

FEV
3-5
2026

I WORKSHOP

MINEIRO

DE

MATEMÁTICA

APLICADA



UFMG - BELO HORIZONTE - MG

Explorando alguns efeitos dos erros de Ponto Flutuante

Ribas, E. R. L. D.^{1,*}, Pazini, D. S.², D'Afonseca, L. A.³, Rocha, L. M.⁴

Departamento de Matemática, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)

*Contato: enzorochaleitedinizribas@gmail.com

INTRODUÇÃO

A aritmética de ponto flutuante é o sistema adotado por computadores para que lidem com números reais utilizando uma notação compacta e eficaz. Essa técnica é utilizada para representar e manipular números reais de forma prática e eficiente. Ela permite representar números de grandezas diversas, que não podem ser armazenados com precisão, utilizando apenas números inteiros. A aritmética de ponto flutuante é amplamente utilizada em diversas áreas, como computação científica, gráficos de computador, simulações numéricas e processamento de sinais. No entanto, é importante compreender suas limitações e os possíveis erros que podem ocorrer durante as operações aritméticas, a fim de garantir resultados precisos e confiáveis em cálculos numéricos.

Sistema de Ponto Flutuante

Um sistema de ponto flutuante F pode ser definido como

$$F(\beta, t, L, U)$$

cuja representação normalizada de um número real N nesse sistema é dada por

$$N = \pm(d_1.d_2...d_t)_\beta \times \beta^e \quad (1)$$

em que

- N é o número real;
- β é a base que a máquina opera;
- t é o número de dígitos na mantissa, tal que $0 \leq d_j \leq \beta - 1$, $j = 1, \dots, t$, $d_1 \neq 0$;
- L é o menor expoente inteiro;
- U é o maior expoente inteiro;
- e é o expoente inteiro no intervalo $[L, U]$.

$$X(\xi_N) = \sum_{i=1}^N \sum_{j=1}^N (x_i - x_j)^2 + \int_{\theta} f(\theta) d\theta.$$

METODOLOGIA

Para investigar os efeitos dos erros de ponto flutuante, foram realizados experimentos computacionais utilizando diferentes configurações de precisão numérica. Utilizando a linguagem de programação Python junto a bibliotecas NumPy e Scipy para melhor confiabilidade das operações e outras bibliotecas gráficas como Matplotlib, seaborn, plotly, foram implementados algoritmos que simulam operações aritméticas em ponto flutuante com diferentes níveis de precisão (simulando simples, dupla e quadrupla precisão). Foram analisados os resultados obtidos em termos de erros numéricos, estabilidade e desempenho computacional.

RESULTADOS

Um erro comum neste sistema é o da perda de significância, ou cancelamento catastrófico, que ocorre quando a subtração de dois números resulta em um valor com menos dígitos significativos do que os números originais. Um exemplo desse comportamento é a função $f(x) = x^{10} + 1 - x^{10}$. Embora para $x \in \mathbb{R}$ o resultado é igual a 1, ao efetuarmos os cálculos usando ponto flutuante, observamos o comportamento ilustrado no gráfico da Figura 1, em que o valor correto é exibido apenas até um certo valor de x . Após esse valor observamos um intervalo em que ocorre uma oscilação caótica no resultado da função. Posteriormente, observamos que a função assume o valor zero.

