# Doing it without Floating 

Real and Solid Computing

Abbas Edalat

## Overview

- The Story of the Decimal System
- Floating Point Computation
- Exact Real Arithmetic
- Solid Modelling \& Computational Geometry
- A New Integration
- The Moral of Our Story


## Decimal System

$$
\begin{aligned}
\text { Digits } & =\{0,1,2,3,4,5,6,7,8,9\} \quad \text { Base }=10 \\
3042 & =3000+40+2=3 \times 10^{3}+4 \times 10+2
\end{aligned}
$$

- Foundation of our computer revolution.
- Imagine computing in the Roman system CCXXXII times XLVIII, i.e. $232 \times 48$.
- Zero was invented by Indian mathematicians, who were inspired by the Babylonian and the Chinese number systems, particularly as used in abacuses.


## The Discovery of Decimal Fractions

- Persians and Arabs invented the representation of decimal fractions that we use today:
- They discovered the rules for basic arithmetic operations that we now learn in school.


## The Long Journey



## Khwarizmi (780-850)

- Settled in the House of Wisdom (Baghdad).
- Wrote three books:
- Hindu Arithmetic
- Al-jabr va Al-Moghabela
- Astronomical Tables
- The established words:
"Algorithm" from "Al-Khwarizmi and "Algebra" from "Al-jabr" testify to his fundamental contribution to human thought.



## The Long Journey



## Adelard of Bath (1080-1160)

- First English Scientist.
- Translated from Arabic to Latin Khwarizmi's astronomical tables with their use of zero.
- After a long rivalry between Algorists and abacists, the decimal system replaced the
 abacus.


## The Long Journey



## Kashani (1380-1429)

- Developed arithmetic algorithms for fractions, that we use today.
- Computed $\pi$ up to 16 decimal places 3.1415926535897932
- Took the unit circle.
- The circumferences of the inscribed and circumscribed polygons with $n$ sides give lower and upper bounds for $2 \pi$.
- He used $n=3 \times 2^{28}=805306368$
- Computed $\sin 1^{\circ}$ up to 16 decimals.


## Kashani (1380-1429)

- Kashani invented the first mechanical special purpose computers:
- to find when the planets are closest,
- to calculate longitudes of planets.
- to predict lunar eclipses.


## Kashani's Planetarium



## Mechanical Computers in Europe

Napier


R 1 .


Oughtred
(1575-1660)

Pascal (1632-1662)


Leibniz
(1646-1716)


Babbage
(1792-1871)

## Modern Computers: Floating Point Numbers

$$
\begin{aligned}
& \pm . \text { NIVM }^{2} \times 1.01 \\
& \begin{array}{c} 
\pm \\
\text { Sign } \\
\pm
\end{array} \\
& \begin{array}{l}
\text { Exponent } \\
\begin{array}{l}
\square \\
\square
\end{array}
\end{array}
\end{aligned}
$$

- Represents only a finite collection of numbers.

Any other number like $\pi$ is rounded or approximated to a close floating point number.

## Floating Point Arithmetic is not sound

- A simple calculation shows:

$$
\begin{aligned}
& 10^{20}+20-10-10^{20}=10 \\
& 10^{20}+20-10^{20}-10=10 \\
& 10^{20}-10-10^{20}+20=10
\end{aligned}
$$

- But using IEEE's standard precision, we get three different results,


## Floating Point Arithmetic is not sound

- A simple calculation shows:

$$
\begin{aligned}
& 10^{20}+20-10-10^{20}=0 \\
& 10^{20}+20-10^{20}-10=-10 \\
& 10^{20}-10-10^{20}+20=20
\end{aligned}
$$

- But using IEEE's standard precision, we get three different results, all wrong.


## Failure of Floating Point Computation

$$
\begin{aligned}
64919121 x_{1}-159018721 x_{2} & =1 \\
41869520.5 x_{1}-102558961 x_{2} & =0
\end{aligned}
$$

- Double precision floating-point arithmetic gives: $\widetilde{x}_{1}=102558961 \quad \widetilde{x}_{2}=41869520.5$
- The correct solution is:

$$
x_{1}=205117922 \quad x_{2}=83739041
$$

## Failure of Floating Point Computation

$$
\begin{aligned}
a_{0} & =1.5100050721318 \\
a_{n+1} & =\frac{3 a_{n}^{4}-20 a_{n}^{3}+35 a_{n}^{2}-24}{4 a_{n}^{3}-30 a_{n}^{2}+70 a_{n}-24}
\end{aligned}
$$

- Depending on the floating point format, the sequence tends to 1 or 2 or 3 or 4 .
- In reality, it oscillates about 1.51 and 2.37.


