

- graphicx
- amsmath amssymb
- hyperref
- babel

consequentemente, não há necessidade de incluí-los no preâmbulo.

**Quem tentar compilar usando a opção pdflatex no Overleaf vai receber uma mensagem de erro solicitando o usuário a ajustar a opção de compilação.**

Para a pergunta *Por que não funciona com o pdflatex?* A resposta é: **fontes!** Pdflatex usa um esquema de codificação de fontes complexo. O fonte academicons<sup>1</sup> não tem as definições necessárias para uso com o pdflatex. Enquanto isso não for realizado, pdflatex não pode ser usado.

## 2 Revisão Bibliográfica

### 2.1 Aritmética de Ponto Flutuante

A norma IEEE 754 é um padrão amplamente adotado para a representação e manipulação de números em ponto flutuante em sistemas de hardware ou software. Ela define formatos de representação, regras de arredondamento, operações aritméticas e tratamento de exceções para garantir consistência e precisão nos cálculos numéricos. A norma especifica diferentes formatos de ponto flutuante, como simples precisão (32 bits) e dupla precisão (64 bits), cada um com sua própria estrutura de bits para representar o sinal, o expoente e a mantissa do número. Além disso, a norma estabelece regras para operações aritméticas, como adição, subtração, multiplicação e divisão, incluindo o tratamento de casos especiais, como overflow, underflow, NaN (Not a Number) e infinitos (IEEE Computer Society [2019]). A norma IEEE 754 é fundamental para garantir a interoperabilidade entre diferentes sistemas computacionais e linguagens de programação, permitindo que cálculos numéricos sejam realizados de forma consistente e previsível.

A representação de um número real em aritmética de ponto flutuante conforme definido pela norma IEEE 754 consistem tem 3 diferentes formatos, no primeiro formato a representação é composta por três componentes principais: o sinal, o expoente e a mantissa (ou significando). O expoente da base utilizada. O bit de sinal indica se o número é positivo ou negativo. A mantissa representa a parte fracionária do número e é normalizada para garantir que o primeiro dígito seja sempre diferente de zero. A combinação desses três componentes permite representar uma ampla gama de números reais, incluindo números muito grandes e muito pequenos, com uma precisão limitada devido ao número finito de bits disponíveis para cada componente. O segundo formato é a representação de infinitos, e por último o terceiro formato é a representação de NaN's, que é usada para indicar resultados indefinidos ou inválidos em operações aritméticas.

- Valor =  $(-1)^{\text{sinal}} \times \text{mantissa} \times \text{base}^{\text{expoente}}$
- $\infty$  e  $-\infty$
- qNaN (quiet NaN) e sNaN (signaling NaN)

qNaN e sNaN são dois tipos de Not-a-Number (NaN)

definidos pela norma IEEE 754, usados para representar resultados indefinidos em aritmética de ponto flutuante.

Um conjunto de finitos pontos flutuantes representáveis a partir de um formato de ponto flutuante é denominado **conjunto de números em ponto flutuante**, e é definido pelos seguintes parâmetros:

- $b = \text{Base}$ : A base do sistema numérico utilizado, 2 ou 10 (geralmente 2 em computadores).
- $p = \text{Precisão}$ : O número de dígitos na mantissa (ou significando) que determina a precisão dos números representados.
- $emax = \text{Expoente máximo}$ : O maior valor do expoente que pode ser representado.
- $emin = (-emax + 1) = \text{Expoente mínimo}$ : O menor valor do expoente que pode ser representado.

A norma IEEE 754 também define regras para arredondamento, que determinam como os números são aproximados quando não podem ser representados exatamente no formato de ponto flutuante. As principais estratégias de arredondamento são:

- roundToIntegralTiesToEven (Arredondar para o inteiro mais próximo, com empates arredondados para o número par mais próximo)
- roundToIntegralTowardsZero (Arredondar para o inteiro mais próximo em direção a zero)
- roundToIntegralTowardPositive (Arredondar para o inteiro mais próximo em direção ao infinito positivo)
- roundToIntegralTowardNegative (Arredondar para o inteiro mais próximo em direção ao infinito negativo)
- roundToIntegralTiesToAway (Arredondar para o inteiro mais próximo, com empates arredondados para longe de zero)
- roundToIntegralExact (Arredondar exatamente)

#### 2.1.1 Units Of Last Place (ULP) e Quantum

Unit Of Last Place (ULP) ou Unidade do Último Lugar é uma medida usada em aritmética de ponto flutuante para quantificar a precisão relativa de um número representado. A ULP representa a menor diferença entre dois números consecutivos que podem ser representados no formato de ponto flutuante. Em outras palavras, a ULP indica o "tamanho do passo" entre números representáveis próximos a um determinado valor como definido por John Harrison em Harrison [1999].

