Análise de Erros em Aritmética de Ponto Flutuante

Márcio Antônio de Andrade Bortoloti

Cálculo Numérico

Departamento de Ciências Exatas e Tecnológicas - DCET
Universidade Estadual do Sudoeste da Bahia

Sumário

Análise de Erros nas Operações de Ponto Flutuante

Erros nas Operações Aritméticas de Ponto Flutuante

Já vimos que o erro relativo no resultado de uma operação (supondo que as parcelas são representadas exatamente) seá

$$
\begin{aligned}
|ER_{op}| &< 10^{-t+1} \quad \text{(Para o caso de Truncamento)} \\
|ER_{op}| &< \frac{1}{2} \times 10^{-t+1} \quad \text{(Para o caso de Arredondamento)}
\end{aligned}
$$

Exemplo:

Considere uma máquina com $t = 4$ e dois números $x = 0.9765 \times 10^4$ e $y = 0.2456 \times 10^2$. Logo

$$
\begin{aligned}
x + y &= 0.9765 \times 10^4 + 0.2456 \times 10^2 \\
&= 0.9765 \times 10^4 + 0.002456 \times 10^4 = 0.978956 \times 10^4 \\
&= \begin{cases} 0.9789 \times 10^4 & \text{Se for usado truncamento} \\ 0.9790 \times 10^4 & \text{Se for usado arredondamento} \end{cases}
\end{aligned}
$$

Análise de Erros - Adição

Sejam x e y tais que

$$\begin{aligned} x &= \overline{x} + EA_x \\ y &= \overline{y} + EA_y \end{aligned}$$

Logo

$$\begin{aligned} x + y &= (\overline{x} + EA_x) + (\overline{y} + EA_y) \\ &= (\overline{x} + \overline{y}) + (EA_x + EA_y) \end{aligned}$$

Assim

$$\boxed{EA_{x+y} = EA_x + EA_y}.$$

Atenção: Aqui discutimos como o erro na representação de cada parcela da soma influencia o resultado final da operação.

Por outro lado, temos

$$
\begin{aligned}
ER_{x+y} &= \frac{EA_{x+y}}{\overline{x}+\overline{y}} \\
&= \frac{EA_x}{\overline{x}}\left(\frac{\overline{x}}{\overline{x}+\overline{y}}\right) + \frac{EA_y}{\overline{y}}\left(\frac{\overline{y}}{\overline{x}+\overline{y}}\right) \\
&= ER_x\left(\frac{\overline{x}}{\overline{x}+\overline{y}}\right) + ER_y\left(\frac{\overline{y}}{\overline{x}+\overline{y}}\right)
\end{aligned}
$$

Análise de Erros - Subtração

É fácil ver que

$$EA_{x-y} = EA_x - EA_y$$

$$\begin{aligned} ER_{x-y} &= \frac{EA_{x-y}}{\overline{x} - \overline{y}} \\ &= \frac{EA_x - EA_y}{\overline{x} - \overline{y}} \\ &= ER_x\left(\frac{\overline{x}}{\overline{x} - \overline{y}}\right) - ER_y\left(\frac{\overline{y}}{\overline{x} - \overline{y}}\right) \end{aligned}$$

Análise de Erros - Multiplicação

Notemos que

$$\begin{aligned} xy &= (\overline{x} + EA_x)(\overline{y} + EA_y) = \overline{x}\,\overline{y} + \overline{x}EA_y + \overline{y}EA_x + EA_xEA_y \\ &\approx \overline{x}\,\overline{y} + \overline{x}EA_y + \overline{y}EA_x \quad \text{pois} \quad EA_xEA_y \to 0 \end{aligned}$$

Logo

$$\boxed{EA_{xy} \approx \overline{x}EA_y + \overline{y}EA_x}$$

E o Erro Relativo ...

