Métodos para a solução numérica de equações diferenciais ordinárias a valores iniciais

Notas de aula em construção

Texto original de 2014

Alexandre Megiorin Roma
roma@ime.usp.br
Joyce da Silva Bevilacqua
joyce@ime.usp.br

Departamento de Matemática Aplicada
Instituto de Matemática e Estatística
Universidade de São Paulo

Rudimar Luiz Nós
rudimarnos@utfpr.edu.br

Departamento Acadêmico de Matemática
Universidade Tecnológica Federal do Paraná

Edição de 2019

Pedro da Silva Peixoto
pedrosp@ime.usp.br
Departamento de Matemática Aplicada
Instituto de Matemática e Estatística
Universidade de São Paulo

"... que a importância de uma coisa não se mede com fita métrica nem com balanças nem com barômetros etc. Que a importância de uma coisa há que ser medida pelo encantamento que a coisa produza em nós."

Manoel de Barros.
Memórias Inventadas, 2006.

Agradecimentos

Aos monitores e aos alunos da disciplina *MAP5725 - Tratamento Numérico de Equações Diferenciais*, disciplina anualmente ministrada dentro do Programa de Verão do IME-USP para a qual tais notas foram escritas, por seu contínuo questionamento, por suas críticas e por sua constante demanda por exemplos e por exercícios resolvidos sem os quais o texto não estaria completo. À CAPES pela bolsa de pós-doutorado para um dos autores durante o período em que o texto foi reorganizado na versão de 2014.

Conteúdo

Prefácio **1**

1 O Problema de Cauchy **3**
1.1 Definição do Problema de Cauchy 5
1.2 Existência e unicidade de solução 6
1.3 Solução do Problema de Cauchy 7
1.3.1 Exercícios 9
1.4 Discretização do Problema de Cauchy 9
1.4.1 Exercícios 15
1.5 Suplemento teórico 15
1.5.1 Fórmula de Taylor de uma função de uma variável 17
1.5.2 Fórmula de Taylor de uma função de duas variáveis 18
1.6 Exercícios resolvidos 18

2 Métodos de passo único **23**
2.1 Erro de discretização local 23
2.2 Consistência 25
2.2.1 Exercícios 27
2.3 Erro de discretização global 27
2.4 Convergência 27
2.5 Expansão do erro de discretização global 30
2.5.1 Estimativa do erro de discretização global 31
2.5.2 Estimativa da ordem de convergência 31
2.5.3 Depuração do código computacional 32
2.5.4 Exercícios 34
2.6 Suplemento teórico 35
2.7 Exercícios resolvidos 36

3 Métodos de passo único de altas ordens **41**
3.1 Métodos da Série de Taylor 41
3.2 Métodos de Runge-Kutta explícitos 42
3.2.1 Métodos de Runge-Kutta explícitos de 2-estágios 43
3.2.2 Métodos de Runge-Kutta explícitos de R-estágios 44
3.2.3 Métodos de Runge-Kutta de ordens mais elevadas 46
3.2.4 Exercícios 48
3.3 Controle automático do passo de integração 48
3.3.1 Exercícios 51
3.4 Suplemento teórico 52
3.5 Exercícios resolvidos 52

4 Estabilidade absoluta dos métodos de passo único **61**
4.1 Estabilidade absoluta 62
4.1.1 Exercícios 64
4.2 Suplemento teórico 66
4.2.1 Instabilidade inerente 66
4.3 Exercícios resolvidos 66

5 Métodos de passo múltiplo lineares **71**
5.1 Caracterização dos métodos de passo múltiplo 71
5.2 Dedução de métodos de passo múltiplo 72
5.3 Erro de discretização local 73
5.4 Consistência 75
5.5 Inicialização 77
5.6 Convergência 79
5.6.1 Exercícios 81
5.7 Exercícios resolvidos 81

6 Zero-estabilidade e convergência dos métodos de passo múltiplo lineares **87**
6.1 Equações de diferenças lineares 87
6.1.1 Exercícios 93
6.2 Exemplo de divergência 94
6.2.1 Consistência e divergência 94
6.2.2 Exercícios 97
6.2.3 Relação com as raízes do polinômio característico 97
6.3 Primeiro e segundo polinômios característicos 99
6.4 Zero-estabilidade 100
6.5 Ordem de convergência 103
6.5.1 Exercícios 104
6.6 Erro de discretização global 105
6.7 Estabilidade absoluta 107
6.7.1 Exercícios 110
6.8 Exercícios resolvidos 111

7 Métodos preditores-corretores **123**
7.0.1 Exercícios 124
7.1 Erro de discretização local 126
7.2 Estratégia de Milne 129
7.3 Estabilidade absoluta 130
7.3.1 Exercícios 133
7.4 Controle do passo de integração 133
7.4.1 Exercícios 134
7.5 Suplemento teórico 135
7.6 Exercícios resolvidos 135

A Exercícios complementares **139**

B Exercícios computacionais **143**

Índice **148**

Prefácio

Por intermédio de equações diferenciais ordinárias, é possível construir relações entre grandezas e suas respectivas variações em relação a um parâmetro independente (e.g. o tempo). Obter tais relações é uma tarefa de modelagem matemática e está fora do escopo destas notas. Aqui, parte-se diretamente de modelos matemáticos já estabelecidos, representados por equações diferenciais ordinárias de primeira ordem com valor inicial. A partir das equações discretizadas, estudam-se os erros de discretização local e global, a consistência de tais discretizações com o modelo matemático, sua estabilidade e a convergência da solução numérica para a solução exata do problema colocado.

Pressupõe-se que o aluno saiba programar em alguma linguagem de alto nível aprendida numa disciplina introdutória de programação científica e que tenha cursado ao menos duas disciplinas de cálculo diferencial e integral. Intimidade com equações diferenciais ordinárias é desejável mas não essencial para o aprendizado. Para uma "leitura suave" dos temas abordados, recomendam-se fortemente revisões prévias de alguns resultados teóricos tais como o desenvolvimento em série de Taylor de uma função de uma e de várias variáveis e de teoremas primeiros como o do confronto e o do valor médio [12]. Os conceitos de limite e de convergência são fundamentais. Há uma grande influência na maneira de como se conduziu o fluxo do texto de excelentes livros na temática abordada, como o de Lambert [16], e de outros, de caráter mais geral, como os de Burden e Faires [3], Stoer e Bulirsch [24], Schwarz [21], dentre muitos outros.

Este texto compreende grande parte das notas de aula revisadas da disciplina *Tratamento Numérico de Equações Diferenciais*, oferecida pelo Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP) em seus programas de Verão há mais de dez anos. Embora tal disciplina tenha como público alvo candidatos ao programa de pós-graduação em Matemática Aplicada, não é incomum alunos de iniciação científica cursarem a disciplina devido à abordagem introdutória dada ao tema. Além disso, ao longo dos anos, o texto original mostrou-se um excelente material de apoio a disciplinas de graduação com foco em métodos numéricos para equações diferenciais ordinárias, oferecidas no IME-USP a alunos do Bacharelado em Matemática Aplicada e Matemática Aplicada e Computacional. Esperam-se resultados ainda melhores com o presente texto.

São Paulo, 30 de setembro de 2014.

Alexandre Megiorin Roma
Joyce da Silva Bevilacqua
Rudimar Luiz Nós

Capítulo 1

O Problema de Cauchy

O corpo humano utiliza a glicose como principal fonte de energia. A glicose é um tipo de açúcar cuja concentração sanguínea deve estar entre 70 e 110 mg/dl em indivíduos normais. A insulina e o glucagon, hormônios produzidos no pâncreas, são liberados sempre que a concentração de glicose aumenta, como ocorre após a ingestão de alimentos.

Quando os níveis de glicose estão acima de 110 ml/dl (hiperglicemia) por um período prolongado, o indivíduo é diagnosticado com *diabetes mellitus*, doença do sistema regulatório da glicose-insulina. Os sintomas do diabetes são sede excessiva (polidipsia), exacerbada produção urinária (poliuria), desidratação, cansaço e perda de peso. O diabetes é classificado em dois tipos: o Tipo I ou juvenil, que é insulino-dependente, e o Tipo II, mais comum em adultos, tratado em casos mais críticos com drogas ou insulina. Independentemente do tipo, o controle da concentração da glicose no sangue é imprescindível para evitar, ou no mínimo retardar, a evolução da doença a qual compromete a microcirculação podendo ocorrer retinopatia, neuropatias e nefropatia, entre outras complicações.

Estudos mostram que a liberação da insulina pelo pâncreas ocorre em diferentes escalas de tempo. As taxas mais rápidas duram em torno de dezenas de segundos e estão vinculadas à concentração de íons de cálcio $\left(Ca^2+\right)$. As taxas médias ocorrem entre 5 e 15 minutos e as liberações lentas entre 50 e 120 minutos. O diagnóstico do diabetes é feito pelo Teste de Tolerância a Glicose (GTT). Neste, o indivíduo ingere em jejum uma dose elevada de glicose e por um período de três a cinco horas colhem-se amostras sucessivas de sangue as quais são utilizadas para avaliar a variação da concentração de glicose no tempo.

