Hoe betrouwbaar zijn onze (grafische) rekenmachines?

Stefan Becuwe Annie Cuyt

Departement wiskunde en informatica
Universiteit Antwerpen

oktober 2002

Vlaamse wetenschapsweek

Motivatie

Vragen:

- bestaat er een x waarvoor $x \times (1/x) \neq 1$?
- hoeveel snijpunten kunnen 2 rechten gemeenschappelijk hebben?
- bereken $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k}$ en $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2}$

Antwoorden m.b.v. gekende wiskundige technieken of gezond verstand

- nee
- 0, 1 of ∞
- ∞ en $\frac{\pi^2}{6}$

Antwoorden m.b.v. een computer

- ja!. . . vb. $x = 1 + 257736490 \times 2^{-52}$
- tja, dat is niet te voorspellen
- machineafhankelijk; beide uitdrukkingen zeker eindig

Ik tel tot 6 . . .

Bereken volgende uitdrukkingen:

$$2 - 1$$

$$\left(\frac{1}{\cos(100\pi + \pi/4)}\right)^2$$

$$3 \times \frac{\tan(\arctan(10000))}{10000}$$

$$\left(\left(\ldots\left(\sqrt{\sqrt{\ldots\sqrt{4}}}\right)^2\ldots\right)^2\right)^2$$

$$5 \times \frac{(1 + e^{-100}) - 1}{(1 + e^{-100}) - 1}$$

$$\frac{\ln(e^{6000})}{1000}$$

Antwoorden m.b.v. een computer

$$\begin{aligned}
\text{een} &= 1.0000000000000000 \\
\text{twee} &= 2.0000000000001110 \\
\text{drie} &= 2.9999999999971618 \\
\text{vier} &= 2.7182818081824731 \\
\text{vijf} &= \texttt{NaN} \\
\text{zes} &= \texttt{Infinity}
\end{aligned}$$

Antwoorden m.b.v. rekenmachines

$$\begin{aligned}
\text{een} &= 1 \\
\text{twee} &= 1.999999998, 1.999999999, 2 \\
&= 2.00000000001, 2.0038707 \\
\text{drie} &= 2.99979, 2.9999989, 3, 3.000000017 \\
\text{vier} &= \ldots \\
\text{vijf} &= \texttt{E}, \texttt{ErrorAri}, \texttt{divisionbyzero} \\
\text{zes} &= \texttt{E}, \texttt{ErrorOFL}, \texttt{overflow}
\end{aligned}$$

Enkele andere voorbeelden

(i) $|3 \times (4/3 - 1) - 1|$

(ii) sin(22 rad)

(iii) 2.5^{125}

(iv) $e^{125 \ln 2.5}$

Antwoorden m.b.v. een computer

(i) `2.22044604925031e-16`

(ii) `-8.85130929040387e-03`

(iii) `5.52714787526044e+49`

(iv) `5.52714787526045e+49`

Antwoorden m.b.v. rekenmachines

(i) `0`, `1 10^-11`, `1.-09`, `1E-13`, `2.2E-016`

(ii) `-0,01`, `-0.0088513`, `-0.00885130929`, `-8.8513092 10^-3`

(iii) `5,5E+049`, `5.52714 10^49`, `5.527147875 49`, `5.52714787526E49`

(iv) `4.00768645579E21`, `4.007686456 21`, `5,5E+049`, `5.52714 49`, `5.52715 49`

Een eenvoudige uitdrukking. . .

Gegeven $a = 77617$ en $b = 33096$, bereken de waarde van volgende uitdrukking:

$$y = 333.75b^6 + a^2(11a^2b^2 - b^6 - 121b^4 - 2) + 5.5b^8 + \frac{a}{2b}$$

Antwoorden te kiezen uit (elk van deze antwoorden kan worden bekomen)

$$1.18059\ldots \times 10^{21}$$

$$-1.18059\ldots \times 10^{21}$$

$$5.76461\ldots \times 10^{17}$$

$$6.33825\ldots \times 10^{29}$$

$$1.1726\ldots$$

$$-0.827396\ldots$$

Het juiste antwoord is. . . $-\frac{54767}{66192} \approx -0.827396\ldots$

Een “klassiek” stelsel. . .

In cryptische vorm:

$$\sum_{j=1}^{n}(1+i)^{j-1}x_j = \frac{(1+i)^n - 1}{i} \qquad 1 \leq i \leq n$$

Vertaald wil dit voor de coëfficiëntenmatrix zeggen:

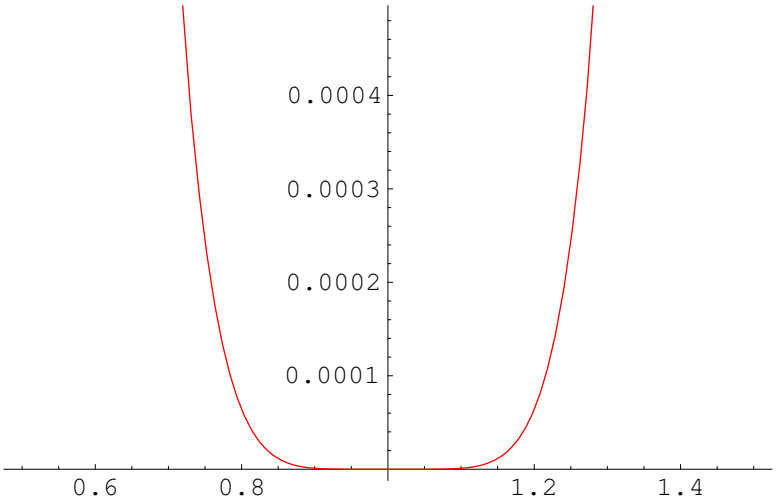
$$\begin{pmatrix} 2^0 & \dots & 2^{n-1} \\ \vdots & & \vdots \\ (n+1)^0 & \dots & (n+1)^{n-1} \end{pmatrix}$$

Als rechterlid nemen we steeds de som van alle elementen op die rij, d.w.z.: de oplossing van dit stelsel is de vector die alleen maar uit 1’tjes bestaat.

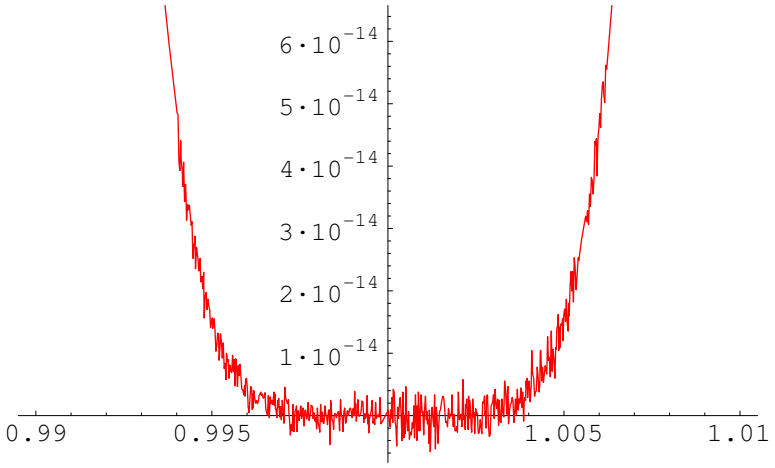
Laten we het rekenwerk echter over aan een computer, dan krijgen we voor $n = 15$ een oplossingsvector met componenten (in absolute waarde) tussen 0 en 10000.

