### Ilum Escola de Ciência

# Notas de Matemática: Análise Numérica

Prof. Dr. Vinícius Francisco Wasques

16 de dezembro de 2023, Campinas

# Sumário

1	Sist	rema de números e estudo de erro	2
	1.1	Representação de números	2
	1.2	Aritmética computacional	4
	1.3	Tipos de erros	5
2	Solu	uções aproximadas de equações	7
	2.1	Método da bissecção	7
	2.2	Outros métodos para se encontrar zero de funções	9
		2.2.1 Método de Newton	9
		2.2.2 Método da secante	10
	2.3	Problema do ponto fixo	11
3	Solu	uções Numéricas para Equações Diferenciais Ordinárias	14
	3.1	Método de Euler	14
	3.2	Método de Runge-Kutta	16
4	Inte	gração Numérica	19
	4.1	Método do trapézio	19
	4.2	Método de Simpson	21
5	Solu	uções numéricas para Equações Diferenciais Parciais	26
	5.1	Método das diferenças finitas	26
		5.1.1 Diferenças finitas para Problema de Valor de Contorno	28

# Introdução

Este texto traz um breve resumo do conteúdo de **Análise Numérica** estudado pela turma 2022 da Ilum Escola de Ciência. Os tópicos abordados aqui corresponderão a uma extensão do conteúdo visto nas disciplinas de Equações Diferenciais e Álgebra Linear Computacional. O foco é explorar alguns recursos computacionais para estudar problemas envolvendo equações diferenciais do ponto de vista numérico.

Inicialmente esse material servirá como guia e também um resumo do conteúdo já visto pelos alunos. Os exemplos e exercícios propostos aqui são em sua maioria os mesmos discutidos em sala de aula.

O texto ficará disponível na plataforma Moodle, na aba Material da disciplina **Análise Numérica**. Este material será atualizado à medida que o curso for avançando.

Bibliografias principais: Conte (1965); Burden and Faires (2003); Ruggiero and Lopes (2000).

## Capítulo 1

## Sistema de números e estudo de erro

A motivação para estudar o conteúdo dessa disciplina é baseada na solução prática obtida por meio de simulações numéricas. Consequentemente, se faz necessário o uso de recursos computacionais, e com isso algumas questões surgem. Na formulação de modelos matemáticos para fenômenos físicos, químicos ou biológicos, se parte do pressuposto que para fornecer uma solução para o problema será necessário resolvê-lo no computador. Uma vez que o modelo tenha sido formulado, é necessário realizar uma *análise numérica* da solução, como exemplo um estudo de erro que pode influenciar no comportamento da solução, ou ainda na acurácia e convergência da mesma. Para isso, é preciso verificar o tamanho dos passos adequados para cada método, o número de iterações necessários, etc.

O curso começará com uma discussão sobre representação de números, aritmética computacional e os erros de arredondamento/truncamento.

### 1.1 Representação de números

A representação de números, em geral, é feita pelo *sistema decimal*. Isto é, o número 2023 pode ser escrito da seguinte forma

$$2023 = 2.1000 + 0.100 + 2.10 + 3.1$$
$$= 2.10^{3} + 0.10^{2} + 2.10 + 3.10^{0},$$

em que 10 é chamada a *base do sistema* e os números 2, 0, 2 e 3 que acompanham essa base são os *coeficientes associados* à base 10 podendo assumir valores entre 0 e 9.

Não há uma razão intrínseca para se considerar a base 10. Outras civilizações utilizaram outras bases, como por exemplo a base 12, 20 ou 60. No entanto, é conveniente considerar a base 2, e consequentemente coeficientes entre 0 e 1, para representação de números no computador Conte (1965). Qualquer número inteiro pode ser escrito tanto na base 10 quanto na base 2 na forma polinomial, como apresentado acima. Para realizar essa conversão, pode-se considerar por exemplo os algorithms 2 Conte (1965).

#### Algorithm 1: Conversão de um número N na base decimal para a binária

Input: Número 
$$N$$
 na base  $10$ ; 
$$b_0=N$$
 
$$b_1=\frac{b_0-a_0}{2};$$
 
$$b_2=\frac{b_1-a_1}{2};$$
 
$$\vdots$$
 
$$b_k=\frac{b_k-a_k}{2};$$
 where 
$$\begin{cases} a_k=1 & \text{se } b_k \text{ \'e impar}\\ a_k=0 & \text{se } b_k \text{ \'e par} \end{cases}$$
 Pare quando  $b_k=0$ ; Output:

**Exemplo 1.1.1** Considere o número 35 na base decimal. Assim, temos que:

 $N=(a_ka_{k-1}a_{k-2}\cdots a_0)_2$  é o número procurado na base 2

k	$b_k$	$a_k$
0	35	1
1	17	1
2	8	0

Tabela 1.1: Decomposição do número 35 da base decimal para a binária

Logo, o número 35 é escrito na base binária como  $35=(011)_2$ . Note que a representação é por  $(a_2a_1a_0)_2$ , e não  $(a_0a_1a_2)_2$ .

#### Algorithm 2: Conversão de um número N na base binária para a decimal

```
Input: Coeficientes do número N na base 2, isto é, N=(a_na_{n-1}\cdots a_1a_0)_2; b_n=a_n for k=n-1,...,0 do | b_k=a_k+2b_{k+1} end Output: N=b_0 é o número procurado na base 10
```

**Exemplo 1.1.2** Considere o número  $(1101)_2$  na base binária. Assim, temos que:

$$\begin{cases} b_3 = 1 \\ b_2 = 1 + 2.1 = 3 \\ b_1 = 0 + 2.3 = 6 \\ b_0 = 1 + 2.6 = 13 \end{cases}$$

Logo, o número  $(1101)_2$  é escrito na base decimal como 13.

Apesar de poder escrever qualquer número inteiro nessas bases, não temos necessariamente que todos os números podem ser escritos dessa forma. Por exemplo, o número racional  $\frac{1}{10}$  não pode ser escrito na base binária em uma forma **finita** como as apresentadas acima. Sua representação binária seria

 $\frac{1}{10} = 00001100110011 \cdots$  Como o foco da nossa discussão será sobre soluções obtidas por meio de computadores, fica claro que tais representações envolverão um erro de aproximação.

#### 1.2 Aritmética computacional

Conseguimos efetuar sem grandes problemas as operações 2+2 e  $(\sqrt{2})^2$ . No primeiro caso o resultado é 4, enquanto que o segundo é 2. Fazendo essa mesma conta no computador, temos que a primeira operação também resultará em 4. Por outro lado, não necessariamente teremos que o resultado final será 2, assim como esperado. Isso ocorre porque o número  $\sqrt{2}$ , que é irracional, possui infinitas casas decimais, significando que para representar esse valor numérico no computador é preciso aproximá-lo, e portanto, comete-se um erro nesse cálculo.

Aqui não vamos focar na representação e armazenamento desses números no computador. Nesse sentido, recomendamos a seguinte referência Ruggiero and Lopes (2000). A ideia é dedicar nossa discussão no conceito do erro, que é chamado de *erro de arredondamento*.

Por muitas vezes esses "erros" passam despercebido, mas em outras eles são propagados, acarretando em alguns comportamentos não esperados na sua solução numérica dos quais falaremos mais adiante. Para "limitar" esses números com infinitas casas decimais, podemos utilizar dois tipos de técnicas, o *truncamento* e o *arredondamento*.

Para exemplificar cada caso, considere o número n=0,45317 e suponha que o computador tenha uma limitação para trabalhar com números com apenas 4 dígitos após a vírgula. O método de truncamento vai simplesmente desconsiderar o último dígito de n, ficando 0,4531. Por outro lado, o método de arredondamento segue o seguinte critério: se o último dígito é maior ou igual a 5 então soma-se 1 ao dígito anterior; se o último dígito é menor que 5, então desconsidera-se o último dígito (assim como é feito no truncamento). O primeiro caso do arredondamento é chamado de arredondamento para cima e o segundo é chamado de arredondamento para baixo. Para o número n deve-se então arredondar para cima, já que o último dígito 7 > 5. Logo, o arredondamento fica 0,4532.

Como foi possível observar acima, dependendo do critério e do número envolvido, pode haver mudanças (mesmo que pequenas) na representação desse número. O erro cometido nessa substituição é chamado de *erro de arredondamento* (independente se o critério foi de arredondamento ou truncamento). Vejamos como medir esses erros de arredondamentos.

Seja n um número e  $n^*$  sua aproximação. O *erro absoluto* é calculado por

$$e_a(n) = |n - n^*|.$$

Já o erro relativo é cálculo por

$$e_r(n) = \frac{|n - n^*|}{n},$$

considerando que  $n \neq 0$ .

A diferença entre esses dois tipos de medidas para o erro é que o erro relativo leva em consideração a magnitude dos valores. Nesse sentido, o erro relativo pode ser mais significativo do que o absoluto. Por exemplo, considere n=0,1 e sua aproximação  $n^*=0,11$ . Assim, temos que  $e_a=0,01$  e  $e_r=0,1$ . Tomando agora, n=0,01 e sua aproximação  $n^*=0,011$ , temos que  $e_a(n)=0,001$  e  $e_r(n)=0,1$ . Por fim, tomando n=0,001 e sua aproximação  $n^*=0,0011$ , temos que  $e_a(n)=0,0001$  e  $e_r(n)=0,1$ . Perceba que o erro relativo foi preservado quando apenas foram adicionadas casas decimais, enquanto que o erro

absoluto não. O exemplo a seguir ilustra a comparação dos erros relativos e absolutos cometidos ao efetuar operações aritméticas em comparação com o valor que deveria ser obtido.

