Conceitos básicos

Irineu Lopes Palhares Junior

FCT/UNESP,
irineu.palhares@unesp.br

Conteúdos

Informações sobre os conteúdos

Representação de um número inteiro

Dada uma base β, onde β é um inteiro ≥ 2, e é escolhido como uma potência de 2.
Assim, dado um número inteiro $n \neq 0$, ele possui uma única representação:

$$n = \pm (n_{-k} n_{-k+1} \ldots n_{-1} n_0) = \pm \left(n_0 \beta^0 + n_{-1} \beta^1 + \ldots + n_{-k} \beta^k \right), \quad (1)$$

onde os n_i, $i = 0, -1, \ldots - k$ são inteiros satisfazendo $\leq n_i < \beta$ e $n_{-k} \neq 0$.

Example

Por exemplo, na base $\beta = 10$, o número 1997 é representado por:

$$1997 = 7 \times 10^0 + 9 \times 10^1 + 9 \times 10^2 + 1 \times 10^3 \quad (2)$$

e é armazenado como $n_{-3} n_{-2} n_{-1} n_0$.

Representação por ponto fixo

Este foi o sistema usado, no passado, em muitos computadores. Assim, dado um número real, $x \neq 0$, ele será representado em ponto fixo por:

$$x = \pm \sum_{i=k}^{n} x_i \beta^i, \tag{3}$$

onde k e n são inteiros satisfazendo $k < n$ e, usualmente, $k <\leq 0$ e $n > 0$ e os x_i são inteiros satisfazendo $0 \leq x_i < \beta$.

Example

$1997.16 = \sum_{i=-3}^{2} x_i \beta^{-i} =$
$1 \times 10^3 + 9 \times 10^2 + 9 \times 10^1 + 7 \times 10^0 + 1 \times 10^{-1} + 6 \times 10^{-2}.$

e é armazenado como $x_{-3}x_{-2}x_{-1}x_0.x_1x_2$.

Representação por ponto fixo: adicional

Link: `https://www.youtube.com/watch?v=8Nev_5ISYOY`
Na representação por ponto fixo dividimos seu armazenamento em três partes: sinal, parte inteira e fracionária.

Example

$3.25 = 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-2} = (11.01)_2$

Na representação por 8 bits, temos:

0	0	0	1	0	1	0	0

Representação por ponto flutuante

Esta representação, que é mais flexível que a representação em ponto fixo, é universalmente utilizada nos dias atuais. Dado um número real, $x \neq 0$, ele será representado em ponto flutuante por:

$$x = \pm d \times \beta^e, \tag{4}$$

onde β é a base do sistema de numeração, d é a mantissa e e é o expoente. A mantissa é um número em ponto fixo, isto é:

$$d = \sum_{i=k}^{n} d_i \beta^{-i}, \tag{5}$$

onde, frequentemente, nos grandes computadores, $K = 1$, tal que se $x \neq 0$, então $d_1 \neq 0$; $0 \leq d_i < \beta$, $i = 1, 2, \ldots, t$, com t a quantidade de dígitos significativos ou precisão do sistema, $\beta^{-1} \leq d < 1$ e $-m \leq e \leq M$.

Observações

1. $d_1 \neq 0$ caracteriza o sistema de números em ponto flutuante normalizado.
2. O número zero pertence a qualquer sistema e é representado com mantissa igual a zero e $e = -m$.

Example

Escrever os números:

$$x_1 = 0.35, \quad x_2 = -5.172 \quad x_3 = 0.0123, \quad x_4 = 5391.3 \quad x_5 = 0.0003, \tag{6}$$

onde todos estão na base $\beta = 10$, em ponto flutuante na forma normalizada.

Notação $F(\beta, t, m, M)$

Agora, para representarmos um sistema de números em ponto flutuante normalizado na base β, com t dígitos significativos e com limites dos expoentes m e M, usaremos a notação: $F(\beta, t, m, M)$.
Assim, um número em $F(\beta, t, m, M)$ será representado por:

$$\pm 0.d_1 d_2 \ldots d_t \times \beta^e, \tag{7}$$

onde $d_1 \neq 0$ e $-m \leq e \leq M$.