Bepaal nulpunten a.d.h.v. een grafiek. . .

Gegeven:



Het is duidelijk dat (een) kandidaat–nulpunt(en) in de buurt van 1 liggen. Daarom vergroten we dat deel uit:



Het antwoord op de vraag is dus. . . heel moeilijk te bepalen!

Ter informatie: dit zijn afbeeldingen van volgende functie:

$$y = (x-1)^6$$

Waarom. . .

- . . . worden de basisregels van onze wiskunde met de voeten getreden?
- . . . gaan wiskundige eigenschappen verloren?
- . . . zien we geen betrouwbare grafieken? Kunnen we wel betrouwbare grafieken maken?
- . . . mogen we niet “zo maar” iets op een computer uitrekenen?
- . . . is dit allemaal nodig? Misschien zijn dit maar “academische” problemen?

Talstelsels

Wij hebben de gewoonte te rekenen in "basis 10", het decimale of tiendelige talstelsel. Een voorbeeld: 181.25 kunnen we opsplitsen als

$$1 \times 10^2 + 8 \times 10^1 + 1 \times 10^0 + 2 \times 10^{-1} + 5 \times 10^{-2}$$

In plaats van machten van 10 te gebruiken, kunnen we ook machten van 2 gebruiken:

$$128 + 32 + 16 + 4 + 1 + \frac{1}{4}$$
$$1 \times 2^7 + 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^2 + 1 \times 2^0 + 1 \times 2^{-2}$$

In "basis 2" of het binaire talstelsel, wordt dit getal geschreven als 10110101.01_2

Andere veel voorkomende talstelsels zijn het octale en hexadecimale. Conversies tussen de talstelsels zijn meestal voorhanden op rekenmachines (bin, oct, hex).

Talstelsels zijn te vergelijken met de eenheden die voor hoeken worden gebruikt: 60–delige graden, radialen, enz.

Gevolgen in het “echte” leven

- hoe “1/10” mensenlevens kan kosten
- verkiezingen: $4.97\% \neq 5.0\%$
- afronden kan ook geld kosten

Help!

Het moment is aangebroken om op zoek te gaan naar antwoorden op de Waarom?–vragen.

Misschien is het ook interessant dat, als we iets op verschillende computers uitrekenen, we ook steeds hetzelfde antwoord te zien krijgen. Er is duidelijk nood aan afspraken, een soort “reglement”.

Wie weet zijn er wel interessante rekenregels die ons veel kunnen verduidelijken. . .

Tijd om de flops achter ons te laten!

FLOPs

FLOP = **Fl**oating-point **Op**eration

Voorbeeld: $(3 + 7)/2 - 12$ "kost" 3 FLOP.

Ter informatie:

computer	snelheid
AMD Athlon, Intel Pentium III, PowerPC G4	$>$ 1 GFLOP

Snelste machine vandaag: NEC Earth Simulator:

- $>$ 35 TFLOP (1 TFLOP $= 10^{12} =$ 1000000000000 FLOP)
- 5120 processoren
- 10 TB geheugen
- 700 TB schijfruimte
- afmetingen: 4 tennisvelden, 3 verdiepingen hoog

Doel: model voor $\cos(x)$ door alleen gebruik te maken van de basisbewerkingen ($+$, $-$, $\times$, $/$).

Is $1 = 1$?

Veronderstel dat we kunnen rekenen met 8 cijfers

$$1/3 \approx 0.3333333 \qquad \text{fout} = 3 \times 10^{-8}$$
$$3 \times (1/3) \approx 0.9999999 \qquad \text{fout} = 1 \times 10^{-7}$$

Elke bewerking *kan* een fout introduceren.

We rekenen *niet noodzakelijk* “exact”.

Getalverzamelingen

natuurlijke getallen ($\mathbb{N}$) 0, 1, 2, . . .

gehele getallen ($\mathbb{Z}$) . . . , -2, -1, 0, 1, 2, . . .

rationale getallen ($\mathbb{Q}$) $-2/7$, $1/3$, . . .

reële getallen ($\mathbb{R}$) $\sqrt{2}$, e, π, . . .

complexe getallen ($\mathbb{C}$) $a + b \cdot \mathtt{i}$

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

Floating-point getallen (voorstelling)

basis β, precisie p en exponent e

$$\pm d_0.d_1 \ldots d_{p-1} \times \beta^e$$

Voorbeeld:

$$\begin{aligned} 0.1 &= 1.00 \times 10^{-1} \\ &= 0.01 \times 10^{1} \\ &\approx 1.10011001100110011001101 \times 2^{-4} \end{aligned}$$

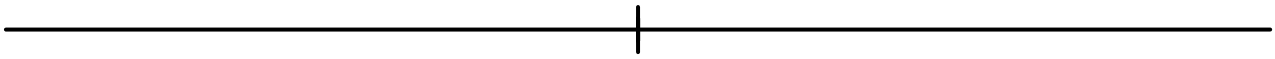
Belangrijke opmerking: 0.1 is niet door een eindig aantal 0'en en 1'en voor te stellen in basis 2. . .

Ter informatie: klassieke parameters op een computer zijn:

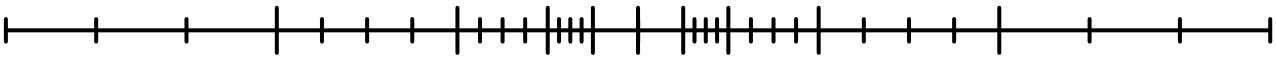
$\beta = 2$ $p = 24$	$-127 < e < 128$	"single precision"
$\beta = 2$ $p = 53$	$-1023 < e < 1024$	"double precision"

Getalverzamelingen (vervolg)

Reële getallen ($\mathbb{R}$): oneindig veel (niet noodzakelijk uit te drukken in een eindig aantal decimale cijfers)



Floating-point getallen ($\mathbb{F}$): beperkt aantal; underflow, overflow
Voorbeeld: $\beta = 2$, $p = 3$, $-1 \leq e \leq 2$



Eigenschappen . . .

