NA01 Aritmetika racunala i pogreske

Ivan Slapničar

11. listopada 2018.

1 Aritmetika računala i pogreške

```
In [1]: using Interact
```

1.1 Apsolutna i relativna pogreška

Neka je α aproksimacija za a. Tada vrijedi

$$err = |a - \alpha| relerr = \frac{err}{|a|} = \frac{|a - \alpha|}{|a|}.$$

```
In [2]: a=5.0
        @manipulate for α=a:0.01:2a
            err=abs(a-α)
            relerr=err/abs(a)
            α, err, relerr
        end
```

```
Out[2]: (7.5, 2.5, 0.5)
```

1.2 Posebne vrijednosti (*special quantities*) 0, -0, `Inf` i `NaN`

Vidi David Goldberg, What Every Computer Scientist Should Know About Floating-Point Arithmetic.

Nula ima predznak:

```
In [3]: a=1.0
        b=0.0
        c=-b
        c,b==c
```

```
Out[3]: (-0.0, true)
```

```
In [4]: d=a/b
        e=a/c
        d==e, 1/d==1/e

Out[4]: (false, true)

In [5]: b/c

Out[5]: NaN

In [7]: # bitstring() u Julia 1.0
        bits(0)

Out[7]: "0000000000000000000000000000000000000000000000000000000000000000"

In [8]: bits(1)

Out[8]: "0000000000000000000000000000000000000000000000000000000000000001"

In [9]: bits(0.0)

Out[9]: "0000000000000000000000000000000000000000000000000000000000000000"

In [10]: bits(-0.0)

Out[10]: "1000000000000000000000000000000000000000000000000000000000000000"

In [11]: bits(1.0)

Out[11]: "0011111111110000000000000000000000000000000000000000000000000000"
```

Zadatak. Objasnite prethodne binarne zapise.

1.3 Točnost stroja ε

je najmanji broj ε takav da je $1 + \varepsilon \neq 1$

```
In [12]: b=1.0
         a=2.0
         while (b+a)!=b
             @show a=a/2
         end
         a
```

```
a = a / 2 = 1.0
a = a / 2 = 0.5
a = a / 2 = 0.25
a = a / 2 = 0.125
a = a / 2 = 0.0625
a = a / 2 = 0.03125
a = a / 2 = 0.015625
a = a / 2 = 0.0078125
a = a / 2 = 0.00390625
a = a / 2 = 0.001953125
a = a / 2 = 0.0009765625
a = a / 2 = 0.00048828125
a = a / 2 = 0.000244140625
a = a / 2 = 0.0001220703125
a = a / 2 = 6.103515625e-5
a = a / 2 = 3.0517578125e-5
a = a / 2 = 1.52587890625e-5
a = a / 2 = 7.62939453125e-6
a = a / 2 = 3.814697265625e-6
a = a / 2 = 1.9073486328125e-6
a = a / 2 = 9.5367431640625e-7
a = a / 2 = 4.76837158203125e-7
a = a / 2 = 2.384185791015625e-7
a = a / 2 = 1.1920928955078125e-7
a = a / 2 = 5.960464477539063e-8
a = a / 2 = 2.9802322387695312e-8
a = a / 2 = 1.4901161193847656e-8
a = a / 2 = 7.450580596923828e-9
a = a / 2 = 3.725290298461914e-9
a = a / 2 = 1.862645149230957e-9
a = a / 2 = 9.313225746154785e-10
a = a / 2 = 4.656612873077393e-10
a = a / 2 = 2.3283064365386963e-10
a = a / 2 = 1.1641532182693481e-10
a = a / 2 = 5.820766091346741e-11
a = a / 2 = 2.9103830456733704e-11
a = a / 2 = 1.4551915228366852e-11
a = a / 2 = 7.275957614183426e-12
a = a / 2 = 3.637978807091713e-12
a = a / 2 = 1.8189894035458565e-12
a = a / 2 = 9.094947017729282e-13
a = a / 2 = 4.547473508864641e-13
a = a / 2 = 2.2737367544323206e-13
a = a / 2 = 1.1368683772161603e-13
a = a / 2 = 5.684341886080802e-14
a = a / 2 = 2.842170943040401e-14
a = a / 2 = 1.4210854715202004e-14
a = a / 2 = 7.105427357601002e-15
```

```
a = a / 2 = 3.552713678800501e-15
a = a / 2 = 1.7763568394002505e-15
a = a / 2 = 8.881784197001252e-16
a = a / 2 = 4.440892098500626e-16
a = a / 2 = 2.220446049250313e-16
a = a / 2 = 1.1102230246251565e-16


Out[12]: 1.1102230246251565e-16

In [13]: 1+a==1.0

Out[13]: true

In [14]: 2a, 1+2a==1.0

Out[14]: (2.220446049250313e-16, false)
```