Vários modelos matemáticos foram desenvolvidos para representar o mecanismo do sistema regulatório glicose-insulina. Segundo Bergman [1, 19], o modelo que contém o menor número de parâmetros, denominado *modelo mínimo*, é representado pelo sistema de equações diferenciais ordinárias

$$\begin{cases} \dfrac{d}{dt}G(t) = -\left[b_1 + X(t)\right]G(t) + b_1 G_b \\ \dfrac{d}{dt}X(t) = -b_2 X(t) + b_3\left[I(t) - I_b\right] \\ \dfrac{d}{dt}I(t) = b_4\left[G(t) - b_5\right]^+ t - b_6\left[I(t) - I_b\right] \end{cases}, \tag{1.0.1}$$

com condições iniciais $G(0) = b_0$, $X(0) = 0$ e $I(0) = b_7 + I_b$, nas quais $G(t)$ e $I(t)$ representam a concentração de glicose e de insulina no sangue, respectivamente,

$X(t)$ é uma função auxiliar que modela o tempo de retardo do consumo da insulina em relação ao consumo de glicose e

$$[G(t) - b_5]^+ = \begin{cases} G(t) - b_5, \text{ se } G(t) > b_5 \\ 0, \text{ caso contrário} \end{cases} .$$

Os parâmetros b_i, $i = 1, \ldots, 7$, são constantes positivas asim como os valores basais de glicose G_b de insulina I_b.

Discutir a modelagem matemática deste processo está fora do escopo deste texto. No caso do mecanismo regulatória glicose-insulina, modelos mais realistas devem incluir as variações das taxas de produção de insulina e do consumo de glicose. O modelo (1.0.1) é aceito como uma boa aproximação para a área médica. Contudo, para uma primeira abordagem, é possível utilizar modelos ainda mais simples. Sejam $G(t)$ e $I(t)$ funções contínuas e definidas num intervalo $[0, T_f]$ que representam a concentração de glicose e de insulina no sangue, respectivamente. O intervalo $[0, T_f]$ representa o tempo de duração do teste que tem inínicio em $t = 0$, após 12 horas de jejum, com a tomada de medidas $G(0)$ e $I(0)$ as quais correspondem a valores de equilíbrio para o paciente. Uma dose de glicose é administrada no paciente e os valores $G(t)$ e $I(t)$ são medidos em vários momentos do tempo até o término do teste em $t = T_f$, que dura aproximadamente 5 horas. As dinâmicas de $G(t)$ e $I(t)$ podem ser representadas pelo sistema de equações diferenciais ordinárias

$$\frac{d}{dt}G(t) = -k_1 G(t) - k_2 I(t) \tag{1.0.2}$$

$$\frac{d}{dt}I(t) = +k_3 G(t) - k_4 I(t) \tag{1.0.3}$$

com condição inicial $G(0) = G_0$ e $I(0) = I_0$, $t \in [0, T_f]$, onde $k_i \geq 0$, $1 \leq i \leq 4$ são constantes positivas.

Quando um aumento na concentração desvia a glicose de seu nível de equilíbrio, parte dela é absorvida pelos tecidos e parte é armazenada no fígado, forçando assim um decréscimo em sua concentração no sangue. Similarmente, um aumento na concentração de insulina também acarreta um decréscimo nos níveis da glicose sanguínea pois o hormônio facilita seu armazenamento nos tecidos e no fígado. Este processo é descrito pela primeira equação, (1.0.2). Por outro lado, um aumento na concentração de glicose faz com que mais insulina seja secretada, aumentando a concentração desta na corrente sanguínea. O organismo, ao identificar este aumento na concentração hormonal, tende a fazê-la baixar novamente. Este mecanismo de autorregulação é descrito pela segunda equação, (1.0.3).

Na prática, elimina-se a função $I(t)$ do modelo matemático derivando-se (1.0.2) e substituindo-se o resultado em (1.0.3). Chega-se assim à equação diferencial ordinária linear de segunda ordem homogênea

$$\frac{d^2}{dt^2}G(t) + 2\alpha\frac{d}{dt}G(t) + \omega^2 G(t) = 0, \tag{1.0.4}$$

onde $\alpha = \dfrac{k_1 + k_4}{2}$ e $\omega^2 = k_1 k_4 + k_2 k_3$ são parâmetros a serem determinados com o auxílio das medidas efetuadas no teste GTT. Neste caso, as condições iniciais passam a ser dadas por $G(0) = G_0$ e $\dfrac{d}{dt}G(0) = \dot{G}_0$, onde $G(0)$ e $\dfrac{d}{dt}G(0)$ são, respectivamente, a concentração de glicose no início do teste e a velocidade com que esta concentração se modifica.

1.1 Definição do Problema de Cauchy

Um problema de valor inicial (p.v.i.) definido por uma equação diferencial ordinária de primeira ordem com condição inicial é chamado *Problema de Cauchy* [5, 10, 16, 22].

Definição 1.1 (Problema de Cauchy)**.**

$$\begin{cases} \dfrac{d}{dt}y(t) = f(t, y(t)), \quad t \in [a, b], \\ \\ y(t_0) = y(a) = y_0 \end{cases} . \tag{1.1.5}$$

Considerando f uma função contínua em ambas variáveis ($f \in C([a, b] \times \mathbb{R}^n, \mathbb{R}^n)$), desejamos encontrar soluções (clássicas) do p.v.i., que são funções continuamente diferenciáveis no tempo ($y \in C^1([a, b], \mathbb{R}^n)$). Para abreviar a notação, consideramos $y'(t) = \dot{y}(t) = \dfrac{d}{dt}y(t)$.

No modelo glicose-insulina (1.0.2)-(1.0.3),

$$y(t) = \begin{bmatrix} G(t) \\ I(t) \end{bmatrix},$$

$$f(t, y(t)) = f(t, G(t), I(t)) = \begin{bmatrix} f_1(t, G(t), I(t)) \\ f_2(t, G(t), I(t)) \end{bmatrix} = \begin{bmatrix} -k_1 G(t) - k_2 I(t) \\ +k_3 G(t) - k_4 I(t) \end{bmatrix}.$$

Além dos modelos descritos por sistemas de equações diferenciais ordinárias de primeira ordem, um equação diferencial ordinária de ordem $m > 1$ pode ser reescrita como um sistema de m equações diferenciais ordinárias de primeira ordem quando tal equação puder ser colocada na forma

$$y^{(m)}(t) = f\big(t, y(t), y^{(1)}(t), y^{(2)}(t), \ldots, y^{(m-1)}(t)\big)$$

com condição inicial dada por $\big(y(t_0), y^{(1)}(t_0), y^{(2)}(t_0), \ldots, y^{(m-1)}(t_0)\big)$. Para isto, faz-se a substituição de variáveis definida por

$$\begin{aligned} x_1(t) &= y(t) \\ x_1'(t) &= x_2(t) = y'(t) \\ x_2'(t) &= x_3(t) = y''(t) \\ &\vdots \\ x_m'(t) &= f\big(x_1, x_2, \ldots, x_{m-1}\big) \end{aligned}$$

e colocam-se $x_1(t_0) = y(t_0)$, $x_2(t_0) = y_2^{(1)}(t_0)$, $\ldots$ $x_m(t_0) = y^{(m-1)}(t_0)$. Como se vê, nas condições anteriores, a definição (1.1.5) é válida para um sistema de equações.

Da teoria de equações diferenciais [2, 8, 22] sabe-se que para uma equação diferencial ordinária de ordem m ou para um sistema de m equações diferenciais ordinárias de primeira ordem são necessárias exatamente m condições para garantir a unicidade de solução. No caso do Problema de Cauchy para sistemas de equações diferenciais, são necessárias m condições iniciais.

1.2 Existência e unicidade de solução

As condições para a existência e unicidade de solução do Problema de Cauchy (1.1.5) são dadas pelo Teorema (1.1), abaixo. Entre tais condições, encontra-se o conceito *condição de Lipschitz*.

Definição 1.2 (Condição de Lipschitz)**.** *Uma função $f(t,y)$ satisfaz a Condição de Lipschitz na variável y, $(t,y) \in \Omega \subset \mathbb{R} \times \mathbb{R}^n$, se e somente se existir uma constante $L > 0$ tal que*

$$||f(t,y_1) - f(t,y_2)|| \leq L||y_1 - y_2|| \tag{1.2.6}$$

para quaisquer par de pontos (t,y_1) e (t,y_2) em Ω e $\|\cdot\|$ é uma norma do $\mathbb{R}^n$.

Teorema 1.1 (Existência e unicidade)**.** *Seja o problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt}y(t) = f(t,y(t)) \\ \\ y(a) = y_0 \end{cases}, \tag{1.2.7}$$

onde $f : [a,b] \times \mathbb{R}^n \to \mathbb{R}^n$ é contínua em t e satisfaz a Condição de Lipschitz em y. Nestas condições, existe uma única solução para o p.v.i. (1.2.7).

A demonstração do teorema de existência e unicidade de solução do Problema de Cauchy pode ser encontrada em Butcher [5]. Note que verificar as condições que garantem a existência e unicidade de solução não implica em encontrar a solução exata propriamente dita. Pode-se, em alguns casos, mostrar a desigualdade de Lipschitz (1.2.6) empregando-se do Teorema do Valor Médio [14], mas isto apenas determina um valor para L, a constante de Lipschitz, como ilustra o Exemplo (1.1).