Example

Considere o sistema $F(10, 3, 2, 2)$. Represente neste sistema os números do exemplo anterior, isto é,

$$x_1 = 0.35, \quad x_2 = -5.172 \quad x_3 = 0.0123, \quad x_4 = 5391.3 \quad x_5 = 0.0003, \tag{8}$$

Digitos significativos

Definition

Seja β a base do sistema de números em ponto flutuante. Dígitos significativos de um número x, são todos os algarismos de 0 a $\beta - 1$, desde que x esteja representado na forma normalizada.

Para exemplificar as limitações da máquina, consideremos agora o seguinte exemplo.

Example

Seja $f(x)$ uma função contínua real definida no intervalo $[a, b]$, $a < b$, e sejam $f(a) < 0$ e $f(b) > 0$. Então, de acordo com o teorema do valor intermediário, existe x, $a < x < b$, tal que $f(x) = 0$. Seja $f(x) = x^3 - 3$. Determinar x tal que $f(x) = 0$.

Solução - exemplo

Example

Solução: Para a função dada, consideremos $t = 10$ e $\beta = 10$. Obtemos então:

$$\begin{aligned} f\left(0.1442249570 \times 10^{1}\right) &= -0.2 \times 10^{-8} \\ f\left(0.1442249571 \times 10^{1}\right) &= 0.4 \times 10^{-8} \end{aligned} \tag{9}$$

Observe que entre $0.1442249570 \times 10^{1}$ e $0.1442249571 \times 10^{1}$ não existe nenhum número que possa ser representado no sistema dado e que a função f muda de sinal nos extremos deste intervalo. Assim, esta máquina não contém o número x tal que $f(x) = 0$ e, portanto, a equação dada não possui solução.

Exercícios

1. Considere o sistema $F(10, 4, 4, 4)$. Represente neste sistema os números:

$$\begin{aligned} x_1 &= 4321.24, \quad x_2 = -0.0013523, \quad x_3 = 125.64, \\ x_4 &= 57481.23, \quad x_5 = 0.00034. \end{aligned} \tag{10}$$

2. Represente no sistema $F(10, 3, 1, 3)$ os números do exercício 1.

Representação por ponto flutuante - adicional

Link: `https://www.youtube.com/watch?v=rx3wA1SkrGc`
Ponto flutuante: desloca-se o ponto da parte fracionária até logo após o primeiro dígito não-nulo, modificando o expoente conforme adequado.

Example

Escrever o número $(25.625)_{10}$ em base 2.

Mudança de base

Como já dissemos anteriormente, a maioria dos computadores trabalha na base β, onde β é um inteiro ≥ 2, e é normalmente escolhido como uma potência de 2. Assim, um mesmo número pode ser representado em mais de uma base. Além disso, sabemos que, através de uma mudança de base, é sempre possível determinar a representação em uma nova base. Veremos então, através de exemplos, como se faz mudança de base.

Exemplos

Example

Mudar a representação dos números:

a) 1101 da base 2 para a base 10,

b) 0.110 da base 2 para a base 10,

c) 13 da base 10 para a base 2,

d) 0.75 da base 10 para a base 2,

e) 3.8 da base 10 para a base 2.

Observação

No exemplo anterior, mudamos a repsentação de números na base 10 para a base 2 e vice-versa. O mesmo procedimento pode ser utilizado para mudar da base 10 para outra base qualquer e vice-versa. A pergunta que surge naturalmente é: qual o procedimento para representar um número que está numa dada base β_1, em uma outra base β_2, onde $\beta_1 \neq \beta_2 \neq 10$? Neste caso, devemos seguir o seguinte procedimento: inicialmente, representamos o número que está na base β_1, na base 10 e, a seguir, o número obtido na base 10, na base β_2.

Example

Dado o número 12.20 que está na base 4, representá-lo na base 3.

Exercícios

1. Considere os seguintes números $x_1 = 34$, $x_2 = 0.125$ e $x_3 = 33.023$ que estão na base 10. Escreva-os na base 2.
2. Considere os seguinte números: $x_1 = 110111$, $x_2 = 0.01011$ e $x_3 = 11.0101$ que estão na base 2. Escreva-os na base 10.
3. Considere os seguintes números: $x_1 = 33$, $x_2 = 0.132$ e $x_3 = 32.013$ que estão na base 4. Escreva-os na base 5.