**Exemplo 1.2.1** Considere  $x = \frac{1}{3}$  e  $y = \frac{4}{7}$ . Na Tabela 1.2 é possível ver os valores das operações computacionais juntamente com os erros associados.

Operação	Valor real	Valor computacional	$e_a$	$e_r$
x+y	$\frac{19}{21}$	0,9047	$0,6190 \times 10^{-4}$	$0,6842 \times 10^{-4}$
x-y	$-\frac{5}{21}$	-0,2381	$0,4761 \times 10^{-5}$	$0,2 \times 10^{-4}$
$x \times y$	$\frac{4}{21}$	0,1905	$0,2857 \times 10^{-4}$	$0,15 \times 10^{-3}$
$x \div y$	$\frac{7}{12}$	0,5833	$0,2916 \times 10^{-4}$	$0.5 \times 10^{-4}$

Tabela 1.2: Tabela com os valores das operações aritméticas feitas pelo computador em relação ao valor esperado

Com isso é possível notar que ao efetuar operações aritméticas no computador se pode cometer erros. Ou seja, ao calcular por exemplo uma soma "na mão" o resultado não será necessariamente o mesmo do que se calcular no computador. Para diferenciar as operações, vamos denotar as operações aritméticas usuais por  $+, -, \times, \div$  e a soma feita pelo computador por  $\oplus, \ominus, \otimes, \oslash$ . Ainda mais, como cada número possui sua representação computacional, vamos denotar a representação de x e y por fl(x) e fl(y), respectivamente fl(x) 1. Isto é, fl(x) = 0.3333 e fl(y) = 0.5714, e assim, como poder ser visto na Tabela 1.2, fl(x) = 1.20, fl(x) = 1.21, fl(x) = 1.21, fl(x) = 1.22, fl(x) = 1.22, fl(x) = 1.22, fl(x) = 1.22, fl(x) = 1.2

**Exercício 1.2.1** Observe a Tabela 1.2. Quais conclusões você poderia obter a partir das informações apresentadas? Ou seja, faça uma análise sobre os erros relativos e absolutos em cada caso.

**Exercício 1.2.2** Considere dois números x e y que sejam "quase" iguais. Por exemplo, x=0,235648431 e y=0,235648432 o que você acha que o computador forneceria como solução ao calcular a subtração entre os dois números? Quais implicações isso teria?

**Exercício 1.2.3** *Considere o polinômio*  $f(x) = x^3 - 6$ ,  $1x^2 + 3$ , 2x + 1, 5.

- 1. Calcule f(4,71) utilizando arredondamento e truncamento. Compare com o valor esperado por meio do erro absoluto e relativo.
- 2. É possível escrever o mesmo polinômio f(x) da seguinte forma f(x) = ((x-6,1)x+3,2)x+1,5. Utilizando essa nova decomposição, calcule f(4,71) utilizando arredondamento e truncamento e obtenha o erro relativo e absoluto.
- 3. O que você observou nos itens 1. e 2.? O que pode ser afirmado?

#### 1.3 Tipos de erros

Como foi possível observar, erros computacionais são praticamente inevitáveis. Sendo assim, em todos os métodos que iremos trabalhar será necessário realizar um estudo sobre o erro cometido. Com isso, estamos interessados em escolher métodos numéricos que sejam confiáveis, ou seja, métodos que não propaguem os erros dos dados iniciais. Um algoritmo que satisfaça essa condição é chamado de *estável*.

 $<sup>^{1}</sup>$ A notação fl está associada ao conceito de *ponto flutuante*. Esse é o nome dado ao formato de representação digital de números racionais no computador.

Caso tal condição não seja satisfeita, chamamos o algoritmo de *instável*. Certos algoritmos são estáveis apenas para alguns valores de dados iniciais. Tais algoritmos são chamados de *condicionalmente estáveis*.

Considerando um erro inicial de magnitude  $E_0>0$ , vamos denotar a magnitude do erro após n operações sucessivas por  $E_n$ . Se  $E_n\approx CnE_0$ , em que C é uma constante que não depende de n, então o crescimento do erro é chamado de *linear*. Por outro lado, Se  $E_n\approx C^nE_0$ , para qualquer C>1, então o crescimento do erro é chamado de *exponencial*.

Os resultados mais aceitos na literatura são aqueles em que os erros cometidos são lineares, cujos valores de C e  $E_0$  são pequenos. Por outro lado, o crescimento exponencial do erro deve ser evitado, já que produz resultados muito distintos do esperado, mesmo para valores pequenos de n. Consequentemente, temos que um algoritmo com comportamento de erro linear é estável, enquanto que o de erro exponencial é instável.

## Capítulo 2

# Soluções aproximadas de equações

Neste capítulo vamos discutir mecanismos numéricos para obter soluções (aproximadas) de equações algébricas que, por muitas vezes, não são simples de se resolver analiticamente. Por exemplo, considere as hipóteses admitidas por Malthus, isto é, durante pequenos períodos de tempo uma população cresce continuamente proporcional a ela mesma. A equação diferencial que descreve esse modelo é dada por  $x'=\lambda x$ , sendo  $\lambda$  a taxa de crescimento. A solução do problema proposto por Malthus é dada por  $x(t)=x_0e^{\lambda t}$ , em que  $x_0$  é o número da população inicial.

Na dinâmica apresentada acima, podemos considerar outros fatores, como por exemplo imigração. Se a imigração de uma população é permitida, digamos a uma taxa constante v, então obtemos a seguinte EDO  $x'=\lambda x+v$ , cuja solução é dada por  $x(t)=x_0e^{\lambda t}+\frac{v}{\lambda}(e^{\lambda t}-1)$ . Sob essa situação, suponha que a população contenha inicialmente 1.000.000 de indivíduos e que 200.000 imigrem para essa comunidade no primeiro ano. Suponha também que ao final desse primeiro ano se tenha 1.350.000 indivíduos. Sendo assim, para determinar a taxa de crescimento dessa população neste ano, devemos encontrar o valor de  $\lambda$  de tal forma que

$$1350000 = 1000000e^{\lambda} + \frac{200000}{\lambda}(e^{\lambda} - 1). \tag{2.1}$$

Resolver este problema do ponto de vista algébrico se torna inviável, uma vez que o termo  $\lambda$  se encontra não somente no argumento da função exponencial, mas também fora dele. Nesse sentido, uma abordagem numérica pode auxiliar na determinação deste parâmetro Burden and Faires (2003). Vamos então dedicar este capítulo a estudar algumas ferramentas numéricas, que podem auxiliar na resolução de problemas como o apresentado acima.

## 2.1 Método da bissecção

Anteriormente comentamos o problema de determinar o parâmetro  $\lambda$  de tal forma que a Equação (2.1) seja válida. Esta mesma equação pode ser reescrita da seguinte forma:

$$1350000 - 1000000e^{\lambda} - \frac{200000}{\lambda}(e^{\lambda} - 1) = 0.$$
 (2.2)

Perceba que o parâmetro  $\lambda$  é a única variável desta equação, portanto, podemos pensar no lado esquerdo da Equação (2.2) como uma função de uma variável que depende de  $\lambda$ , isto é:

$$f(\lambda) = 1350000 - 1000000e^{\lambda} - \frac{200000}{\lambda}(e^{\lambda} - 1).$$
 (2.3)

Sendo assim, encontrar a solução de (2.2) é equivalente a encontrar o valor  $\lambda$  que zera a função f. Este problema é chamado de *encontrar a raiz da função* ou *zeros da função* f. Para resolver esse tipo de problema, podemos utilizar a técnica do *método da Bissecção*. Tal método funciona da seguinte forma, considere uma função f contínua em um intervalo [a,b], de tal forma que existem dois valores  $p_1$  e  $p_2$  tais que  $f(p_1) > 0$  e  $f(p_2) < 0$ , conforme ilustrado na Figura 2.1.

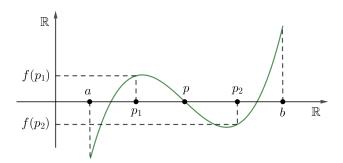


Figura 2.1: Ilustração das condições necessárias para o método da bissecção.

Existe um teorema na área da Matemática chamado de *Teorema do Valor Intermediário (TVI)*<sup>1</sup> que garante o seguinte, se f é contínua e existem  $p_1$  e  $p_2$  tais que  $f(p_2) < 0 < f(p_1)$ , então existe um número p tal que f(p) = 0 (veja a Figura 2.1 para se convencer disso). Esse resultado é importante para garantir a existência desse valor p para que possamos estimá-lo numericamente.

Pois bem, o método da bissecção consiste em uma contínua diminuição do intervalo original [a,b] em outros menores, até encontrar o valor p. Para iniciar o processo de obtenção do valor p, começamos tomando  $\overline{a}=a$  e  $\overline{b}=b$  e  $\overline{p}$  como sendo o ponto médio do intervalo  $[\overline{a},\overline{b}]$ , ou seja,  $\overline{p}=\frac{\overline{a}+\overline{b}}{2}$ . Se  $f(\overline{p})=0$ , então o processo está finalizado e obtemos  $p=\overline{p}$ . Caso contrário, segue que ou  $f(\overline{p})$  tem o mesmo sinal que  $f(\overline{a})$  ou o mesmo sinal que  $f(\overline{b})$ . Se  $f(\overline{p})$  e  $f(\overline{a})$  tem o mesmo sinal, então escolhemos  $\overline{a}=\overline{p}$  e  $\overline{b}=\overline{b}$ . Se  $f(\overline{p})$  e  $f(\overline{b})$  tem o mesmo sinal, então escolhemos  $\overline{a}=\overline{a}$  e  $\overline{b}=\overline{p}$ . Com isso, aplicamos o método novamente ao intervalo  $[\overline{a},\overline{b}]$ . Caso o ponto médio seja o valor procurado, então terminamos o processo, caso contrário o algoritmo continua rodando. Esse processo como um todo pode ser visualizado na Figura 2.2.

