Problems with Floating Point Operations

dkfz.
GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION

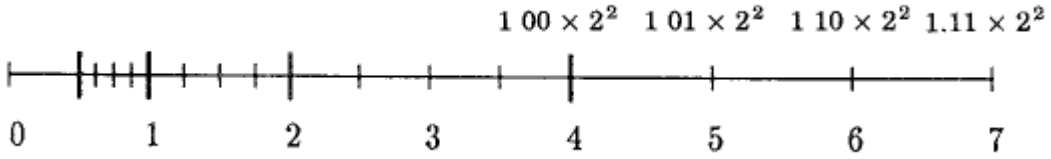
Is there any difference?

- $x_{i+1} = (R+1)x_i - R(x_i x_i)$
- $x_{i+1} = (R+1)x_i - (Rx_i)x_i$
- $x_{i+1} = ((R+1) - Rx_i)x_i$
- $x_{i+1} = Rx_i + (1 - Rx_i)x_i$
- $x_{i+1} = x_i + R(x_i - x_i x_i)$

What is a floating point

$$sssssssss * B^{eeee}$$

Significant Base Exponent



- Wide range of values possible
- Not all values can be represented
- Today: Mostly IEEE-754
- Significant decimal digits:
  - Single Precision (Float) : 6 – 9 digits
  - Double Precision (Double): 15 – 17 digits

1,2574849984 8...

Comparison

- Floating Point is subject to rounding error!
  - True Value is in a range

- How to compare float A and float B?

```
A == B
```

```
epsilon= 0.0001
std::fabs(A-B) < epsilon
```

```
epsilon = 10^-5
std::fabs(|A|-|B|) < |A|*epsilon
```

(|y| := std::fabs(y))

Addition

```
float time = 0.0 f;
for (int i = 0; i < 20000;++i)
    time += 0.1 f;
```

- Expected result: 2000
- Real-Life-Result: 1999,6588

- $10{,}2 + 0{,}145 = 10{,}3$
  - Error: $0{,}045$

- Serious for:
  - Summation, Mean-Calculation, ln(1+x) for small x ….

Cumulativity

Example for Significant = 3 and Base = 10

- $5 + 5 + 0{,}04 + 0{,}03 + 0{,}03 = 10{,}1$
- $\Big(\big((5 + 5) + 0{,}04\big) + 0{,}03\Big) + 0{,}03$
  $= 10{,}0 + 0{,}04 + 0{,}03 + 0{,}03 = 10{,}0 + 0{,}03 = 10{,}0$ 
- $5 + \Big(5 + \big(0{,}04 + (0{,}03 + 0{,}03)\big)\Big)$
  $= 5 + 5 + 0{,}04 + 0{,}06 = 10{,}0 + 0{,}1 = 10{,}1$ 

- $10 + 0{,}05 + 0{,}05 = 10{,}1$
- $(10 + 0{,}05) + 0{,}05 = 10{,}1 + 0{,}05 = 10{,}2$ 
- $10 + (0{,}05 + 0{,}05) = 10{,}0 + 0{,}1 = 10{,}1$ 

Subtraction

- Example:
- $a = 3{,}34; b = 1{,}22; c = 2{,}28$
- $b^2 - 4ac$

Exact result: $0{,}0292$

Rounded $b^2 = 11{,}2$
Rounded $4ac = 11{,}1$
Computer-result: $0{,}1$

- Subtraction of numbers in the same range cancels out significant digits!

AVOID Subtraction

Problem solver

- Know what you are doing
  - What Every Computer Scientist Should Know About Floating-Point Arithmetic (David Goldberg)

- Think about what your doing
- Avoid subtraction
- Addition goes from small to big
- Think about comparison

- Use double

Is there any difference?

- (1): $x_{i+1} = (R+1)x_i - R(x_i x_i)$
- (2): $x_{i+1} = (R+1)x_i - (Rx_i)x_i$
- (3): $x_{i+1} = ((R+1) - Rx_i)x_i$
- (4): $x_{i+1} = Rx_i + (1 - Rx_i)x_i$
- (5): $x_{i+1} = x_i + R(x_i - x_i x_i)$

$$R = 3{,}0; x_0 = 0{,}5$$

Iterations	1	2	3	4	5
0	0,5	0,5	0,5	0,5	0,5
100	0.2908	1.298	1.22	1.317	0.0498
200	0.4478	1.331	0.1807	0.0001793	0.00967
300	0.1247	0.1883	1.309	1.072	0.3747
400	1.2	0.337	1.015	0.9295	0.4422
500	1.301	1.206	0.8349	0.9845	0.05571
1000	1.331	0.3915	0.2754	0.4863	0.1421