Representação no sistema $F(\beta, t, m, M)$

Sabemos que os números reais podem ser representados por uma reta contínua. Entretanto, em ponto flutuante, podemos representar apenas pontos discretos na reta real.
Para ilustra este fato, consideremos o seguinte exemplo.

Example

Considre o sistema $F(2, 3, 1, 2)$. Quantos e quais números podem ser representados neste sistema?

Example

Considerando o mesmo sistema do exemplo anterior, represente os números: $x_1 = 0.38$, $x_2 = 5.3$ e $x_3 = 0.15$ dados na base 10.

Arredondamento

Todas as operações num computador são arredondadas. Para ilustrar este fato, consideremos o seguinte exemplo.

Example

Calcular o quociente entre 15 e 7.

Solução: Temos três representações alternativas:

$$x_1 = \frac{15}{7}, \qquad x_2 = 2\frac{1}{7}, \qquad x_3 = 2.142857. \tag{11}$$

No que x_1 e x_2 são representações exatas, e x_3 é uma aproximação do quociente. Suponha agora que só dispomos de quatro dígitos para representar o quociente entre 15 e 7. Daí, $\frac{15}{7} = 2.142$. Mas não seria melhor aproximarmos $\frac{15}{7}$ por 2.143? A resposta é sim, e isto significa que o número foi arredondado. Mas o que significa arredondar um número?

Arredondamento em Ponto Flutuante

Definition

Arredondar um número x, por outro com um número menor de dígitos significativos, consiste em encontrar um número $\bar{x}$, pertencente ao sistema de numeração, tal que $|\bar{x} - x|$ seja o menor possível.

Assim, para o exemplo dado: $|2.142 - x_3| = 0.000857$ e $|2.143 - x_3| = 0.000143$. Logo, 2.143 representa a melhor aproximação para $\frac{15}{7}$, usando quatro dígitos significativos.

Arredondamento em ponto flutuante

Daremos a seguir a regra de como arredondar um número.
Dado x, seja $\bar{x}$ sua representação em $F(\beta, t, m, M)$ adotando arredondamento. Se $x = 0$, então $\bar{x} = 0$. Se $x \neq 0$, então escolhemos s e e tais que:

$$|x| = s \times \beta^e, \text{ onde } \beta^{-1}\left(1 - \frac{1}{2}\beta^{-t}\right) \leq s < 1 - \frac{1}{2}\beta^{-t}. \tag{12}$$

Se e está fora do intervalo $[-m, M]$ não temos condições de representar o número no sistema. Se $e \in [-m, M]$ então calculamos:

$$s + \frac{1}{2}\beta^{-t} = 0.d_1 d_2 \ldots d_t d_{t+1} \tag{13}$$

e truncamos em t dígitos. Assim, o número arredondado será:

$$\bar{x} = (sinal\ x)\,(0.d_1 d_2 \ldots d_t) \times \beta^e. \tag{14}$$

Exemplo

Example

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 $x_5 = -0.0000001$.

Assim, em linhas gerais, para arredondar um número na base 10, devemos apenas observar o primeiro dígito a ser descartado. Se este dígito é menor que 5 deixamos os dígitos inalterados; e se é maior ou igual a 5 devemos somar 1 ao último dígito remanescente.

Exercícios

1. Considere o sistema $F(10, 4, 4, 4)$.
   - Qual o intervalo para s neste caso?
   - Represente os números do exemplo passado, $x_1 = 1234.56$, $x_2 = -0.00054962$, $x_3 = 0.9995$, $x_4 = 123456.7$ e $x_5 = -0.0000001$, neste sistema.

Operações aritméticas em ponto flutuante

Considere uma máquina qualquer e uma série de operações aritméticas. Pelo fato do arredondamento ser feito após cada operação, temos, ao contrário do que é válido para números reais, que as operações aritméticas (adição, subtração, divisão e multiplicação) não são nem associativas e nem distributivas. Ilustraremos este fato através de exemplos.
Nos exemplos a seguir, considere o sistema com base $\beta = 10$ e três dígitos significativos.

Example

Efetue as operações indicadas:

a) $(11.4 + 3.18) + 5.05$ e $11.4 + (3.18 + 5.05)$

b) $\frac{3.18 \times 11.4}{5.05}$ e $\left(\frac{3.18}{5.05}\right) \times 11.4$

c) $3.18 \times (5.05 + 11.4)$ e $3.18 \times 5.05 + 3.18 \times 11.4$.