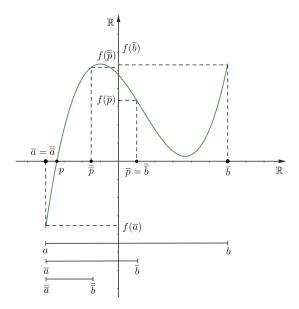


Figura 2.2: Representação gráfica do funcionamento do método da bissecção.

 $<sup>^1</sup>$ O Teorema do Valor Intermediário, como é mais conhecido na literatura brasileira, é um resultado estabelecido por Bernard Bolzano e Augustin Louis Cauchy (Teorema Bolzano-Cauchy), cujo enunciado é: Seja f uma função contínua em um intervalo fechado [a,b] e K um valor entre f(a) e f(b), então existe um número c no intervalo aberto (a,b), tal que f(c)=K.

Exercício 2.1.1 Utilize o método da bissecção para resolver a equação (1)

**Exercício 2.1.2** Utilize o método da bissecção para encontrar uma raiz do polinômio  $p(x)=x^3+x-2$ . Após encontrar tal raiz, faça a decomposição  $q(x)=(x-\overline{x})r(x)$ , sendo  $\overline{x}$  a raiz encontrada numericamente. O que pode ser dito sobre o polinômio r(x) quanto as suas raízes?

**Exercício 2.1.3** Utilize o método da bissecção para encontrar uma raiz do polinômio  $f(x) = x^3 + x^2 - 10x + 8$  no intervalo [-5,5]. Como o polinômio tem grau três, é possível que exista outras raízes reais, sendo assim considere outro intervalo inicial de modo a tentar buscar outras possíveis raízes. O que obteve?

### 2.2 Outros métodos para se encontrar zero de funções

O método da bissecção visto na seção anterior não é a única forma de se obter os zeros de uma função. Vamos aqui apresentar brevemente mais dois métodos, apenas para curiosidade. Nosso objetivo não é explorar todos os métodos de uma forma aprofundada, mas sim um panorama geral dos mecanismos que podem ser utilizados.

#### 2.2.1 Método de Newton

O primeiro que iremos comentar é o *método de Newton* que é um dos mais conhecidos e mais utilizados para obter as raízes de uma função. Algebricamente, o método é definido por meio da extensão via polinômio de Taylor, isto é, considerando f uma função de classe  $C^2$  e  $\overline{p}$  uma aproximação de p, de tal forma que  $f'(\overline{p}) \neq 0$  e  $|p-\overline{p}|$  é "pequeno", então

$$f(x) = f(\overline{p}) + (x - \overline{p})f'(x) + \frac{(x - \overline{p})^2}{2}f''(\xi(x)),$$

em que  $\xi(x)$  assume valores entre  $x \in \overline{p}$ .

Assim, tomando x=p e assumindo o fato de que p é uma raíz da função f, temos que

$$0 = f(\overline{p}) + (p - \overline{p})f'(p) + \frac{(p - \overline{p})^2}{2}f''(\xi(p)).$$

Assumimos anteriormente que  $|p-\overline{p}|$  é "pequeno", logo  $(p-\overline{p})^2$  é menor ainda. Portanto, podemos concluir que

$$0 \approx f(\overline{p}) + (p - \overline{p})f'(p),$$

e consequentemente,

$$p pprox \overline{p} - rac{f(\overline{p})}{f'(p)}.$$
 (2.4)

A estimativa (2.4) é chamada de método de Newton. Com isso, a partir de um chute inicial  $p_0$ , se estabelece um método iterativo com o intuito de estimar o valor p, que é dado por

$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}. (2.5)$$

Geometricamente, o método de Newton tem o seguinte significado, primeiro considera-se um chute

 $<sup>^2</sup>$ Uma função é dita de classe  $\mathbb{C}^n$  se suas n derivadas forem continuas.

inicial  $p_0$ . A partir disso, calcula-se a tangente da função f no ponto  $p_0$ . Essa reta tangente cruzará o eixo-x em algum ponto<sup>3</sup>, digamos  $p_1$ . Novamente calcula-se a tangente da função f, mas agora no ponto  $p_1$ . A reta tangente cruzará o eixo-x em algum ponto  $p_3$ , e o método continua até que o ponto que a tangente cruzar for exatamente a raiz p. Para visualização, veja a Figura 2.3.

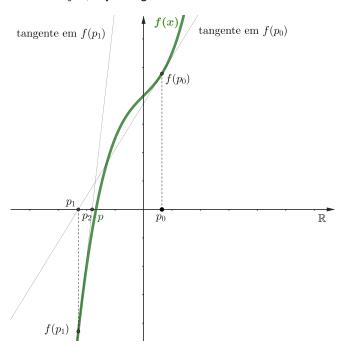


Figura 2.3: Interpretação geométrica do método de Newton para obter raízes de uma função.

#### 2.2.2 Método da secante

O método de Newton é uma ferramenta extremamente poderosa e converge rápido para a raiz da função (ou certas circunstâncias, fica evidente que a convergência é improvável). No entanto, existe um grande problema com este método que é saber/estimar a derivada da função no ponto, que do ponto de vista prático gera tantos problemas quanto analisar a função f. Para evitar este problema, fazemos uma variação do método de Newton chamado de *método da secante*.

O método da secante é introduzido a partir da definição de derivada, isto é,

$$f'(p_{n-1}) = \lim_{x \to p_{n-1}} \frac{f(x) - f(p_{n-1})}{x - p_{n-1}}.$$

Tomando  $x=p_{n-2}$ , temos então que

$$f'(p_{n-1}) \approx \frac{f(p_{n-2}) - f(p_{n-1})}{p_{n-2} - p_{n-1}}$$
$$= \frac{f(p_{n-1}) - f(p_{n-2})}{p_{n-1} - p_{n-2}}.$$

Substituindo a aproximação da derivada no método de Newton, obtemos:

$$p_n = p_{n-1} - \frac{f(p_{n-1})(p_{n-1} - p_{n-2})}{f(p_{n-1}) - f(p_{n-2})}.$$
(2.6)

 $<sup>^3</sup>$ Perceba que isso só não será verdade se a reta tangente for paralela ao eixo-x, isto é, se a derivada da função no ponto for igual a zero. Por esse motivo estamos supondo que f'(p) é diferente de zero para todo p.

A equação (2.6) é chamada de método da secante. Geometricamente o método da secante consiste em considerar dois chutes iniciais  $p_0$  e  $p_1$ , e a partir deles construir a reta secante que passa pela imagem dos dois pontos. A reta secante vai cruzar o eixo-x em um ponto  $p_2$ . Com esse ponto  $p_2$  constrói-se a reta secante que passa por  $f(p_2)$  e  $f(p_0)$ . Essa nova reta vai cruzar o eixo-x em um novo ponto  $p_3$ . O método continua dessa forma até obter o valor p que é o zero da função f. Veja a Figura 2.4 para visualizar esse processo.

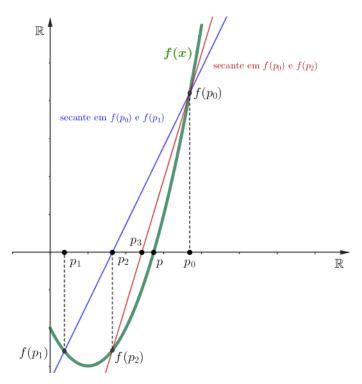


Figura 2.4: Interpretação geométrica do método da secante para obter raízes de uma função.

É importante observar que quando computamos o método iterativo de Newton (ou da secante), estamos estudando o seguinte problema  $p_n=g(p_{n-1})$ , em que  $g(p_{n-1})=p_{n-1}-\frac{f(p_{n-1})}{f'(p_{n-1})}$ , com  $n\geq 1$ . Este problema é conhecido como problema do ponto fixo, que discutiremos a seguir.

**Exercício 2.2.1** Considere o problema apresentado no início do capítulo sobre o crescimento populacional de Malthus considerando imigração constante, isto é,  $x'=\lambda x+\upsilon$ , cuja solução é dada por  $x(t)=x_0e^{\lambda t}+\frac{\upsilon}{\upsilon}(e^{\lambda t}-1)$ .

1. Utilize o método da secante para determinar o parâmetro  $\lambda$  de tal modo que

$$1350000 = 1000000e^{\lambda} + \frac{200000}{\lambda}(e^{\lambda} - 1).$$

2. Interprete o valor  $\lambda$  encontrado.

### 2.3 Problema do ponto fixo

Um número p é chamado de *ponto fixo* para uma função g, se g(p)=p. Este problema está associado ao métodos iterativos que estudamos na seção anterior no seguinte sentido, determinar o ponto fixo de uma função g é equivalente a buscar um valor p de tal forma que g(p)-p=0. Em outras palavras,

o problema se resume a encontrar o zero da função f(x)=g(x)-x. Por outro lado, buscar o zero de uma função f, isto é, determinar p tal que f(p)=0 é equivalente a buscar o ponto fixo p de g, isto é, sendo p um ponto fixo de g tem-se que f(p)=g(p)-p=0.

Geometricamente, determinar o ponto fixo de uma função f é buscar os pontos de intersecção da função f com a reta identidade h(x)=x. Note na Figura 2.5 que a função  $f(x)=x^2+x-1$  é interceptada pela função h em dois pontos, em  $p_1=-1$  e  $p_2=1$ . De fato, note que algebricamente temos:

$$f(-1) = (-1)^2 + (-1) - 1 = -1 = g(-1)$$
 e  $f(1) = (1)^2 + (1) - 1 = 1 = g(1)$ .

