UNIVERSIDADE DO MINHO

Licenciatura em Ciências da Computação

Análise Numérica

Duração: 2 horas — 29 de outubro de 2018

TESTE 1 (COM CONSULTA)

1. **a)** No Matlab executa

   ```
   >> fracdectobin (2^-1+2^-3+2^-7+2^-11 )
   ```

   e explica o resultado obtido (nota: a função fracdectobin foi desenvolvida ns aulas).

   b) Determina a representação decimal do número que na base 3 tem a representação

   $$21.1022001$$

   Na tua folha de respostas escreve o(s) comando(s) executados no Matlab e o resultado obtido.

2. Em aritmética exata, as condições

   $$x > 0$$

   e

   $$0.25 + x > 0.25$$

   são equivalentes. Usa um valor de x à tua escolha para mostrar que tal equivalência não é verdadeira no Matlab.

3. Seja x um número real tal que

   $$realmin \leq x \leq realmax$$

   e seja

   $$fl(x) = (1.b_{-1}b_{-2}\cdots b_{-52})_2 \times 2^E$$

   a representação normalizada de x. Diz, justificando, se é verdadeira a condição

   $$\frac{|x - fl(x)|}{x} < 2^{-52},$$

   independentemente do expoente E.

4. Para $|x| < 1$, tem-se

   $$log(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} x^{n+1}$$

   onde $log(1+x)$ é o logaritmo natural de $1+x$.

   a) Se quisermos usar a série de potências anterior para aproximar o valor de $log(1.5)$ com um erro de truncatura que é garantidamente menor do que 0.001, qual é o último termo que teremos que adicionar? Justifica a tua resposta.

b) No Matlab, soma os termos necessários para calcular $log(1.5)$ com erro de truncatura inferior a 0.001. Na tua folha de respostas escreve o(s) comando(s) executado(s) e o resultado obtido.

5. Seja $\tilde{x}$ uma aproximação de $x = 0.001$ tal que

$$|x - \tilde{x}| \approx 10^{-6}.$$

Será de esperar que

$$|log(x) - log(\tilde{x})| \approx 10^{-6}?$$

Justifica a tua resposta.

6. No Matlab define a função dada por

$$f(x) = \frac{(1 - cos(x))(1 + cos(x))}{sin^2(x)}$$

e executa $>> f(pi + 1e - 7)$. Calcula o erro relativo do resultado obtido e diz qual é a causa deste erro.

7. Seja g uma função contínua e derivável num certo intervalo [a,b] que contem uma e uma só raiz da equação $g(x) = 0$. Diz, justificando, se concordas com a seguinte afirmação: *o número k de iterações do método da bisseção necessárias para produzir um intervalo* $[a_k, b_k]$ *tal que*

$$b_k - a_k < tol$$

(tol é uma tolerância fixada pelo utilizador) não depende dos valores da derivada $g'(x)$ *no intervalo [a,b].*

8. A equação polinomial

$$x^3 + x - 1 = 0$$

tem uma raíz próxima de 0.7

a) Será possível usar a fórmula iterativa

$$x_{k+1} = 1 - x_k^3$$

para calcular essa raíz? Porquê?

b) Reescreve a equação dada na forma $x = \varphi(x)$, com uma função φ que produza uma sequência de aproximações $x_{k+1} = \varphi(x_k)$ que seja convergente para a raíz.

c) Usa a função de iteração que escolheste na alínea anterior para calcular a raíz com pelo menos 5 algarismos significativos corretos. Deves apresentar na tua folha de respostas todas as iterações produzidas.

questão	1a	1b	2	3	4a	4b	5	6	7	8a	8b	8c	Total
cotação	1.5	1.5	1.5	1.5	1.5	2	1.5	2	1.5	2	2	1.5	20

RESOLUÇÃO

1. **a)**

```
>> fracdectobin(2^-1+2^-3+2^-7+2^-11)

ans =

     1     0     1     0     0     0     1     0     0     0     1
```

A função *fracdectobin* converte a representação decimal de um número fracionário puro para a base 2. Os bits produzidos corrrepondem portanto à representação binária

$$0.b_{-1}b_{-2}b_{-3}b_{-4}b_{-5}b_{-6}b_{-7}b_{-8}b_{-9}b_{-10}b_{-11}$$

com $b_{-1} = b_{-3} = b_{-7} = b_{-11} = 1$ e os restantes bits iguais a 0.

b) Trata-se do número

$$2 \times 3^1 + 1 \times 3^0 + 1 \times 3^{-1} + 2 \times 3^{-3} + 2 \times 3^{-4} + 1 \times 3^{-7}$$

que pode calcular-se simplesmente executando

```
>> 2*3^1+1*3^-1+2*3^-3+2*3^-4+1*3^-7
```

que dá 7.4326. O método mais eficiente está implementado na função horner. Com

```
>> horner([2,1],3)+horner([1,0,0,2,2,0,1,0],1/3)
```

obtemos o mesmo resultado sem ser necessário calcular potências.

2. Uma vez que

$$0.25 = 1 \times 2^{-2}$$

o número que lhe sucede no conjunto dos números de ponto flutuante no formato duplo da norma IEEE é

$$0.25 + 2^{-54} = \left(1 + 2^{-52}\right) \times 2^{-2}$$

e para qualquer número x tal que $0 < x \leq 2^{-55}$, o valor aredondado de $(0.25+x)$ será 0.25.

