Cálculo Numérico
Aula 1

Fabricação – 2 sem

Prof Luis Carlos

Tópico 1: Erros

1 Introdução

Os fenômenos da natureza são extremamente complexos. Com o objetivo de melhor compreendê-los, desenvolvem-se modelos matemáticos mais simples do que a realidade, fornecendo desta forma resultados aproximados dos fenômenos originais.

Embora estes modelos matemáticos sejam simplificações dos fenômenos da realidade, ainda assim, com freqüência, são muito complexos para serem resolvidos analiticamente. Logo, existe a necessidade de resolvê-los numericamente.

Os métodos numéricos buscam soluções aproximadas para estes modelos.

Nos problemas reais, os dados são obtidos através de medições físicas e, como tais, não são exatas. Desta maneira a presença do erro é inevitável devido à imprecisão das medidas.

É próprio dos métodos numéricos trabalhar com características tais como: a aproximação, o erro e o desvio.

2 Erros

2.1 Existência

Como vimos anteriormente, os dados obtidos provenientes de medições físicas não são exatos, além disso, as operações associadas a estes dados não exatos propagam esses erros a resultados futuros. Por outro lado, os próprios métodos numéricos, buscam resultados o mais próximo possível do que seriam valores exatos de tal forma a **minimizar** estes erros.

Algumas questões devem ser salientadas:

- Cada medida é um intervalo e não um número. Isso decorre do processo de medição, do erro do medidor, da incerteza do valor verdadeiro. Dessa forma, um comprimento não é de 56.7 cm mas, possivelmente, (56.7 ± 0.2) cm, isto é, algo no intervalo 56.5 cm a 56.9 cm.
- Quando se executam operações matemáticas com este valor, sua incerteza é propagada para o resultado das operações. Chama-se a esse processo, *propagação de erro*.
- Os métodos numéricos, freqüentemente iterativos, podem ou não chegar a resultados exatos num número finito de iterações. Buscam obter valores aproximados, diminuindo o erro a cada iteração, num processo de aproximação sucessiva.

- O computador representa números reais com um número finito de dígitos, sendo forçado a aproximá-los quando os números reais exigem mais dígitos de que ele está capacitado para usar. Como exemplo, ao representarmos o número exato π , ele deverá ser forçosamente arredondado, pois seus infinitos dígitos não poderão ser integralmente representados no computador.
- Quando se representa um valor da forma $m \pm \varepsilon$, ε positivo, vamos sempre admitir que o valor de ε seja bem inferior ao valor absoluto de m, para se supor que a medida tenha sido bem feita. Assim, o valor m é expressivo diante de ε . A medida $23.537\text{m} \pm 0.002\text{m}$, significa que o valor está entre 23.535m e 23.539m. Essa medida tem boa precisão, boa aproximação do valor, embora com certa margem de erro.
- No entanto, se um comprimento é dado por $5\text{m} \pm 4\text{m}$, observamos que se sabe muito pouco sobre este valor, uma vez que ele pode variar desde 1 m até 9 m. Essa medida não tem boa precisão.
- Chama-se desvio absoluto, ou **erro absoluto**, ao valor de ε . (Muitas vezes usamos o desvio padrão para o valor de ε).
- Chama-se desvio relativo, ou **erro relativo**, à relação: $\dfrac{\varepsilon}{|m|}$,

  em que $|m|$ é o valor absoluto de m.

2.2 Propagação de erros

Vejamos alguns exemplos:

Dados $a = 50 \pm 3$ e $b = 21 \pm 1$, calcular a soma $a + b$, a subtração $a - b$ e o produto $a \,.\, b$

Menor valor de $a = 50 - 3 = 47$ Maior valor de $a = 50 + 3 = 53$

Menor valor de $b = 21 - 1 = 20$ Maior valor de $b = 21 + 1 = 22$

Então:

Menor valor de $a + b = (47 + 20) = 67$

Maior valor de $a + b = (53 + 22) = 75$

$$\Rightarrow (a+b) = (50+21) \pm 4 = 71 \pm 4$$

Menor valor de $a - b = (47 - 22) = 25$

Maior valor de $a - b = (53 - 20) = 33$

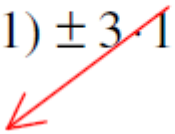
$$\Rightarrow (a-b) = (50-21) \pm 4 = 29 \pm 4$$

Observe que na subtração, os erros absolutos se somam, pois sempre se admite o pior caso; nunca se subtraem erros, contando com a sorte; prevê-se, sempre, o caso mais desfavorável.