Exemplo 1.1 (Constante de Lipschitz)**.** *Sejam*

$$\begin{aligned} \Omega &= \{(t,y);\ 1 \leq t \leq 2,\ y \in \mathbb{R}\} e \\ f(t,y) &= t|y|. \end{aligned}$$

Assim:

$$\begin{aligned} ||f(t,y_1) - f(t,y_2)|| &= ||\,t|y_1| - t|y_2|\,|| \\ &= |t|\ ||\,|y_1| - |y_2|\,|| \\ &\leq 2||y_1 - y_2||. \end{aligned}$$

Portanto, $f(t,y)$ satisfaz a Condição de Lipschitz com constante $L = 2$.

Quando $f\,(t,y(t))$ é contínua, o Problema de Cauchy (1.1.5) também pode ser colocado em uma *forma integral*, ou seja,

$$y(t) - y(t_0) = \int_{t_0}^{t} f(s,y(s))ds. \tag{1.2.8}$$

A forma (1.2.8) é obtida aplicando-se o Teorema Fundamental do Cálculo a (1.1.5).

Outro resultado importante, que pode ser encontrados em livros clássicos de equações diferenciais (e.g. [10]), diz respeito à dependência contínua da solução em relação a condição inicial. Mais precisamente, considerando o problema de valor inicial,

$$\begin{cases} y'(t) = f(t, y(t)), \quad t \in [a, b] \\ y(a) = s \end{cases}, \tag{1.2.9}$$

sob as mesmas hipóteses para f do teorema de existência e unicidade, e portanto com solução única denotada por $y(t; s)$, temos que

$$\|y(t; s_1) - y(t, s_2)\| \leq e^{L(t-a)}\|s_1 - s_2\|, \tag{1.2.10}$$

para $t \in [a, b]$. Consequentemente, as soluções dos p.v.i., únicas para cada s, variam continuamente com as condições iniciais (s).

1.3 Solução do Problema de Cauchy

Em algumas situações especiais é possível determinar a solução exata da equação diferencial sendo considerada. Existem várias técnicas para tal como a *separação de variáveis* e o uso de *fatores de integração*, dentre outras. Contudo são pouquíssimas as equações para as quais se sabe que tais técnicas podem ser empregadas.

Exemplo 1.2. *Solucione o p.v.i.*

$$\begin{cases} \dfrac{d}{dt}y(t) = 3ty(t), \quad t > 0 \\ y(0) = 1 \end{cases}. \tag{1.3.11}$$

Em (1.3.11), tem-se uma equação diferencial ordinária linear, de primeira ordem, homogênea.

$$\begin{aligned} \frac{d}{dt}y(t) &= 3ty(t) \\ \frac{1}{y(t)}\frac{d}{dt}y(t) &= 3t \\ \int \frac{1}{y(t)}\frac{d}{dt}y(t)dt &= \int 3tdt \\ ln|y(t)| &= \frac{3}{2}t^2 + C_1 \\ y(t) &= Ce^{\frac{3}{2}t^2} \end{aligned} \tag{1.3.12}$$

Usando a condição inicial em (1.3.12), tem-se que

$$y(0) = C \Rightarrow C = 1.$$

Logo,

$$\boxed{y(t) = e^{\frac{3}{2}t^2}.} \tag{1.3.13}$$

Exemplo 1.3. *Analise a existência e unicidade do p.v.i.*

$$\begin{cases} \dfrac{d}{dt}y(t) = 3(y(t))^{2/3}, \quad t > 0 \\ \\ y(0) = 0 \end{cases} . \tag{1.3.14}$$

Primeiro note que $y(t) = 0$ é solução do p.v.i. Além disso, o p.v.i. (1.3.14) pode ser resolvido pelo método da separação de variáveis para obtermos que $y(t) = t^3$ também é solução do problema. Verifique que $f(t, y) = 3(y(t))^{2/3}$ não é Lipschitz na variável y em $t = 0$, e portanto não podemos usar o Teorema da Existência e Unicidade 1.1.

Exemplo 1.4. *Solucione o p.v.i.*

$$\begin{cases} \dfrac{d}{dt}y(t) = 2y(t) + e^{3t}, \quad t > 0 \\ \\ y(0) = 1 \end{cases} . \tag{1.3.15}$$

Em (1.3.15), tem-se uma equação diferencial ordinária linear, de primeira ordem, não homogênea.
Fator integrante: $e^{\int -2dt} = e^{-2t}$

$$\begin{aligned} e^{-2t}\left[\frac{d}{dt}y(t) - 2y(t)\right] &= e^{-2t}e^{3t} \\ e^{-2t}\frac{d}{dt}y(t) - 2e^{-2t}y(t) &= e^{t} \\ \frac{d}{dt}\left[e^{-2t}y(t)\right] &= e^{t} \\ \int \frac{d}{dt}\left[e^{-2t}y(t)\right]dt &= \int e^{t}dt \\ e^{-2t}y(t) &= e^{t} + C \\ y(t) &= e^{3t} + Ce^{2t} \end{aligned} \tag{1.3.16}$$

Usando a condição inicial em (1.3.16), tem-se que

$$y(0) = 1 + C \Rightarrow C = 0.$$

Logo,

$$\boxed{y(t) = e^{3t}.} \tag{1.3.17}$$

Exemplo 1.5. *Reescreva o p.v.i.*

$$\begin{cases} y''(t) + 2y'(t) - 3y(t) = 0, \quad t > 0 \\ y(0) = k_1 \\ y'(0) = k_2 \end{cases} \tag{1.3.18}$$

como um sistema de duas equações diferenciais ordinárias de primeira ordem.
Convenções:

$$y(t) = y_1(t); \quad y'(t) = y_1'(t) = y_2(t); \quad y''(t) = y_2'(t). \tag{1.3.19}$$

Substituindo (1.3.19) em (1.3.18), tem-se que:

$$\begin{cases} y_1'(t) = y_2(t) \\ y_2'(t) = -2y_2(t) + 3y_1(t) \\ y_1(0) = k_1 \\ y_2(0) = k_2 \end{cases} .$$

1.3.1 Exercícios

Exercício 1.1. *Verifique que a solução (1.3.13) satisfaz de fato o p.v.i. (1.3.11).*

Exercício 1.2. *Verifique que a solução (1.3.17) satisfaz de fato o p.v.i. (1.3.15).*

Exercício 1.3. *Obtenha (1.0.4) a partir de (1.0.2) e (1.0.3).*

Exercício 1.4. *Reescreva o modelo (1.0.4) como um sistema de equações diferenciais ordinárias de primeira ordem.*
Sugestão: Use $y_1(t) = g(t)$ *e* $y_2(t) = \dfrac{d}{dt}g(t)$.

1.4 Discretização do Problema de Cauchy

Com as técnicas matemáticas conhecidas, pode-se determinar a solução exata de um pequeno número de equações diferenciais. Para a esmagadora maioria é possível apenas calcular soluções aproximadas. Contudo, estas não podem ser determinadas em todos os pontos do domínio. Isto é fácil de entender quando se pensa nas restrições impostas pelo uso do computador, tais como aritmética de ponto flutuante e limitações de tempo e memória.

Ao solucionar numericamente o Problema de Cauchy (1.1.5), objetiva-se determinar uma aproximação da solução exata em um *conjunto discreto finito de pontos* definido por

$$a = t_0,\, t_1 = t_0 + \Delta t_1,\, t_2 = t_1 + \Delta t_2, \cdots, b = t_n = t_{n-1} + \Delta t_n,$$

onde cada Δt_k, $1 \leq k \leq n$, denota o *passo de integração* necessário para ir de t_{k-1} a t_k. A menos de menção em contrário, adotar-se-á neste texto um passo de integração uniforme, isto é,

$$\Delta t_1 = \Delta t_2 = \cdots = \Delta t_n = \Delta t = h = \frac{b-a}{n},$$
$$t_k = a + k\Delta t = a + kh,$$

onde n é o número de partições uniformes efetuadas no intervalo $[a, b]$. Uma vez obtida a solução aproximada nos pontos t_k, pode-se utilizar as técnicas de interpolação polinomial, aproximação de funções por splines ou pelo método dos mínimos quadrados [3, 11, 21, 24, 25, 26] para se obter uma aproximação contínua válida em todo o intervalo de estudo.