Exemplos

Example

Somar $\frac{1}{3}$ dez vezes consecutivas, usando arredondamento.
Solução: Temos

$$0.333 + 0.333 + \ldots + 0.333 = 3.31. \tag{15}$$

Entretanto, podemos obter um resultado melhor se multiplicarmos 0.333 por 10, obtendo assim 3.33.

Example

Avaliar o polinômio:

$$P(x) = x^3 - 6x^2 + 4x - 0.1 \tag{16}$$

no ponto 5.24 e comparar com o resultado exato.

Observações

Observando os três últimos exemplos, vemos que erros consideráveis podem ocorrer durante a execução de um algoritmo. Isto se deve ao fato de que existem limitações da máquina e também porque os erros de arredondamento são introduzidos a cada operação efetuada. Em consequência, podemos obter resultados diferentes mesmo utilizando métodos numéricos matematicamente equivalentes.
Assim devemos ser capazes de conseguir desenvolver um algoritmo tal que os efeitos da aritmética discreto do computador permaneçam inofensivos quando um grande número de operações são executadas.

Exercícios

1. Considere o sistema $F(10, 3, 5, 5)$. Efetue as operações indicadas:
   a) $(1.386 - 0.987) + 7.6485$ e $1.386 - (0.987 - 7.6485)$,
   b) $\frac{1.338 - 2.038}{4.577}$ e $\frac{1.338}{4.5777} - \frac{2.038}{4.577}$.
2. Seja
$$x = \frac{17.678}{3.471} + \frac{(9.617)^2}{3.716 \times 1.85}. \tag{17}$$
   a) Calcule x com todos os algarismos da sua calculadora, sem efetuar arredondamento.
   b) Calcule x considerando o sistema $F(10, 3, 4, 3)$. Faça arredondamento a cada operação efetuada.
3. Seja $P(x) = 2.3x^3 - 0.6x^2 + 1.8x - 2.2$. Deseja-se obter o valor de $P(x)$ para $x = 1.61$.
   a) Calcule $P(1.61)$ com todos os algarismos da sua calculadora, sem efetuar arredondamento.
   b) Calcule $P(1.61)$ considerando sistema $F(10, 3, 4, 3)$. Faça arredondamento a cada operação efetuada.

Efeitos numéricos

Além dos problemas dos erros causados pelas operações aritméticas, das fontes de erros citadas ao longo desta apresentação, existem certos efeitos numéricos que contribuem para que o resultado obtido não tenha crédito. Alguns dos mais frequêntes são:

- Cancelamento
- Propagação do erro
- Instabilidade numérica
- Mal condicionamento

Cancelamento

O cancelamento ocorre na subtração de dois números quase iguais. Vamos supor que estamos operando com aritmética de ponto flutuante. Sejam x e y dois números com expoente e. Quando formamos a diferença $x - y$, ela também terá o expoente e. Se normalizarmos o número obtido, veremos que devemos mover os dígitos para a esquerda de tal forma que o primeiro dígito seja diferente de zero. Assim, uma quantidade de dígitos iguais a zero apaerece no final da mantissa do número normalizado. Estes zeros não possuem significado algum.
Veremos este fato através de exemplos, onde iremos considerar que estamos trabalhando com o sistema $F(10, 10, 10, 10, 10)$.

Exemplos

Example

Calcular: $\sqrt{9876} - \sqrt{9875}$

Solução: Temos que:

$$\sqrt{9876} = 0.9937806599 \times 10^2 \quad \text{e} \quad \sqrt{9875} = 0.9937303457 \times 10^2. \tag{18}$$

Portanto: $\sqrt{9876} - \sqrt{9875} = 0.0000503142 \times 10^2$.

A normalização muda este resultado para: $0.5031420000 \times 10^{-4}$. Assim, os quatro zeros no final da mantissa não têm significado e assim perdemos quatro casas decimais. A pergunta que surge naturalmente é: podemos obter um resultado mais preciso? Neste caso, a resposta é sim. Basta considerarmos a identidade:

$$\sqrt{x} - \sqrt{y} = \frac{x - y}{\sqrt{x} + \sqrt{y}} \tag{19}$$

Continuação - exemplo

Example

e assim, no nosso caso, obtemos:

$$\sqrt{9876} - \sqrt{9875} = \frac{1}{\sqrt{9876} + \sqrt{9875}} = 0.5031418679 \times 10^{-4} \qquad (20)$$

que é um resultado com todos os dígitos corretos.