Menor valor de $a.b = (47.20) = 940$

Maior valor de $a.b = (53\ .\ 22) = 1166$

$$\Rightarrow (a \cdot b) = (50 \pm 3) \cdot (21 \pm 1) = 1050 \pm 50 \cdot 1 \pm 3 \cdot 21 \pm 3 \cdot 1 =$$
$$= 1050 \pm (50 \cdot 1 + 3 \cdot 21) \pm 3 \cdot 1 \approx 1050 \pm 113$$



Assim, o produto ficaria entre 937 e 1163, ligeiramente diferente do verdadeiro intervalo, exatamente pelo abandono do produto $3 \cdot 1$, considerado desprezível.

Quando efetuamos operações sobre números sujeitos a erro, esses erros se propagam aos resultados das operações, que vão refletir a incerteza dos números que compõem a operação.

Assim:

$$(a \pm e_a) + (b \pm e_b) = a + b \pm (e_a + e_b)$$

$$(a \pm e_a) - (b \pm e_b) = a - b \pm (e_a + e_b)$$

$$(a \pm e_a) \cdot (b \pm e_b) \approx a \cdot b \pm (a \cdot e_b + b \cdot e_a)$$

Estamos admitindo a, b, e_a, e_b sempre positivos. No caso de valores negativos tomaremos $-a$, $-b$, etc…

3 Aritmética de ponto flutuante

Em um computador ou calculadora, um número real
é representado por um sistema denominado **aritmética de ponto flutuante**.

Neste sistema, o número a será representado por:

$$a = \pm(.\, d_1 d_2 d_3 ... d_t) \times \beta^e$$

em que :

β : base em que a máquina trabalha

t : número de dígitos na mantissa; $0 \leq d_j \leq (\beta - 1)$, $j = 1, \ldots t$, $d_1 \neq 0$.

e : expoente no intervalo $[i\ ,\ s]$

Exemplo 1.1

Considere uma máquina que opera no sistema $\beta = 10$, $t = 3$, $e \in [-5 , 5]$. Representar nesta máquina os números 43.2 , 235.89 , 0.00000034 e 2135789.

Número	Representação por arredondamento	Representação por truncamento	Situação
43.2	0.432×10^{2}	0.432×10^{2}	*OK*
235.89	0.236×10^{3}	0.235×10^{3}	*OK*
0.00000034	Não há (0.34×10^{-6})	Não há (0.34×10^{-6})	*Underflow* (expoente menor do que –5)
2135789	Não há (0.214×10^{7})	Não há (0.213×10^{7})	*Overflow* (expoente maior do que 5)

Obs : Algumas linguagens de programação permitem que as variáveis sejam declaradas com *precisão dupla* de modo a controlar os problemas gerados por *underflow/overflow*. Neste caso, estas variáveis serão representadas no sistema de aritmética de ponto flutuante da máquina, mas com aproximadamente o dobro de dígitos disponíveis na mantissa. No entanto, o tempo de execução e o requerimento de memória aumentam de forma significativa.

Exemplo 1.2

Represente os mesmos números do exemplo anterior em uma máquina que opera no sistema $\beta = 10, t = 5, e \in [-10 , 10]$.

Número	Representação por arredondamento	Representação por truncamento	Situação
43.2	0.43200×10^{2}	0.43200×10^{2}	*OK*
235.89	0.23589×10^{3}	0.23589×10^{3}	*OK*
0.00000034	0.34000×10^{-6}	0.3400×10^{-6}	*OK*
2135789	0.21358×10^{7}	0.21357×10^{7}	*OK*