CI1164 – Introdução à Computação Científica

Prof. Guilherme Derenievicz
Prof. Armando Delgado

Exercícios de Revisão para Prova 01

Questão 1

Considere um equipamento cujo sistema de ponto flutuante **normalizado** é SPF(2,4,-5,5), ou seja, de **base 2**, possui **4 dígitos na mantissa**, **menor expoente -5** e **maior expoente 5**. Os números abaixo são fornecidos a este sistema:

(a) 0.1011×2^4 (b) 0.1101×2^{-1} (c) 0.1110×2^1

Qual é o resultado das seguintes operações, considerando que a máquina efetua o truncamento dos resultados. Calcule também para cada item o valor exato (sem considerar truncamento) .

i. $(a+b)+c$

ii. $a+(b+c)$

iii. Explique a diferença de resultados verificadas nos itens (1) e (2)

Questão 2

Considere um equipamento cujo sistema de ponto flutuante **normalizado** é SPF(10,4,-5,5), ou seja, de **base 10**, possui **4 dígitos na mantissa**, **menor expoente -5** e **maior expoente 5**. Os números abaixo são fornecidos a este sistema:

(a) 0.4523×10^4 (b) 0.2116×10^{-1} (c) 0.2583×10^1

Qual é o resultado das seguintes operações, considerando que a máquina efetua o truncamento dos resultados. Calcule os erros absolutos e relativos destas aproximações:

i. $(a+b)+c$

ii. $a+(b+c)$

Questão 3

Considere os métodos numéricos estudados nesta disciplina e o cálculo de erros apresentados abaixo.

i) Dada uma função $f(x)$ definida e contínua no intervalo I , chamamos de zero (ou raiz) da função a todo $\alpha \in I \mid f(\alpha)=0$. Considere $x_i \approx \alpha$ o resultado da i-ésima iteração de algum método numérico para o cálculo de α . Indique em que situações cada uma das formas de cálculo de erro abaixo é mais adequada:

(a) $|f(x_i)|$

(b) $\dfrac{|x_i - x_{i-1}|}{|x_i|}$

(c) $|x_b - x_a|$, onde $x_a \le \alpha \le x_b$

ii) Seja um Sistema de Equações Lineares da forma $Ax=b, A \in \mathbb{R}^{n \times n}, \{x, b\} \in \mathbb{R}^n$ e $\overline{x}^{(k)} \approx x$ o resultado da k-ésima iteração de algum método para a solução de sistemas lineares. Lembrando que o resíduo é definido $r = b - A\overline{x}^{(k)}$, indique em que situações cada uma das formas de cálculo de erro abaixo é mais adequada:

(a) $\|x\|_\infty = max\left(|\overline{x}_i^{(k)} - \overline{x}_i^{(k-1)}|\right), i=1,2,\ldots,n$

(b) $\|r\|_2 = \sqrt{r_1^2 + r_2^2 + \ldots + r_n^2}$

iii) Sejam `xd` e `raiz` a representação em ponto flutuante IEEE754 (`double`) dos valores de x_i e α do item (i), respectivamente. Considerando ainda que o método numérico esteja convergindo, explique porque o laço a seguir (em linguagem C) não é uma opção para testar a convergência do método.

```
while( fabs(xd - raiz) > 0.0) )
{
    ...
}
```

Questão 4

O algoritmo abaixo pode ser utilizado para calcular com quantos dígitos um computador trabalha.

```
ε ← 1.0
j = 1
Enquanto (1.0 + ε > 1.0) faça
   ε ←  ε / 2.0
   j ← j + 1

Escreva o valor de j
```

Explique por que o algoritmo não entra em laço infinito.

Questão 5

Considere as duas expressões equivalentes abaixo para calcular a abcissa da interseção da reta que passa pelos pontos $(x_{0,} y_0)$ e (x_1, y_1) com o eixo x :

(a) $x=\frac{x_0 y_1 - x_1 y_0}{y_1 - y_0}$ (b) $x=x_0-\frac{(x_1-x_0)y_0}{y_1-y_0}$

i. Usando os pontos $(0.131\times10^1, 0.324\times10^1)$ e $(0.193\times10^1, 0.476\times10^1)$ calcule o valor de x em um SPF(10,3,-5,5) **normalizado**. Calcule os erros absolutos e relativos destas aproximações.

ii. Qual dos dois métodos é melhor? Justifique.

Questão 6

Considere um equipamento cujo sistema de ponto flutuante **normalizado** é SPF(2,3,-4,4). Responda:

(a) Qual o menor número positivo exatamente representável, em base 2?

(b) Qual o próximo positivo, depois do menor positivo representável, em base 2?

(c) Verifique se existem números reais entre o menor e o próximo positivo. Comente as implicações de sua verificação.

Questão 7

Um paralelepípedo retangular **tem** dimensões $x=3\,\text{cm},\ y=4\,\text{cm e } z=5\,\text{cm}$. Ele foi medido com um paquímetro com precisão de $\pm 0{,}1\,\text{cm}$.

a) Calcule o erro absoluto máximo e o erro relativo máximo no volume do paralelepípedo.
b) Este erro é Real ou Aproximado? Justifique.