. . . zoek de verschillen

$\mathbb{R}, +, \cdot$	$\mathbb{F}, +, \cdot$
$x + y \in \mathbb{R}$	$x + y \overset{?}{\in} \mathbb{F}$
	$O(x + y) \doteq x \oplus y \in \mathbb{F}$
$x + (y + z) = (x + y) + z$	$x \oplus (y \oplus z) \neq (x \oplus y) \oplus z$
$0 \in \mathbb{R}$	$\pm 0 \in \mathbb{F}$
$x + 0 = x = 0 + x$	$x \oplus 0 = x = 0 \oplus x$
$-x \in \mathbb{R}$	$-x \in \mathbb{F}$
$x + (-x) = 0$	$x \oplus (-x) = 0$
$x + y = y + x$	$x \oplus y = y \oplus x$

$\mathbb{R}, +, \cdot$	$\mathbb{F}, +, \cdot$
$x \cdot y \in \mathbb{R}$	$x \cdot y \overset{?}{\in} \mathbb{F}$
	$O(x \cdot y) \doteqdot x \odot y \in \mathbb{F}$
$x \cdot (y \cdot z) = (x \cdot y) \cdot z$	$x \odot (y \odot z) \neq (x \odot y) \odot z$
$1 \in \mathbb{R}$	$1 \in \mathbb{F}$
$x \cdot 1 = x = 1 \cdot x$	$x \odot 1 = x = 1 \odot x$
$1/x \in \mathbb{R}$	$1/x \overset{?}{\in} \mathbb{F}$
	$O(1/x) \doteqdot 1 \oslash x \in \mathbb{F}$
$x \cdot (1/x) = 1$	$x \odot (1 \oslash x) \neq 1$
$x \cdot y = y \cdot x$	$x \odot y = y \odot x$
$x \cdot (y + z) = x \cdot y + x \cdot z$	$x \odot (y \oplus z) \neq x \odot y \oplus x \odot z$
	$\beta \odot (y \oplus z) = \beta \odot y \oplus \beta \odot z$

$\Rightarrow$ . . . problemen met wiskundige identiteiten!

Identiteitsproblemen

Volgende uitspraken gelden niet (altijd) meer in $\mathbb{F}$:

$$a \times (b + c) = a \times b + a \times c$$

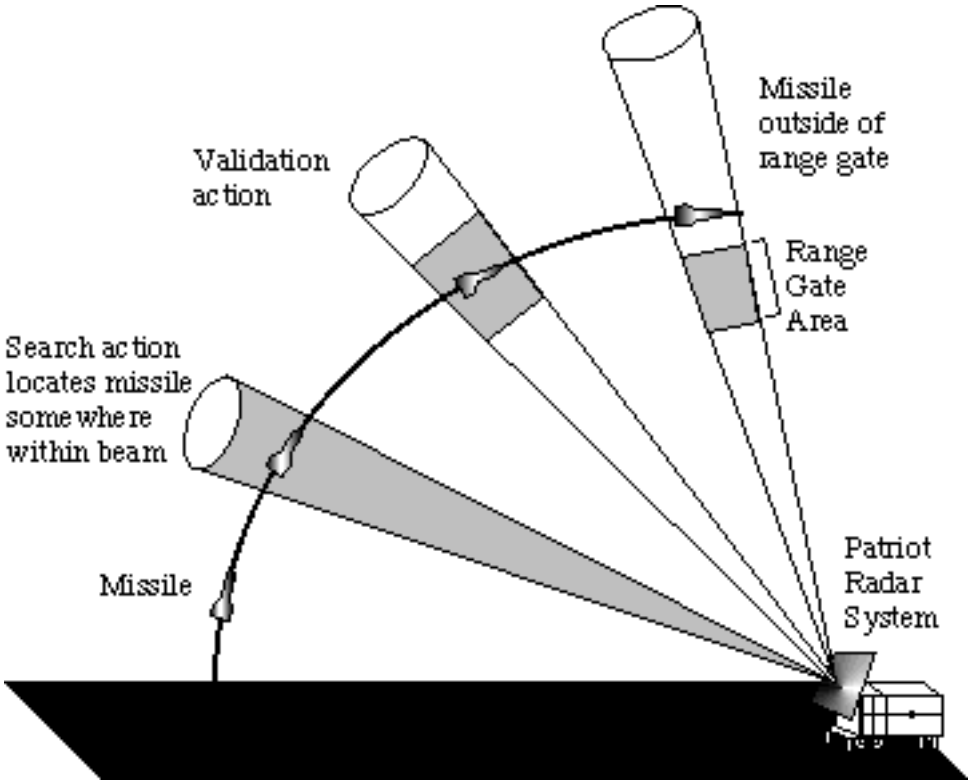
$$x - y = 0 \Rightarrow x = y$$

$$\sqrt{x}^2 = x$$

$$x/y = x \times (1/y)$$

Hoe “1/10” mensenlevens kan kosten

uur	berekende tijd (s)	fout (s)	afstand (m)
0	0	0	0
1	3599.9966	.0034	7
8	28799.9725	.0275	55
20	71999.9313	.0687	137
48	172799.8352	.1648	330
72	259199.7528	.2472	494
100	359999.6567	.3433	687



Verkiezingen: $4.97\% \neq 5.0\%$

5 april 1992

Deelstaatverkiezingen in Schleswig–Holstein (Duitsland)

Er is een kiesdrempel van 5%, die de groene partij blijkbaar juist haalt: op de afdruk staat immers "5.0%". Iemand ontdekte toevallig dat die 5.0% eigenlijk 4.97% was: het programma drukte immers slechts 1 cijfer na de komma af, en rondde af. Dit programma was jaren gebruikt, zonder dat iemand zich om de afronding had bekommerd. . .

Afronden kan ook geld kosten

Beursindex van Vancouver (Canada)

- januari 1982: index wordt gestart op 1000
- november 1983: index staat blijkbaar op 520. . .

. . . en toch scheen de beurs het niet slecht te doen?!

Verklaring: de index werd altijd naar beneden afgerond op 3 cijfers, en dit bij elke berekening, vb. $678.35 \rightarrow 678$. De fout werd steeds in dezelfde "richting" gemaakt en zo werden vele kleine foutjes héél groot.

Na een correcte herberekening, bleek de index eigenlijk rond 1040 te staan.

Fibonacci–getallen

De Fibonacci–getallen zijn gedefinieerd door

$$f_0 = 1$$
$$f_1 = 1$$
$$f_n = f_{n-1} + f_{n-2}$$

Het begin van deze bekende rij is 1, 1, 2, 3, 5, 8, 13.