A discretização do Problema de Cauchy não é única, podendo ser obtida pela expansão da solução exata $y(t)$ em Série de Taylor, pela integração numérica da equação integral, por extrapolação e por interpolação. Para ilustrar a primeira dessas abordagens para a discretização do Problema de Cauchy, considere sua forma diferencial (1.1.5) com $y : [a, b] \to \mathbb{R}$. Supondo que a solução $y(t)$ seja única e que tenha pelo menos m derivadas contínuas, pode-se escrever seu polinômio de Taylor com resto de Lagrange

$$y(t_k + h) = y(t_k) + hy^{(1)}(t_k) + \frac{h^2}{2!}y^{(2)}(t_k) + \cdots + \frac{h^m}{m!}y^{(m)}(\xi), \qquad (1.4.20)$$

para algum ponto ξ, $t_k < \xi < t_{k+1}$. Lembrando que $y(t_k+h) = y(t_{k+1})$, reescreve-se (1.4.20) como

$$\frac{y(t_{k+1}) - y(t_k)}{h} = y^{(1)}(t_k) + \frac{h}{2!}y^{(2)}(t_k) + \cdots + \frac{h^{m-1}}{m!}y^{(m)}(\xi)\,. \qquad (1.4.21)$$

Para valores de h suficientemente pequenos, é razoável truncar a expansão (1.4.21) no primeiro termo do lado direito da igualdade. A aproximação que se obtém define o Método de Euler:

$$\frac{y(t_{k+1}) - y(t_k)}{h} \approx y^{(1)}(t_k) = f(t_k, y(t_k))\,. \tag{1.4.22}$$

Método 1.1 (Euler Explícito)**.**

$$\begin{cases} y_0 &= y(t_0) \\ y_{k+1} &= y_k + h\,\Phi(t_k, y_k, h) \end{cases} \tag{1.4.23}$$

com $\quad t_{k+1} = t_k + h, \quad 0 \le k \le n-1, \quad h = \dfrac{b-a}{n} \quad e \quad \Phi(t_k, y_k, h) = f\,(t_k, y_k).$

No Método de Euler, calcula-se a aproximação y_{k+1} da solução no instante t_{k+1} conhecendo-se apenas o valor da aproximação no instante de tempo t_k imediatamente anterior. Métodos numéricos com esta propriedade são denominados de *métodos de passo único* ou *métodos de um passo*, traduções livres dos termos *single step methods* e *one step methods*, respectivamente. Além disso, o Método de Euler é um *método explícito*, isto é, não é necessário solucionar qualquer tipo de equação algébrica para se determinar a aproximação y_{k+1}. Tal aproximação, neste caso, pode ser calculada diretamente pela soma das parcelas do lado direito de (1.4.23).

Geometricamente, o Método de Euler (1.4.23) fornece a aproximação y_{k+1} por meio da reta tangente à curva que define a solução exata. A equação da reta tangente ao gráfico da solução exata no ponto (t_0, y_0) é dada por

$$r(t) - y_0 = (t - t_0)y'(t_0) = (t - t_0)f(t_0, y_0).$$

Quando $t = t_0 + h = t_1$, tem-se que $r(t_0 + h) = y_0 + h\,f(t_0, y_0) = y_1$. A Figura (1.1) ilustra a interpretação geométrica.

Outros exemplos que se enquadram na classe dos métodos de passo único são o Método de Euler Implícito, o Método de Euler Aprimorado e o Método do Trapézio, descritos a seguir.

Método 1.2 (Euler Implícito)**.**

$$\begin{cases} y_0 &= y(t_0) \\ y_{k+1} &= y_k + h\,\Phi(t_k, y_{k+1}, h) \end{cases} \tag{1.4.24}$$

com $\quad t_{k+1} = t_k + h, \quad 0 \le k \le n-1, \quad h = \dfrac{b-a}{n} \quad e$

$$\Phi(t_k, y_{k+1}, h) = f(t_k + h, y_{k+1}).$$

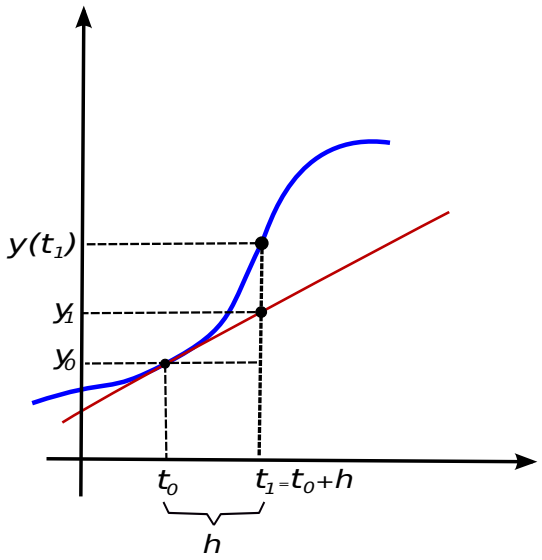


Figura 1.1: Interpretação geométrica do Método de Euler: solução exata (curva) e solução aproximada (reta).

Método 1.3 (Euler Aprimorado, do Trapézio explícito ou de Heun)**.**

$$\begin{cases} y_0 &= y(t_0) \\ y_{k+1} &= y_k + h\,\Phi(t_k, y_k, h) \end{cases} \tag{1.4.25}$$

com $t_{k+1} = t_k + h, \quad 0 \le k \le n-1, \quad h = \dfrac{b-a}{n}, \quad \Phi(t_k, y_k, h) = \dfrac{1}{2}(\kappa_1 + \kappa_2) \quad e$

$$\begin{cases} \kappa_1 = f(t_k, y_k) \\ \kappa_2 = f(t_k + h, y_k + h\,\kappa_1) \end{cases} .$$

Método 1.4 (Trapézio Implícito)**.**

$$\begin{cases} y_0 &= y(t_0) \\ y_{k+1} &= y_k + h\,\Phi(t_k, y_k, y_{k+1}, h) \end{cases} \tag{1.4.26}$$

com $t_{k+1} = t_k + h, \quad 0 \le k \le n-1, \quad h = \dfrac{b-a}{n} \quad e$

$$\Phi(t_k, y_k, y_{k+1}, h) = \frac{1}{2}\left(f(t_k, y_k) + f(t_k + h, y_{k+1})\right).$$

O Método de Euler Aprimorado (1.4.25), assim como o Método de Euler (1.4.23), é um método explícito. Já nos Métodos de Euler Implícito (1.4.24) e do Trapézio

(1.4.26), para avançar a solução no tempo, é necessário solucionar uma equação algébrica cuja incógnita é y_{k+1}. Por exemplo, se $f(t,y) = \sin(y + \pi t)$, reescreve-se o Método do Trapézio (1.4.26) como

$$y_{k+1} - \frac{h}{2}\sin(y_{k+1} + \pi t_{k+1}) = y_k + \frac{h}{2}\sin(y_k + \pi t_k). \tag{1.4.27}$$

Em (1.4.27), para cada instante de tempo é necessário empregar um método como o de Newton [3, 11, 21, 24] para determinar a raiz y_{k+1}. Um método como o Método de Euler Implícito ou o Método do Trapézio é chamado *método implícito*.

Exemplo 1.6. *Considere o Problema de Cauchy dado por*

$$\begin{cases} \dfrac{d}{dt}y(t) = e^{2t}y(t), \quad t \in [0,1], \\ \\ y(0) = 1. \end{cases} \tag{1.4.28}$$

Pedem-se:

1. *Discretize-o usando o Método de Euler;*
2. *Calcule a solução aproximada para a partição $h_1 = 1/2$;*
3. *Sabendo que a solução exata é*

$$y(t) = e^{\frac{e^{2t}-1}{2}}$$

calcule o erro absoluto, $E_A(t) = y(t) - y_2$, em $t = 1$.

<u>Solução</u>*:*

1.

$$\begin{cases} y(0) = 1 \\ y_{k+1} = y_k + h\,\Phi(t_k, y_k, h) \end{cases}$$

$t_{k+1} = t_k + h$, $0 \le k \le n-1$, $h = 1/n$ *e* $\Phi(t_k, y_k, h) = e^{2t_k}y_k$. *Aplicação:*

$$k = 0 \Rightarrow y_1 = \left(1 + he^{2t_0}\right) y_0$$

$$k = 1 \Rightarrow y_2 = \left(1 + he^{2t_1}\right) y_1$$

$$k = 2 \Rightarrow y_3 = \left(1 + he^{2t_2}\right) y_2$$

$$\vdots$$

$$y_{k+1} = \left(1 + he^{2t_k}\right) y_k$$

2. *Para $h = 1/2$, tem-se*

$$k = 0 \Rightarrow y_1 = \left(1 + \frac{1}{2}e^0\right)(1) = \frac{3}{2}$$

$$k = 1 \Rightarrow y_2 = \left(1 + \frac{1}{2}e^1\right)\left(\frac{3}{2}\right) = \frac{3}{4}(e+2) \approx 3,539$$

3. $E_A(1) = y(1) - y_2 = e^{\frac{e^2-1}{2}} - 3,539 = 24,399 - 3,539.$

Em uma outra abordagem, empregam-se na discretização do Problema de Cauchy técnicas de quadratura numérica ao invés de expansões em Série de Taylor. Essa abordagem é uma das formas naturais de se obter *métodos de passo múltiplo* (*multistep methods*). Nestes, para determinar y_{k+1}, devem-se conhecer aproximações da solução em p instantes anteriores y_k, y_{k-1}, y_{k-2}, ..., y_{k-p+1}. São por este motivo conhecidos também como *métodos de p-passos*. Como exemplo, apresentam-se a seguir os Métodos de Adams-Bashforth e de Adams-Moulton de quatro passos.

