen bogen we ons over voor die tijd zeer grote stelsels vergelijkingen: duizenden vergelijkingen met even zovele onbekenden. Deze ontstonden bij het modelleren van olie-stroming door poreuze aardlagen. We wierpen ons op twee verschillende aanpakken: een variant van de directe eliminatietechniek van Gauss: de zogenaamde Choleskimethode, en een variant van de Lanczosmethode: de zogenaamde Geconjugeerde Gradiënten methode. De Choleskimethode was voor onze problemen veel te duur (in termen van rekentijd en benodigde geheugenruimte), de Geconjugeerde Gradiënten methode was verdacht omdat deze zich onbegrijpelijk gedroeg in de eindige rekenprecisie van een computer. We hadden opgemerkt dat de Choleskimethode bij het rekenen aan ons soort grote stelsels al snel tot kleine getallen leidde als tussenresultaat. Die kleine tussenresultaten lieten we dus maar gewoon weg in de hoop dat dat het totale resultaat niet al te zeer zou beïnvloeden. We noemden dat, terecht, de Incomplete Choleskimethode. Die aanpak leverde geen goede benaderingen voor de ge-

zochte oplossing op. Echter, als je met de gevonden benaderingen voor de oplossing weer volgens het recept van Krylov een ruimte opbouwde en daarbij te werk ging met de zojuist genoemde Geconjugeerde Gradiënten methode, dan kreeg je opeens een zeer snel werkende methode (ICCG): de Incomplete Choleski Conjugate Gradients-methode [5].

Incomplete Choleski Conjugate Gradients

In 1972 publiceerden we ons eerste rapport over deze methode. Werd dat een onmiddellijk succes? Ik herinner me dat we bij onze eerste voordrachten over deze methode een beetje wantrouwend werden bekeken: we voerden immers het Choleskiproces beroerd uit en we combineerden dat dan nog eens met methoden waarvan men in Zürich al had laten zien dat die uiterst verdacht waren. Beroerd en verdacht lijken nu niet precies de ideale ingrediënten voor wat een zeer succesvolle methode zou blijken te zijn. Het publiceren werd een martelgang: opeenvolgende tijdschriften weigerden publicatie. We zouden de rekenvoorbeelden te gunstig gekozen hebben en in feite zouden we niets nieuws te zeggen hebben. Toen de gezaghebbende Stanfordiaan Golub in 1975 toevallig op doorreis was in Utrecht, vertelde Van der Sluis over onze resultaten. Golub toonde zich zeer sceptisch: er was al zoveel onderzoek op dit gebied geweest in Amerika en hij kon zich dan ook niet voorstellen dat er nu plotseling iets nieuws uit de hoge hoed getoverd kon worden. Toch nieuwsgierig geworden nodigde hij mij uit voor een bezoek aan Stanford voor de zomer van 1975. Daar raakte hij overtuigd van de kracht van ICCG, onder meer doordat hij een eigen promovendus de zaak ook eens had laten doorrekenen en toen kwam er schot in de zaak. Ons artikel werd eindelijk gepubliceerd in 1977.

De Amerikaanse fysicus David Kershaw was een van de eersten die de ICCG-methode gebruikte voor het doorrekenen van stelsels die voortvloeiden uit de laser-techniek en publiceerde enthousiast de resultaten in 1978. Dat was het feitelijke begin van de zegetocht van deze methode. Als er eenmaal één schaap over de dam is, volgen er meer en al gauw kwamen er de nodige citaties. De groep rond Simon Polak bij het Natlab van Philips was de eerste die, buiten Shell Exploratie waar Koos Meijerink werkte, de ICCG-methode oppikte voor het rekenen aan halfgeleiders. Op numeriek wiskundige congressen werden incomplete decompositietechnieken gemeengoed en ook de geconjugeerde gradiëntenmethode werd plotseling zeer populair (daar had ook

de Engelsman John Reid behoorlijk aan bijgedragen). Opmerkelijk dat de geconjugeerde gradiëntenmethode, in 1952 gepubliceerd door Hestenes en Stiefel [6], er ongeveer vijftientig jaar voor nodig had om door te breken als grensverleggende rekentechniek. Onze incomplete Choleskitechniek had ongeveer een jaar of tien nodig om definitief door te breken. Dat zijn geen unieke situaties; het geldt voor veel nu populaire rekenmethoden. In 1986 bedachten de Amerikanen Saad en Schultz de GMRES-methode, die geschikt was voor nog weer een bredere klasse van problemen dan de Geconjugeerde Gradiëntenmethode (voor de laatste methode moesten de coëfficiënten van het stelsel aan speciale eisen voldoen). Ook voor GMRES duurde het ongeveer een jaar of tien voordat het de method of choice zou worden.