Exemplo

Example

Resolver a equação

$$x^2 - 1634x + 2 = 0 \tag{21}$$

Solução: Temos

$$x = \frac{1634 \pm \sqrt{(1634)^2 - 4\,(2)}}{2} = 817 \pm \sqrt{667487}. \tag{22}$$

Assim:

$$\begin{aligned} x_1 &= 817 + 816.9987760 = 0.1633998776 \times 10^3 \\ x_2 &= 817 - 816.9987760 = 0.1224000000 \times 10^{-2} \end{aligned} \tag{23}$$

Os seis zeros da mantissa de x_2 são resultado do cancelamento e, portanto, não têm significado algum. Uma pergunta que surge naturalmente é: podemos obter um resultado mais preciso? Neste caso, a resposta é sim. Basta lembrar que o produto das raízes é igual ao termo independente da equação, ou seja:

$$x_1 \times x_2 = 2 \rightarrow x_2 = \frac{2}{x_1}. \tag{24}$$

Logo: $x_2 = 0.1223991125 \times 10^{-2}$, onde agora todos os dígitos estão corretos.

Observação

Nos exemplos dados, foi razoavelmente fácil resolver o problema do cancelamento. Entretanto, cabe salientar que nem sempre existe uma maneira trivial de resolver problemas ocasionados pelo cancelamento.

Propagação do Erro

O cancelamento não ocorre somente quando dois números quase iguais são subtraídos diretamente um do outro. Ele também ocorre no cálculo de uma soma, quando uma soma parcial é muito grande se comparada com o resultado final. Para exemplificar, consideremos que:

$$s = \sum_{k=1}^{n} a_k, \tag{25}$$

seja a soma a ser calculada, onde os a_k podem ser positivos ou negativos. Vamos supor que o cálculo seja feito através de uma sequência de somas parciais, da seguinte forma:

$$s_1 = a_1, \qquad s_k = s_{k-1} + a_k, \qquad k = 2, 3, \ldots, n, \tag{26}$$

tal que $s = s_n$.

Propagação do erro

Se a soma é calculada em aritmética de ponto fixo, então cada a_k está afetado de algum erro, os quais são limitados por algum ϵ para todo k. Se nenhum *overflow* ocorre, o erro na soma final s será de no máximo $n\epsilon$. Agora, devido ao fato de nem todos os a_k terem o mesmo sinal, então o erro será menor do que $n\epsilon$.
Mas se a soma é calculada em aritmética de ponto flutuante, um novo fenômeno pode ocorrer. Vamos supor que uma das somas intermediárias s_k é consideravelmente grande em relação à soma final s, no sentido que o expoente de s_k excede o expoente de s em, digamos, p unidades. É claro que isso só pode ocorrer se nem todos os a_k possuem o mesmo sinal. Se simularmos tal soma em aritmética de ponto fixo (usando para todas as somas parciais o mesmo expoente de s), então devemos trocar os últimos p dígitos de s_k por zeros. Estes dígitos influenciam os últimos dígitos de s e como, em geral, estão errados, não podemos falar que o erro final será pequeno.

A perda de algarismos significativos devido a uma soma intermediária grande é chamada de propagação do erro. Veremos este fato através de exemplos.

Exemplos

Example

Calcule $e^{-5.25}$ utilizando cinco dígitos significativos em todas as operações.
Solução: O seguinte resultado matemático é bem conhecido: para todo número real x

$$e^{-x} = \sum_{k=0}^{\infty} (-1)^k \frac{x^k}{k!}. \tag{27}$$

Se e^{-x} é calculado usando esta fórmula, a série deve ser truncada. Assim, já estaremos introduzindo um erro de truncamento. Vamos considerar os primeiros vinte termos da série anterior para avaliar $e^{-5.25}$.