Método 1.5 (Adams-Bashforth)**.**

$$\begin{cases} y_0 &= y(t_0)\,, \quad y_p \ \textit{pré-determinado}\,, \ 1 \le p \le 3 \\ y_{k+1} &= y_k + \frac{h}{24}(55f_k - 59f_{k-1} + 37f_{k-2} - 9f_{k-3}) \end{cases} \tag{1.4.29}$$

com $3 \le k \le n-1, \quad h = \dfrac{b-a}{n}, \quad f_{k-m} = f(t_{k-m}, y_{k-m}) \quad e \quad 0 \le m \le 3.$

Método 1.6 (Adams-Moulton)**.**

$$\begin{cases} y_0 &= y(t_0)\,, \quad y_p \ \textit{pré-determinado}\,, \ 1 \le p \le 3 \\ y_{k+1} &= y_k + \frac{h}{720}(251f_{k+1} + 646f_k - 264f_{k-1} + 106f_{k-2} - 19f_{k-3}) \end{cases} \tag{1.4.30}$$

com $3 \le k \le n-1, \quad h = \dfrac{b-a}{n}, \quad f_{k-m} = f(t_{k-m}, y_{k-m}) \quad e \quad -1 \le m \le 3.$

O método (1.4.29) é apenas um dos muitos métodos pertencentes à classe dos métodos de Adams-Bashforth. Todos eles são explícitos e envolvem um número maior ou menor de passos no tempo. O mesmo se pode dizer do método (1.4.30), com a diferença de que é um método implícito.

Nos métodos de passo múltiplo são necessárias as aproximações da solução em alguns dos instantes iniciais. Por exemplo, para empregar (1.4.29) tem que se conhecer as aproximações y_1, y_2 e y_3, além da condição inicial y_0. Estas aproximações podem ser determinadas aplicando-se um método de passo único.

O Método de Simpson, outro método de passo múltiplo, pode ser obtido através da forma integral (1.2.8) do Problema de Cauchy. Considerando o intervalo de tempo $t_{k-1} \le t \le t_{k+1}$, chega-se a

$$y(t_{k+1}) - y(t_{k-1}) = \int_{t_{k-1}}^{t_{k+1}} \frac{d}{ds} y(s)\, ds = \int_{t_{k-1}}^{t_{k+1}} f(s, y(s))\, ds. \tag{1.4.31}$$

Aplicando-se quadratura numérica [3, 11, 21, 24], aproxima-se a integral em (1.4.31) pelo Método de Simpson. Dessa maneira, integra-se o polinômio de segundo grau $p_2(s)$ que interpola $f(s, y(s))$ nos pontos s_{n-1}, s_n e s_{n+1}. O polinômio $p_2(s)$ é escrito na forma de Lagrange como

$$\begin{aligned} p_2(s) &= f(t_{k+1}, y(t_{k+1}))L_{k+1}(s) \\ &\quad + f(t_k, y(t_k))L_k(s) + f(t_{k-1}, y(t_{k-1}))L_{k-1}(s), \end{aligned} \tag{1.4.32}$$

onde $L_j(s)$ são os polinômios de Lagrange

$$L_j(s) \quad = \quad \Pi_{i=k-1}^{k+1} \frac{s-t_i}{t_j-t_i}\,, \quad i \neq j\,, \tag{1.4.33}$$

para $j = k-1$, k e $k+1$.

Substituindo-se (1.4.33) em (1.4.32) e admitindo-se um passo de integração constante no tempo, chega-se a

$$\begin{aligned} p_2(s) \quad &= \quad f(t_{k+1}, y(t_{k+1}))\, \frac{(s-t_{k-1})(s-t_k)}{2h^2} - f(t_k, y(t_k))\, \frac{(s-t_{k-1})(s-t_{k+1})}{h^2} \\ &+ \quad f(t_{k-1}, y(t_{k-1}))\, \frac{(s-t_k)(s-t_{k+1})}{2h^2}, \end{aligned} \tag{1.4.34}$$

o qual é integrado no intervalo $[t_{k-1}, t_{k+1}]$. Como resultado, tem-se o Método de Simpson

$$y(t_{k+1}) - y(t_{k-1}) \quad \approx \quad \frac{h}{3}\left[f(t_{k+1}, y(t_{k+1})) + 4f(t_k, y(t_k)) + f(t_{k-1}, y(t_{k-1}))\right], \tag{1.4.35}$$

um método de dois passos implícito.

Método 1.7 (Simpson Implícito)**.**

$$\begin{cases} \quad y_0 \quad = \quad y(t_0)\,, \quad y_1 \ \textit{pré-determinado} \\ \\ y_{k+1} \quad = \quad y_{k-1} + \frac{h}{3}(f_{k+1} + 4f_k + f_{k-1}) \end{cases} \tag{1.4.36}$$

com $\quad 1 \leq k \leq n-1, \quad h = \dfrac{b-a}{n}, \quad f_{k-m} = f(t_{k-m}, y_{k-m}) \quad e \quad -1 \leq m \leq 1.$

Exemplo 1.7. *Discretize o Problema de Cauchy*

$$\begin{cases} \dfrac{d}{dt}y(t) = \sin(y) \\ \\ y(t_0) = y_0 \end{cases}$$

usando:

1. *o Método de Euler Aprimorado (1.4.25);*

$$\begin{cases} y_0 = y(t_0) \\ y_{k+1} = y_k + \frac{h}{2}\left[\sin(y_k) + \sin\left(y_k + h\sin(y_k)\right)\right] \end{cases}$$

2. *o Método do Trapézio (1.4.26).*

$$\begin{cases} y_0 = y(t_0) \\ y_{k+1} = y_k + \frac{h}{2}\left(\sin(y_k) + \sin(y_{k+1})\right) \end{cases}$$

1.4.1 Exercícios

Exercício 1.5. *Discretize o p.v.i.* $y' = y\,, y(0) = 1$, *com* $t \in [0,1]$ *e* $y(t) \in \mathbb{R}$, *usando o Método de Euler e o Euler Aprimorado. Escreva um algoritmo (ou um programa em linguagem C [15, 20, 25] ou Fortran [26]) que forneça a solução aproximada* y_{k+1} *dados* t_0, t_{final}, n *e* $y(t_0)$. *Prepare um arquivo de saída do tipo*

$$\begin{array}{ll} t_0 & y_0 \\ t_1 & y_1 \\ \vdots & \\ t_n & y_n \end{array}$$

que será utilizado para traçar o gráfico da solução aproximada.

Exercício 1.6. *Discretize o problema do exercício anterior utilizando o Método de Adams-Bashforth (1.4.29).*

Exercício 1.7. *Empregue o Método do Trapézio*

$$\left\{ \begin{array}{rcl} y_0 & = & y(t_0) \\ y_{k+1} & = & y_k + h\,\Phi(t_k, y_k, y_{k+1}, h) \end{array} \right.$$

com $t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1, \quad h = \dfrac{b-a}{n}, \quad e$

$$\Phi(t_k, y_k, y_{k+1}, h) = \frac{1}{2}\left(f\left(t_k, y_k\right) + f\left(t_k + h, y_{k+1}\right)\right)$$

para discretizar o modelo biológico presa-predador definido por

$$\left\{ \begin{array}{rcl} \dot{x} & = & 1,2x - x^2 - \dfrac{xy}{x+0,2} \\ \dot{y} & = & \dfrac{1,5xy}{x+0,2} - y \end{array} \right. .$$

1.5 Suplemento teórico

Apresentamos aqui alguns resultados clássicos de Cálculo Diferencial e Integral e Equações Diferenciais facilmente encontrados na maioria dos livros texto do assunto (e.g. [12, 22, 10]).

Teorema 1.2 (Teorema do Valor Médio)**.** *Se* $f(t)$ *for uma função contínua em* $[a,b]$ *e derivável em* $]a,b[$ *então existirá pelo menos um* c *em* $]a,b[$ *tal que*

$$f(b) - f(a) = f'(c)(b-a).$$

Definição 1.3 (Conjunto convexo)**.** *Um conjunto* $D \subset \mathbb{R}^2$ *é convexo se e somente se para quaisquer pontos* (t_1, y_1) *e* (t_2, y_2) *pertencentes a* D *o ponto*

$$\left((1-\lambda)\,t_1 + \lambda t_2, (1-\lambda)\,y_1 + \lambda y_2\right)$$

também pertence a D *para todo* $\lambda \in [0,1]$.

Teorema 1.3. *Se $f(t,y)$ for definida em um conjunto convexo $D \subset \mathbb{R}^2$ e se existir uma constante $L > 0$ tal que*

$$\left|\frac{\partial}{\partial y} f(t,y)\right| \leq L \ \forall \ (t,y) \in D,$$

então $f(t,y)$ satisfaz a Condição de Lipschitz em D na variável y com constante de Lipschitz L.