ICCG en zaken daaromheen leidden er ook toe dat ik, nog steeds als medewerker bij het Academisch Computer Centrum Utrecht (ACCU), dan eindelijk in 1982 zou promoveren en twee jaar later werd ik benoemd tot hoogleeraar in Delft (bij de groep rond Piet Wesseling). Daar ontmoette ik Peter Sonneveld: een vat vol ideeën maar met merkwaardig weinig publikatiedrift. Zulke personen zijn in de huidige academische cultuur moeilijk te handhaven, maar naar mijn overtuiging zou elke groep zich de handen moeten dichtknijpen zo'n persoon in hun midden te hebben als brandpunt van de discussies en als aanjager van ideeën. Jaren later zou er gepleit worden voor wiskunde-onderzoeksgroepen van voldoende omvang om een beter klimaat te scheppen voor funderend en toepasbaar onderzoek in samenhang. Zulke groepen zouden beter in staat moeten zijn om diverse vormen van talent te bundelen. Daarover later nog.

Combineren van methodes

Sonneveld bestookte me met zijn gedachten over varianten van de geconjugeerde gradiënten methode die bruikbaar zouden moeten zijn voor algemenere stelsels vergelijkingen. Zo'n variant was begin jaren zeventig door Fletcher al voorgesteld, maar deze zogenaamde Bi-Conjugate Gradientmethode was niet echt doorgebroken. Menig toegepast wiskundige stond uiterst sceptisch tegenover Bi-CG: een methode met menige valkuil. Sonneveld liet zien hoe je deze methode met vrijwel dezelfde inspanning (rekentijd en geheugenruimte) in feite twee keer herhaald kon uitvoeren: in feite dus twee keer zo snel bij de oplossing, als de methode zich tenminste een beetje netjes gedroeg. Het gedrag van Bi-CG

was zeer wispelturig en Sonneveld kreeg met zijn presentatie voor de Nederlandse numeriek wiskundigen geen poot aan de grond. Zijn CGS methode was in feite al bekend begin jaren 80, maar werd in 1989 na een moeizaam proces gepubliceerd [7]. De methode werd een groot succes en het betreffende artikel kon zich al snel temidden van meest geciteerde wiskundepublicaties scharen. Toen ik eindelijk zo'n beetje begreep welk gedachtengoed er achter CGS schuil ging, kon ik zelf de Bi-CGSTAB-methode presenteren in 1992: in feite een goedkope, maar zeer effectieve, combinatie van BiCG en GMRES (in plaats van twee keer BiCG). Onderwijl verruilde ik in 1990 mijn aanstelling in Delft voor een aanstelling aan de Utrechtse universiteit.



Presentatie van ICCG

Het betrekkelijke van een citatie index

De publicatie over Bi-CGSTAB werd in 2000 door het ISI (Institute for Scientific Information) aangemerkt als het wereldwijd meest geciteerde wiskundeartikel dat in de jaren negentig gepubliceerd was: 379 citaties op dat moment. Sonneveld's CGS-publicatie kon toen ook al op 338 citaties bogen. Ter vergelijking: het beroemde artikel van Andrew Wiles uit 1995 met het bewijs voor de laatste stelling van Fermat had toen nog geen honderd citaties. Dit geeft overigens ook aan dat je met enige reserve naar citatietellingen moet kijken. Deze citatieaantallen zijn ontleend aan de gezaghebbende Science Citation Index. Het wapenfeit leverde veel publiciteit op: onder meer een groot artikel in de Volkskrant en zelfs het gratis reizigersblad Metro gaf een klein berichtje over deze opmerkelijke citatiescore.