Gegeven is nu

$$f_{78} = 14472334024676221$$
$$f_{77} = 8944394323791464$$

Kan je zo de startwaarden van de rij terugvinden?

	exact	afgerond
f_{78}	1447233402467622 1	1447233402467622 0
f_{77}	894439432379146 4	894439432379146 4
f_{76}	552793970088475 7	552793970088475 6
f_{75}	341645462290670 7	341645462290670 8
f_{74}	211148507797805 0	211148507797804 8
f_{73}	130496954492865 7	130496954492866 0
f_{72}	80651553304939 3	80651553304938 8
f_{71}	4984540118792 64	4984540118792 72
f_{70}	3080615211701 29	3080615211701 16
$\vdots$	$\vdots$	$\vdots$
f_{41}	267914296	282844648
f_{40}	165580141	141422324
f_{39}	102334155	141422324
f_{38}	63245986	0
f_{37}	39088169	141422324
f_{36}	24157817	-141422324
f_{35}	14930352	282844648
$\vdots$	$\vdots$	$\vdots$
f_4	5	$-80651553304938 8$
f_3	3	130496954492866 0
f_2	2	$-211148507797804 8$
f_1	1	341645462290670 8
f_0	1	$-552793970088475 6$

Wortels van een vierkantsvergelijking

De vergelijking $ax^2 + bx + c = 0$ heeft 2 oplossingen:

$$x_{\pm} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Wiskundig zeer eenvoudig, maar numeriek. . .
Voorbeeld: $b^2 \gg |4ac|$

Bereken de wortels van

$$x^2 - 1000000x + 1 = 0$$

De echte oplossingen zijn:

$$x_1 = 500000 + 127\sqrt{15500031} \approx 999999.99999900000$$
$$x_2 = 500000 - 127\sqrt{15500031} \approx 0.000001000076145$$

Gebruik van de "klassieke" formule geeft:

$x_1 = 999999.99897600$	zeer goede benadering
$x_2 = 0.00102400$	totaal verkeerd

Bereken je echter x_1 a.d.h.v. de formule

$$x_1 = -\frac{b + \text{sign}(b)\sqrt{b^2 - 4ac}}{2a}$$

en x_2 uit

$$x_2 = \frac{c}{ax_1}$$

dan krijg je wel een correcte(re) waarde voor x_2:

$$x_1 = 999999.99897600$$
$$x_2 = 0.00000100$$

Is $0 = 1$?

Veronderstel dat we rekenen met 8 cijfers. Bereken de variantie van $x_1, \ldots, x_n$ op 2 manieren:

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

$$s_n^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2 \right) \tag{2}$$

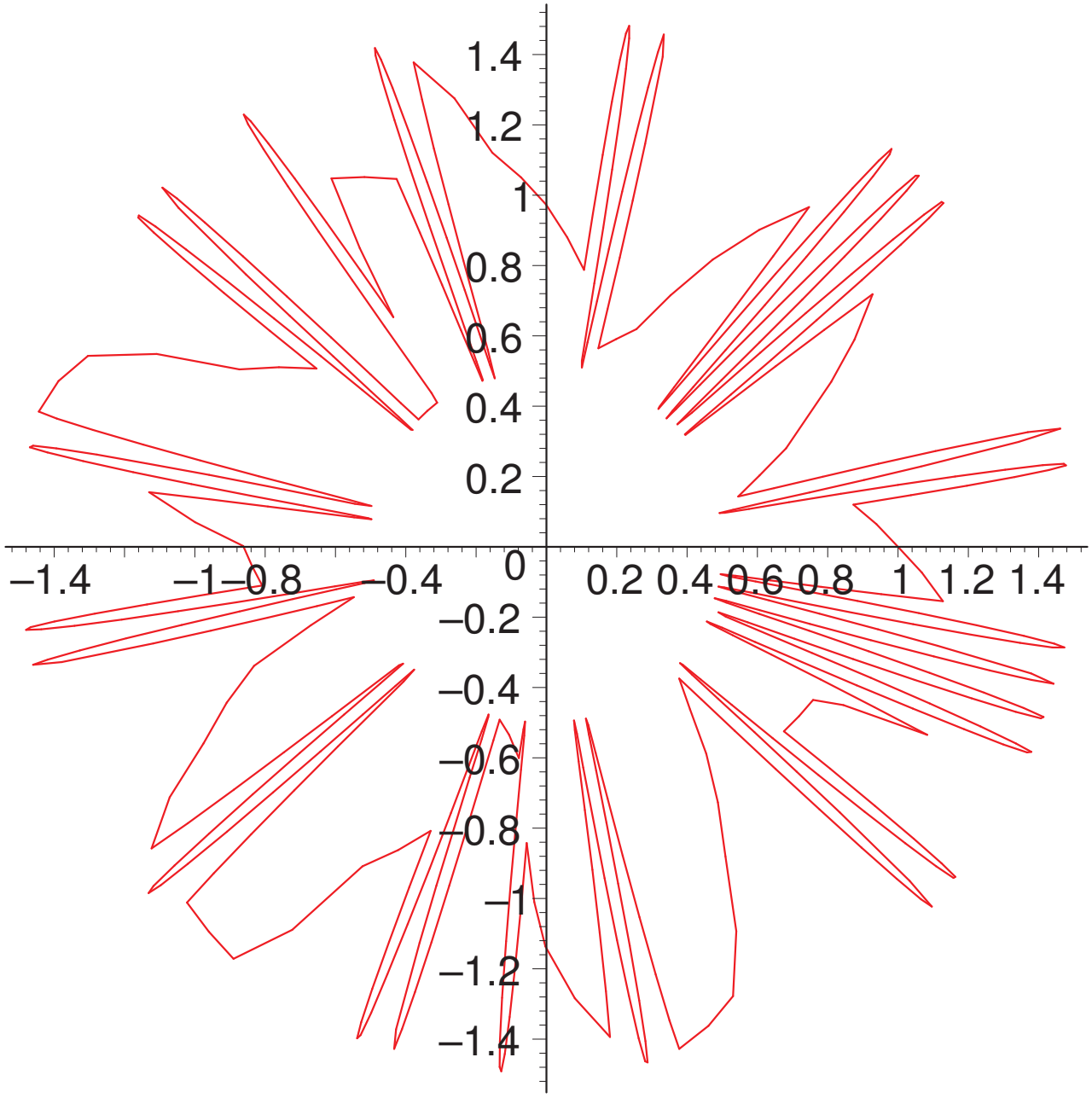
Beschouw data: $[10000, 10001, 10002]$.
Formule (1) geeft 1, formule (2) geeft 0.

Schijn bedriegt. . .