Na prática podemos entender a ULP, segundo Muller [2016] como: Se  $x$  é exatamente representável em um formato de ponto flutuante e não é uma potência inteira da base  $\beta$ , o termo  $ulp(x)$  (unidade do último lugar) denota a magnitude do último dígito do significando de  $x$ . Ou seja, se

$$x = \pm d_0.d_1d_2 \dots d_{p-1} \times \beta^{e_x},$$

então a unidade do último lugar de  $x$  é dada por

$$ulp(x) = \beta^{e_x - p + 1}.$$

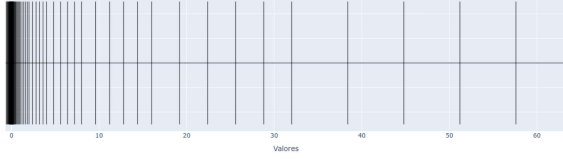
onde  $p$  é a precisão (número de dígitos no significando) e  $e_x$  é o expoente de  $x$ .

A figura 1 ilustra a densidade dos números representáveis em ponto flutuante ao longo da reta numérica. Note que a

<sup>1</sup>Do fonte em questão usa-se apenas o glifo associado ao Orchid.

densidade é maior próximo de zero e diminui à medida que nos afastamos de zero, refletindo a natureza relativa da precisão em aritmética de ponto flutuante.

Escala de Ponto Flutuante: Densidade Próxima a Zero



**Figura 1.** Densidade dos números em ponto flutuante ao longo da reta numérica.

Quando um número real  $x$  não tem representação exata em um formato de ponto flutuante, a ULP pode ser usada para medir o erro de arredondamento ao aproximar  $x$  pelo número representável mais próximo. O erro de arredondamento é frequentemente expresso em termos de ULPs, indicando quantas unidades do último lugar o valor aproximado difere do valor real.

Demonstração do Arredondamento entre Pontos de Ponto Flutuante



**Figura 2.** Erro de arredondamento de Pontos Flutuantes.

O conceito de Quantum em aritmética de ponto flutuante é semelhante ao de ULP, mas com uma definição ligeiramente diferente. O Quantum é definido pela (IEEE Computer Society [2008]) como: O Quantum de uma representação finita em ponto flutuante é o valor de uma unidade na última posição do seu significando. Isso é igual à base (radix) elevada ao expoente  $q$ , que é usado quando o significando é considerado como um inteiro.

Tanto o *Quantum* quanto a *ULP* medem a diferença entre dois números consecutivos de ponto flutuante, mas usam critérios ligeiramente diferentes nas fronteiras dos blocos de expoente.

Seja a aritmética binária (base  $b = 2$ ) com precisão  $p$  (número de bits do significando, incluindo o bit implícito). Para números com expoente  $e$  (isto é,  $x \in [2^e, 2^{e+1})$ ) o Quantum é constante e dado por

$$\text{Quantum}(e) = 2^{e-p+1},$$

ou seja, o espaçamento entre representáveis dentro desse mesmo intervalo de expoente.

A definição de ULP usada por John Harrison (e por várias definições formais de precisão) é a distância entre os dois representáveis adjacentes que envolvem um valor real  $x$ . Em particular:

- se  $x$  não é exatamente uma potência de dois, então a ULP de  $x$  coincide com o Quantum do seu intervalo de expoente;

- se  $x = 2^e$  (fronteira entre blocos de expoente), o Quantum observa o bloco acima ( $[2^e, 2^{e+1})$ ) enquanto a ULP de Harrison usa o espaçamento imediatamente abaixo (do bloco  $[2^{e-1}, 2^e)$ ). Assim

$$\text{ULP}_{\text{Harrison}}(2^e) = 2^{e-p} = \frac{1}{2} \text{Quantum}(e).$$

Para análise de erro de arredondamento, é habitual usar ULP: o erro de arredondamento em uma operação corretamente arredondada é no máximo 0.5 ULP.

### 2.1.2 Exemplo numérico

Para  $e = 2$ ,  $p = 4$  e  $b = 2$  temos:

$$\text{Quantum}(e) = 2^{e-p+1} = 2^{2-4+1} = 2^{-1} = \frac{1}{2} = 0.5,$$

$$\text{ULP}_{\text{Harrison}}(2^e) = 2^{e-p} = 2^{2-4} = 2^{-2} = \frac{1}{4} = 0.25 = \frac{1}{2} \text{Quantum}(e).$$

Para  $x_{\text{real}} = 4.3$  e  $x_{\text{float}} = 4.5$ , temos:

$$\text{erro absoluto} = |4.3 - 4.5| = 0.2,$$

$$\frac{\text{erro}}{\text{Quantum}} = \frac{0.2}{0.5} = 0.4 \text{ Quantum},$$

$$\frac{\text{erro}}{\text{ULP}_H} = \frac{0.2}{0.25} = 0.8 \text{ ULP}_H.$$

Erro de Arredondamento: Quantum vs ULP de Harrison



**Figura 3.** Comparação entre Quantum e ULP na representação de números em ponto flutuante.

## 3 Exemplo de Título Nível 1 (Seção)

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

### 3.1 Exemplo de Título Nível 2 (Subseção)

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.