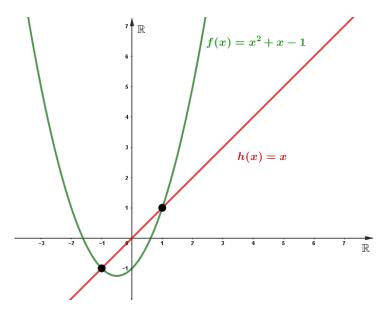


Figura 2.5: Representação gráfico do problema do ponto fixo. A função f admite dois pontos fixos,  $p_1=-1$  e  $p_2=1$ .

No exemplo acima, utilizamos recursos gráficos para determinar tais pontos, partindo do pressuposto que os pontos fixos existiam. O teorema abaixo fornece as condições suficientes para a existência e a singularidade de um ponto fixo.

**Teorema 2.3.1** Seja g uma função contínua em um intervalo [a, b].

- 1. Se a imagem da g satisfizer  $g(x) \in [a,b]$  para todo  $x \in [a,b]$ , então a função g possui um ponto fixo.
- 2. Se além disso, a função g admitir derivada em (a,b) e existir alguma constante positiva k<1 de tal modo que  $|g'(x)| \le k$ , para todo  $x \in (a,b)$ , então o ponto fixo em [a,b] é único.

Os pontos fixo de uma função podem ser, em algumas vezes, determinados algebricamente. Por exemplo, considere a função  $g(x)=\frac{(x^2-1)}{3}$  no intervalo [-3,3]. A função g satisfaz as condições 1. e 2. do Teorema 2.3.1 no intervalo  $[-3,3]^4$ , assim podemos calcular o seguinte:

$$g(p) = p \Leftrightarrow \frac{(p^2 - 1)}{3} = p$$
$$\Leftrightarrow p^2 - 1 = 3p$$
$$\Leftrightarrow p^2 - 3p - 1 = 0.$$

 $<sup>^4</sup>$ Para verificar se as hipóteses do teorema do ponto fixo são satisfeitas, recomenda-se utilizar recursos da Análise Real. Neste caso particular, pode-se utilizar o *Teorema de Weierstrass* que diz o seguinte, toda função contínua em um intervalo compacto [a,b] de  $\mathbb R$  possui um valor máximo e um mínimo nesse intervalo.

Aplicando a fórmula quadrática, obtemos dois pontos fixo  $p_1=\frac{1}{2}(3-\sqrt{13})$  e  $p_2=\frac{1}{2}(3+\sqrt{13})$ . Note que apenas  $p_1$  pertence ao intervalo [-1,1], e portanto é o procurado. É importante notar que se estivéssemos olhando para o intervalo [3,4], ou seja, o intervalo que contém o segundo ponto fixo  $p_2$ , não teríamos que as condições do Teorema 2.3.1 seriam satisfeitas, pois  $g'(4)=\frac{8}{3}>1$ . Isso significa que as condições colocadas no teorema do ponto fixo são suficientes, mas não necessárias.

Caso não seja possível obter os pontos fixos de uma forma similar a esta, pode-se obter uma aproximação deles, por meio do que chamamos de *método iterativo do ponto fixo*. A ideia consiste em criar uma sequência de valores da seguinte forma,  $p_n=g(p_{n-1})$ . Se esta sequência convergir para p e g for uma função contínua, então obtemos uma solução para g(x)=x. A Figura 2.6 ilustra esse mecanismo geometricamente.

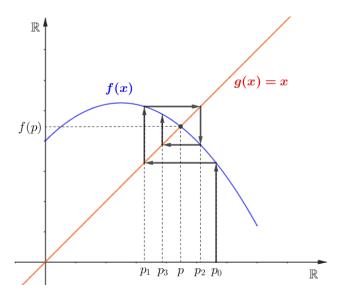


Figura 2.6: Interpretação geométrica do método iterativo do ponto fixo.

**Exercício 2.3.1** A partir da equação  $x^3 + 4x^2 - 10 = 0$ , determine o problema do ponto fixo associado. Este novo problema satisfaz as condições do Teorema 2.3.1 no intervalo [1,2]? Se sim, determine seu ponto fixo.

## Capítulo 3

# Soluções Numéricas para Equações Diferenciais Ordinárias

Neste capítulo a discussão será sobre a resolução de *Equações Diferenciais Ordinárias* (EDOs) do ponto de vista numérico. Em particular, trataremos de *Problemas de Valores Iniciais* (PVIs) e *Problemas de Valores de Contorno (ou Fronteira)* (PVCs). Os PVIs são regidos por uma EDO, digamos por exemplo x'(t) = f(t,x), que está associada a uma condição inicial  $x(t_0) = x_0^{-1}$  Sob algumas condições  $x'(t) = x_0^{-1}$  Sob algumas condições

Aqui serão tratados três métodos numéricos: *método de Euler*, *método de Runge-Kutta* e o *método de Diferenças Finitas*, sendo que este último também pode ser aplicado para *Equações Diferenciais Parciais* (EDPs).

#### 3.1 Método de Euler

Considere uma EDO de primeira ordem

$$x'(t) = f(t, x) \tag{3.1}$$

com condição inicial dada por  $x(t_0) = x_0$ .

No lado esquerdo da EDO (3.1) temos a derivada de x em relação à variável t, isto é, a derivada é dada por

$$x'(t) = \lim_{h \to 0} \frac{x(t+h) - x(t)}{h}.$$
(3.2)

<sup>&</sup>lt;sup>1</sup>Lembrando que se tivéssemos uma EDO de segunda ordem - ou maior - seriam necessárias também as condições iniciais das derivadas de ordem superior, ou seja,  $x(t_0)=x_0^{(0)}, x'(t_0)=x_0^{(1)}, \ldots, x^{(n-1)}(t_0)=x_0^{(n-1)}$ , para pudesse existir solução única para o problema.

 $<sup>^2</sup>$ Se a função f(t,x) é limitada e contínua em t, e Lipschitz contínua em x, então existe solução para o PVI e a solução é única  $^3$ Uma função real g é Lipschitz contínua se existir L>0 tal que  $|g(x)-g(y)|\leq L|x-y|$ , para todo  $x,y\in\mathbb{R}$ . O conceito de função Lipschitziana pode ser estendido para qualquer espaço métrico (X,d).

Veja que o limite (3.2) é uma variação do limite

$$x'(t_p) = \lim_{t \to t_p} \frac{x(t) - x(t_p)}{t - t_p},$$
(3.3)

chamando  $h = t - t_p$ .

Perceba que o limite (3.3) especifica qual é o ponto que esta sendo tratado na derivada. Por outro lado, o primeiro limite (3.2) é mais geral, no sentido de não deixar fixo o ponto que esta sendo trabalhado. O valor  $h=t-t_p$  é chamado de *espaçamento* ou *malha*, e para fins de facilitar a implementação computacional, consideramos que o espaçamento h é sempre de mesmo tamanho, ou seja,  $h=t_{i+1}-t_i$  para todo i. Quaisquer dos limites apresentados acima podem ser utilizados, a pessoa leitora só deve se atentar ao fato de que h é o espaçamento entre t e  $t_p$ . Dito isso, vamos para o método.

Pela expansão de Taylor, de ordem 2, da função x no ponto  $t_n$  tem-se:

$$x(t) = x(t_p) + (t - t_p)x'(t_p) + (t - t_p)^2 \frac{x''(t_p)}{2},$$

e dividindo ambos os lados por  $t-t_p$  (note que estamos considerando até o momento  $t \neq t_p$ , e portanto essa divisão esta bem definida), obtêm-se

$$\frac{x(t) - x(t_p)}{t - t_p} = x'(t_p) + (t - t_p) \frac{x''(t_p)}{2}.$$

A expressão acima nos diz que a derivada de uma função no ponto  $t_p$  pode ser aproximada por  $\frac{x(t)-x(t_p)}{t-t_p}$  cometendo-se um erro  $(t-t_p)\frac{x''(t_p)}{2}$ . Perceba que  $\frac{x''(t_p)}{2}$  é um número fixo, enquanto  $t-t_p=h$  é um número que varia, pois t está variando. Isso significa que o erro cometido na aproximação da derivada é de  $\operatorname{ordem} h$ , também denotado por  $\mathcal{O}(h)$ .

Assim o método de Euler é definido da seguinte forma, considere  $\mathcal{P}=\{a=t_0,t_1,\ldots,t_n=b\}$  uma partição para [a,b] da qual a solução do PVI estará definida, da qual é igualmente espaçada, isto é,  $h=t_i-t_{i-1}$  para todo  $i=\{1,\ldots,n\}$ . Considerando a aproximação mencionada acima, tem-se que

$$x'(t_i) = \frac{x(t_i) - x(t_i - h)}{h}$$

por outro lado, pela EDO (3.1) tem-se que  $x'(t_i) = f(t_i, x(t_i))$ , assim:

$$f(t_i, x(t_i)) = \frac{x(t_i) - x(t_i - h)}{h}$$

$$x(t_i) - x(t_i - h) = hf(t_i, x(t_i))$$

$$x(t_i) = x(t_i - h) + hf(t_i, x(t_i)).$$

Portanto, o método de Euler é definido como

$$x_{n+1} = x_n + h f(t_n, x_n). (3.4)$$

Para estimar o erro cometido pelo método, vamos comparar a expressão (3.4) com a expansão de Taylor de ordem 2. Obtemos o seguinte:

$$x(t+h) = x(t) + hx'(t) + h^2 \frac{x''(t)}{2}$$
$$x(t+h) = x(t) + hf(t,x) + h^2 \frac{f'(t,x)}{2}.$$

Note que x(t+h)=x(t)+hf(t,x) é exatamente o método de Euler. Assim, o erro cometido na aproximação do método é de ordem  $h^2$ , isto é,  $\mathcal{O}(h^2)$ .

