## Cálculo Numérico - IME/UERJ

## Lista de Exercícios 1

## Aritmética de ponto flutuante

1. Converter para decimal os seguintes números binários:

| (a)    | 10011  |
|--------|--|
| (b)    | 11100010   |
| (c)    | 1000001  |
| (d)    | 1,1  |
| (e)    | 1100,01  |
| (f)    | 1000,001   |
| 2. Con | verter para binário os seguintes números decimais:   |
| (a)    | 23   |
| (b)    | 2615   |
| (c)    | 2,5  |
| (d)    | 0,1  |
| (e)    | 3,8  |
| (f)    | 10,05  |
| bits   | computador armazena números reais utilizando 1 bit para o sinal do número, para o expoente e 8 bits para a mantissa. Admitindo que haja arredondamento o ficariam armazenados os seguintes números decimais? |
| (a)    | 265  |
| (b)    | 12, 5  |
| (c)    | -445, 25   |
| (d)    | -0,1   |
| (e)    | -12, 8   |
| (f)    | 2500,05  |
|        |  |

(a) Qual o maior e o menor número positivo nele representável?

(b) Qual o menor número maior que 100 nele representável?

4. Ainda sobre o computador do exercício anterior:

- (c) Qual o maior número menor que 20, nele representável?
- (d) Quais os erros absoluto e relativo ao se tentar nele representar os números: m = 25, 5; n = 120, 25; p = 2, 5; a = 460, 25; b = 24,005.
- (e) Usando os valores acima, trabalhando em binário, qual o resultado das operações a seguir, bem como os erros absoluto e relativo:

(e-1) 
$$m + n$$
; (e-2)  $m \times p$ ; (e-3)  $n \times p$ ; (e-4)  $a + b$ ; (e-5)  $a - b$ ; (e-6)  $a/n$ .

- 5. Seja um computador binário, cujo sistema de ponto flutuante tenha 1 bit para o sinal do número, 5 bits para o expoente e 6 bits para a mantissa num total de 12 bits. Responda:
  - (a) Qual o menor número positivo e o maior número positivo nele representável?
  - (b) Qual o maior e > 0, tal que 4,25 + e = 4,25?
  - (c) Qual o menor número maior que 4,25 nele representável?
  - (d) Qual o maior número menor que 80, nele representável?
  - (e) Efetue nele a multiplicação  $0.8 \times 5$  e indique o resultado
- 6. (Trabalho extra Valendo 1,0 ponto) Seja um computador binário de precisão simples, ou seja, de 32 bits, cujo sistema de ponto flutuante armazena 1 bit para o sinal do número, 8 bits para o expoente e 23 bits para a mantissa. Responda:
  - (a) Qual o maior número positivo nele representável?
  - (b) Qual o menor número positivo nele representável?
  - (c) Qual o erro relativo máximo considerando que houve truncamento ao aproximar um certo número?
  - (d) Qual o valor representado por 12,8 neste computador?
  - (e) Qual o valor representado por 28,8 neste computador?