Geciteerd worden is voor wetenschappers een belangrijke zaak: het toont min of meer objectief hun zichtbaarheid en belang aan. Men kijkt daarbij meestal naar de citaties die gedaan worden in de eerste paar jaar na het verschijnen van een artikel. Of dat het hele verhaal vertelt is zeer de vraag. Een voorbeeld: in 1995 werden

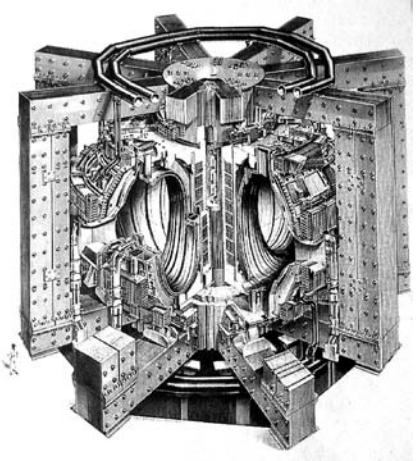
de citatiescores van Utrechtse wetenschappers geanalyseerd, als vingeroefening voor komende VSNU-visitaties. Mijn Bi-CGSTAB-publicatie (1992) kon toen bogen op 75 citaties en dat was veel. De meeste artikelen komen niet verder dan slechts een paar citaties. Zoals gemeld, na de eeuwwisseling was het het wereldwijd meest geciteerde artikel met 379 citaties. Op dit moment ligt haar citatiescore ruim boven de duizend. De meeste citaties, dat wil zeggen het grootste succes, kwamen dus pas tien jaar na de publikatiedatum. Die aantallen worden evenwel bij VSNU-visitaties niet meer meegeteld: daar sluiten de citatievensters na tien jaar. Op die manier zouden een behoorlijk aantal succesverhalen uit de (toegepaste) wiskunde nauwelijks, of helemaal niet, zichtbaar zijn geworden. De Lanczos-methode werd pas populair ongeveer dertig jaar na haar verschijning, voor de geconjugeerde gradiënten methode zou de bloei na vijftig jaar komen.

Het belang van geschiedenis van wiskunde

Soms komen methoden op een wonderlijke en schijnbaar toevallige manier tot stand. Weer een voorbeeld uit de eigen praktijk. Midden jaren negentig nam onze numerieke groep deel aan het NWO-project Massaal Parallel Rekenen. Dit was een door de Utrechtse plasmafysicus Hans Goedbloed geleid project, waar verschillende onderzoeksgroepen in participeerden: plasmafysici, geologen, astronomen, oceanografen, informatici, en numeriek wiskundigen. De gemeenschappelijke noemer van het project was dat de deelnemers tegen enorm grote stelsels vergelijkingen aanliepen die allemaal voortvloeiden uit vrijwel hetzelfde type wiskundige modelering.

De informatici en numeriek wiskundigen moesten helpen om het mogelijk te maken deze enorme stelsels op te kunnen lossen met de moderne grootste computers (die inmiddels parallel waren geworden). Goedbloed had het project intelligent opgezet. Er werden ons kleine stelseltjes als worstjes voor de neus gehangen, er waren grote stelsels waarvan het denkbaar was dat we die ook zouden kunnen helpen oplossen en aan de horizon waren de zeer grote stelsels geplaatst als grote uitdaging. Een van de uitdagingen was het doorrekenen van de modellen voor de stabiliteit van het plasma in een Tokamak-fusioreactor.

Rond 1994 kon dat gemodelleerd worden met rekenmodellen met ongeveer maximaal 20.000 vrijheidsgraden (onbekenden). Dat gebeurde met de Lanczosmethode en de