## Failure of Floating Point Computation

$a_{0}=\frac{11}{2}$
$a_{1}=\frac{61}{11}$
$a_{n+1}=111-\frac{1130}{a_{n}}+\frac{3000}{a_{n-1} a_{n}}$

- In any floating point format, the sequence converges to 100.
- In reality, it converges to 6 .

Floating Point

$$
a_{3}=5.633431
$$

$$
a_{3}=5.633431
$$

## Failure of Floating Point Computation

$a_{.}=\frac{11}{r} \quad a_{1}=\frac{\pi}{11} \quad a_{n+1}=11-\frac{11 r \cdot}{a_{n}}+\frac{r \ldots}{a_{n-1} a_{n}}$

- In any floating point format, the sequence converges to 100.
- In reality, it converges to 6.

Floating Point

$$
\begin{aligned}
& a_{13}=6.563994 \\
& a_{15}=68.463424
\end{aligned}
$$

## Failure of Floating Point Computation

$a_{0}=\frac{11}{2}$
$a_{1}=\frac{61}{11}$
$a_{n+1}=111-\frac{1130}{a_{n}}+\frac{3000}{a_{n-1} a_{n}}$

- In any floating point format, the sequence converges to 100.
- In reality, it converges to 6 .

Floating Point

$$
a_{20}=99.999964
$$

$$
a_{40}=100.000000
$$

$$
a_{80}=100.000000
$$

Exact Arithmetic
$a_{20}=5.974579$
$a_{40}=5.999320$
$a_{80}=5.999999$

## Banker's Example

- A banker offers a client a 25 year investment scheme.

- The client will invest £e, i.e. £2.71828...
- Initially, there is a bank fee of $£ 1$.


## Banker's Example



- After 1 year, the money is multiplied by 1 , and $£ 1$ bank fee is subtracted.


## Banker's Example



- After 2 years, the money is multiplied by 2, and $£ 1$ bank fee is subtracted.


## Banker's Example



- After 3 years, the money is multiplied by 3 , and $£ 1$ bank fee is subtracted.
- And so on . . .


## Banker's Example



- Finally, after 25 years, the money is multiplied by 25 , and $£ 1$ bank fee is subtracted. The final balance is returned to the client.


## Banker's Example

- The client calculates his final balance after 25 year

$$
25!\left(e-1-1-\frac{1}{2!}-\frac{1}{3!}-\ldots \ldots-\frac{1}{25!}\right)
$$

with floating point numbers on his computer.

- He finds out that he would have an overdraft of £2,000,000,000.00 !!


## Banker's Example

- Suspicious about this astonishing result, he buys a better computer.
- This time he calculates that after 25 years he would have a credit of £4,000,000,000.00 !!



## Banker's Example

- He is delighted and makes the investment.
- 25 years later, the banker, using correct arithmetic, computes
 the value of
$25!\left(e-1-1-\frac{1}{2!}-\frac{1}{3!}-\ldots \ldots .-\frac{1}{25!}\right)$
- The client's balance is: $4 p$


## Pilot's dilemma



On February 25, 1991, during the Gulf War, an American Patriot Missile battery in Dharan, Saudi Arabia, failed to intercept an incoming Iraqi Scud missile, due to failure of floating point computation. The Scud missile struck an American Army barracks and killed 28 soldiers.

## Exact Real Arithmetic

- Evaluate numerical expressions correctly up to any given number of decimal places.
- Real numbers have in general an infinite decimal expansion.
$\forall \pi=3.1415 \ldots$ gives a shrinking sequence of rational intervals.



## Exact Real Arithmetic

- Evaluate numerical expressions correctly up to any given number of decimal places.
- Real numbers have in general an infinite decimal expansion.
$\forall \pi=3.1415 \ldots$ gives a shrinking sequence of rational intervals.



## Exact Real Arithmetic

- Evaluate numerical expressions correctly up to any given number of decimal places.
- Real numbers have in general an infinite decimal expansion.
$\forall \pi=3.1415 \ldots$ gives a shrinking sequence of rational intervals.



## Exact Real Arithmetic

- A computation is possible only if any output digit can be calculated from a finite number of the input digits.


Conclusion: Multiplication is not computable in the decimal system.