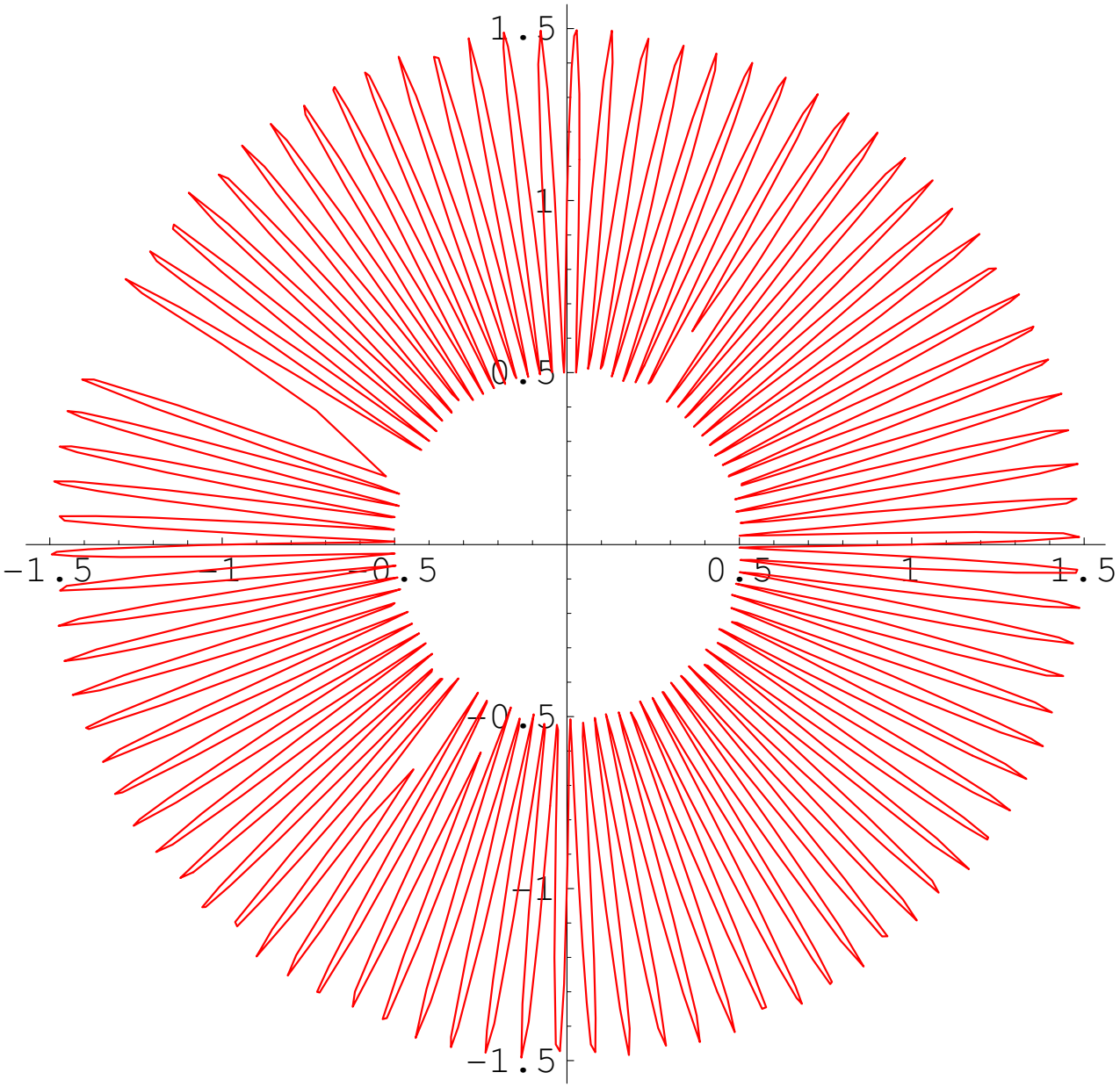
Hierna volgen drie plots van dezelfde functie. Ze zijn gemaakt in verschillende programma's door gebruik te maken van de standaard instellingen. . .

Het functievoorschrift luidt:

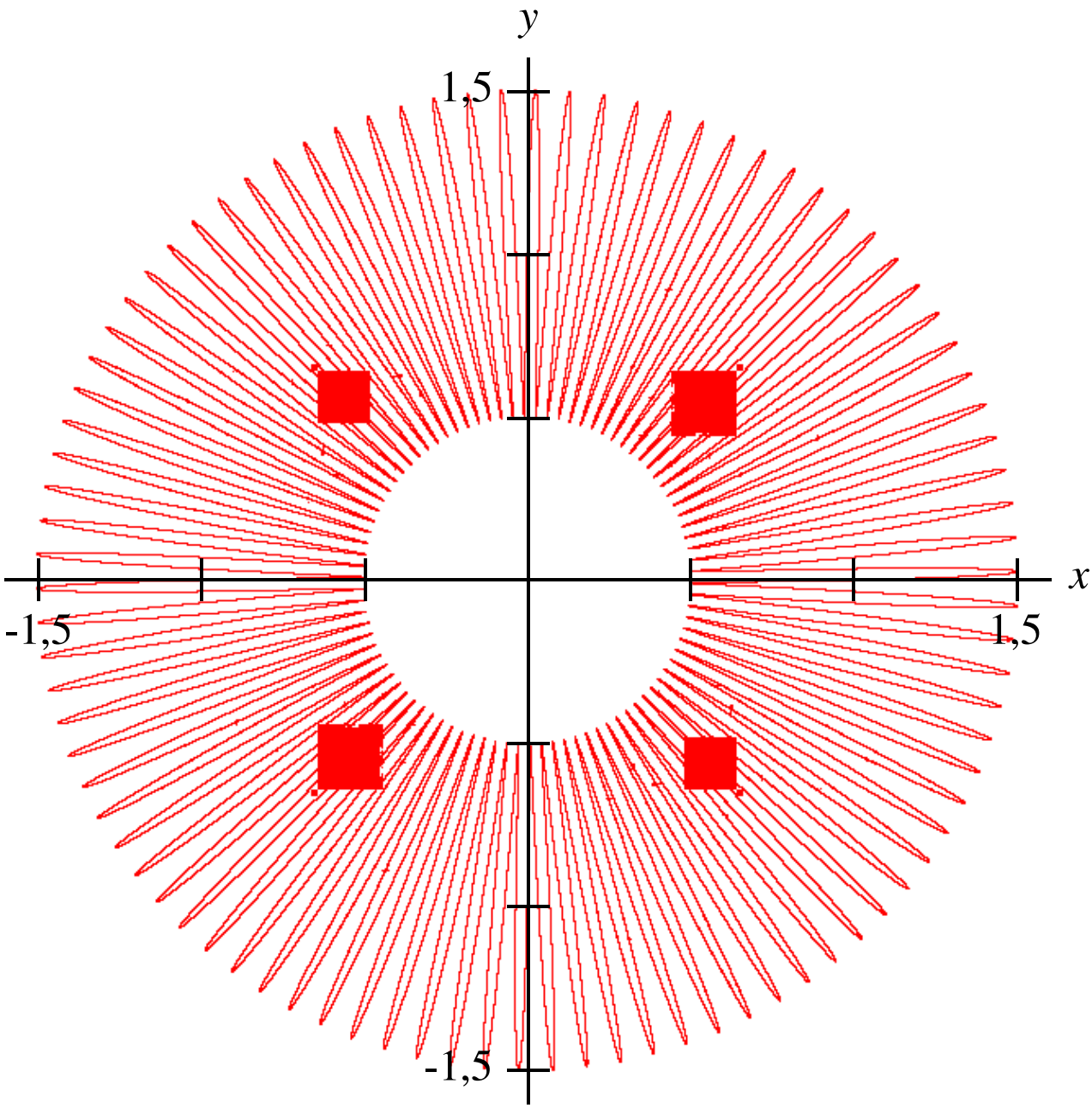
$$r = \frac{1}{2}\sin(90t)$$
$$x = (1 + r)\cos(t)$$
$$y = (1 + r)\sin(t) \qquad t = 0 \ldots 2\pi$$



