Primeira Lista de Cálculo Numérico
Primeiro trimestre 2012
Rodrigo Fresneda

13 de fevereiro de 2012

1. Considere o sistema $F(2,5,3,1)$
   (a) Quantos números podemos representar neste sistema?
   (b) Qual o maior número na base 10 que podemos representar neste sistema (sem fazer arredondamento)?
2. Mude da base 10 para base 2 os seguintes números:
   (a) 78.5
   (b) 0.1
   (c) 0.125
   (d) 34
   (e) 33.023
3. Considere o sistema $F(10,3,5,5)$. Represente neste sistema os números: $x_1 = 1234.56$, $x_2 = -0.00054962$, $x_3 = 0.9995$, $x_4 = 123456.7$ e $x_5 = -0.0000001$.
4. Considere os seguintes números: $x_1 = 110111$, $x_2 = 0.01011$ e $x_3 = 11.0101$ que estão na base 2. Escreva-os na base 10.
5. Considere a representação de ponto flutuante $F(2,3,1,2)$.
   (a) escreva todos os números representáveis e converta-os para base 10.
   (b) os números $x_1 = 0.38$, $x_2 = 5.3$ e $x_3 = 0.15$ são representáveis?
6. Efetue as operações de ponto flutuante indicadas, levando em conta que $\beta = 10$ e $t = 3$:
   (a) $(11.4 + 3.18) + 5.05$ e $11.4 + (3.18 + 5.505)$,
   (b) $(3.18 \times 11.4)/5.05$ e $(3.18/5.05) \times 11.4$,
   (c) $3.18 \times (5.05 + 11.4)$ e $3.18 \times 5.05 + 3.18 \times 11.4$
   (d) Quais os erros relativos nas operações acima?

7. Avaliar o polinômio $p(x) = x^3 - 6x^2 + 4x - 0.1$ no ponto 5.24 nos seguintes casos:

   (a) Calcule sem arredondar;

   (b) Calcule com arredondamento (verifique se vale associatividade para soma);

   (c) Desta vez utilize a expressão $p(x) = x(x(x-6)+4) - 0.1$ e compare o resultado com item b.

8. Considere uma representação de ponto flutuante com $t = 10$, $\beta = 10$ e $|e| \leqq 10$. Qual dos procedimentos abaixo é melhor? Comente.

   (a) Calcule diretamente $\sqrt{9876} - \sqrt{9875}$

   (b) Faça a conta acima utilizando a identidade

$$\sqrt{x} - \sqrt{y} = \frac{x - y}{\sqrt{x} + \sqrt{y}}$$

9. Deseja-se calcular

$$S = \sum_{k=1}^{10} \frac{2}{k^2}$$

   no sistema $F(10, 3, 4, 5)$ usando arredondamento em todas as operações. Assim, efetue a soma:

   (a) da direita para a esquerda,

   (b) da esquerda para a direita,

   (c) os valores obtidos em a) e b são iguais? Explique.

10. Considere o sistema $F(2, 8, 4, 4)$ e os números $x_1 = 0.10110011 \times 2^2$ e $x_2 = 0.10110010 \times 2^2$. Qual dos dois números representa melhor $(2.8)_{10}$?

11. A representação de ponto flutuante de precisão dupla (64bits) no padrão IEE754 é caracterizada por 1 bit de sinal S, 11 bits para o expoente E, e 52 bits para a fração

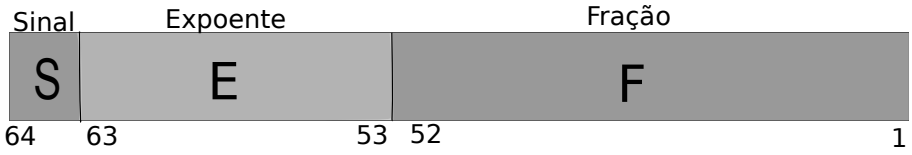

Figura 1: Representação de ponto flutuante em IEE754-64bits

Nessa representação, números normalizados são dados pela fórmula

$$x = (-1)^S * 2^{E-1023} * 1.F$$

em que $0 < E < 2^{11} - 1$ e $0 \leqslant F < 1$. Em relação a esse padrão, responda:

(a) Qual o maior número normalizável (base 10)?

(b) Qual o menor número normalizável (base 10)?

(c) Quanto vale eps?

(d) Represente 0.1 (determine os inteiros S,E e F na base 2).

12. Mostre que a representação de números normalizados no IEE754-64bits pode ser escrita também na forma

$$x = M * 2^{e-52}$$

em que e e M são inteiros satisfazendo $-1022 \leq e \leq 1023$ e $2^{52} \leq |M| < 2^{53}$.

13. Na representação descrita na questão anterior, mostre que M e e podem ser escritos como função de x como[1]:

$$e = \lfloor \log_2 |x| \rfloor \ , \ M = \frac{x}{2^{e-52}}$$

Represente 0.1 e compare com o resultado de 11d.

(a) Assumindo que $a \neq 0$ e $b^2 - 4ac > 0$, considere a equação $ax^2 + bx + c = 0$. As raízes podem ser calculadas com o auxílio das fórmulas

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \ , \ x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{1}$$

Mostre que essas raízes podem ser equivalentemente calculadas usando as fórmulas

$$x_1 = \frac{-2c}{b + \sqrt{b^2 - 4ac}} \ , \ x_2 = \frac{-2c}{b - \sqrt{b^2 - 4ac}} \tag{2}$$

Dica: racionalize os numeradores em (1). Nos casos em que $|b| \approx \sqrt{b^2 - 4ac}$, deve-se tomar cuidado para evitar cancelamento catastrófico. Se $b > 0$, então x_1 deve ser calculado pela fórmula (2) e x_2 pela fórmula (1). Se $b < 0$, então $x_1$deve ser calculado por (1) e x_2 por (2).

14. Use a fórmula apropriada para x_1 e x_2 como mencionado no exercício anterior para encontrar as raízes dos seguintes polinômios:

(a) $x^2 - 1000,001x + 1 = 0$

(b) $x^2 - 10.000,0001x + 1 = 0$

(c) $x^2 - 100.000,00001x + 1 = 0$

(d) $x^2 - 1.000.000,000001x + 1 = 0$

15. Mostre que se x é um número no sistema $F(\beta, t, m, M)$ então $x = x(1 + \delta)$ onde $|\delta| \leq \frac{1}{2}\beta^{1-t}$.

[1] O operador $\lfloor x \rfloor$ extrai o maior inteiro menor ou igual x, i.e., $\lfloor 2.9 \rfloor = 2$.