The Signed Decimal System

## The Signed Decimal System $\{\overline{9}, \overline{8}, \overline{7}, \overline{6}, \overline{5}, \overline{4}, \overline{3}, \overline{2}, \overline{1}, 0,1,2,3,4,5,6,7,8,9\}$

- Gives a redundant representation.

$$
1 \overline{2} 3.5 \overline{6}=100+(-20)+3+\frac{5}{10}+\left(-\frac{6}{100}\right)=83.44
$$

- We can now compute:


Numbers as Sequences of Operations

- Signed binary system:

$$
\begin{gathered}
\text { Digits }=\{\overline{1}, 0,1\} \quad \text { Base }=2 \\
\text { Base Interval }=[-1,1]
\end{gathered}
$$

$$
-1\left[\frac{0 .---}{\frac{[. \overline{1}--}{}}\right] 1
$$

Numbers as Sequences of Operations

- Signed binary system:

$$
\begin{gathered}
\text { Digits }=\{\overline{1}, 0,1\} \quad \text { Base }=2 \\
\text { Base Interval }=[-1,1]
\end{gathered}
$$



Numbers as Sequences of Operations

- Signed binary system:

$$
\text { Digits }=\{\overline{1}, 0,1\} \quad \text { Base }=2
$$

Base Interval $=[-1,1]$


Numbers as Sequences of Operations

- A number such as corresponds to:
 $L=$ Left half $\quad M=$ Middle half $R=$ Right half $-1$

Numbers as Sequences of Operations

- A number such as corresponds to:
 $L=$ Left half $\quad M=$ Middle half $R=$ Right half


Numbers as Sequences of Operations

- A number such as corresponds to:
 $L=$ Left half $\quad M=$ Middle half $R=$ Right half



## Numbers as Sequences of Operations

- A number such as corresponds to:
 $L=$ Left half $\quad M=$ Middle half $R=$ Right half



## Numbers as Sequences of Operations

- A number such as corresponds to: $L=$ Left half $\quad M=$ Middle half $R=$ Right half

- Mathematically:

$$
L: x \cdot \frac{x-1}{2} \quad M: x \cdot \frac{x}{2} \quad R: x \square \frac{x+1}{2}
$$

## Numbers as Sequences of Operations

- $L, M, R$ are affine maps, special case of linear fractional transformations of the form:

$$
x \quad \frac{a x+c}{b x+d}
$$

represented by: $\quad\left(\begin{array}{ll}a & c \\ b & d\end{array}\right)$

- Sequences of these operations give a general representation for numbers.

Numbers as Sequences of Operations

$$
\sqrt{2}=\left(\begin{array}{ll}
-1 & 7 \\
-1 & 5
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right) \ldots
$$

$$
\sqrt{2}
$$

Numbers as Sequences of Operations

$$
\sqrt{2}=\left(\begin{array}{ll}
-1 & 7 \\
-1 & 5
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right) \ldots
$$



## Numbers as Sequences of Operations

$$
\sqrt{2}=\left(\begin{array}{ll}
-1 & 7 \\
-1 & 5
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right) \ldots
$$



## Numbers as Sequences of Operations

$$
\sqrt{2}=\left(\begin{array}{ll}
-1 & 7 \\
-1 & 5
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
-1 & 6
\end{array}\right) \ldots
$$



Numbers as Sequences of Operations

$$
\pi=\left(\begin{array}{cc}
-2 & 2 \\
0 & 1
\end{array}\right)\left(\begin{array}{cc}
-1 & -3 \\
0 & 4
\end{array}\right)\left(\begin{array}{cc}
-5 & -13 \\
3 & 15
\end{array}\right) \ldots
$$



Numbers as Sequences of Operations

$$
\pi=\left(\begin{array}{cc}
-2 & 2 \\
0 & 1
\end{array}\right)\left(\begin{array}{cc}
-1 & -3 \\
0 & 4
\end{array}\right)\left(\begin{array}{cc}
-5 & -13 \\
3 & 15
\end{array}\right) \ldots
$$



Numbers as Sequences of Operations

$$
\pi=\left(\begin{array}{cc}
-2 & 2 \\
0 & 1
\end{array}\right)\left(\begin{array}{cc}
-1 & -3 \\
0 & 4
\end{array}\right)\left(\begin{array}{cc}
-5 & -13 \\
3 & 15
\end{array}\right) \ldots
$$