$$\begin{aligned} ER_{xy} &= \frac{EA_{xy}}{\overline{x}\,\overline{y}} \\ &\approx \frac{\overline{x}EA_y + \overline{y}EA_x}{\overline{x}\,\overline{y}} = \frac{EA_x}{\overline{x}} + \frac{EA_y}{\overline{y}} = ER_x + ER_y \end{aligned}$$

Logo

$$\boxed{ER_{xy} = ER_x + ER_y}$$

Análise de Erros - Divisão

$$
\begin{aligned}
\frac{x}{y} &= \frac{\overline{x} + EA_x}{\overline{y} + EA_y} \\
&= \frac{\overline{x} + EA_x}{\overline{y}(1 + \frac{EA_y}{\overline{y}})} \\
&= \frac{\overline{x} + EA_x}{\overline{y}}\left(\frac{1}{1 + \frac{EA_y}{\overline{y}}}\right)
\end{aligned}
$$

Agora note que

$$
\begin{aligned}
\frac{1}{1 + \frac{EA_y}{\overline{y}}} &= \sum_{n=0}^{\infty}(-1)^n\left(\frac{EA_x}{\overline{y}}\right)^n \quad \text{(Note que } EA_x \to 0\text{)} \\
&\approx 1 - \frac{EA_x}{\overline{y}}
\end{aligned}
$$

Análise de Erros - Divisão

Logo

$$
\begin{aligned}
\frac{x}{y} &\approx \frac{\overline{x} + EA_x}{\overline{y}}\Big(1 - \frac{EA_x}{\overline{y}}\Big) \\
&\approx \frac{\overline{x}}{\overline{y}} + \frac{EA_X}{\overline{y}} - \frac{\overline{x}EA_y}{\overline{y}^2} - \frac{EA_x\, EA_y}{\overline{y}^2} \\
&\approx \frac{\overline{x}}{\overline{y}} + \frac{EA_X}{\overline{y}} - \frac{\overline{x}EA_y}{\overline{y}^2} \quad (EA_xEA_y \to 0)
\end{aligned}
$$

Logo

$$EA_{x/y} \approx \frac{EA_X}{\overline{y}} - \frac{\overline{x}EA_y}{\overline{y}^2} = \frac{\overline{y}EA_x - \overline{x}EA_y}{\overline{y}^2}$$

Assim,

$$\boxed{EA_{x/y} \approx \frac{\overline{y}EA_x - \overline{x}EA_y}{\overline{y}^2}}$$

Análise de Erros - Divisão

E finalmente,

$$\begin{aligned} ER_{x/y} &= \frac{EA_{x/y}}{\overline{x}/\overline{y}} \\ &\approx \Big(\frac{\overline{y}EA_x - \overline{x}EA_y}{\overline{y}^2}\Big)\frac{\overline{y}}{\overline{x}} \\ &\approx \frac{EA_x}{\overline{x}} - \frac{EA_y}{\overline{y}} \\ &\approx ER_x - ER_y \end{aligned}$$

Logo

$$\boxed{ER_{x/y} \approx ER_x - ER_y}$$

Exemplo

Sejam x, y, z, t representados exatamente. Qual o erro total cometido no cálculo de

$$u = (x + y)z - t \quad ?$$

$$\begin{aligned} ER_{x+y} &= ER_x\left(\frac{\overline{x}}{\overline{x}+\overline{y}}\right) + ER_y\left(\frac{\overline{y}}{\overline{x}+\overline{y}}\right) + RA \\ &= 0 + 0 + RA \end{aligned}$$

Assim,

$$\begin{aligned} ER_{(x+y)z} &= ER_{x+y} + ER_z + RA \\ &= RA + 0 + RA \\ &= 2RA \end{aligned}$$

RA é o erro relativo cometido na aproximação após a operação.