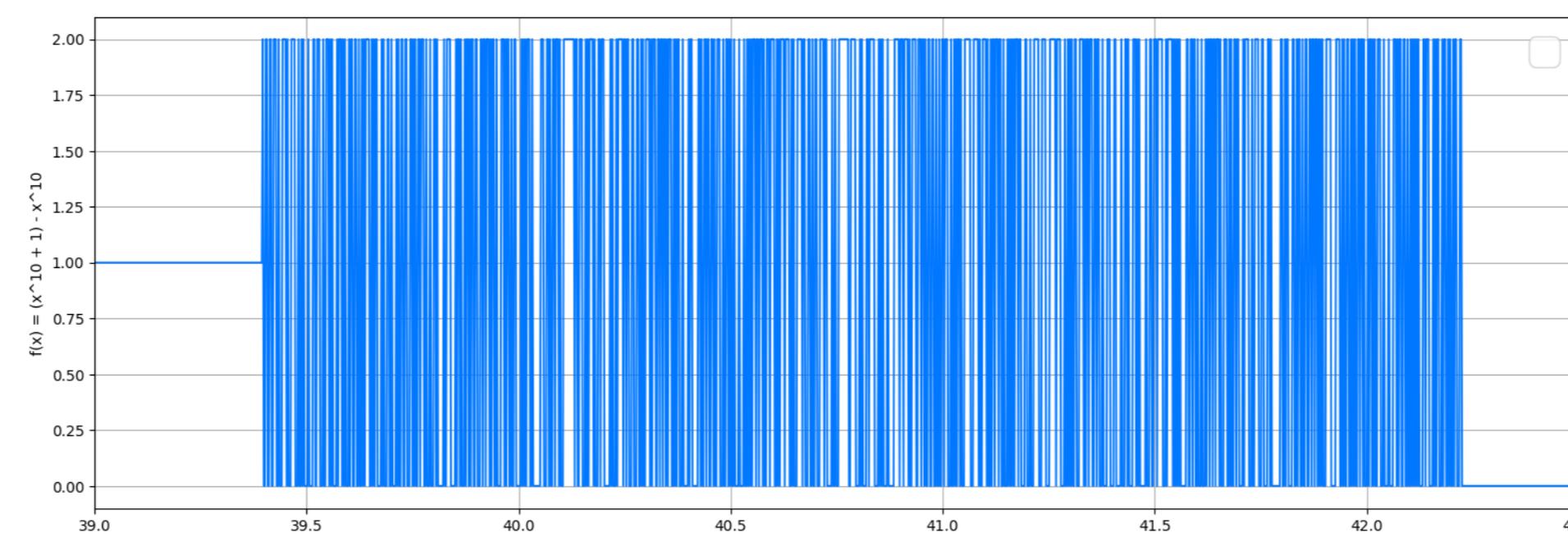


Figura 1: Perda de significância.

Um outro erro é o não cancelamento adequado de termos em expressões matematicamente equivalentes. A Figura 2 apresenta a comparação entre os resultados obtidos ao calcular $p(x) = (x - 1)^6$ e sua forma expandida $q(x) = x^6 - 6x^5 + 15x^4 - 20x^3 + 15x^2 - 6x + 1$. Nesta figura, o gráfico de p exibe os resultados obtidos pelo cálculo da expressão fatorada utilizando ponto flutuante obtendo a curva esperada. Entretanto, o gráfico da expressão expandida q , calculado no mesmo sistema de ponto flutuante, produz resultados caóticos. Note que, apesar de possivelmente surpreendente, esse fenômeno não invalida a utilidade do ponto flutuante, pois o erro observado é da ordem de 10^{-14} .

Um erro comum é o da Derivada Numérica.

Utilizando Diferenças finitas para aproximar a derivada de uma função, ao diminuir o valor de h , espera-se que a aproximação melhore. Mas devido aos erros de arredondamento em ponto flutuante, para h muito pequeno, o erro na aproximação começa a aumentar. A Figura 3 ilustra esse comportamento, mostrando o erro absoluto na aproximação da derivada da função f .