Tokamak fusioreactor

daarmee verwante Arnoldimethode. Dit aantal vrijheidsgraden was volstrekt onvoldoende om een realistisch beeld te krijgen. Omdat wij goed op de hoogte waren met beide methoden leek onze groep geschikt om de modellen te helpen opschalen. Goedbloed had rekenmodellen met meer dan 500.000 vrijheidsgraden aan onze gezichtseinder geplaatst, ver buiten het bereik van beide methoden. Dit aantal was nodig om instabiliteiten in het plasma, die inmiddels in de praktijk geconstateerd waren (met veel schade), te kunnen voorspellen. Rond diezelfde tijd hadden we contact met de Utrechtse chemicus Joop van Lenthe, die ook grote stabiliteitsproblemen doorrekende. Die haalde zijn neus op voor 'onze' Lanczosmethoden en wees op de successen die de chemici behaalden met de zogenaamde Davidsonmethode. Dit was een in 1975 door de Amerikaanse chemicus Davidson gepubliceerde rekenmethode [8]. Deze methode had bepaalde eigenaardigheden: hoe meer het stelsel leek op een uiterst gemakkelijk op te lossen stelsel, hoe beroerder het convergeerde en voor diagonaalstelsels, stelsels waarin je de oplossing in één oogopslag kon aflezen, was de methode helemaal niet in staat die oplossing te berekenen. Numeriek wiskundigen liepen met een boogje om de methode heen; het aantal publicaties uit numeriek wiskundige kring over Davidson's methode was tot rond 1992 op de vingers van één hand te tellen.

Mijn medewerker Gerard Sleijpen en ik bestudeerden deze methode, daartoe min of meer uitgedaagd door van Lenthe, maar we konden er evenmin als anderen onze vingers achter krijgen. Op dat moment bleek de kracht van een breed geschakeerd Mathematisch Instituut. Dat instituut herbergt een klein, maar uiterst fijn, groepje Geschiedenis van de Wiskunde, toenmaals onder leiding van Henk

Bos. Henk Bos had een afstudeerstudente, Anjet den Boer, die met hulp van Gerard Sleijpen oude artikelen van de astronoom-wiskundige Jacobi (1804-1851) bestudeerde: artikelen uit de periode 1845-1850. Jacobi bestudeerde lineaire stelsels, maar kon destijds nog niet over moderne matrixnotatie beschikken. Dat maakt zijn artikelen moeilijk leesbaar. Hij ontwierp een paar zeer krachtige methoden die echter later in de vergetelheid raakten. Een van die methoden, de Jacobi-diagonalisatiemethode, werd rond 1950 herontdekt door de wiskundige John von Neumann en is nog immer populair. Anjet den Boer bestudeerde de geschiedenis rond deze artikelen en herschreef de oude technieken van Jacobi in moderne matrixnotatie. Dat leverde ons een schat van gegevens op. Zo werden wij een door Jacobi beschreven uitbreiding gewaar die hij gebruikte om zijn diagonalisatiemethode te versnellen. We hadden toen tamelijk snel in de gaten dat deze techniek ook gebruikt kon worden om de Davidson's methode te corrigeren. Dit leidde in 1996 tot publicatie van de inmiddels ook al weer bekende Jacobi-Davidsonmethode [9]. Een aantal elementen kwam in deze methode elegant samen: de reductie van een probleem tot een kleiner probleem in de geest van Krylov, de modificatie van het probleem van Davidson en de correctie-aanpak van Jacobi. De mogelijkheid om tot deze methode te komen werd geschapen door de omgeving waarin we verkeerden: een uitstekende afdeling Theoretische Chemie, die zeer goed op de hoogte was van moderne rekenmethoden, de aanwezigheid van een groepje Geschiedenis van de Wiskunde (uniek in Nederland) en de intellectuele druk geleverd door uitdagende problemen in het NWO-project. De verstrekte mogelijkheden van de nieuwe Jacobi-Davidson methode konden verder uitgebuit worden door de beschikbaarheid van een stel uitstekende promovendi. Bovendien zaten we in die tijd nog vrij ruim in de afstudeerstudenten. Met deze methode konden uiteindelijk rond de eeuwwisseling de zeer grote rekenmodellen rond de Tokamak-stabiliteit worden opgelost. Anjet de Boer, die een zeer leesbare scriptie had geschreven, koos voor een andere voortzetting van haar loopbaan. Onder het pseudoniem Anjet Daanje schreef zij onder meer de roman *De blinde Fotograaf* (1997).