## Numbers as Sequences of Operations

$$
\pi=\left(\begin{array}{cc}
-2 & 2 \\
0 & 1
\end{array}\right)\left(\begin{array}{cc}
-1 & -3 \\
0 & 4
\end{array}\right)\left(\begin{array}{cc}
-5 & -13 \\
3 & 15
\end{array}\right) \cdots
$$



## Numbers as Sequences of Operations

$$
e=\left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right)\left(\begin{array}{ll}
0 & 1 \\
1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
1 & 10
\end{array}\right) \ldots
$$

## Numbers as Sequences of Operations

$$
e=\left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right)\left(\begin{array}{ll}
0 & 1 \\
1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
1 & 10
\end{array}\right) \ldots
$$



## Numbers as Sequences of Operations

$$
e=\left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right)\left(\begin{array}{ll}
0 & 1 \\
1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
1 & 10
\end{array}\right) \ldots
$$



## Numbers as Sequences of Operations

$$
e=\left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right)\left(\begin{array}{ll}
0 & 1 \\
1 & 6
\end{array}\right)\left(\begin{array}{cc}
0 & 1 \\
1 & 10
\end{array}\right) \ldots
$$



## Basic Arithmetic Operations

- Use linear fractional transformations with two entries

$$
(x, y) \div \frac{a x y+c x+e y+g}{b x y+d x+f y+h}
$$

represented by:

$$
\left(\begin{array}{llll}
a & c & e & g \\
b & d & f & h
\end{array}\right)
$$

- For example, addition $(x, y) \quad x+y$ uses:

$$
\left(\begin{array}{llll}
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right)
$$

## Addition

$$
\begin{array}{r}
\left.\pi+e=\begin{array}{llll}
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right) \\
\pi \\
\pi
\end{array}
$$

$$
\pi+e
$$

## Addition

$$
\begin{aligned}
& \pi+e=\quad\left(\begin{array}{llll}
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right) \\
& \left(\begin{array}{cc}
-2 & 2 \\
0 & 1
\end{array}\right) \quad\left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right) \\
& \pi+e \\
& \begin{aligned}
\left(\begin{array}{ll}
0 & 1 \\
1 & 6
\end{array}\right) & \\
& \left(\begin{array}{cc}
0 & 1 \\
1 & 10
\end{array}\right)
\end{aligned}
\end{aligned}
$$

## Addition

$$
\left.\begin{array}{ccc}
\pi+e= & \left(\begin{array}{llll}
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right) \\
& \left(\begin{array}{cc}
-2 & 2 \\
0 & 1
\end{array}\right) & \left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right) \\
\\
& \left(\begin{array}{cc}
-1 & -3 \\
0 & 4
\end{array}\right) & \\
/ & \left(\begin{array}{ll}
0 & 1 \\
1 & 6
\end{array}\right) & \pi+e
\end{array}\right\}
$$

## Addition

$$
\begin{aligned}
& \pi+e=\quad\left(\begin{array}{llll}
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right) \\
& \left(\begin{array}{ll}
1 & 3 \\
1 & 1
\end{array}\right)
\end{aligned}
$$

## Addition

## Elementary Functions

- $\sin x, \cos x, \tan x, e^{x}, \log x, e t c$.
- Each of them is computed by a composition of Linear Fractional Transformations presented as a binary tree.
- A C-library for computing elementary functions is on the WWW.


## Elementary Functions



## Elementary Functions

$$
\left.\begin{array}{c}
\tan \frac{1}{\pi}=\left(\begin{array}{cccc}
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 1
\end{array}\right) \\
\frac{1}{\pi} \\
\frac{1}{\pi} \\
\frac{1}{\pi} \\
\frac{1}{\pi} \\
-1
\end{array}\right)
$$

## Elementary Functions



## Elementary Functions

$$
\begin{aligned}
& \left.\tan \frac{1}{\pi}=\begin{array}{cccc}
0 & 1 & 0 & 0 \\
-1 & 0 & 0 & 1
\end{array}\right)
\end{aligned}
$$

$$
\begin{aligned}
& \left(\begin{array}{cc}
-5 & -11 \\
4 & 12
\end{array}\right)
\end{aligned}
$$

## Elementary Functions




## Domain of Intervals



More information

- Dana Scott introduced domain theory in 1970 as a mathematical model of programming languages.
- Domain theory found applications in numerical computation in 1990's.


## Solid Modelling / Computational Geometry

- Manufactured objects are generally modelled with CAD, a package for solid and geometric modelling.
- Correct geometric algorithms become unreliable when implemented in floating
 point.