Continuação - exemplo

Example

Temos então:

$$
\begin{aligned}
e^{-5.25} &= (0.10000 - 0.52500)\,10^{1} \\
&+ (0.13781 - 0.24117 + 0.31654 - 0.33236 + 0.29082 - 0.21811 + 0.14314)\,10^{2} \\
&+ (-0.83497 + 0.43836 - 0.20922)\,10^{1} + (0.91532 - 0.36965 + 0.13862)\,10^{0} \\
&+ (-0.48516 + 0.15919)\,10^{-1} + (-0.49164 + 0.14339)\,10^{-2} + (-0.39620 + 0.10401)\,10^{-3} \\
&+ (-0.26003)\,10^{-4} + (0.62050 - 0.14163)\,10^{-5} + (0.30982)\,10^{-6}
\end{aligned}
\tag{28}
$$

Efetuando os cálculos, obtemos: $e^{-5.25} = 0.65974 \times 10^{-2}$. Observe que, usando uma calculadora, o resultado de $e^{-5.25}$ é 0.52475×10^{-2}. Essa diferença entre os valores obtidos ocorreu porque, na expressão anterior, temos parcelas da ordem de 10^{2} que desprezam toda grandeza interior a 10^{-3}, enquanto que o resultado real de $e^{-5.25}$ é constituído quase que exclusivamente de grandezas desta ordem.

Exemplo

Example

Deseja-se determinar numericamente o valor exato da integral:

$$y_n = \int_0^1 \frac{x^n}{x+a} dx, \tag{29}$$

para um valor fixo de $a >> 1$ e, $n = 0, 1, \ldots, 10$.

Solução: Não fazer em sala de aula (olha o livro da Neide Franco). O objetivo deste exercício é mostrar como valores grandes no somatório podem resultar em soluções totalmente erradas.

Instabilidade numérica

Se um resultado intermediário de um cálculo é contaminado por um erro de arredondamento, este erro pode influenciar todos os resultados subsequentes que dependem deste resultado intermediário. Os erros de arredondamento podem propagar-se mesmo que todos os cálculo subsequentes sejam feitos com precisão dupla. Na realidade, cada novo resultado intermediário introduz um novo erro de arredondamento. É de se esperar, portanto, que todos esses erros influenciem o resultado final. Numa situação simples como o caso de uma soma, o erro final pode ser igual à soma de todos os erros intermediários.

Instabilidade numérica

Entretanto, os erros intermediários podem, algumas vezes, cancelar-se uns com os outros no mínimo parcialmente. Em outros casos (tal como em processos iterativos), os erros intermediários podem ter um efeito desprezível no resultado final. Algoritmos com essa propriedade são chamados estáveis.
A instabilidade numérica ocorre se os erros intermediários têm uma influência muito grande no resultado final. Veremos este fato através de exemplos.

Exemplos

Example

Resolver a integral:

$$I_n = e^{-1} \int_0^1 x^n e^x dx. \tag{30}$$

Solução: Construir uma fórmula recursiva a partir da integração por partes:

$$I_n = 1 - nI_{n-1}, \ \ n = 1, 2 \ldots. \tag{31}$$

O valor de I_0 é calculado diretamente: $I_0 = 0.6321$ (Usar apenas 4 casas decimais). Assim, temos:

$$\begin{aligned} &I_0 = 0.6321, \quad I_1 = 0.3679, I_2 = 0.2642, \quad I_3 = 0.2074, \\ &I_4 = 0.1704, \quad I_5 = 0.1480, I_6 = 0.1120, \quad I_7 = 0.216. \end{aligned} \tag{32}$$

Continuação - exemplo

Example

O resultado obtido para I_7 está claramente errado, desde que:

$$I_7 < e^{-1} \max_{0 \leq x \leq 1} e^x \int_0^1 x^n dx < \frac{1}{n+1}. \tag{33}$$

Além disso, a sequência I_n é uma sequência decrescente.

Discussão do exemplo anterior

Para ver que a instabilidade existe, vamos supor que o valor de I_0 esteja afetado de um erro ϵ_0. Vamos supor ainda que todas as operações aritméticas subsequentes são calculadas exatamente. Denotando por I_n o valor exato da integral, e por $\tilde{I}_n$ o valor calculado, assumindo que só existe erro no valor inicial, obtemos:

$$\tilde{I}_0 = I_0 + \epsilon_0 \tag{34}$$

e assim:

$$\tilde{I}_n = 1 - n\tilde{I_{n-1}}, \;\; n = 1, 2, \ldots \tag{35}$$

Seja r_n o erro, isto é:

$$r_n = \tilde{I}_n - I_n. \tag{36}$$

Após algumas manipulações algébricas chagamos que $r_n = (-n)^n \epsilon_0$.