Zo aan het einde van een actieve loopbaan komt ook het moment om de eigen publicatielijst nog eens te overzien. Wetenschappers informeren het publiek op uiteenlopende wijze: de publicaties voor vakgenoten, voor toegepaste wetenschappers ook de publikaties

voor wetenschappers in andere vakgebieden en de industrie, lezingen voor uiteenlopende gezelschappen (van uiterst wetenschappelijk tot vrij populair), populariserende bijdragen, Patenten, en wat dies meer zij. Ik heb zelf met veel plezier meegewerkt aan een paar boeken op mijn vakgebied. Enkele daarvan hebben redelijk hoge oplagen bereikt, echter voordat men daar verkeerde conclusies aan verbindt: doorgaans komen de revenuen daarvan niet direct bij de auteurs terecht. In ons geval werd een deel van de opbrengst besteed aan een fonds om studenten te kunnen laten deelnemen aan congressen in Amerika.

Het afrekenen op resultaat

Ik heb u geprobeerd uit te leggen met wat voor soort problemen ik mij lange tijd heb bezig gehouden. Die problemen vond ik zelf ontzettend belangrijk, maar dat vindt elke wiskundige van zijn problemen. Daarin hebben ze meestal dan ook gelijk. Ik had het voordeel dat mijn problemen van direct belang waren voor de toepassende wereld; men stond als het ware te trappelen van ongeduld om deze toegang tot de wiskunde te kunnen gebruiken. Desondanks moet je constateren dat oplossingen die achteraf zeer succesvol bleken toch de nodige tijd vergden om werkelijk ingang te vinden in de gebruikerswereld. Daar kun je je over verbazen, maar het is minder vreemd dan het lijkt. De computerprogramma's die gebruikt worden voor ingewikkelde simulaties zijn groot en complex. Men is met de hebbelikheden, mogelijkheden en beperkingen daarvan vertrouwd. Men leert zich als het ware aan te passen aan die beperkte mogelijkheden. Dan komt er opeens een nieuwe rekenmethode waarmee grotere en moeilijker stelsels kunnen worden uitgerekend. Wordt dat meteen omarmd? Natuurlijk niet. De eerste fase bestaat uit het zich er van te vergewissen dat die nieuwe methode ook inderdaad beter is. Niemand van de kennissen gebruikt hem immers en dat rechtvaardigt een zeker wantrouwen. Men moet er dus eerst wat ervaring mee op doen. Als dan blijkt dat er inderdaad enige gerechtvaardigde hoop bestaat op betere en sneller resultaten, dan komt een lastiger situatie. Niet alleen moet er, vaak ingrijpend, in een grote computercode worden gesneden (de vervanging van het rekenhart), echter die code moet nu ook geschikt gemaakt worden voor die grotere problemen. De onderdelen van die codes zijn vaak gemaakt door verschillende programmeurs, velen daarvan zijn al verdwenen naar andere werkplekken, en men moet dus inbreken in het gedachtengoed van anderen.

Men moet wel heel zeker zijn van zijn zaak om aan deze klus te beginnen. Men moet immers de inspanningen ook met enig succes kunnen rapporteren aan de werkleider. Nieuwe wiskundige technieken, zelfs als ze al aangeboden worden als hapklare in programmeerbare vorm gegoten brokken voor problemen die door iedereen in deze programmeerwereld begrepen worden, vinden dus maar zeer langzaam ingang. Ik bevind mij met mijn problemen, zoals al eerder opgemerkt, aan de uiterst toepasbare kant van de wiskunde. Via dit soort rekentechnieken wordt vaak een venster geopend op de onderliggende nog weer verder van de directe toepassing gelegen, maar o zo essentiële diepere wiskunde. Men kan zich dan een beetje voorstellen hoe lang het duurt voordat die algemeen toepassing vindt. Dat brengt wiskundigen in een lastig parket in deze op direct resultaat afrekenende samenleving. Organisaties die het fundamentele onderzoek financieren worden door hun geldgevers ook weer afgerekend op het al dan niet zichtbare resultaat. Het mooiste is het wanneer dat onderzoek meteen een grens verlegt, een beroerde ziekte oplost, de televisie nog weer platter en goedkoper maakt, de auto schoner, het verkeer soepeler laat doorstromen, en het klimaat op korte termijn verbeterd of schonere energie tegen lagere prijs oplevert. Dat dwingt de onderzoekers min of meer om hun geldschietters gouden bergen voor te spiegelen. Zij willen hun doen geloven in een toekomst waar ze zelf een klein radertje in zijn. Dat is lang niet altijd een valse voorstelling van zaken, maar de aangesprokenen, zowel als de boodschapper, vergissen zich meestal in de tijdschaal waarmee nieuwe inzichten de samenleving penetreren. De zichtbare resultaten van het fundamentele onderzoek zijn meestal niet binnen een jaar of twintig te verwachten. Echter als dat onderzoek twintig of meer jaren geleden niet had plaats gevonden, dan hadden we met elkaar hier niet gezeten zoals we nu zitten. De economie was gaan achter lopen, de welvaart zou gedaald zijn, onze gezondheidszorg zou zijn blijven steken, enfin, vult u zelf maar verder in.

