

# Aritmética de Ponto Flutuante

Prof. Americo Cunha

Universidade do Estado do Rio de Janeiro – UERJ

[americo.cunha@uerj.br](mailto:americo.cunha@uerj.br)

[www.americocunha.org](http://www.americocunha.org)




# Um experimento computacional

$$a = 4/3, \quad b = a - 1, \quad c = 3b, \quad e = 1 - c$$

Em aritmética exata

$$e = 1 - c = 1 - 3b = 1 - 3(a - 1) = 1 - 3(4/3 - 1) = 0$$

Agora vamos fazer essa conta no GNU Octave:



```
> > format long
> > a = 4/3
> > b = a - 1
> > c = 3*b
> > e = 1 - c
```



# As operações em ponto flutuante

Sejam  $f1(x)$  e  $f1(y)$  as representação dos reais  $x$  e  $y$  em  $F(\beta, t, m, M)$ , respectivamente. As *operações aritméticas em ponto flutuante*, são definidas por:

$$x \oplus y = f1(f1(x) + f1(y))$$

$$x \ominus y = f1(f1(x) - f1(y))$$

$$x \otimes y = f1(f1(x) \times f1(y))$$

$$x \oslash y = f1(f1(x) / f1(y))$$

Ao final de cada operação *arredonda-se (ou trunca-se)*.



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$x \oplus y$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$x \oplus y = \text{fl}(\text{fl}(x) + \text{fl}(y))$$





# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$\begin{aligned} x \oplus y &= \text{fl}(\text{fl}(x) + \text{fl}(y)) \\ &= \text{fl}(0,9370 \times 10^4 + 0,1272 \times 10^2) \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$\begin{aligned} x \oplus y &= \text{fl}(\text{fl}(x) + \text{fl}(y)) \\ &= \text{fl}(0,9370 \times 10^4 + 0,1272 \times 10^2) \\ &= \text{fl}((0,9370 + 0,001272) \times 10^4) \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$\begin{aligned} x \oplus y &= \text{fl}(\text{fl}(x) + \text{fl}(y)) \\ &= \text{fl}(0,9370 \times 10^4 + 0,1272 \times 10^2) \\ &= \text{fl}((0,9370 + 0,001272) \times 10^4) \\ &= \text{fl}(0,938272 \times 10^4) \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$fl(x) = 0,9370 \times 10^4 \quad fl(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$\begin{aligned} x \oplus y &= fl(fl(x) + fl(y)) \\ &= fl(0,9370 \times 10^4 + 0,1272 \times 10^2) \\ &= fl((0,9370 + 0,001272) \times 10^4) \\ &= fl(0,938272 \times 10^4) \\ &= 0,9382 \times 10^4 \text{ (truncamento)} \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$fl(x) = 0,9370 \times 10^4 \quad fl(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Soma em ponto flutuante:

$$\begin{aligned} x \oplus y &= fl(fl(x) + fl(y)) \\ &= fl(0,9370 \times 10^4 + 0,1272 \times 10^2) \\ &= fl((0,9370 + 0,001272) \times 10^4) \\ &= fl(0,938272 \times 10^4) \\ &= 0,9382 \times 10^4 \text{ (truncamento)} \\ &= 0,9383 \times 10^4 \text{ (arredondamento)} \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Produto em ponto flutuante:

$$x \otimes y$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Produto em ponto flutuante:

$$x \otimes y = \text{fl}(\text{fl}(x) \times \text{fl}(y))$$





# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Produto em ponto flutuante:

$$\begin{aligned} x \otimes y &= \text{fl}(\text{fl}(x) \times \text{fl}(y)) \\ &= \text{fl}(0,9370 \times 10^4 \times 0,1272 \times 10^2) \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Produto em ponto flutuante:

$$\begin{aligned} x \otimes y &= \text{fl}(\text{fl}(x) \times \text{fl}(y)) \\ &= \text{fl}(0,9370 \times 10^4 \times 0,1272 \times 10^2) \\ &= \text{fl}(0,1191864 \times 10^6) \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$\text{fl}(x) = 0,9370 \times 10^4 \quad \text{fl}(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Produto em ponto flutuante:

$$\begin{aligned} x \otimes y &= \text{fl}(\text{fl}(x) \times \text{fl}(y)) \\ &= \text{fl}(0,9370 \times 10^4 \times 0,1272 \times 10^2) \\ &= \text{fl}(0,1191864 \times 10^6) \\ &= 0,1191 \times 10^4 \text{ (truncamento)} \end{aligned}$$



# Um exemplo simplista em ponto flutuante

Um sistema de ponto flutuante e dois reais dados:

$$F(10, 4, -1022, 1023) \quad x = 9370 \quad y = 12,72$$

Representação em ponto flutuante:

$$fl(x) = 0,9370 \times 10^4 \quad fl(y) = 0,1272 \times 10^2$$

→ A representação é exata!