Maple



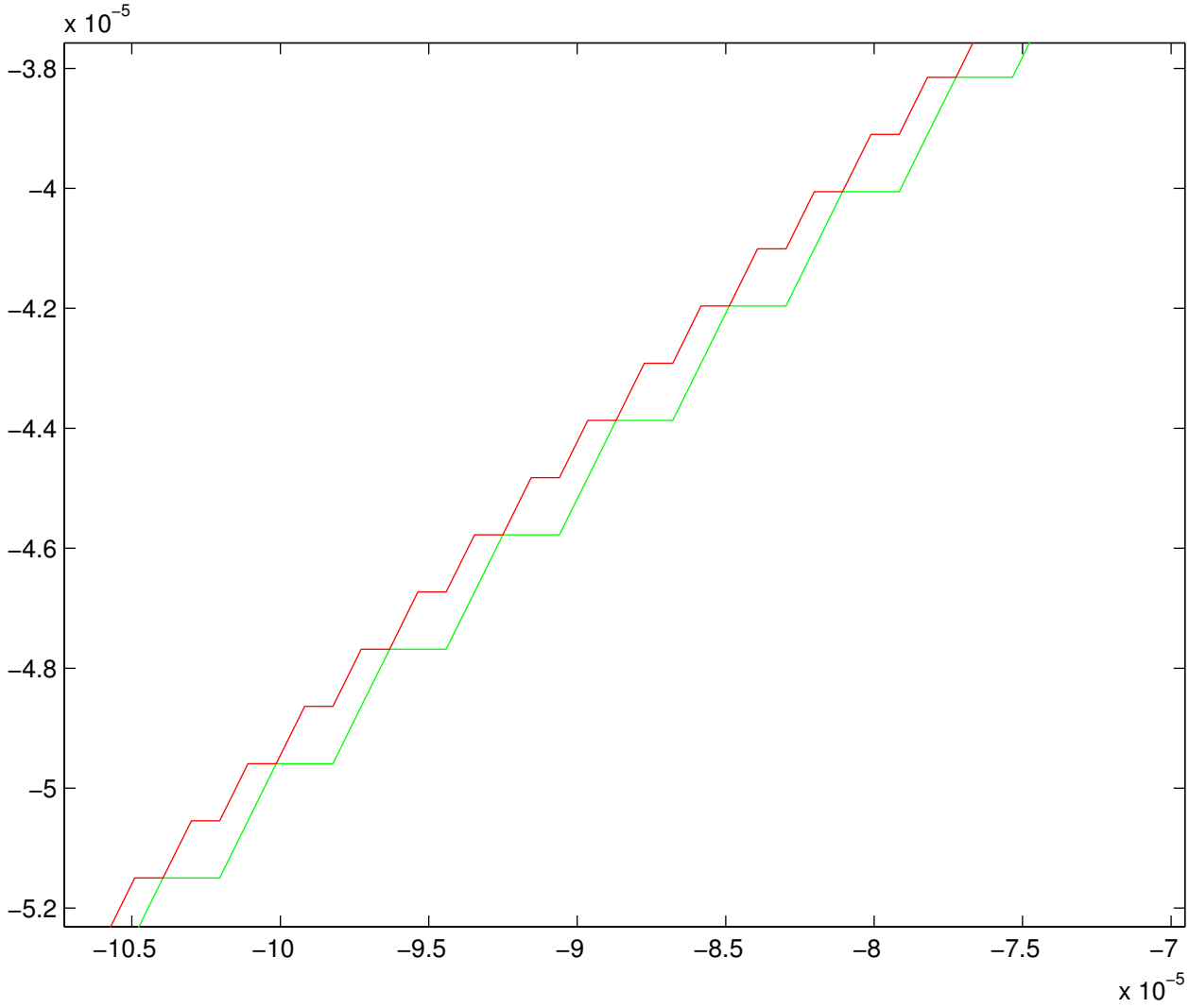
Mathematica


y
1,5
x
-1,5
1,5
-1,5
GrafEq

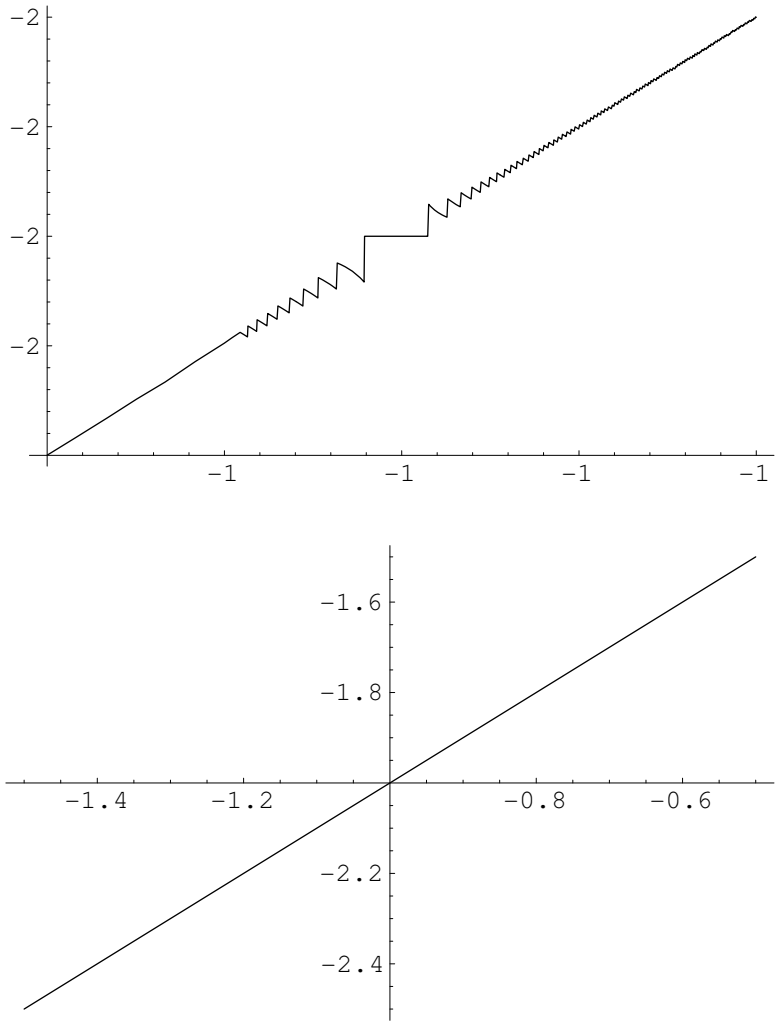
Twee rechten, véél snijpunten. . .