Erros nas Operações Aritméticas de Ponto Flutuante

Para simplificar a notação, faremos $m = (x+y)z$.

$$\begin{aligned} ER_{m-t} &= ER_m\Big(\frac{\overline{m}}{\overline{m}-\overline{t}}\Big) - ER_t\Big(\frac{\overline{t}}{\overline{m}-\overline{t}}\Big) + RA \\ &= ER_m\Big(\frac{\overline{m}}{\overline{m}-\overline{t}}\Big) - 0 + RA \\ &= 2RA\Big(\frac{\overline{m}}{\overline{m}-\overline{t}}\Big) + RA \end{aligned}$$

Logo

$$\begin{aligned} |ER_{m-t}| &= \Big|2RA\Big(\frac{\overline{m}}{\overline{m}-\overline{t}}\Big) + RA\Big| \\ &\leq 2|RA|\Big|\frac{\overline{m}}{\overline{m}-\overline{t}}\Big| + |RA| \\ &\leq 10^{-t+1}\Big(\Big|\frac{\overline{m}}{\overline{m}-\overline{t}}\Big| + \frac{1}{2}\Big) \end{aligned}$$

Efeitos Numéricos

O erro relativo na subtração é dado por

$$ER_{x-y} = ER_x\Big(\frac{\overline{x}}{\overline{x}-\overline{y}}\Big) - ER_y\Big(\frac{\overline{y}}{\overline{x}-\overline{y}}\Big)$$

Exemplo

Considere $\overline{x} = 0.2357 \times 10^3$, $\overline{y} = 0.2353 \times 10^3$ e $\overline{r} = 0.4537 \times 10^3$. Vamos analisar $w = (x-y)r$ em uma máquina com quatro algarismos significativos. Temos $\overline{x} - \overline{y} = 0.0004 \times 10^3$. Note que

$$\begin{aligned}
|ER_{x-y}| &= \Big|ER_x\Big(\frac{\overline{x}}{\overline{x}-\overline{y}}\Big) - ER_y\Big(\frac{\overline{y}}{\overline{x}-\overline{y}}\Big)\Big| \\
&< \Big(\Big|\frac{\overline{x}}{\overline{x}-\overline{y}}\Big| + \Big|\frac{\overline{y}}{\overline{x}-\overline{y}}\Big|\Big) \times \frac{1}{2} \times 10^{-3} \quad (t=4) \\
&< \Big(\frac{0.2357 \times 10^3 + 0.2353 \times 10^3}{0.0004 \times 10^3}\Big) \times \frac{1}{2} \times 10^{-3} \\
&< 0.5888 \approx 59\%
\end{aligned}$$

Efeitos Numéricos - Cancelamento Subtrativo

Agora, note que

$$\begin{aligned} |ER_w| &= |ER_{x-y} + ER_r| \\ &\leq |ER_{x-y}| + |ER_r| \\ &\leq 0.59 + \frac{1}{2} \times 10^{-3} \\ &\leq 0.5905 \approx 59\% \end{aligned}$$

Efeitos Numéricos - Cancelamento Subtrativo

Exemplo

Calcular

$$\sqrt{9876} - \sqrt{9875}.$$

Temos que

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

Portanto

$$\begin{aligned} \sqrt{9876} - \sqrt{9875} &= 0.0000503142 \times 10^2 \\ &= 0.5031420000 \times 10^{-4} \end{aligned}$$

Podemos obter um resultado mais preciso? Neste caso a resposta é sim! Basta notar que

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

Logo

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

Instabilidade Numérica
Ocorre quando um resultado intermediário, contaminado com um erro, influencia todos os resultados subsequentes, mesmo que todos os cálculos subsequentes sejam feitos com exatidão.