Definição 1.4. *O problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt} y(t) = f(t, y(t)), \quad t \in [a,b] \\ \\ y(a) = \alpha \end{cases} \tag{1.5.37}$$

é dito bem posto se e somente se:

1. *Existir uma única solução $y(t)$;*
2. *$\forall \ \epsilon > 0 \ \exists \ \kappa(\epsilon) > 0$ com a propriedade de que, sempre que $|\epsilon_0| < \epsilon$ e $\delta(t)$ for contínua com $|\delta(t)| < \epsilon$ em $[a,b]$, a solução única $z(t)$ para o problema*

$$\begin{cases} \dfrac{d}{dt} z(t) = f(t, z(t)) + \delta(t), \quad t \in [a,b] \\ \\ z(a) = \alpha + \epsilon_0 \end{cases} \tag{1.5.38}$$

existe com

$$|z(t) - y(t)| < \kappa(\epsilon)\epsilon, \ \forall \ t \in [a,b].$$

O problema de valor inicial (1.5.38) é chamado *problema perturbado* associado ao problema de valor inicial original (1.5.37).

Teorema 1.4 (Problema de valor inicial bem posto)**.** *Seja $D = \{(t,y) | \, a \leq t \leq b$ e $y \in (-\infty, \infty)\}$. Se $f(t,y)$ for uma função contínua que satisfaz a Condição de Lipschitz na variável y em D então o problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt} y(t) = f(t, y(t)) \\ \\ y(a) = \alpha \end{cases}$$

é bem posto.

Teorema 1.5 (Teorema de Picard)**.** *Se $f(t,y)$ for contínua e de Lipschitz em $\Omega = I_a \times B_b$, onde $I_a = \{t; |t-t_0| \leq a\}$ e $B_b = \{y; |y-y_0| \leq b\}$, e se satisfizer $||f(t,y)|| \leq M$ em Ω então existe uma única solução de*

$$\begin{cases} \dot{y}(t) = \dfrac{d}{dt}y(t) = f(t, y(t)) \\ \\ y(t_0) = y_0 \end{cases}$$

em I_α, onde $\alpha = min\{a, \frac{b}{M}\}$.

A demonstração do Teorema de Picard pode ser encontrada em Doering [8] e em Sotomayor [22].

Teorema 1.6 (Teorema Fundamental do Cálculo)**.** *Se $f(x)$ for contínua em $[a,b]$ então*

$$\int_a^b f(x)dx = F(b) - F(a),$$

onde $F(x)$ é uma primitiva de $f(x)$, isto é, $F(x)$ é uma função tal que $F^{'}(x) = f(x)$.

1.5.1 Fórmula de Taylor de uma função de uma variável

Seja f derivável até ordem n no intervalo I e seja $t_0 \in I$. O polinômio

$$p_n(t) = \sum_{k=0}^{n} \frac{f^{(k)}(t_0)}{k!}(t-t_0)^k \tag{1.5.39}$$

denomina-se polinômio de Taylor de ordem n, de f, em torno de t_0. Se f tiver $n+1$ derivadas então

$$f(t) = p_n(t) + R_n(t), \tag{1.5.40}$$

onde $R_n(t)$ é o resto (de Lagrange) o qual tem a forma

$$R_n(t) = f(t) - p_n(t) = \frac{f^{(n+1)}(\varepsilon)}{(n+1)!}(t-t_0)^{n+1},$$

com ε pertencente ao intervalo formado por t_0 e t.

Considerando-se $n = 0$ em (1.5.40), tem-se que

$$f(t) = f(t_0) + R_0(t) = f(t_0) + f'(\varepsilon)(t-t_0), \tag{1.5.41}$$

ou seja,

$$f(t) - f(t_0) \quad = \quad f'(\varepsilon)(t-t_0), \tag{1.5.42}$$

que é exatamente o que afirma o Teorema do Valor Médio.

1.5.2 Fórmula de Taylor de uma função de duas variáveis

Seja $f(t,y)$ uma função de classe C^{n+1} definida num conjunto aberto $\Omega \in \mathbb{R}^2$ a valores reais e sejam (t_0,y_0) e (t_0+h,y_0+k), $h \neq 0$, $k \neq 0$, pontos tais que o segmento que os une esteja completamente contido em Ω. Nestas condições,

$$\begin{aligned} f(t_0+h,y_0+k) &= f(t_0,y_0) + \\ &+ \sum_{r=1}^{n} \frac{1}{r!}\Big[\sum_{p=0}^{r} \frac{r!}{p!(r-p)!}\frac{\partial^r f}{\partial t^{(r-p)}\partial y^p}(t_0,y_0)\, h^{(r-p)}k^p\Big] + \\ &+ R_n(h,k), \end{aligned} \tag{1.5.43}$$

onde

$$R_n(h,k) = \frac{1}{(n+1)!}\sum_{p=0}^{n+1}\frac{(n+1)!}{p!(n+1-p)!}\frac{\partial^{(n+1)} f}{\partial t^{(n+1-p)}\partial y^p}(\zeta,\eta)\, h^{(n+1-p)}k^p$$

para algum $\zeta \in [t_0,t_0+h]$ e $\eta \in [y_0,y_0+k]$.

Corriqueiramente, será necessário o desenvolvimento acima para $n=2$ na teoria a ser apresentada nos capítulos a seguir. Em tal caso, a Fórmula de Taylor assume a forma

$$\begin{aligned} f(t_0+h,y_0+k) &= f(t_0,y_0) + \Big[\frac{\partial}{\partial t}f(t_0,y_0)\,h + \frac{\partial}{\partial y}f(t_0,y_0)\,k\Big] + \\ &+ \Big[\frac{\partial^2}{\partial t^2}f(t_0,y_0)\,\frac{h^2}{2!} + \frac{\partial^2}{\partial t\partial y}f(t_0,y_0)\,hk + \frac{\partial^2}{\partial y^2}f(t_0,y_0)\,\frac{k^2}{2!}\Big] + \\ &+ R_2(\zeta,\eta). \end{aligned}$$

1.6 Exercícios resolvidos

Exercício Resolvido 1.1. *Mostre que o Método de Euler é um método linear explícito e de passo único.*

Solução:

O Método de Euler pode ser reescrito como

$$y_{k+1} = y_k + h\Phi(t_k,y_k) \Rightarrow y_{k+1} = y_k + hf_k \Rightarrow \underbrace{y_{k+1} - y_k - hf_k}_{F(y_k,y_{k+1},f_k)} = 0,$$

com

$$F(y_k,y_{k+1},f_k) = y_{k+1} - y_k - hf_k.$$

O operador F, acima, define o Método de Euler. Para mostrar que o método é linear, basta mostrar que o operador F é linear. Assim sendo, deve-se verificar que

$$F(y_k+\alpha\tilde{y}_k, y_{k+1}+\alpha\tilde{y}_{k+1}, f_k+\alpha\tilde{f}_k) = F(y_k,y_{k+1},f_k) + \alpha F(\tilde{y}_k,\tilde{y}_{k+1},\tilde{f}_k).$$

$$\begin{aligned} F(y_k + \alpha\tilde{y}_k, y_{k+1} + \alpha\tilde{y}_{k+1}, f_k + \alpha\tilde{f}_k) = \\ &= y_{k+1} + \alpha\tilde{y}_{k+1} - y_k - \alpha\tilde{y}_k - hf_k - \alpha h\tilde{f}_k \\ &= y_{k+1} - y_k - hf_k + \alpha\tilde{y}_{k+1} - \alpha\tilde{y}_k - \alpha h\tilde{f}_k \\ &= y_{k+1} - y_k - hf_k + \alpha(\tilde{y}_{k+1} - \tilde{y}_k - h\tilde{f}_k) \\ &= F(y_k, y_{k+1}, f_k) + \alpha F(\tilde{y}_k, \tilde{y}_{k+1}, \tilde{f}_k). \end{aligned}$$

Portanto, o Método de Euler é linear. Tem-se ainda que:

- *O método é explícito porque Φ, sua função de discretização, não depende de y_{k+1};*
- *Ele é de passo único pois para calcular a aproximação no instante t_{k+1} são necessárias somente as informações do instante imediatamente anterior, t_k.*

Exercício Resolvido 1.2. *Seja*

$$\ddot{x}(t) + 0,12\dot{x}(t) + 2x(t) = 0 \tag{1.6.44}$$

com condições iniciais $x(0) = 1$ e $\dot{x}(0) = 0$.

(a) Reescreva (1.6.44) como um sistema de equações diferenciais ordinárias de primeira ordem;

(b) Discretize o sistema obtido no item (a) usando o Método de Euler.