Hoeveel snijpunten hebben volgende rechten gemeenschappelijk?

$$\frac{x}{2} = \frac{32x}{65}$$



Is dit nu een rechte of niet?



$$\frac{x^2 - 1}{x + 1}$$

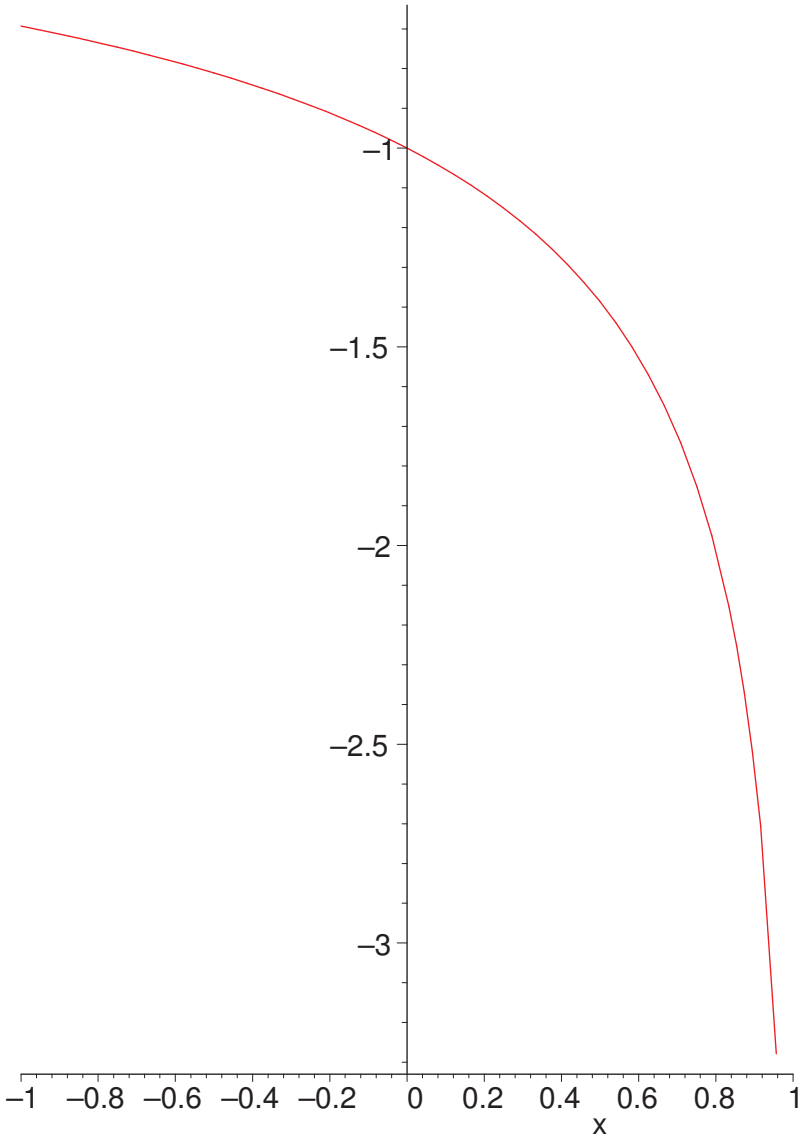
Waar ligt de limiet?

Gevraagd: de limietwaarde van de functie

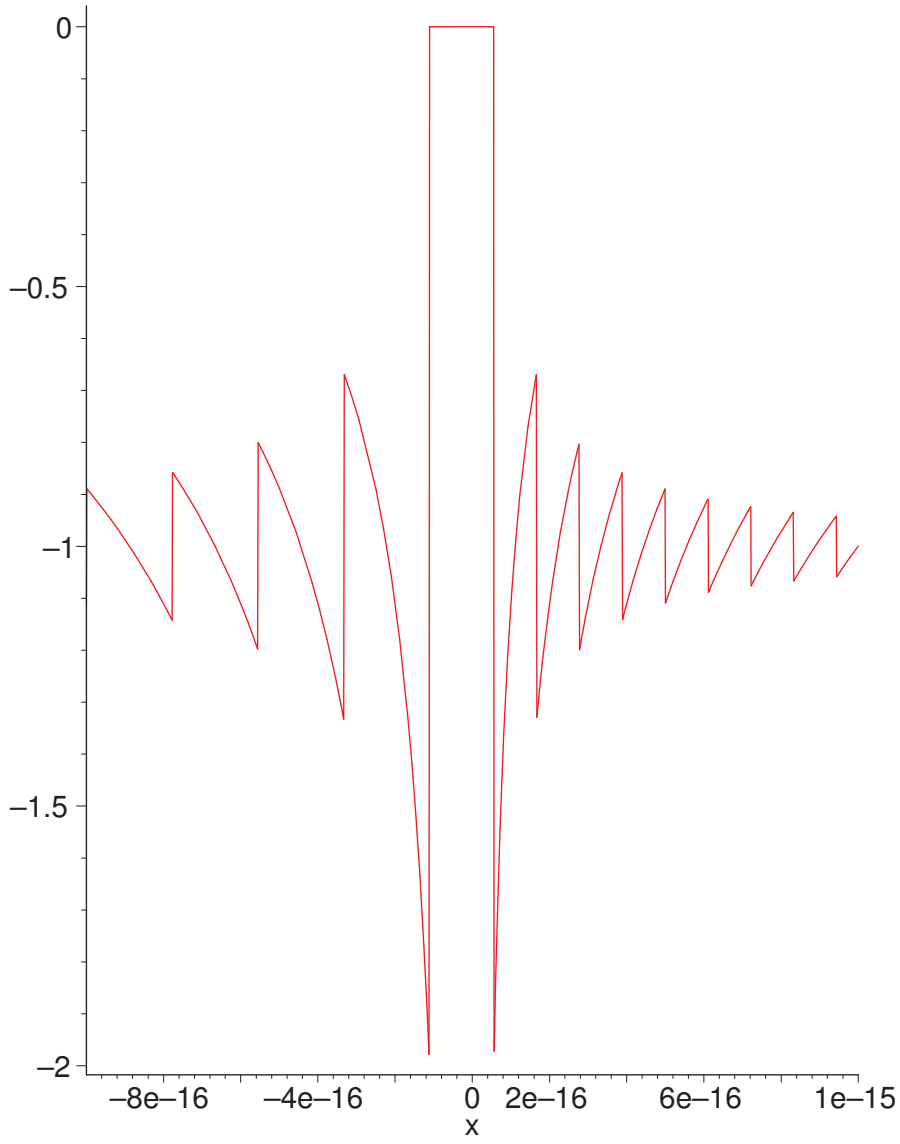
$$\frac{\ln(1-x)}{x}$$

in $x = 0$. Probeer dit ook te controleren a.d.h.v. een grafiek.

Dit is een globaal overzicht. . .