Mal condicionamento

A maioria dos processos numéricos segue a seguinte linha geral:

- Dados são fornecidos.
- Os dados são processados de acordo com um plano pré-estabelecido (algoritmo)
- Resultados são produzidos.

Analisaremos a seguir problemas onde os resultados dependem continuamente dos dados. Este tipo de problema é chamado de problema bem posto. Um problema que não depende continuamente dos dados é chamado de problema mal posto.
Vamos então analisar como perturbações nos dados podem ou não influenciar os resultados.

Exemplos

Example

Resolver o sistema linear:

$$\begin{cases} x + y = 2 \\ x + 1.01y = 2.01 \end{cases} \tag{37}$$

Após isto, resolver novamente substituindo o valor 2.01 por 2.02.

Continuação - exemplo

O ponto de interseção é muito sensível a pequenas perturbações em cada uma dessas retas, desde que elas são praticamente paralelas. De fato, se o coeficiente de y na segunda equação é 1, as duas retas são exatamente paralelas e o sistema linear não tem solução. Isto é típico de problemas mal condicionados. Eles são também chamados de problemas críticos, pois ou possuem infinitas soluções ou não possuem nenhuma.

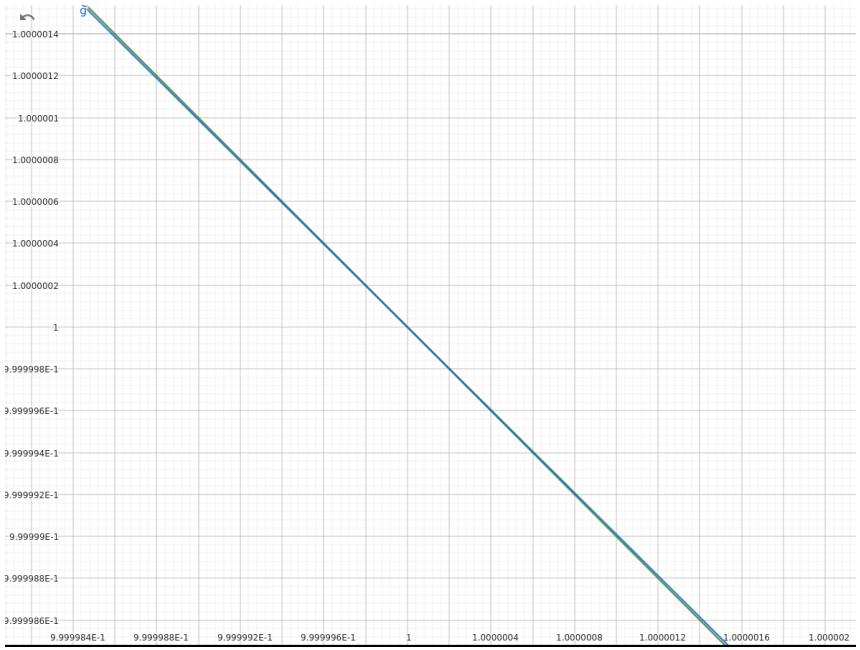

Figura 1: Gráfico das funções $y = 2 - x$ e $y = (2.01 - x)/1.01$.

Exemplo

Example

Determinar a solução do problema de valor inicial:

$$\begin{cases} y'' = y \\ y(0) = a \\ y'(0) = b \end{cases} \tag{38}$$

onde a e b são dados. Resolver o sistema considerando $a = 1$ e $b = -1$. Após isto, resolver novamente acrescentando uma perturbação δ no valor de b, isto é, $b = -1 + \delta$.

Problema mal condicionado

Torna-se então necessário introduzir uma medida para o grau de continuidade de um problema. Tal medida é essencial em muitas definições de continuidade.
Seja X o espaço dos dados, os elementos x e X podem ser números, pontos de um espaço euclidiano, vetores, matrizes, funções etc. Podemos então falar em continuidade se pudermos ser capazes de medir a distância entre os elementos de X. Suponhamos que o espaço X está dotado com uma função distância $d(x, y)$ que mede a distância entre os elementos x e y de X. Se, por exemplo, X é o espaço dos números reais, a função distância é definida por: $d(x, y) = |x - y|$.