Geometricamente, o método de Euler produz uma solução numérica que aproxima a solução analítica por meio das derivadas. Com isso, quanto menor o valor do espaçamento h, melhor deve ser a aproximação. A Figura 3.1 ilustra essa afirmação.

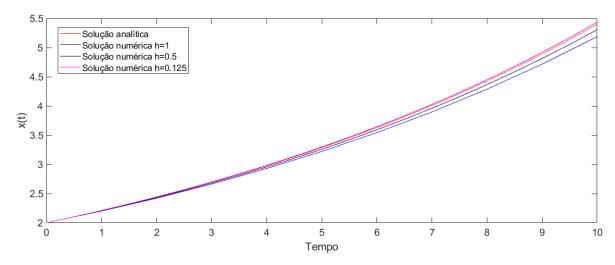


Figura 3.1: Aproximação da solução analítica da EDO  $x'(t) = \lambda x(t)$  pelo método de Euler, considerando condição inicial x(0) = 2.

No entanto, o método de Euler apresenta problemas com precisão, isto é, normalmente precisamos escolher um passo h muito pequeno para obter soluções numéricas de "boa qualidade" (veja novamente a Figura 3.1), o que implica um número elevado de passos e, consequentemente, um maior custo computacional. A partir dessa observação, vamos introduzir um segundo método numérico chamado de *método de Runge-Kutta*.

### 3.2 Método de Runge-Kutta

No final do século XIX, Carl Runge deduziu diversas fórmulas para encontrar uma aproximação de solução para PVIs. Martin Wilhelm Kutta generalizou muitos dos métodos fornecidos por Carl Runge. Aqui ilustraremos um desses métodos chamados de método de Runge-Kutta de ordem 4. Este método é apenas um dos chamados de método de Runge-Kutta de ordem superior.

O método de ordem 1 é também conhecido como método de Euler. Já o método de ordem 2 é chamado de método de Euler aperfeiçoado/modificado ou também método de Heun<sup>4</sup>. Aqui focaremos no método de Runge-Kutta de ordem 4, que assim como os anteriores pode ser deduzido da expansão de Taylor para funções de mais de uma variável. Não faremos a dedução aqui por se tratar de uma construção bastante técnica (para mais detalhes da construção recomenda-se a leitura dos livros Burden and Faires (2003); Conte (1965)), apresentaremos apenas sua fórmula.

Considere o PVI x'=f(t,x) cuja condição inicial é dada por  $x(0)=x_0$ . Assim, a partir de uma partição  $\mathcal P$  com tamanho h, o método de Runge-Kutta de ordem 4 é definido por

$$x_{n+1} = x_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4), \tag{3.5}$$

 $<sup>^4</sup>$ Karl Heun introduziu a técnica de aproximação de solução para um PVI de ordem  $\mathcal{O}(h^3)$ , que não focaremos aqui mas é muito utilizado na literatura e pode ser consultado no artigo Heun (1900)

em que

$$k_1 = f(t_n, x_n),$$

$$k_2 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_1),$$

$$k_3 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_2),$$

$$k_4 = f(t_n + h, x_n + hk_3).$$

Para ilustrar a solução numérica via método de Runge-Kutta, consideraremos o mesmo PVI de Malthus da subseção anterior, isto é,  $x'=\lambda x$ . A Figura 3.2 mostra graficamente a solução obtida.

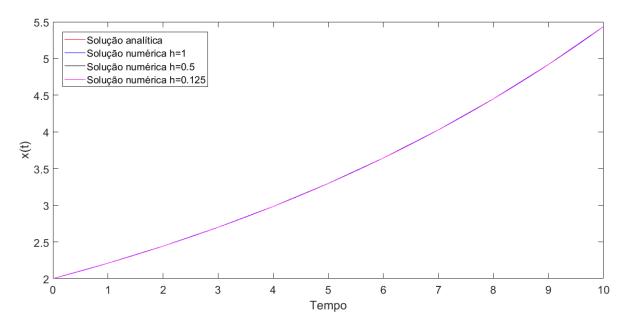
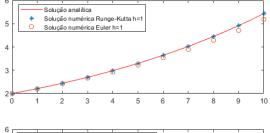


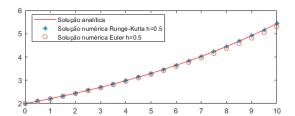
Figura 3.2: Aproximação da solução analítica da EDO  $x'(t) = \lambda x(t)$  pelo método de Runge-Kutta de ordem 4, considerando a condição inicial x(0) = 2.

Perceba que independente do tamanho do espaçamento h, a solução numérica produzida é praticamente a mesma que a solução analítica  $x(t)=x_0e^{\lambda t}$ . Para mostrar a precisão deste método em relação ao método de Euler, apresentamos a Figura 3.3 que apresenta ambas as soluções numérica em três diferentes cenários.

A Figura 3.3 ilustra como o método de Runge-Kutta é mais preciso que a de Euler. Perceba que a solução numérica produzida pelo método de Euler para h=0.125 é tão "eficiente" quanto a de Runge-Kutta para h=1. No entanto, os esforços computacionais são distintos em cada caso. O principal esforço computacional na aplicação do método de Runge-Kutta de ordem 4 é o cálculo do campo f, isto é, o método requer quatro cálculos da f por passo, cujo erro cometido é  $\mathcal{O}(h^4)$ . Em comparação com o método de Euler, para cada passo o método de Runge-Kutta de ordem 4 requer quatro cálculos e a de Euler apenas um.

Nesse capítulo estudamos métodos de aproximação de derivada para resolver numericamente equações diferenciais ordinárias, a partir de uma condição inicial. Os métodos considerados neste estudo foram os de Euler e Runge-Kutta que dependem de passos anteriores, a fim de realizar estimativas dos próximos. Tais métodos são conhecidos como *métodos de passo único* e são os mais simples na área de Análise Numérica. O objetivo aqui não é tratar de métodos mais elaborados para lidar com EDOs numericamente, e sim ter um panorama geral de como abordar tais problemas do ponto de vista numérico. Para obter uma melhor aproximação da solução outros métodos mais robustos podem ser considerados, como





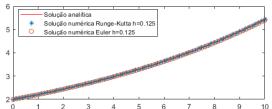


Figura 3.3: Aproximação da solução analítica da EDO  $x'(t) = \lambda x(t)$  pelo método de Runge-Kutta de ordem 4 e Euler, considerando a condição inicial x(0) = 2 e três diferentes espaçamentos h = 1, h = 0.5 e h = 0.125.

por exemplo os métodos de *passos múltiplos*. Para a leitora que tiver interesse em aprofundar seu estudo sobre o assunto, recomendam-se as referências Burden and Faires (2003); Conte (1965).

No próximo capítulo abordaremos métodos numéricos para aproximar a solução de integrais definidas.

**Exercício 3.2.1** Forneça a solução numérica de Runge-Kutta de ordem 4 para o modelo populacional de Verhulst. Faça o mesmo para o método de Euler e compare as soluções. Para evidenciar as soluções obtidas, faça diferentes testes para valores de h distintos.

**Exercício 3.2.2** O modelo epidemiológico do tipo SI (Suscetível-Infectado) é dado por um sistema de EDOs acopladas, isto é, uma EDO depende da outra. A dinâmica para cada população é dada por  $S' = -\lambda SI$  e  $I' = \lambda SI$ , sendo  $S(0) = S_0$  e  $I(0) = I_0$  as condições iniciais da população (aqui não está sendo considerado dinâmica vital). Forneça a solução numérica para esse sistema, utilizando os métodos de Euler e Runge-Kutta. Compare os resultados.

## Capítulo 4

# Integração Numérica

Alguns problemas práticos nas áreas de engenharia, biologia, física e química recaem em problemas que são modelados por integrais. Por exemplo, para se determinar o comprimento de uma chapa com formato ondulatório, digamos descrita por uma função de onda senoidal f(x) = sen(x), é preciso considerar a fórmula de comprimento de curva Guidorizzi (2001), que é dada pela seguinte integral  $L = \int_a^b \sqrt{1+cos^2(x)} dx$  Burden and Faires (2003). Esse tipo de problema é conhecido como problema de *integral elíptica* Kaplan (1972) que não possui uma solução explicita. Assim como esse problema, existem várias outras integrais que não podem ser resolvidas explicitamente. Nosso objetivo então é fornecer uma aproximação numérica para os valores que tais integrais definidas assumem.

Os métodos que discutiremos aqui serão os do trap'ezio e de Simpson. Ambos os métodos são obtidos por variações do método de integração chamado de  $quadradura\ num\'erica$ , em que se utiliza uma soma  $\sum_{i=0}^n a_i f(x_i)$  para aproximar o valor de  $\int_a^b f(x)dx$ . Basicamente, consideramos um conjunto de pontos  $\{x_0,x_1,\ldots,x_n\}$  do intervalo [a,b] e para eles construímos o polinômio interpolador de Lagrange, isto é,  $P_n(x) = \sum_{i=0}^n f(x_i) L_i(x)$ . A partir disso, a constante  $a_i$  procurada na quadradura é dada por  $a_i = \int_a^b L_i(x)dx$ , para cada  $i=1,\ldots,n$ . A dedução desse fato pode ser encontrada em Burden and Faires (2003). Ao considerar os polinômios de Lagrange de graus 1 e 2, obteremos os métodos do trapézio e de Simpson, respectivamente. Vamos então discutir a especificidade de cada um deles.