Mijn betoog moet nu ook weer niet uitgelegd worden als een pleidooi om elke zichzelf funderend noemend onderzoeker zonder voorbehoud maar in de watten te leggen. Natuurlijk moeten er kwaliteitscriteria aangelegd worden, we moeten echter voorzichtig zijn met het afrekenen op direct economisch of maatschappelijk nut. Dat is er meestal niet. Wat goed onderzoek in elk geval wel oplevert zijn uitstekende jonge onderzoekers die ge-

traint zijn in zich het hoofd te breken over ingewikkelde materie. Zulke denkkraft is in onze maatschappij op tal van plaatsen broodnodig en gewenst. Er is dus wel degelijk ook een direct nut op de korte termijn.

Het belang van clustervorming

Daarmee kom ik op een zorgelijk punt. Eerder noemde ik de beschikbaarheid van afgestudeerde studenten als een der sleutels voor succesrijk onderzoek. Helaas zijn in de afgelopen decennia de studentenaantallen in de wiskunde in Nederland zorgelijk gedaald. Men kan dat toeschrijven aan de talrijke verbeteringen in ons middelbaar onderwijs, zoals ik een kunsthistoricus ooit cynisch hoorde opmerken, of aan veranderingen in de maatschappij, feit blijft dat de aantallen veel te laag zijn geworden voor een samenleving die het van innovatieve slagkracht moet hebben om haar welvaart op peil te houden. De wiskundigen hebben deze ontwikkelingen niet rustig afgewacht, integendeel, ze hebben geprobeerd op allerlei wijzen het tij te keren: intensieve voorlichting, aanpassingen in de aangeboden studiepakketten en meer contact met middelbare scholen.

Omdat de financiering van het universitaire onderzoek mede afhangt van de studentenaantallen, kwam ook de omvang van het onderzoek onder druk te staan. De KNAW bracht de bedreigingen voor het onderzoek in het TWON-rapport duidelijk in kaart [10]. De wiskundigen bundelden hun krachten in het zogenaamde OOW en in een unieke samenwerking met NWO, en zeker door de niet-aflatende inzet van de NWO-ers Annejet Meijler en Lex Zandee, werd een steunplan voor de wiskunde tot wasdom gebracht, zoals neer-

gelegd in de notitie *Nieuwe Dimensies Ruimer Bereik* [11]. Samen met collega Rien Kaashoek (emeritus VU), Chris Zaai en de eerder genoemde ruime NWO-steun, heb ik hier aan mee kunnen werken.

Alhoewel het allemaal beter had gekund, stemt het toch tot tevredenheid dat inmiddels een drietal clusters van wiskunde-onderzoek van start kon gaan met, voor de wiskunde, ruimere onderzoeksfinanciering. De financiën hiervoor werden geleverd door de ministeries van Economische Zaken (EZ) en Onderwijs, Cultuur en Wetenschap (OCW) en door NWO. Inmiddels lijkt het tij gunstig voor het opstarten van een vierde en mogelijk zelfs een vijfde cluster.

Clustering van het Master-onderwijs

Parallel hieraan loopt een unieke landelijke samenwerking van de Nederlandse wiskundigen bij het Master-onderwijs. Door de dalende aantallen studenten en de daardoor ook in omvang dalende staf, dreigde met name het onderwijsaanbod in de masteropleidingen te verschrompelen in diversiteit. De wiskundigen, onder lichte dwang van OCW en door de VSNU aangemoedigd, sloegen de handen ineen en ontwierpen een aantrekkelijke serie colleges die centraal gegeven worden.