$$f(x) = \frac{x^3}{3} - 3x + 3 \quad f'(x) = x^2 - 3$$

$$D_f(x, h) = \frac{f(x + h) - f(x - h)}{2h}$$

O erro é calculado como

$$\text{Erro}(h) = |f'(x) - D_f(x, h)|$$

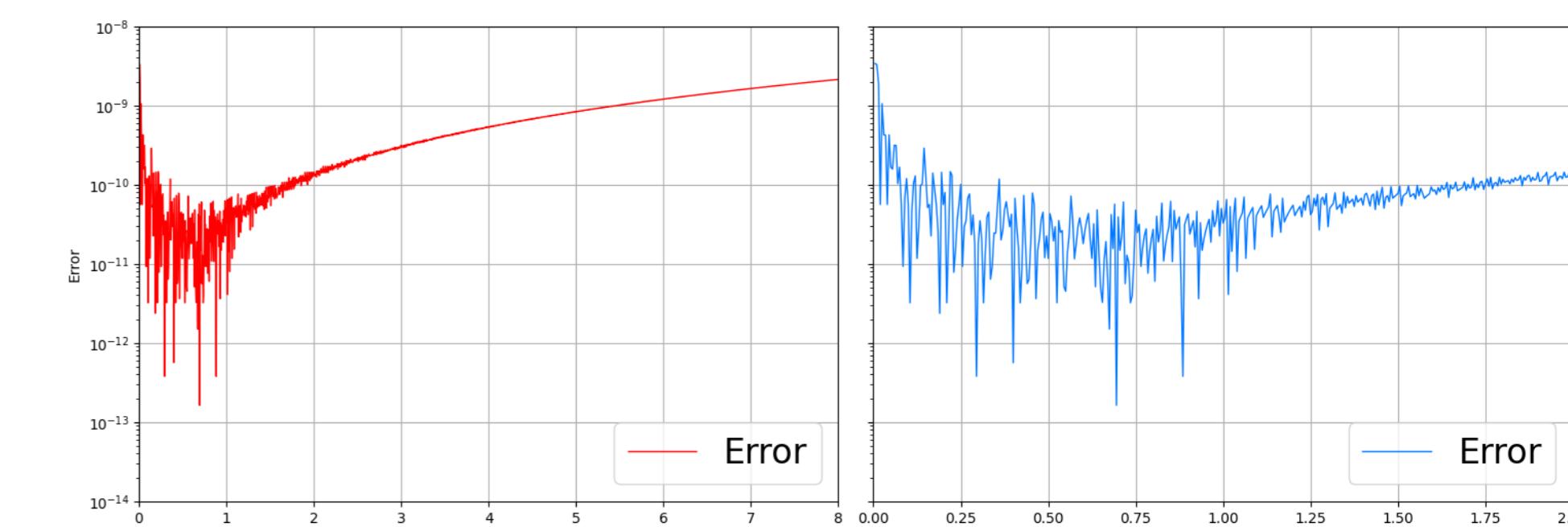


Figura 3: Outra legenda ilustrativa descrevendo uma variação em duas condições genéricas.

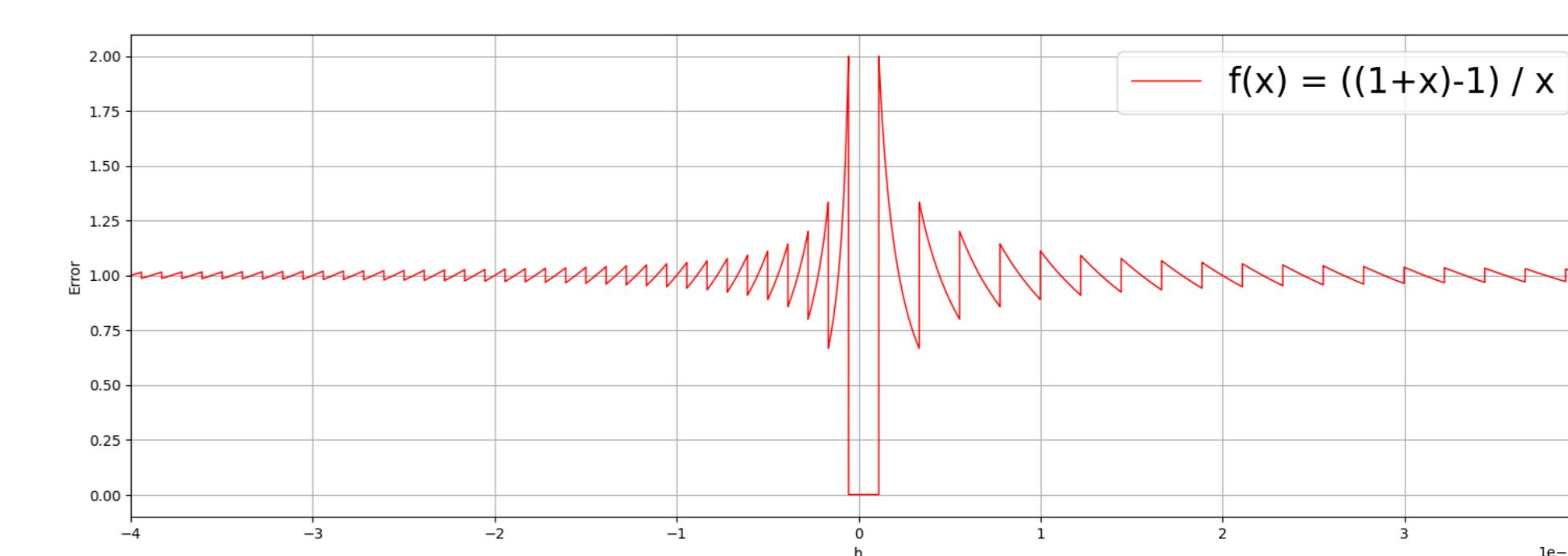


Figura 4: Outra legenda ilustrativa descrevendo uma variação em duas condições genéricas.

CONCLUSÃO

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum aliquam augue et convallis accumsan. Sed vitae tortor nec sapien volutpat venenatis. Integer congue nisi sed magna aliquet, vel faucibus est malesuada.

REFERÊNCIAS

- [1] Chapra, S. C., Canale, R. P. Numerical Methods for Engineers. McGraw-Hill International Editions, 1985.
- [2] IEEE Standard for Floating-Point Arithmetic. IEEE Std 754-2008, 2008.
- [3] IEE Standard for Floating-Point Arithmetic. IEEE Std 754-2019, 2019.

AGRADECIMENTOS

Agradecemos ao CEFET-MG pelo apoio institucional disponibilizada para a realização deste trabalho. Agradecemos aos orientadores e colegas do grupo de pesquisa em Métodos Numéricos por suas contribuições e discussões enriquecedoras.