Exemplo:

Resolver a integral

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

Integrando por partes,

$$\begin{aligned} I_n &= e^{-1} \left\{ \left[x^n e^x \right]_0^1 - \int_0^1 n x^{n-1} e^x \, dx \right\} \\ &= 1 - n e^{-1} \int_0^1 x^{n-1} e^x \, dx = 1 - n I_{n-1} \end{aligned}$$

Efeitos Numéricos - Instabilidade Numérica

Assim, obtemos uma fórmula de recorrência

$$\begin{cases} I_0 = e^{-1}\int_0^1 e^x\,dx = e^{-1}(e-1) = 0.6321 \\ I_n = 1 - nI_{n-1}, \quad n = 1, 2, \cdots \end{cases}$$

n	0	1	2	3	4	5	6	7
I_n	0.6321	0.3679	0.2642	0.2074	0.1704	0.1480	0.1120	0.2160

Por outro lado, observe que

$$\begin{aligned} I_n &= e^{-1}\int_0^1 x^n e^x\,dx < e^{-1}\max_{0\le x\le 1}\{e^x\}\int_0^1 x^n\,dx \\ &= e^{-1}\max_{0\le x\le 1}\{e^x\}\frac{1}{n+1} \\ &< \frac{1}{n+1} \end{aligned}$$

Assim, $I_7 < \dfrac{1}{8} = 0.1250$.

Analisando a instabilidade ...

Considere

$$\tilde{I}_0 = I_0 + \epsilon_0,$$

onde ϵ_0 representa um erro. Considere todas as operações seguintes realizadas de modo exato.

Assim

$$\tilde{I}_n = 1 - n\tilde{I}_{n-1}, \quad n = 1, 2, \cdots$$

Vamos definir o erro no passo n por $r_n = \tilde{I}_n - I_n$.

Assim

$$r_n = -nr_{n-1}, \quad n = 1, 2, \cdots$$

Assim,

$$r_n = (-1)^n n!\, \epsilon_0$$

Analisando a instabilidade ...

Exercício
Encontre uma forma mais precisa de calcular I_7, no exemplo anterior.

- Observa-se que o erro cresce na direção de n mas decresce na direção oposta.
- Nota-se que
$$I_{n-1} = \frac{1 - I_n}{n}$$
- Para calcularmos I_7 temos que ter um valor inicial para a a sequência I_n, o que não é fácil encontrar.
- Mas, neste caso, nota-se que $I_n \to 0$.
- Assim, se tomarmos $I_{20} = 0$ e construir a sequência para $n = 20, 19, 18, \cdots$, vamos obter $I_7 = 0.1123835$ com todos os dígitos corretos.

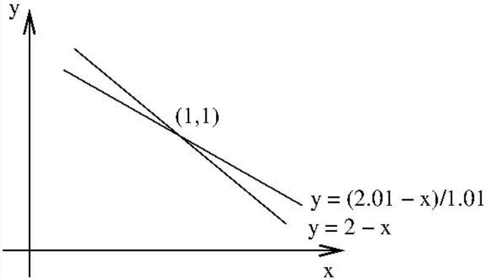
Efeitos Numéricos - Mal - Condicionamento

Considere o sistema

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

cuja solução é $(1, 1)$.

O que aconteceria se alterarmos o número 2.01 para 2.02 ? A solução agora seria $(0, 2)$.



Outro Exemplo

Considere o seguinte problema de valor inicial

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

onde a, b são dados. A solução é dada por

$$y(x) = C_1 e^x + C_2 e^{-x}$$

Se $a = 1$ e $b = -1$ então, das condições iniciais

$$\begin{cases} y(0) = C_1 + C_2 = 1 \\ y'(0) = C_1 - C_2 = -1 \end{cases}$$

cuja solução é $C_1 = 0$ e $C_2 = 1$. De onde obtem-se $y(x) = e^{-x}$.

Efeitos Numéricos - Mal - Condicionamento

Mas se fizermos $a = 1$ e $b = -1 + \delta$ com $|\delta|$ pequeno teremos

$$\begin{cases} y(0) = C_1 + C_2 = 1 \\ y'(0) = C_1 - C_2 = -1 + \delta \end{cases}$$

cuja solução é $C_1 = \delta/2$ e $C_2 = 1 - \delta/2$. Assim a solução desse novo problema é

$$y(x) = e^{-x} + \delta \text{senh} x.$$