## 4.1 Método do trapézio

Consideremos um conjunto de dois pontos  $\{x_0, x_1\}$  de tal forma que  $a=x_0$  e  $b=x_1$ . Considere o polinômio de Lagrange de grau 1 dado por:

$$P_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1).$$

A partir do polinômio de Lagrange e a expansão da função f até grau 2, obtemos  $^1$ :

$$\int_{a}^{b} f(x)dx = \int_{x_{0}}^{x_{1}} \frac{x - x_{1}}{x_{0} - x_{1}} f(x_{0}) + \frac{x - x_{0}}{x_{1} - x_{0}} f(x_{1}) dx + \frac{1}{2} \int_{x_{0}}^{x_{1}} f''(\xi)(x - x_{0})(x - x_{1})$$

$$= \left( \frac{(x - x_{1})^{2}}{2(x_{0} - x_{1})} f(x_{0}) + \frac{(x - x_{0})^{2}}{2(x_{1} - x_{0})} f(x_{1}) \right) \Big|_{x_{0}}^{x_{1}} + \frac{f''(\xi)}{2} \left( \frac{x^{3}}{3} - \frac{(x_{0} + x_{1})}{2} x^{2} + x_{0} x_{1} x \right) \Big|_{x_{0}}^{x_{1}}$$

$$= \frac{h}{2} (f(x_{0}) + f(x_{1})) - \frac{h^{3}}{6} f''(\xi),$$

sendo  $h=x_1-x_0$  o espaçamento considerado no método.

Portanto, o método do trapézio ou também chamado de regra do trapézio é dado por

$$\int_{a}^{b} f(x)dx \approx \frac{h}{2}(f(a) + f(b)),\tag{4.1}$$

sendo  $h = x_1 - x_0$  com erro da ordem  $\mathcal{O}(h^3)$ .

A Figura 4.1 ilustra graficamente o método do trapézio para aproximar a integral de uma função.

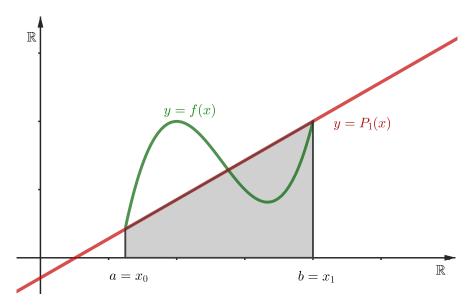


Figura 4.1: Método do trapézio para aproximar a integral de uma função no intervalo [a,b]. A função f é dada pela curva verde, o polinômio de Lagrange de ordem 1 é dada pela função em vermelho e a região em cinza é o trapézio definido pelo método.

É possível perceber na imagem acima que a aproximação desta área não foi tão precisa. Uma forma de tornar o método mais eficaz é considerar uma partição  $\{x_0, x_1, \ldots, x_n\}$  com mais pontos e aplicar o método do trapézio para cada intervalo  $[x_{i-1}, x_i]$ . Essa abordagem é conhecida como *método do trapézio composto* e sua fórmula é dada por

$$\int_{a}^{b} f(x)dx \approx \sum_{i=0}^{n-1} \frac{h}{2} (f(x_i) + f(x_{i+1})), \tag{4.2}$$

sendo  $h = x_i - x_{i-1}$  o tamanho de cada subintervalo igualmente espaçado.

Na Figura 4.2 é possível perceber que com uma partição um pouco maior foi possível obter uma melhor aproximação.

<sup>&</sup>lt;sup>1</sup>A existência do valor  $\xi$  constante é garantida pelo teorema do valor médio para integrais, já que  $(x-x_0)(x-x_1)$  não muda de sinal

É interessante notar que na dedução do método do trapézio aparece um termo que depende da segunda derivada da função f, que está associada ao erro cometido na aproximação. Assim, qualquer função f que tiver a segunda derivada igual a zero, teremos que o método do trapézio fornecerá o valor exato da integral. Para se convencer disso, considere por exemplo uma função que descreve uma reta, digamos f(x)=2x+1. Veja que a região determinada por essa função será exatamente o trapézio obtido pelo método numérico.

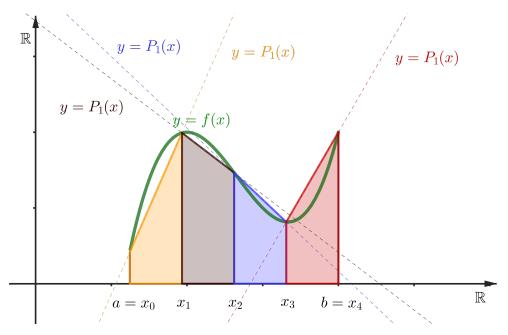


Figura 4.2: Método do trapézio composto para aproximar a integral de uma função no intervalo [a,b], considerando uma partição com 5 pontos  $\{x_0,x_1,x_2,x_3,x_4\}$ . A curva verde representa a função f, e os trapézios laranja, preto, azul e vermelho são os obtidos a partir da aplicação do método do trapézio para os intervalos  $[x_0,x_1]$ ,  $[x_1,x_2]$ ,  $[x_2,x_3]$  e  $[x_3,x_4]$ , respectivamente.

**Exercício 4.1.1** Considere a função  $f(x) = x^2$  no intervalo [1,2]. Utilize o método do trapézio para estimar a área da região estabelecida pela função.

**Exercício 4.1.2** Considere agora uma partição de [1,2] com 5 pontos. Utilize o método do trapézio composto para estimar a área da função  $f(x) = x^2$  no intervalo [1,2].

**Exercício 4.1.3** Compare os resultados obtidos nos dois exercícios anteriores com o real valor da integral da função  $f(x) = x^2$  no intervalo [1, 2].

**Exercício 4.1.4** Utilize o método do trapézio para calcular a integral definida no problema introduzido no início do capítulo, para a=0 e b=48 (Esse exercício foi retirado de Burden and Faires (2003))

**Exercício 4.1.5** Considere uma partição com 4 pontos e utilize o método do trapézio composto para calcular a integral do mesmo problema do exercício anterior.

### 4.2 Método de Simpson

O *método de Simpson* é muito similar ao método do trapézio que vimos na seção anterior. O que difere um do outro é que no de Simpson consideramos o polinômio de Lagrange de ordem 2. Pois bem, vamos ao método.

Consideremos um conjunto de três pontos  $\{x_0, x_1, x_2\}$  de tal forma que  $a = x_0$  e  $b = x_2$ . Considere o polinômio de Lagrange de grau 2 dado por:

$$P_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2).$$

A partir do polinômio de Lagrange e a expansão da função f até grau 3, obtemos:

$$\int_{a}^{b} f(x)dx = \int_{x_{0}}^{x_{2}} \left( \frac{(x-x_{1})(x-x_{2})}{(x_{0}-x_{1})(x_{0}-x_{2})} f(x_{0}) + \frac{(x-x_{0})(x-x_{2})}{(x_{1}-x_{0})(x_{1}-x_{2})} f(x_{1}) + \frac{(x-x_{0})(x-x_{1})}{(x_{2}-x_{0})(x_{2}-x_{1})} f(x_{2}) \right) dx + \int_{a}^{b} \frac{(x-x_{0})(x-x_{1})(x-x_{2})}{6} f^{(3)}(\xi) dx.$$

Resolvendo as integrais acima, obtemos:

$$\int_{a}^{b} f(x)dx = \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2)) + \mathcal{O}(h^4)$$

sendo  $h = x_i - x_{i+1}$  o espaçamento considerado no método em cada subintervalo.

Portanto, o método de Simpson ou também chamado de regra de Simpson é dado por

$$\int_{a}^{b} f(x)dx \approx \frac{h}{3}(f(x_0) + 4f(x_1) + f(x_2)),\tag{4.3}$$

sendo h o tamanho do espaçamento com erro da ordem  $\mathcal{O}(h^4)$ .

A Figura 4.3 ilustra graficamente o método de Simpson para aproximar a integral de uma função.

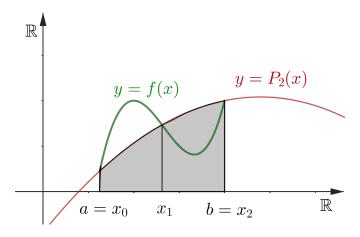


Figura 4.3: Método de Simpson para aproximar a integral de uma função no intervalo [a,b]. A função f é dada pela curva verde, o polinômio de Lagrange de ordem 2 é dada pela função em vermelho e a região em cinza é a região definida pelo método.

O método de Simpson é significantemente melhor que a do trapézio. Para ilustrar essa afirmação, fornecemos a Tabela 4.1 com alguns valores<sup>2</sup>.

Podemos fazer o mesmo processo feito anteriormente para o método dos trapézios compostos, isto é, podemos considerar uma partição P para o intervalo [a,b] e aplicar o método de Simpson em cada subintervalo desta partição. No entanto, é importante reforçar alguns aspectos dessa abordagem.

O método de Simpson trabalha com polinômio de Lagrange de ordem 2, sendo assim precisamos

<sup>&</sup>lt;sup>2</sup>a tabela foi "recortada" do livro Burden and Faires (2003).

	$x^2$	sen(x)	$e^x$
Valor exato	2.667	1.416	6.389
Trapézio	4	0.909	8.389
Simpson	2.667	1.425	6.421

Tabela 4.1: Aproximações de integrais para algumas funções no intervalo [0,2] a partir dos métodos do trapézio e Simpson (simples).

sempre de três pontos para definir uma parábola. Logo, para aproximar a área da função f no intervalo [a,b] precisamos então de um terceiro ponto, que no caso é o ponto médio já que os subintervalos serão considerados de mesmo tamanho. Então tomemos a partição  $\{a,x_1,b\}$ , em que  $x_1=\frac{(a+b)}{2}$ .