Problema mal condicionado

Seja P o processo no qual os dados x são transformados no resultado y, isto é, $y = P(x)$. Se o processo P é contínuo num ponto x, então a definição de continuidade (matemática) exige que para cada $\epsilon > 0$, $\exists \delta(\epsilon) > 0$, tais que:

$$|P(\tilde{x}) - P(x)| < \epsilon \text{ sempre que } |\tilde{x} - x| < \delta(\epsilon) \tag{39}$$

Quanto maior a função $\delta(\epsilon)$ pode ser escolhida, mais contínuo é o processo P. No caso em que grandes mudanças nos dados produzem somente pequenas mudanças nos resultados, ou se $\delta(\epsilon)$ pode ser escolhida grande, a condição do problema é boa, e o problema é chamado bem condicionado. Por outro lado, se pequenas mudanças nos dados produzem grandes mudanças nos resultados, ou se $\delta(\epsilon)$ deve ser escolhida pequena, a condição do problema é má, e o problema é chamado mal condicionado.

Exemplos

Example

Analisar o problema de valor inicial do exemplo anterior:

$$\begin{cases} y'' = y \\ y(0) = a \\ y'(0) = b \end{cases} \tag{40}$$

onde a e b são dados.

Solução: Não fazer em sala de aula (olhar livro da Neide Franco). O objetivo deste problema é demonstrar o mal condicionamento a partir da relação de continuidade.

Número de condição

Podemos também verificar se um problema é ou não mal condicionado analisando o número de condição do problema. O problema será bem condicionado se o número de condição for pequeno, e será mal condicionado se o número de condiçõa for grande. Entretanto, a definição de número de condição depende do problema.
Seja $y = P(x)$, com P diferenciável. Então a mudança em y causada pela mudança em x pode ser aproximada (no sentido do cálculo diferencial) pelo diferencial de y, isto é: $dy = P'(x)dx$. Assim, o comprimento de $|P'(x)|$ do operador linear $P(x)$ representa o número de condição do problea num ponto x.
O número de condição relativa é definido por:

$$c_r = \frac{|P'(x)|}{|P(x)|}. \tag{41}$$

Assim, se $c_r \leq 1$ dizemos que o problema é relativamente bem condicionado.

Exemplo

Example

Analisar o problema de calcular:

$$f(x) = \left(\ln \frac{1}{x}\right)^{-\frac{1}{8}} \tag{42}$$

num ponto x qualquer. Determine o número de condição e número de condição relativa. Para quais valores de x o problema é bem condicionado?
Solução: Para $x = 0$ e $x = 1$, tanto o número de condição como o número de condição relativa são infinitos, e assim nestes pontos o problema é extremamente mal condicionado.
Para aproximadamente $0.1537 \leq x \leq 0.5360$, $c_r \leq 1$. Portanto, neste intervalo o problema de calcular f é bem condicionado.

Observações

O problema de resolver um sistema linear é um outro exemplo de problema onde pequenas perturbações nos dados podem alterar de modo significativo o resultado.
Teoricamente, o termo mal condicionado é usado somente para modelos matemáticos ou problemas, e o termo instabilidade somente para algoritmos. Entretanto, na prática, os dois termos são usados sem distinção.

Exercícios

1. Considere a integral

$$y_n = \int_0^1 \frac{x^n}{x+a} dx, \tag{43}$$

para um valor fixo de $a >> 1$ e, $n = 0, 1, \ldots, 10$. Tomemos $a = 10$.

   a) Calcule y_0 usando a integral.

   b) Mostre que uma relação de recorrência para y_n é dada por:

$$y_n = \frac{1}{n} - a y_{n-1}. \tag{44}$$

   c) Calcule y_n, $n = 1, 2, \ldots, 10$, usando a relação de recorrência acima. Os valores obtidos são confiáveis?

2. Considere agora a relação de recorrência do exercício anterior escrita na forma:

$$y_{n-1} = \frac{1}{a}\left(\frac{1}{n} - y_n\right). \tag{45}$$

Considere ainda que $y_{20} = 0$. Usando este dado e a relação de recorrência acima, obtenha os valores de $y_{10}, y_9, \ldots, y_1$. Os resultados agora são melhores? Como você explica isso?