Produto em ponto flutuante:

$$\begin{aligned} x \otimes y &= fl(fl(x) \times fl(y)) \\ &= fl(0,9370 \times 10^4 \times 0,1272 \times 10^2) \\ &= fl(0,1191864 \times 10^6) \\ &= 0,1191 \times 10^4 \text{ (truncamento)} \\ &= 0,1192 \times 10^4 \text{ (arredondamento)} \end{aligned}$$



# Qual o resultado desses cálculos?



# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$



# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$

Truncamento:

$$x \oplus y = 0,9382 \times 10^4$$

$$x \otimes y = 0,1191 \times 10^6$$



# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$

Truncamento:

$$x \oplus y = 0,9382 \times 10^4$$

$$x \otimes y = 0,1191 \times 10^6$$

$$“x + y = 9382” \neq 9382,72 \quad “x \times y = 119100” \neq 119186,4$$





# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$

Truncamento:

$$x \oplus y = 0,9382 \times 10^4$$

$$x \otimes y = 0,1191 \times 10^6$$

$$“x + y = 9382” \neq 9382,72 \quad “x \times y = 119100” \neq 119186,4$$

Arredondamento:

$$x \oplus y = 0,9383 \times 10^4$$

$$x \otimes y = 0,1192 \times 10^6$$



# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$

Truncamento:

$$x \oplus y = 0,9382 \times 10^4$$

$$x \otimes y = 0,1191 \times 10^6$$

$$"x + y = 9382" \neq 9382,72 \quad "x \times y = 119100" \neq 119186,4$$

Arredondamento:

$$x \oplus y = 0,9383 \times 10^4$$

$$x \otimes y = 0,1192 \times 10^6$$

$$"x + y = 9383" \neq 9382,72 \quad "x \times y = 119200" \neq 119186,4$$



# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$

Truncamento:

$$x \oplus y = 0,9382 \times 10^4$$

$$x \otimes y = 0,1191 \times 10^6$$

$$“x + y = 9382” \neq 9382,72 \quad “x \times y = 119100” \neq 119186,4$$

Arredondamento:

$$x \oplus y = 0,9383 \times 10^4$$

$$x \otimes y = 0,1192 \times 10^6$$

$$“x + y = 9383” \neq 9382,72 \quad “x \times y = 119200” \neq 119186,4$$

Em geral, operações em ponto flutuante não são exatas.



# Qual o resultado desses cálculos?

Aritmética exata:

$$x + y = 9382,72$$

$$x \times y = 119186,4$$

Truncamento:

$$x \oplus y = 0,9382 \times 10^4$$

$$x \otimes y = 0,1191 \times 10^6$$

$$“x + y = 9382” \neq 9382,72 \quad “x \times y = 119100” \neq 119186,4$$

Arredondamento:

$$x \oplus y = 0,9383 \times 10^4$$

$$x \otimes y = 0,1192 \times 10^6$$

$$“x + y = 9383” \neq 9382,72 \quad “x \times y = 119200” \neq 119186,4$$

Em geral, operações em ponto flutuante não são exatas.

Até quando  $x$  e  $y$  tem representação exata!



# Erros nas operações em ponto flutuante

Se  $\text{fl}(x), \text{fl}(y) \in F(\beta, t, m, M)$  então

$$x \oplus y = (\text{fl}(x) + \text{fl}(y))(1 + \varepsilon_1)$$

$$x \ominus y = (\text{fl}(x) - \text{fl}(y))(1 + \varepsilon_2)$$

$$x \otimes y = (\text{fl}(x) \times \text{fl}(y))(1 + \varepsilon_3)$$

$$x \oslash y = (\text{fl}(x) / \text{fl}(y))(1 + \varepsilon_4)$$

onde  $|\varepsilon_j| \leq \epsilon_M$  ( $\epsilon_M$  é precisão da máquina).

Essas operações:

- tem erro relativo menor que  $\epsilon_M$ ;
- não são associativas e nem distributivas.

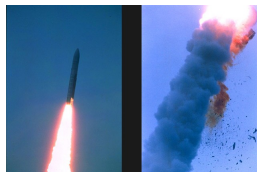


# Desastres envolvendo ponto flutuante

- Bolsa de Vancouver  
Jan. 1982: ↑ 1000,000 pontos  
Nov. 1983: ↓ 524,811 pontos  
Corrigido: ↑ 1098,892 pontos
- Míssil Patriota (25/2/1991)  
28 vítimas fatais e +100 feridos
- Foguete Ariane 5 (4/6/1996)  
US\$ 500M em prejuízo



Patriota ©



Ariane 5\*

Detalhes em:

[http://shodor.org/succeed-1.0/programs/sspa2002/pages/math\\_disasters.html](http://shodor.org/succeed-1.0/programs/sspa2002/pages/math_disasters.html)

\* Figura obtida em <https://5g.security/timeline/ariane-5-rocket-explosion/>

## Para pensar em casa ...

### Exercício teórico:

Considere o sistema de ponto flutuante

$F(\beta, t, m, M) = F(10, 3, -3, 4)$ , que trunca a parcela que não pode ser incorporada à mantissa.

Nesse sistema de ponto flutuante, qual o resultado de  $10^3 + 1 - 10^3$ ?

### Exercício teórico:

Sabendo que o sistema de ponto flutuante

IEEE 754 de precisão dupla é da forma

$F(\beta, t, m, M) = F(2, 53, -1022, 1023)$ , explique porque a maioria dos dispositivos eletrônicos calcula  $(1 + 2^{53}) - 2^{53} = 0$ .



## Como citar esse material?

A. Cunha, *Aritmética de Ponto Flutuante*, Universidade do Estado do Rio de Janeiro – UERJ, 2020.

Essas notas de aula podem ser compartilhadas nos termos da licença Creative Commons BY-NC-ND 3.0, com propósitos exclusivamente educacionais.