O método de Simpson para três pontos (que é a versão simples), assim como vimos acima, é dada por

$$\left(\frac{h}{3}\right)(f(a) + 4f(x1) + f(b)).$$

Agora consideremos um intervalo com 4 pontos, digamos  $\{a,x_1,x_2,b\}$ , então iremos aplicar a fórmula de Simpson duas vezes, uma para os pontos da partição  $\{a,x_1,x_2\}$  e outra para a partição  $\{x_1,x_2,b\}$ . O problema é que ao fazer esse processo a região entre  $x_1$  e  $x_2$  será somada duas vezes, fornecendo assim um resultado não compatível com o problema. A Figura 4.4 ilustra um problema desse tipo, considerando uma partição com 5 pontos  $\{a,x_1,x_2,x_3,b\}$ .

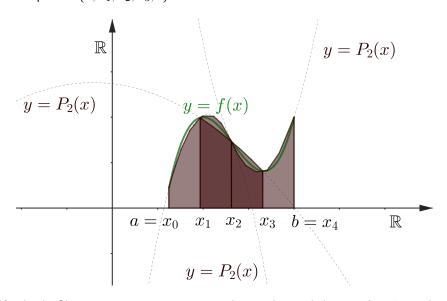


Figura 4.4: Método de Simpson composto para aproximar a integral de uma função no intervalo [a,b]. A função f é dada pela curva verde, o polinômio de Lagrange de ordem 2 é dada pelas funções em preto e a região em vinho é a região definida pelo método. As regiões em vinho escuro estão sendo consideradas mais de uma vez no método.

Assim, para evitar sobreposição da região, nós consideramos uma partição com uma quantidade ímpar de pontos, e seguimos com o método realizando uma espécie de deslocamento. Por exemplo, considerando a partição  $\{a, x_1, x_2, x_3, b\}$ , vamos aplicar o método de Simpson duas vezes, ao invés de três (assim como apresentado na Figura 4.4. Primeiro aplicamos o método para os pontos  $\{a, x_1, x_2\}$  e outra nos pontos  $\{x_2, x_3, b\}$ .

Dessa forma, na partição  $\{a,x_1,x_2\}$  a fórmula fornece

$$\left(\frac{h}{3}\right)\left(f(a) + 4f(x_1) + f(x_2)\right)$$

e na partição  $\{x_2, x_3, b\}$  a fórmula fornece

$$\left(\frac{h}{3}\right)(f(x_2)+4f(x_3)+f(b)).$$

E assim, o valor final da integral numérica é dada pela soma dessas duas expressões, isto é,

$$\left(\frac{h}{3}\right)\left(f(a) + 4f(x_1) + f(x_2)\right) + \left(\frac{h}{3}\right)\left(f(x_2) + 4f(x_3) + f(b)\right).$$

Note que o termo  $f(x_2)$  aparece duas vezes nas fórmulas, e assim, obtemos:

$$\left(\frac{h}{3}\right)(f(a) + 2f(x_2) + 4f(x_1) + 4f(x_3) + f(b))$$

ou ainda, colocando o 4 em evidência:

$$\left(\frac{h}{3}\right)(f(a)+2f(x_2)+4[f(x_1)+f(x_3)]+f(b)).$$

Em outras palavras, os elementos com índices pares são acompanhados do termo 2 enquanto os elementos de índices ímpares são acompanhados de 4. O 2 aparece por conta que o ponto final de uma partição sempre será o inicial da outra, e portanto, aparecerá duas vezes. Já o 4 aparece por conta da fórmula de Simpson. Note que ao fazer esse procedimento a região que era dada pela Figura 4.4, agora é dada pela região ilustrada na Figura 4.5.

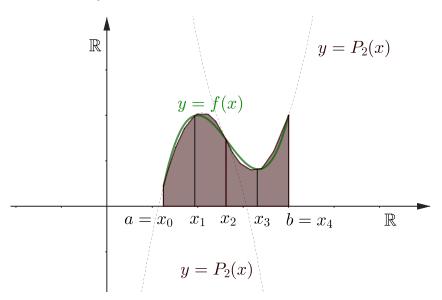


Figura 4.5: Método de Simpson composto para aproximar a integral de uma função no intervalo [a,b]. A função f é dada pela curva verde, o polinômio de Lagrange de ordem 2 é dada pelas funções em preto e a região em vinho é a região definida pelo método.

O teorema a seguir nos fornece a expressão geral do método de Simpson composto.

**Teorema 4.2.1** Seja f uma função de classe $^3$   $C^4$ , n um número par $^4$ ,  $h=\frac{b-a}{n}$  com  $x_j=a+jh$ , sendo

<sup>&</sup>lt;sup>3</sup>função que possui até a quarta derivada contínua

 $<sup>^4</sup>n$  é o número de intervalos e j o número de pontos escolhidos. Neste caso, como n é par, então o número de pontos escolhidos deve ser ímpar, já que j possui n+1 pontos.

 $j=0,\ldots,n$ . Então a aproximação do método de Simpson composto é dado por

$$\int_{a}^{b} f(x)dx \approx \frac{h}{3} \left( f(a) + 2 \sum_{j=1}^{(n/2)-1} f(x_{2j}) + 4 \sum_{j=1}^{(n/2)} f(x_{2j-1}) + f(b) \right)$$
(4.4)

e mais, a ordem do erro cometido é de  $\mathcal{O}(h^4)$ 

Dem: A demonstração deste resultado pode ser encontrada em Burden and Faires (2003).

Em questões de terminologia, se considerarmos o intervalo [a,b] particionado em três pontos igualmente espaçados, isto é,  $\{a=x_0,x_1,x_2=b\}$ , onde  $x_2-x_1=x_1-x_0$ , então o método de Simpson é também conhecido como  $\frac{1}{3}$  de Simpson. Ainda mais, se considerarmos o intervalo [a,b] e aplicarmos o método para Lagrange de terceiro grau, (de forma que são necessários 4 pontos para obter esse polinômio por interpolação) e com espaçamento  $h=\frac{(b-a)}{3n}$ , então o método de Simpson é também conhecido como  $\frac{3}{8}$  de Simpson.

**Exercício 4.2.1** Considere a função  $f(x) = e^x$  no intervalo [0,2].

- 1. Determine o valor exato da área determinada pela função f(x) no intervalo [0,2].
- 2. Utilize o método do Trapézio para estimar o valor desta área.
- 3. Utilize o método de Simpson para estimar o valor desta área.
- 4. Compare os métodos numéricos com o real valor da área.

**Exercício 4.2.2** Considere agora a função  $g(x)=e^{-x^2}$  no intervalo [-1,1]. Utilize o método que você considerou mais eficiente no exercício anterior e utilize sua versão composta para estimar a área da região determinada por g(x). Utilize quantos pontos desejar na partição.

## Capítulo 5

# Soluções numéricas para Equações Diferenciais Parciais

Até este momento estudamos método numéricos para aproximar as soluções exatas (também chamadas de soluções analíticas) de equações diferenciais e integrais que dependem de apenas uma variável. Isto é, para cada valor de t (geralmente referindo-se ao tempo) discutimos métodos que produzem valores  $x_t$  que sejam próximos do valor exato x(t) para cada instante t.

Nosso próximo objetivo é discutir aproximações de funções que não dependam só do tempo, mas também do espaço. Por exemplo, considere o fluxo de calor em um objeto feito de algum material condutor de calor, de tal modo que esteja sujeito a alguma fonte externa de calor ao longo de seu comprimento e algumas condições de contorno em cada extremidade. Sob algumas hipóteses, podemos descrever o processo de transferência de calor desse objeto pela seguinte equação diferencial

$$u_t(x,t) = (\kappa(x)u_x(x,t))_t + \psi(x,t) \tag{5.1}$$

em que  $\kappa(x)$  representa o coeficiente de condução de calor, que varia com respeito a posição x, e  $\psi(x,t)$  é a fonte de calor.

Neste caso, o valor u(x,t) representa a temperatura do objeto na posição x e tempo t. Sendo assim, como a Equação Diferencial (5.1) depende de mais de uma variável, chamamos-a de *Equação Diferencial Parcial* (EDP) que são extremamente complexas de se trabalhar, principalmente de uma forma analítica. Com isso, recursos numéricos são essenciais para que se faça um estudo do comportamento de tais fenômenos.

A Equação (5.1) é também chamada de *Equação de Difusão* já que modela um processo de difusão de um modo geral, e não somente a equação do calor LeVeque (1955). Para fins didáticos, vamos utilizar o método de *diferenças finitas* para estudar as equações de *Advecção*, *Difusão* e de *Onda*, nas suas formas mais simplificadas. O método de diferenças finitas pode ser utilizado para resolver EDOs, mas sua maior "importância" é na resolução de EDPs.

#### 5.1 Método das diferenças finitas

O objetivo principal deste método é aproximar soluções para EDPs, ou seja, estimar os valores de uma função que satisfaz uma dada relação (determinada pelo campo) que dependem de suas derivadas

parciais com respeito ao espaço e/ou tempo, juntamente com algumas condições de contorno ao longo das bordas deste domínio.

O método de *diferenças finitas* procede substituindo as derivadas nas equações diferenciais por aproximações de diferenças finitas. Essa abordagem resulta em um grande sistema algébrico no lugar da equação diferencial, o que torna o problema mais fácil de ser resolvido numericamente/computacionalmente.