Solução:

(a) Convenções adotadas:

$$y_1(t) = x(t) \Rightarrow \dot{y}_1(t) = \dot{x}(t), \tag{1.6.45}$$

$$y_2(t) = \dot{y}_1(t) = \dot{x}(t) \Rightarrow \dot{y}_2(t) = \ddot{x}(t). \tag{1.6.46}$$

Substituindo-se (1.6.45) e (1.6.46) em (1.6.44), obtém-se:

$$\dot{y}_2(t) + 0,12y_2(t) + 2y_1(t) = 0 \Rightarrow \dot{y}_2(t) = -0,12y_2(t) - 2y_1(t). \tag{1.6.47}$$

Assim, tem-se o sistema de EDOs a seguir.

$$\boxed{\begin{cases} \dot{y}_1(t) = y_2(t) \\ \dot{y}_2(t) = -0,12y_2(t) - 2y_1(t) \end{cases}} \tag{1.6.48}$$

Como

$$x(0) = 1 \Rightarrow y_1(0) = 1 \text{ e } \dot{x}(0) = 0 \Rightarrow y_2(0) = 0, \tag{1.6.49}$$

os dados iniciais do sistema (1.6.48) são

$$\boxed{\begin{cases} y_1(0) = 1 \\ y_2(0) = 0 \end{cases}}. \tag{1.6.50}$$

(b) Discretização do sistema de EDOs pelo Método de Euler.

$$\begin{bmatrix} y_1^{k+1} \\ y_2^{k+1} \end{bmatrix} = \begin{bmatrix} y_1^k \\ y_2^k \end{bmatrix} + h \begin{bmatrix} y_2^k \\ -0,12y_2^k - 2y_1^k \end{bmatrix}$$

ou

$$\begin{bmatrix} y_{1,k+1} \\ y_{2,k+1} \end{bmatrix} = \begin{bmatrix} y_{1,k} \\ y_{2,k} \end{bmatrix} + h \begin{bmatrix} y_{2,k} \\ -0,12y_{2,k} - 2y_{1,k} \end{bmatrix}.$$

Exercício Resolvido 1.3. *Obtenha o Método do Trapézio*

$$\begin{cases} y_0 = y(t_0), \\ y_{k+1} = y_k + h\Phi(t_k, y_k, y_{k+1}, h), \quad t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1, \end{cases}$$

$$\Phi(t_k, y_k, y_{k+1}, h) = \frac{1}{2}\left(f\left(t_k, y_k\right) + f\left(t_{k+1}, y_{k+1}\right)\right),$$

usando quadratura numérica partindo da forma integral do Problema de Cauchy.
Solução:

Da forma integral do Problema de Cauchy,

$$y_{k+1} = y_k + \int_{t_k}^{t_{k+1}} f(s, y)ds, \tag{1.6.51}$$

inicialmente aproxima-se $f(s, y)$ pelo polinômio interpolador de Lagrange de primeiro grau nos pontos t_k e t_{k+1}. Depois disso, integra-se esse polinômio no intervalo $[t_k, t_{k+1}]$. Para simplificar, denota-se $f(t_k, y(t_k))$ por f_k.

Pontos de interpolação: (t_k, f_k), (t_{k+1}, f_{k+1}).

$$\begin{aligned} p_1(s) &= f_k L_k(s) + f_{k+1} L_{k+1}(s) \\ &= f_k \frac{s - t_{k+1}}{t_k - t_{k+1}} + f_{k+1} \frac{s - t_k}{t_{k+1} - t_k} \\ &= -\frac{f_k}{h}(s - t_{k+1}) + \frac{f_{k+1}}{h}(s - t_k) \end{aligned}$$

$$\begin{aligned} \int_{t_k}^{t_{k+1}} p_1(s)ds &= -\frac{f_k}{h}\int_{t_k}^{t_{k+1}} (s - t_{k+1})ds + \frac{f_{k+1}}{h}\int_{t_k}^{t_{k+1}} (s - t_k)ds \\ &= -\frac{f_k}{h}\left[\frac{s^2 - 2st_{k+1}}{2}\right]_{t_k}^{t_{k+1}} + \frac{f_{k+1}}{h}\left[\frac{s^2 - 2st_k}{2}\right]_{t_k}^{t_{k+1}} \\ &= \frac{f_k}{h}\left[\frac{t_{k+1}^2 - 2t_k t_{k+1} + t_k^2}{2}\right] + \frac{f_{k+1}}{h}\left[\frac{t_{k+1}^2 - 2t_k t_{k+1} + t_k^2}{2}\right] \\ &= \frac{f_k}{2h}(t_{k+1} - t_k)^2 + \frac{f_{k+1}}{2h}(t_{k+1} - t_k)^2 \\ &= \frac{f_k}{2h}h^2 + \frac{f_{k+1}}{2h}h^2 \\ &= h\left[\frac{1}{2}\left(f_k + f_{k+1}\right)\right] \end{aligned} \tag{1.6.52}$$

Substituindo-se (1.6.52) em (1.6.51), tem-se que:

$$\begin{aligned} y_{k+1} &= y_k + h\left[\frac{1}{2}\left(f_k + f_{k+1}\right)\right] \\ &= y_k + h\underbrace{\left[\frac{1}{2}\left(f(t_k, y_k) + f(t_{k+1}, y_{k+1})\right)\right]}_{\Phi(t_k, y_k, y_{k+1}, h)} \\ &= y_k + h\Phi(t_k, y_k, y_{k+1}, h). \end{aligned}$$

Portanto, o método do Trapézio é definido por

$$y_{k+1} = y_k + h\Phi(t_k, y_k, y_{k+1}, h)$$

onde

$$\Phi(t_k, y_k, y_{k+1}, h) = \frac{1}{2}\left(f\left(t_k, y_k\right) + f\left(t_{k+1}, y_{k+1}\right)\right).$$

Note que o nome do método é herdado da técnica de quadratura empregada, o Método do Trapézio.

Exercício Resolvido 1.4. *Obtenha o Método de Simpson*

$$\begin{cases} y_0 = y(t_0), \\ y_{k+2} = y_k + \dfrac{h}{3}(f_{k+2} + 4f_{k+1} + f_k), \quad t_{k+1} = t_k + h, \quad 0 \le k \le n-1, \end{cases}$$

$$f_{k+i} = f(t_{k+i}, y_{k+i}), \quad i = 0, 1, 2,$$

usando quadratura numérica partindo da forma integral do Problema de Cauchy.
Solução:

Da forma integral do Problema de Cauchy,

$$y_{k+2} - y_k = \int_{t_k}^{t_{k+2}} f(s, y)ds. \tag{1.6.53}$$

Aproxima-se $f(s, y)$ em (1.6.53) pelo polinômio interpolador de Lagrange de grau 2 nos pontos t_k, t_{k+1} e t_{k+2}, sendo estes igualmente espaçados. Feito isso, integra-se no intervalo $[t_k, t_{k+2}]$ o polinômio obtido.

Pontos de interpolação: $\left(t_k, f_k\right)$, $\left(t_{k+1}, f_{k+1}\right)$, $\left(t_{k+2}, f_{k+2}\right)$.

$$\begin{aligned} p_2(s) &= f_k L_k(s) + f_{k+1}L_{k+1}(s) + f_{k+2}L_{k+2}(s) \\ &= f_k\frac{(s - t_{k+1})(s - t_{k+2})}{(t_k - t_{k+1})(t_k - t_{k+2})} + \\ &+ f_{k+1}\frac{(s - t_k)(s - t_{k+2})}{(t_{k+1} - t_k)(t_{k+1} - t_{k+2})} + \\ &+ f_{k+2}\frac{(s - t_k)(s - t_{k+1})}{(t_{k+2} - t_k)(t_{k+2} - t_{k+1})} \\ &= \frac{f_k}{2h^2}(s - t_{k+1})(s - t_{k+2}) + \\ &- \frac{f_{k+1}}{h^2}(s - t_k)(s - t_{k+2}) + \\ &+ \frac{f_{k+2}}{2h^2}(s - t_k)(s - t_{k+1}) \end{aligned}$$

$$\begin{aligned}\int_{t_k}^{t_{k+2}} p_2(s)ds &= \frac{f_k}{2h^2}\int_{t_k}^{t_{k+2}}(s-t_{k+1})(s-t_{k+2})ds + \\ &- \frac{f_{k+1}}{h^2}\int_{t_k}^{t_{k+2}}(s-t_k)(s-t_{k+2})ds + \\ &+ \frac{f_{k+2}}{2h^2}\int_{t_k}^{t_{k+2}}(s-t_k)(s-t_{k+1})ds \qquad (1.6.54)\end{aligned}$$

$$\begin{aligned}\int_{t_k}^{t_{k+2}} p_2(s)ds &= \frac{f_k}{2h^2}\frac{2h^3}{3} + \frac{f_{k+1}}{h^2}\frac{4h^3}{3} + \frac{f_{k+2}}{2h^2}\frac{2h^3}{3} \qquad (1.6.55)\\ &= \frac{f_k h}{3} + \frac{4f_{k+1}h}{3} + \frac{f_{k+2}h}{3} \\ &= \frac{h}{3}\left(f_k + 4f_{k+1} + f_{k+2}\right) \qquad (1.6.56)\end{aligned}$$

Observação*: A passagem de (1.6.54) para (1.6.55) não é trivial. Ela exige considerável manipulação algébrica. Você consegue propor uma substituição adequada que simplifique essa manipulação algébrica?*

Substituindo-se (1.6.56) em (1.6.53), obtém-se o Método de Simpson

$$\boxed{y_{k+2} = y_k + \frac{h}{3}\left(f_k + 4f_{k+1} + f_{k+2}\right)}.$$

Capítulo 2

Métodos de passo único

Para (1.1.5), um Problema de Cauchy bem posto com solução exata $y(t)$, um método de passo único (ou método de um passo) assume a seguinte forma.