Programi imaju ugrađenu naredbu koja daje ε

```
In [15]: eps()

Out[15]: 2.220446049250313e-16

In [16]: # Što je ovo?
         eps(200.0)

Out[16]: 2.842170943040401e-14

In [17]: methods(eps)

Out[17]: # 9 methods for generic function "eps":
         eps(t::Base.Dates.Time) in Base.Dates at dates/types.jl:331
         eps(dt::Date) in Base.Dates at dates/types.jl:330
         eps(dt::DateTime) in Base.Dates at dates/types.jl:329
         eps() in Base at float.jl:715
         eps(x::AbstractFloat) in Base at float.jl:711
         eps(::Type{Float16}) in Base at float.jl:712
         eps(::Type{Float32}) in Base at float.jl:713
         eps(::Type{Float64}) in Base at float.jl:714
         eps(::Type{BigFloat}) in Base.MPFR at mpfr.jl:854

In [18]: eps(Float64), 2.0^(-52)

Out[18]: (2.220446049250313e-16, 2.220446049250313e-16)
```

```
In [19]: eps(Float32), 2.0^(-23)
```

```
Out[19]: (1.1920929f-7, 1.1920928955078125e-7)
```

```
In [20]: eps(Float16), 2.0^(-10)
```

```
Out[20]: (Float16(0.000977), 0.0009765625)
```

```
In [21]: eps(BigFloat), 2.0^(-255)
```

```
Out[21]:
(1.72723371101888892507727037256007991422320007288725627700474069
4033718360632485e-77, 1.727233711018889e-77)
```

1.4 Katastrofalno kraćenje (*catastrophic cancellation*)

U egzaktnoj aritmetici kvadratna jednadžba

$$ax^2 + bx + c = 0$$

ima rješenja

$$\begin{aligned} x_1 &= \frac{-b - \sqrt{b^2 - 4ac}}{2a} \\ x_2 &= \frac{-b + \sqrt{b^2 - 4ac}}{2a} \equiv \frac{-b + \sqrt{b^2 - 4ac}}{2a} \cdot \frac{-b - \sqrt{b^2 - 4ac}}{-b - \sqrt{b^2 - 4ac}} \\ &= \frac{2c}{-b - \sqrt{b^2 - 4ac}} = x_3 \end{aligned}$$

```
In [22]: a=2.0
         b=123456789.0
         c=4.0

         x1=(-b-sqrt(b*b-4*a*c))/(2.0*a)
         x2=(-b+sqrt(b*b-4*a*c))/(2.0*a)
         x3=(2*c)/(-b-sqrt(b*b-4*a*c))
         x1,x2,x3
```

```
Out[22]: (-6.172839449999997e7, -3.3527612686157227e-8, -3.240000029484002e-8)
```

Provjerimo s `BigFloat`:

```
In [23]: a=BigFloat(a)
         b=BigFloat(b)
         c=BigFloat(c)
         x2=(-b+sqrt(b*b-4*a*c))/(2.0*a)
```

```
Out[23]:
-3.240000029484001968915648868258452417675753633383540
09951677951071299216719687 18e-08
```

Još jedan primjer:

```
In [24]: x=1e-10
         tan(x)-sin(x)
```

```
Out[24]: 0.0
```

Međutim, trignometrijski indentiti daju:

$$\begin{aligned}\tan x - \sin x &= \tan x(1-\cos x) = \tan x(1-\cos x)\frac{1+\cos x}{1+\cos x}\\ &= \tan x\frac{1-\cos^2 x}{1+\cos x}\\ &= \tan x \sin^2 x\frac{1}{1+\cos x},\end{aligned}$$

a Taylorova formula daje:

$$\begin{aligned}\tan x &= x + \frac{x^3}{3} + \frac{2x^5}{15} + O(x^7)\\ \sin x &= x - \frac{x^3}{6} + \frac{x^5}{120} + O(x^7)\\ \tan x - \sin x &= \frac{x^3}{2} + \frac{7x^5}{120} + O(x^7)\end{aligned}$$

Obe formule daju potpuno točan rezultat:

```
In [25]: tan(x)*sin(x)^2/(1+cos(x)), x^3/2+7*x^5/120
```

```
Out[25]: (5.0e-31, 5.0e-31)
```