Dit had meerdere voordelen. Zo konden onderwerpen aan bod komen die anders slechts op een publiek van slechts een paar studenten hadden kunnen rekenen (weinig effectief dus), de topwiskundigen konden zich zo presenteren aan een groter studentenpubliek en, ook niet onbelangrijk, studenten uit verschillende universiteiten ontmoeten elkaar in grotere groepen rond onderwerpen van wederzijds belang. Dit landelijke

masteronderwijs vult het lokale masteronderwijs substantieel aan en loopt nu al weer een paar jaar met groot succes.

Wiskunde staat midden in de maatschappij

Ik kom aan het eind van mijn beschouwingen. Heel veel is niet aan de orde gekomen, maar ik hoop u een blik op het leven van een toegepast wiskundige geboden te hebben. Wiskundigen hebben de reputatie enigszins van de wereld afgekeerde wetenschappers te zijn die in hun kamertjes eenzaam op hun problemen zitten te broeden. Die voorstelling, mocht zij ooit al karakteristiek zijn geweest, klopt heden ten dage zeker niet. Wiskunde wordt in interactie in grotere groepen bedreven met veel uitwisseling van ideeën. Het is een uitermate prettig en schoon bedrijf: ik ben dan ook mijn hele loopbaan met veel plezier naar mijn werkplek gegaan. Naast de jaren in Petten (Reactor Centrum) en bij het Computercentrum ACCU in Utrecht, heb ik zes jaar als hoogleraar in Delft gewerkt. De contacten daar opgedaan hebben mijn verdere loopbaan ingrijpend (en positief) beïnvloed. Zij hebben onder meer geleid tot mijn lidmaatschap van de huidige Netherlands Academy of Technology and Innovation (voorheen FORUM). Dat heeft mij een gevarieerd beeld gegeven van de impact van wetenschappelijk onderzoek op onze samenleving. De langste tijd heb ik doorgebracht in Utrecht bij het Mathematisch Instituut: een bijzonder collegiale werkomgeving van zeer hoog wetenschappelijk niveau. Toen ik nog bij het ACCU werkte kon ik vanuit de lift af en toe een glimp daarvan opvangen, nog niet bevroedende dat ik daar een substantieel deel van mijn actieve loopbaan zou doorbrengen. ←

References

- 1 R.S. Varga, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
- 2 Zie voor meer informatie over Krylov: www.math.uu.nl/people/vorst/kryl.html
- 3 C. Lanczos, 'Solution of systems of linear equations by minimized iterations', *J. Res. Natl. Bur. Stand.* **49** (1952), pp. 33–53.
- 4 M. Engeli, T. Ginsburg, H. Rutishauser, E. Stiefel, *Refined Iterative Methods for the Computation of the Solution and the Eigenvalues of Self-Adjoint Boundary value Problems*, Birkhäuser, Basel, 1959
- 5 J.A. Meijerink, H.A. van der Vorst, 'An Iterative Solution Method for Linear Systems of Which the Coefficient Matrix is a Symmetric M-Matrix', *Math. Of Comp.* **31** (1977), pp. 148–162.
- 6 M.R. Hestenes, E. Stiefel, 'Methods of Conjugate Gradients for Solving Linear Systems', *J. Res. Natl. Bur. Stand.* **49** (1952), pp. 409–436.
- 7 P. Sonneveld, 'CGS: A fast Lanczos-type Solver for nonsymmetric linear Systems', *SIAM. J. Sci. Statist. Comp.* **10** (1989), pp. 36–52.
- 8 E.R. Davidson, 'The iterative Calculation of a Few of the lowest Eigenvalues and corresponding Eigenvectors of large real symmetric matrices', *J. Comp. Phys.* **17** (1975), pp. 87–94.
- 9 G.L.G. Sleijpen, H.A. van der Vorst, 'A Jacobi-Davidson iteration Method for linear Eigenvalue Problems', *SIAM J. Matrix Anal. Appl.* **17** (1996), pp. 401–425.
- 10 *De toekomst van het wiskunde-onderzoek in Nederland*, Verkenningen, deel 1, Koninklijke Nederlandse Akademie van Wetenschappen, 1999.
- 11 *Nieuwe dimensies, ruimer bereik – Een nationale strategie voor wiskundeonderzoek en gerelateerde master opleidingen*, Overleg Onderzoekscholen Wiskunde, NWO, 2002.