Consideremos uma função u(x) que vamos assumir ser de classe  $C^{\infty}$ . Para aproximar a derivada desta função em um determinado ponto, podemos calcular dois tipos de aproximações:

$$D_{+}u(x) = \frac{u(x+h) - u(x)}{h}$$
 (5.2)

ou

$$D_{-}u(x) = \frac{u(x) - u(x - h)}{h}. (5.3)$$

As aproximações  $D_+$  e  $D_-$  são chamadas de *diferença finita à direita* e *à esquerda*, respectivamente<sup>1</sup>. Ambas aproximam as derivadas de primeira ordem e cometem um erro de aproximação da ordem  $\mathcal{O}(h)$  LeVeque (1955). Uma outra forma de aproximar a primeira derivada de uma função é por meio da *diferença finita centrada*, que é obtida "dando dois passos" na discretização do domínio, isto é,

$$D_c u(x) = \frac{u(x+h) - u(x-h)}{2h}. (5.4)$$

A diferença finita centrada também é conhecida como a média entre as diferenças finitas à direita e à esquerda, uma vez que

$$\frac{D_{+}u(x) + D_{-}u(x)}{2} = \frac{\frac{u(x+h) - u(x)}{h} + \frac{u(x) - u(x-h)}{h}}{2}$$

$$= \frac{u(x+h) - u(x-h)}{2h}$$

$$= D_{c}u(x)$$

Geometricamente falando, as diferenças finitas são interpretadas pelas inclinações das respectivas retas secantes, assim como a Figura 5.1 ilustra.

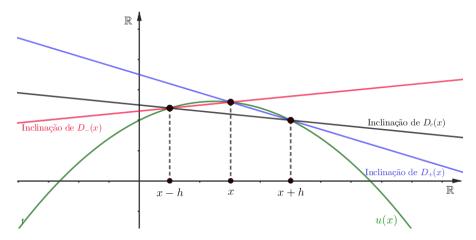


Figura 5.1: Aproximações da derivada da função u(x) por meio das inclinações das retas secantes para cada tipo de diferença finita. As retas vermelha, azul e preta representas as secantes obtidas pela diferença à esquerda, à direita e a centrada, respectivamente

<sup>&</sup>lt;sup>1</sup>Essas diferenças também são conhecidas como diferença progressiva e diferença regressiva

Também é possível realizar estimativas para as derivadas de ordem superior, como por exemplo a derivada de segunda ordem. Para isso, podemos considerar a expansão de Taylor até ordem 3 em duas versões:

$$u(x+h) = u(x) + u'(x)h + \frac{u''(x)h^2}{2!} + \frac{u'''(x)h^3}{3!}$$
(5.5)

е

$$u(x-h) = u(x) - u'(x)h + \frac{u''(x)h^2}{2!} - \frac{u'''(x)h^3}{3!}.$$
 (5.6)

Somando as Equações (5.5) e (5.6), obtemos:

$$u(x-h) + u(x+h) = 2u(x) + \frac{2u''(x)h^2}{2!},$$

e portanto, uma estimativa para a segunda derivada é dada por

$$D_2 u(x) = \frac{u(x-h) - 2u(x) + u(x+h)}{h^2},$$
(5.7)

em que o erro cometido nessa aproximação é da ordem  $\mathcal{O}(h^2)$ . Tal aproximação é conhecida como *diferença finita centrada de segunda ordem*.

Uma outra forma de obter a aproximação da segunda derivada é aplicando a diferença finita à esquerda, e em seguida à direita. Por exemplo,

$$D_{+}u(D_{-}u(x)) = \frac{1}{h}(D_{-}u(x+h) - D_{-}u(x))$$

$$= \frac{1}{h}\left(\frac{u(x+h) - u(x)}{h} - \frac{u(x) - u(x-h)}{h}\right)$$

$$= D_{2}u(x)$$

De modo similar, podemos obter  $D_2u(x)$  fazendo  $D_-u(D_+u(x))$  ou  $D_cu(D_cu(x))$ . Desse modo, podemos obter qualquer aproximação das derivadas superiores repetindo processos análogos a estes. Apenas para ilustrar o método das diferenças finitas, a próxima subseção fornece um exemplo de aplicação deste método para resolver EDOs.

#### 5.1.1 Diferenças finitas para Problema de Valor de Contorno

Considere uma EDO dada por

$$\begin{cases} x''(x) = f(x), & 0 \le x \le 1 \\ x(0) = a \\ x(1) = b \end{cases}$$
 (5.8)

Um problema descrito dessa forma é chamado de *Problema de Valor de Contorno* (PVC) Boyce and DiPrima (2009).

Considere agora  $\mathcal P$  uma partição do intervalo (0,1), por exemplo,  $\mathcal P=\{x_0,x_1,\dots,x_{m+1}\}$ . Para simplificar a notação, vamos denotar  $U_j$  a aproximação obtida pelo método das diferenças finitas, que aproxima o valor exato  $u(x_j)$ . Nesse contexto, estamos considerando que cada  $x_j$  é dado por  $x_j=jh$ , sendo  $h=\frac{1}{m+1}$ .

As condições x(0) = a e x(1) = b, chamadas de *condições de fronteira*, estabelecem os valores

de  $U_0$  e  $U_{m+1}$ , respectivamente, isto é,  $U_0=a$  e  $U_{m+1}=b$ , faltando determinar  $U_1,\ldots,U_m$ .

Considerando a aproximação por diferença finita centrada de ordem 2 na equação (5.8), obtemos:

$$\frac{1}{h^2}(U_{j-1} - 2U_j + U_{j+1}) = f(x_j), \tag{5.9}$$

onde  $U_{j-1} = u(x-h)$ ,  $U_{j+1} = u(x+h)$  e  $U_j = u(x)$ , para j = 1, ..., m.

Assim, para entendermos um pouco melhor essa expressão, consideremos j=1, a fim de determinar  $U_1$ . Isto é, tomando j=1, temos que

$$\frac{1}{h^2}(U_0 - 2U_1 + U_2) = f(x_1).$$

No entanto, o termo  $U_1$  também aparecerá na fórmula (5.9) quando fizermos j=2, pois:

$$\frac{1}{h^2}(U_1 - 2U_2 + U_3) = f(x_2).$$

Em outras palavras, o termo  $U_1$  aparecerá em duas equações. O mesmo vale para o termo  $U_m$ , que aparecerá na última e penúltima equação associada ao índice j, ou seja,

$$\frac{1}{h^2}(U_{m-2} - 2U_{m-1} + U_m) = f(x_{m-1}) \qquad (j = m - 1)$$

е

$$\frac{1}{h^2}(U_{m-1} - 2U_m + U_{m+1}) = f(x_m) \qquad (j = m).$$

Já os termos intermediários aparecem em três equações, por exemplo, o termo  $U_2$  aparece em:

$$\frac{1}{h^2}(U_0 - 2U_1 + U_2) = f(x_1) \qquad (j = 1),$$

$$\frac{1}{h^2}(U_1 - 2U_2 + U_3) = f(x_2) \qquad (j = 2)$$

е

$$\frac{1}{h^2}(U_2 - 2U_3 + U_4) = f(x_3) \qquad (j = 3).$$

Desse modo, para obter todos os  $U_j$  devemos resolver um sistema linear AU=F, sendo

$$A = \frac{1}{h^2} \begin{pmatrix} -2 & 1 & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ \vdots \\ U_{m-1} \\ U_m \end{pmatrix} \text{ e } F = \begin{pmatrix} f(x_1) - \frac{a}{h^2} \\ f(x_2) \\ f(x_3) \\ \vdots \\ f(x_{m-1}) \\ f(x_m) - \frac{b}{h^2} \end{pmatrix}.$$

#### Referências

Boyce, W.E. and DiPrima, R.C., 2009. *Elementary differential equations and boundary value problems*. John Wiley & Sons.

Burden, R.L. and Faires, J.D., 2003. Análise numérica. Thomson.

Conte, S., 1965. Elementary numerical analysis. McGraw Hill.

Guidorizzi, H.L., 2001. Um curso de cálculo, vol. 1. LTC.

Heun, K., 1900. Neue methoden zur approximativen integration der differentialgleichungen einer unabhängigen. *Veranderlichen. z. math. phys*, 45, pp.23–38.

Kaplan, W., 1972. Cálculo avançado, vol 1. Universidade de São Paulo.

LeVeque, R.J., 1955. Finite difference methods for ordinary and partial differential equations. Siam Philadelphia.

Ruggiero, M.A.G. and Lopes, V.L.R., 2000. *Cálculo numérico: Aspectos teóricos e computacionais*. Pearson Universidades.

# **Índice Remissivo**

algoritmo condicionalmente estável, 6 algoritmo estável, 5 algoritmo instável, 6 arredondamento, 4 arredondamento para baixo, 4 arredondamento para cima, 4
base do sistema, 2
condições de fronteira, 28 crescimento exponencial do erro, 6 crescimento linear do erro, 6
diferença finita centrada, 27 diferença finita centrada de segunda ordem, 28 diferença finita à direita, 27 diferença finita à esquerda, 27
equação de advecção, 26 equação de difusão, 26 equação de onda, 26 Equação Diferencial Parcial, 26 equações diferenciais ordinárias, 14 Equações Diferenciais Parciais, 14 erro absoluto, 4 erro de arredondamento, 4 espaçamento, 15
malha, 15 método $\frac{1}{3}$ de Simpson, 25 método $\frac{3}{8}$ de Simpson, 25 método da Bissecção, 8 método da secante, 10 método de diferenças finitas, 14, 26, 27 método de Euler, 14 método de Newton, 9 método de passo múltiplo, 18 método de Runge-Kutta, 14, 16 método de Simpson, 19, 21, 22

método de Simpson composto, 24

```
método do trapézio, 19, 20
método do trapézio composto, 20
método iterativo do ponto fixo, 13
métodos de passo único, 17
ordem do erro, 15
ponto fixo, 11
ponto flutuante, 5
problema de valor de contorno, 14, 28
problema de valor inicial, 14
problema integral elíptica, 19
quadradura numérica, 19
regra de Simpson, 22
regra do trapézio, 20
sistema decimal, 2
Teorema de Weierstrass, 12
Teorema do Valor Intermediário, 8
truncamento, 4
zero da função, 8
```