## Solid Modelling / Computational Geometry



- With floating point arithmetic, find the point $P$ of the intersection of $L_{1}$ and $L_{2}$. Then:

minimum_distance $\left(P, L_{1}\right)>0$<br>minimum_distance $\left(P, L_{2}\right)>0$

## The Convex Hull Algorithm

$A, B \& C$ nearly collinear
With floating point we can get:


## The Convex Hull Algorithm

$A, B \& C$ nearly collinear
With floating point we can get:
(i) $A C$, or


## The Convex Hull Algorithm

$A, B \& C$ nearly collinear
With floating point we can get:
(i) $A C$, or
(ii) just $A B$, or


## The Convex Hull Algorithm

$A, B \& C$ nearly collinear
With floating point we can get:
(i) $A C$, or
(ii) just $A B$, or
(iii) just $B C$, or

B

## The Convex Hull Algorithm

$A, B \& C$ nearly collinear
With floating point we can get:
(i) $A C$, or
(ii) just $A B$, or
(iii) just $B C$, or
(iv) none of them.

## A

$$
B
$$

The quest for robust algorithms is the most fundamental unresolved problem in solid modelling and computational geometry.

## A Fundamental Problem

- The basic building blocks of classical geometry are not continuous and hence not computable.
- Example: The point $x$ is in the box.



## True

## A Fundamental Problem

- The basic building blocks of classical geometry are not continuous and hence not computable.
- Example: The point $x$ is in the box.



## False

$x$

## A Fundamental Problem

- There is a discontinuity if $x$ goes through the boundary.

- This predicate is not computable:

If $x$ is on the boundary, we cannot determine if it is in or out at any finite
 stage.

## Intersection of Two Cubes



## Intersection of Two Cubes



## This is Really Ironical!

- Topology and geometry have been developed to study continuous functions and transformations on spaces.
- The membership predicate and the intersection operation are the fundamental building blocks of topology and geometry.
- Yet, these basic elements are not continuous in classical topology and geometry.


## Foundation of a Computable Geometry

- Reconsider the membership predicate:



## A Three-Valued Logic

- A domain

with its Scott topology. It is called $B_{\perp}$


## Computing a Solid Object

In this model, a solid object is represented by its interior and exterior, each approximated by a nested sequence of rational polyhedra.
Mathematically, a solid object is given by a continuous function from the Euclidean space to $B_{\perp}$


## Computing a Solid Object

- Kashani's computation of $\pi$



## Computable Predicates \& Operations

- This gives a model for geometry and topology in which all the basic building blocks (membership, intersection, union) are continuous and computable.
- In practice, a geometric object is approximated by two rational polyhedra, one inside and one outside, so that the area between them is as small as desired.


## The Convex Hull Algorithm



## The Convex Hull Algorithm



## The Convex Hull Algorithm



## The Convex Hull Algorithm



The inner and outer convex hulls can be computed by a robust $\mathrm{N} \log \mathrm{N}$ algorithm i.e. with the same complexity as the non-robust classical algorithms.

## Calculating the Number of Holes

- For a computable solid with computable volume, one can calculate the number of holes with volume greater than any desired value.


1 hole


2 holes


19 holes

- In mathematical terms, this model enables us to study the computability or decidability of various homotopic properties of solids.


## The Riemann Integral



## The Riemann Integral



## The Riemann Integral



## The Riemann Integral



## The Riemann Integral



This method can be extended using domain theory to more general distributions on more general spaces.

## The Generalized Riemann Integral

- The generalized Riemann integral has been applied to compute physical quantities in chaotic systems:
- Feigenbaum map on the route to chaos:

- The physical quantities of the 1 -dimensional random field Ising model.


## The Real and Solid <br> People



## The Long Journey



## The Moral of Our Story

- The ever increasing power of computer technology enables us to perform exact computation efficiently, in the spirit of Kashani.
- People from many nations have contributed to the present achievements of science and technology.
- History has imposed a reversal of fortune: Nations who developed the foundation of our present computer revolution in the very dark ages of Europe, later experienced a much stifled development.
- The Internet can be a global equaliser if, and only if, we make it available to the youth of the developing countries.

Empowering the Youth in the Developing World


- Science and Arts Foundation was launched in March 1999 at Imperial College:
- To provide Computer/Internet Sites for school children and students in the Developing World.
- To establish Internet incubators.

THE END