. . . en zo ziet de grafiek er uit in de omgeving van 0



Wiskundig model

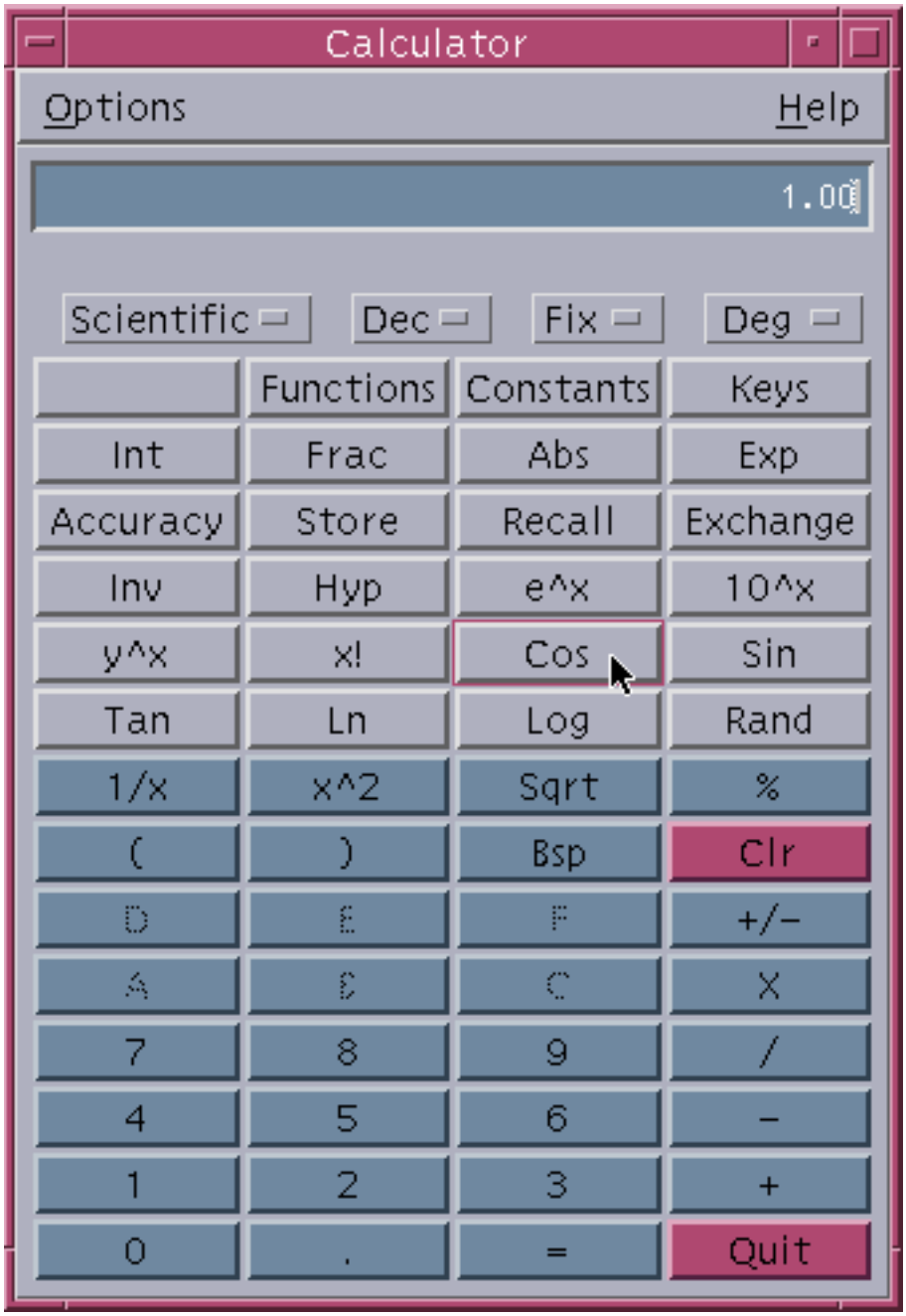
Vraag: welk type functies kunnen we construeren a.d.h.v. de basisbewerkingen ($+$, $-$, $\times$, $/$)?

Antwoord: Veeltermen of Rationale functies

$$c_l x^l + \cdots + c_0 x^0 \tag{V}$$

$$\frac{a_m x^m + \cdots + a_0 x^0}{b_n x^n + \cdots + b_0 x^0} \tag{R}$$

Wat gebeurt er als ik op cos druk?



- Welke eigenschappen heeft cos?

- Wat is het verband met sin?

Vermits $\cos(-x) = \cos(x)$, kunnen we de implementatie al beperken tot positieve x'en. Bovendien is de functie periodisch. En we kunnen natuurlijk de helft van het rekenwerk doorgeven aan de sinus functie!

Als x zeer klein is ($x < 2^{-27}$), geef dan x terug.

Zorg voor een goede benadering op $[0, \pi/4]$, d.m.v. een veelterm van slechts graad 14!

$$p(x) = 1 - \frac{1}{2}x^2 + \sum_{i=1}^{6} c_i \, x^{2i+2}$$

$$|\cos(x) - p(x)| \leq 2^{-58}$$

Referenties

[1] Mike Cowlishaw. Decimal arithmetic – FAQ. `http://www2.hursley.ibm.com/decimal/decifaq.html`, 2002.

[2] Annie Cuyt, Brigitte Verdonk, Stefan Becuwe, and Peter Kuterna. A remarkable example of catastrophic cancellation unraveled. *Computing*, 66:309–320, 2001.

[3] James Demmel. Basic issues in floating point arithmetic and error analysis. 1995.

[4] Alan Edelman. When is $x * (1/x) \neq 1$? December 1994.

[5] George E. Forsythe, Michael A. Malcolm, and Cleve B. Moler. Solving quadratic equations. In *Computer Methods for Mathematical Computations*, section 2.6, pages 20–23. Prentice Hall Inc., 1977.

[6] Ilse Geulig and Walter Krämer. Intervallrechnung in Maple – Die Erweiterung `intpakx` zum Paket `intpak` der Share-Library. Preprint 99/2, Universität Karlsruhe, IWRMM, 1999.

[7] David Goldberg. What every computer scientist should know about floating-point arithmetic. *ACM Comput. Surveys*, 23:5–48, 1991.

[8] Nicholas J. Higham. *Accuracy and stability of numerical algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[9] Nick Higham. Pitfalls in floating point computation—and how to avoid them. 1995.

[10] Nick Higham. Can you "count" on your computer. Public lecture for Science Week, 1998.

[11] Donald E. Knuth. *Seminumerical Algorithms*, volume 2 of *The Art of Computer Programming*. Addison–Wesley, third edition, 1998.

[12] Jacques-Louis Lions. ARIANE 5: Flight 501 failure, July 1996.

[13] Miles Murdocca. Data representation. In *Principles of Computer Architecture*, chapter 2. Prentice-Hall, 2000.

[14] Siegfried M. Rump. Algorithms for verified inclusions – theory and practice. In R.E. Moore, editor, *Reliability in Computing*, volume 19 of *Perspectives in Computing*, pages 109–126. Academic Press, 1988.

[15] Jeff A. Tupper. Graphing equations with generalized interval arithmetic. Master's thesis, University of Toronto, Department of Computer Science, 1996.

[16] United States General Accounting Office. Patriot missile software problem. GAO/IMTEC 92–26, 4 February 1992.

[17] Debora Weber-Wulff. Rounding error changes parliament makeup. *The Risks Digest*, 13(37), 1992.