3. A condição é verdadeira. Para o erro absoluto de arredondamento de um número positivo x tem-se

$$|x - fl(x)| < 2^{E-52}$$

donde resulta, atendendo a $x \geq 2^E$ (por ser normalizada a representação),

$$\frac{|x - fl(x)|}{x} < \frac{2^{E-52}}{2^E} = 2^{-52}.$$

4. **a)** Numa série alternada o erro de truncatura é inferior ao valor absoluto do primeiro termo que se despreza. Uma vez que

$$\frac{0.5^7}{7} = 0.0011... > 0.001$$

e

$$|-\frac{0.5^8}{8}| = 4.8...e - 04 < 0.001$$

concluímos que teremos que adicionar ainda o termo $\frac{0.5^7}{7}$ para garantir um erro de truncatura inferior a 0.001.

b) Executando no Matlab

```
>> x=0.5; x-x^2/2+x^3/3-x^4/4+x^5/5-x^6/6+x^7/7
```

obtemos o resultado 0.4058. O mesmo resultado é obtido de forma mais eficiente com

```
>> horner([1/7 -1/6 1/5 -1/4 1/3 -1/2 1 0],0.5)
```

5. Se f é uma função que admite primeira e segunda derivadas contínuas num intervalo que contem os pontos x e $\tilde{x}$, tem-se (fórmula de Taylor com resto de 2ª ordem na forma de Lagrange)

$$f(x) = f(\tilde{x}) + f'(x)(x - \tilde{x}) + \frac{f''(\theta)}{2}(x - \tilde{x})^2$$

onde θ está entre x e $\tilde{x}$. Para $\tilde{x}$ suficientemente próximo de x, o termo de 2ª ordem (isto é, que contem $(x - \tilde{x})^2$) pode desprezar-se e resulta

$$|f(x) - f(\tilde{x})| \approx |f'(x)| \times |x - \tilde{x}|$$

e o erro na variável independente será multiplicado por $|f'(x)|$ (o número de condição absoluto da função f no ponto x) para dar (aproximadamente) o erro no valor calculado $f(\tilde{x})$. Com $f(x) = log(x)$ é $f'(x) = 1/x$ e para $x = 0.001$ é $f'(x) = 1000$. Portanto, tem-se neste caso

$$|log(x) - log(\tilde{x})| \approx 1000 \times 10^{-6}.$$

6.
```
>> f=inline('(1-cos(x))*(1+cos(x))/sin(x)^2')

f =

     Inline function:
     f(x) = (1-cos(x))*(1+cos(x))/sin(x)^2

>> f(pi+1e-7)

ans =

    0.9992
```

O resultado correto é um, como resulta da fórmula fundamental da trigonometria

$$sin^2(x) + cos^2(x) = 1.$$

Assim, o erro relativo é

$$\frac{1-0.9992}{1} = 7.9927e-04.$$

O erro é causado pelo cancelamento subtrativo que ocorre no cálculo de $1+cos(x)$ uma vez que $cos(\pi+10^{-7}) = -0.99999999999995...$ e são perdidos muitos algarismos significativos na operação

$$1+(-0.99999999999995).$$

7. A afirmação é verdadeira. O método da bisseção não usa os valores da derivada da função (com efeito, para usar este método basta que g seja uma função contínua, não precisa de ser derivável). Uma vez que

$$b_k - a_k = \frac{b-a}{2^k},$$

o número de iterações necessárias para satisfazer a condição

$$b_k - a_k < tol$$

só depende da amplitude $b-a$ do intervalo inicial e da tolerância tol.

8. **a)** Seja α tal que $\alpha = \varphi(\alpha)$ (isto é, α é ponto fixo de φ). Para a sequência produzida com $x_{k+1} = \varphi(x_k)$ tem-se

$$x_{k+1} - \alpha = \varphi'(\xi)(x_k - \alpha)$$

onde ξ é um ponto entre α e x_k. Daqui se conclui que para que a sucessão x_k convirja para α é necessário ter-se $|\varphi'(x)| < 1$ num intervalo que contem a raíz. Com

$$\varphi(x) = 1 - x^3$$

tem-se $\varphi'(x) = -3x^2$ e para valores próximos de 0.7 a condição $|\varphi'(x)| < 1$ não é cumprida e o método diverge.

b) Uma possibilidade é escrever a equação na forma

$$x = (1-x)^{1/3}$$

uma vez que a derivada

$$\varphi'(x) = \frac{1}{3}(1-x)^{-2/3}$$

satisfaz a condição de convergência (embora seja $\varphi'(0.7) = 0.74...$ e a convergência será lenta).

c)
```
>> phi=inline('(1-x)^(1/3)')

phi =

     Inline function:
     phi(x) = (1-x)^(1/3)

>> x=0.7
```
A repetição do comando

```
>> x=phi(x)
```

produz a seguinte sucessão de aproximações

```
0.66...
0.69...
0.67...
0.68...
0.678...
0.684...
0.680...
0.683...
 0.681...
0.682...
0.6818...
 0.6826...
 0.6820...
0.6824...
0.6822...
0.68241...
0.68226...
0.68237...
 0.68229...
0.68235...
 0.68231...
 0.68233...
0.68231...
0.68233...
0.68232...
0.68233...
0.682325...
```

e com 5 algarismos significativos corretos a raíz é 0.68233.