Questão 8

Observe o trecho de código a seguir, considere as variáveis **soma1** e **soma2** e responda:

```
float soma1=0.0f, soma2=0.0f;

for (int i=1; i<=200; ++i)
   soma1 += 1.0f / (i*i);

for (int i=200; i>=1; --i)
   soma2 += 1.0f / (i*i);
```

a) Qual variável terá o valor mais exato? Por que isso ocorre?
b) A precisão das variáveis é a mesma? Justifique sua resposta.

Questão 9

Considere um equipamento cujo sistema de ponto flutuante (SPF) **normalizado** de **base 2** possui **4 dígitos na mantissa**, **menor expoente -5** e **maior expoente 5** (SPF(2,4,-5,5)). Para este sistema:

a) Qual a diferença entre o menor número positivo representável e o próximo número, imediatamente maior (em base 2)?
b) Qual a diferença entre o maior número positivo representável e o número anterior, imediatamente menor (em base 2)?
c) Qual é o maior número inteiro ímpar que este sistema pode representar (em base 2 ou base 10)?
d) Explique uma implicação de suas respostas anteriores.

Questão 10

Escreva uma função em linguagem C que receba como parâmetros de entrada os limites de um intervalo (a,b) e um valor de tolerância, e calcule uma raiz da equação $sen(x)-x^3+x+1=0$ utilizando duas iterações do método da Bisseção e em seguida o método da Secante até que o erro aproximado absoluto em x seja menor do que a tolerância estipulada.

DICA: a função de iteração do método da Secante é dada por

$$x_{k+1}=\frac{x_{k-1}f(x_k)\ -\ x_k f(x_{k-1})}{f(x_k)\ -\ f(x_{k-1})}\quad .$$

Questão 11

Escreva uma função em linguagem C que receba como parâmetros de entrada as diagonais de uma matriz tri-diagonal, o vetor de termos independentes $\vec{b}$, o vetor com os valores iniciais para $\vec{x}$, a ordem **n** da matriz, e um valor de tolerância tol, e devolva sua solução $\vec{x}$ usando o método de Gauss-Seidel até 5 (cinco) iterações. Se existir alguma condição ou propriedade para os vetores das diagonais e termos independentes, isto deve ser indicado na resposta.

Questão 12

Seja um Sistema de Equações Lineares da forma $Ax=b, A\in \mathbb{R}^{n\times n}, \{x,b\}\in \mathbb{R}^n$. Para a solução deste sistema, quantas operações aritméticas são realizadas EM CADA ITERAÇÃO dos Métodos de Jacobi e de Gauss-Seidel quando:

a) **A** é uma matriz $n\times n$ *?*
b) **A** é uma matriz de banda *k*-diagonal?
c) Em quais condições estes métodos iterativos serão competitivos com o Método da Eliminação Gaussiana (sem considerar pivotamento)?

Questão 13

a) Explique como a inversa de uma matriz $A, n \times n$, pode ser obtida através da resolução de n sistemas lineares $n \times n$.

b) Entre o método da Eliminação de Gauss e a decomposição LU, qual o mais indicado para este caso? Justifique.

Questão 14

Matrizes tridiagonais são aquelas em que apenas os elementos da diagonal principal, e os elementos das diagonais imediatamente acima e abaixo são não nulos. Sistemas lineares com matrizes de coeficientes tridiagonais são bastante comuns na solução de problemas de computação científica.

a) Elabore uma estrutura de dados em linguagem C para armazenar um sistema linear com matriz de coeficientes tridiagonal, que seja eficiente para resolução pelo método de Gauss-Seidel;

b) Implemente uma função em C que resolva pelo método de Gauss-Seidel um sistema linear tridiagonal;

Questão 15

Dado um sistema linear com N equações, da forma $Ax=b$, a decomposição de Cholesky fatoriza a matriz de coeficientes A em uma matriz triangular superior e sua conjugada transposta $A=RR^T$. Os elementos de $R=r_{i,j}$ $i \leq j \leq N$ são calculados da seguinte maneira:

Elementos da diagonal principal:

$$r_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} (r_{ki})^2}$$

Elementos acima da diagonal principal $(i<j)$:

$$r_{ij} = \sqrt{a_{ij} - \sum_{k=1}^{i-1} (r_{ki})(r_{kj})}$$

Implemente uma função em linguagem C, utilizando o cabeçalho definido abaixo, para calcular a matriz R da decomposição de Cholesky.

```
/* A: matriz de coeficientes  de um S.L. de ordem 'n'
   R: decomposição de Cholesky
   n: ordem das matrizes A e R */
void cholesky( double A[][], double R[][], uint n )
{   }
```

Dica: A decomposição é calculada uma coluna de cada vez, iniciando pelo elemento

$r_{11} = \sqrt{a_{11}}$

Questão 16

Defina o que é um sistema linear bem condicionado (estável) e o que é um sistema linear mal condicionado.

Questão 17

Quando a decomposição LU é vantajosa computacionalmente se comparada ao Método da Eliminação de Gauss?