Definição 2.1 (Método de Passo Único)**.**

$$\begin{cases} y_0 &= y(t_0) \\ y_{k+1} &= y_k + h\,\Phi(t_k, t_{k+1}, y_k, y_{k+1}, h) \end{cases} \tag{2.0.1}$$

com $t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1 \quad e \quad h = \dfrac{b-a}{n}.$

Se em (2.0.1) a função Φ não depende de nenhuma informação no instante t_{k+1} então (2.0.1) é um método de passo único explícito. Caso contrário, o método de passo único (2.0.1) é implícito. Dentre os métodos de passo único apresentados até o momento, os Métodos de Euler (1.4.23) e de Euler Aprimorado (1.4.25) são explícitos, enquanto que os Métodos de Euler Implícito (1.4.24) e do Trapézio (1.4.26) são implícitos.

Embora as definições e resultados teóricos a serem apresentados a seguir também sejam válidos para métodos implícitos, por simplicidade, eles serão colocados apenas no contexto de métodos explícitos. Será conveniente também que um método de passo único explícito seja reescrito como

$$\begin{cases} y_0 &= y(t_0) \\ \dfrac{y_{k+1} - y_k}{h} - \Phi(t_k, y_k, h) &= 0 \end{cases}. \tag{2.0.2}$$

2.1 Erro de discretização local

Definição 2.2 (Erro local)**.** *Dado um método numérico de passo único explícito, associado a um problema de valor inicial com solução única $y(t)$, o erro de discretização local, associado ao instante $t = t_k$, é definido por*

$$\alpha_k \doteq \frac{y(t_{k+1}) - y(t_k)}{h} - \Phi(t_k, y(t_k), h). \tag{2.1.3}$$

É importante destacar que na definição do erro de discretização local (2.1.3) usa-se $y(t)$, a solução exata do Problema de Cauchy, e não uma sua aproximação como é feito em (2.0.2). Além disso, dado um instante de tempo t, fixo, assumimos $t = t_k = kh + t_0$. Portanto, α_k depende implicitamente de t, t_0 e h. Omitimos essa dependência apenas para simplificar a notação.

Multiplicando-se (2.1.3) pelo passo de integração h chega-se a

$$h\,\alpha_k \doteq d_k \doteq y(t_{k+1}) - [y(t_k) + h\,\Phi(t_k, y(t_k), h)] \doteq y(t_{k+1}) - y_{k+1}. \qquad (2.1.4)$$

Em (2.1.4), a diferença $y(t_{k+1}) - y_{k+1}$, a diferença entre a solução exata e a solução aproximada, pode ser interpretada como sendo *o erro produzido em uma única aplicação do método numérico partindo-se de valores exatos no instante de tempo anterior*. Isto confere a (2.1.4) seu *caráter local* pois parte-se da solução exata no instante anterior e, por isso mesmo, nenhum erro é cometido anteriormente (note em (2.1.4), entre os colchetes, a aplicação do método considerado). Na Figura (2.1), pode-se visualizar a interpretação geométrica do erro de discretização local $d_k = h\alpha_k = y(t_{k+1}) - y_{k+1}$ quando a solução numérica é calculada pelo Método de Euler.

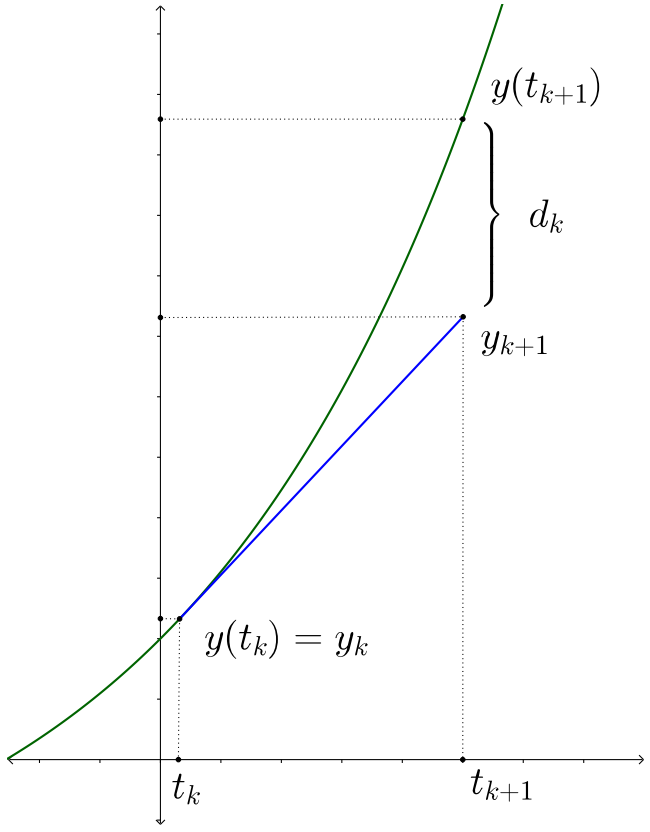


Figura 2.1: Interpretação geométrica do erro de discretização local do Método de Euler: solução exata (curva em verde) e solução aproximada (reta em azul).

2.2 Consistência

Supondo que a função de discretização Φ seja contínua como função do passo de integração h e lembrando que $y(t)$ é a solução exata do Problema de Cauchy (1.1.5), podemos calcular o limite do erro de discretização local (2.1.3) quando o passo de integração tende a zero. Porém, ao fazermos $h \to 0$, como $t_k = hk + t_0$, teríamos $t_k \to 0$. Para que a noção de erro local de discretização assintótico esteja definida em todo intervalo $t \in [a, b]$, fixamos t no intervalo e mantemos $hk = t - t_0$ fixo na convergência de α_k com $h \to 0$. Assim, formalmente $k \to \infty$ quando $h \to 0$, e $\alpha_k \to \alpha(t)$, representando o erro local de discretização no tempo fixado t. Sob essa noção de convergência, com $hk = t - t_0$ fixado, podemos deduzir que

$$\begin{aligned} \lim_{h\to 0} \alpha_k &= \lim_{h\to 0} \left[\frac{y(t_k + h) - y(t_k)}{h} - \Phi(t_k, y(t_k), h) \right] \\ &= y^{'}(t) - \Phi(t, y(t), 0) \\ &= f(t, y(t)) - \Phi(t, y(t), 0). \end{aligned}$$

À medida que o passo de integração h diminui, é razoável supor que o erro local seja cada vez menor (!) e que, no limite para h tendendo a zero, ele venha a se anular. Assim, no limite, tem-se $\Phi(t, y, 0) = f(t, y)$. Em palavras, à medida que o passo de integração tende a zero, a função de discretização deve representar cada vez melhor a equação diferencial que define o Problema de Cauchy. No limite, a função de discretização deve ser *consistente com a equação diferencial* (ou, para encurtar, *consistente*).

Definição 2.3 (Consistência)**.** *Supondo f e Φ contínuas, um método de passo único explícito é dito consistente com um problema de valor inicial bem posto se e somente se*

$$\Phi(t, y, 0) = f(t, y)$$

ou, equivalentemente, se e somente se

$$\lim_{h\to 0} \alpha_k = 0, \ \forall t \in [a, b], \tag{2.2.5}$$

com $hk = t - t_0$ fixado.

O conceito de consistência está sempre associado a um dado problema de valor inicial bem posto. Diremos que o método é consistente, de forma genérica, quando o método for consistente para qualquer problema de valor inicial bem posto.

Exemplo 2.1. *O Método de Euler Aprimorado é consistente (com qualquer problema de valor inicial bem posto):*

$$\begin{cases} y_0 = y(t_0) \\ y_{k+1} = y_k + \frac{h}{2} \left[f(t_k, y_k) + f(t_{k+1}, y_k + h\, f(t_k, y_k)) \right] \end{cases}$$

$$\begin{aligned} \Phi(t, y, h) &= \frac{1}{2} \left[f(t, y) + f(t + h, y + h f(t, y)) \right], \\ \Phi(t, y, 0) &= \frac{1}{2} \left[f(t, y) + f(t, y) \right] = f(t, y). \end{aligned} \tag{2.2.6}$$

É interessante lembrar que, por hipótese, $f(t, y)$ é contínua e Lipschitziana na variável y. Assim, $\Phi(t, y, h)$ é contínua em seu terceiro argumento pois é dada por uma soma de funções contínuas (a segunda parcela do lado direito de (2.2.6) é uma função composta de funções contínuas e, portanto, contínua).

Definição 2.4 (Ordem de consistência)**.** *Se existirem constantes positivas C, h_0 e q, independentes do tamanho do passo de integração h e do subíndice temporal k, tais que o erro de discretização local satisfaça*

$$\max_k \|\alpha_k\| \leq Ch^q, \quad 0 < h \leq h_0, \tag{2.2.7}$$

então o método numérico tem ordem de consistência q, sendo a ordem de consistência atrelada a norma utilizada $\|\cdot\|$.

A desigualdade (2.2.7) expressa o quão rapidamente o erro de discretização local vai a zero à medida que h diminui, ou seja, quão próxima a soluçõ aproximada está da solução exata.

Notação: Nas condições da Definição 2.4, escreve-se

$$\alpha_k = O(h^q),$$

que se lê "o erro de discretização local no instante t_k tem ordem q".

Exemplo 2.2. *Qual é a ordem de consistência e uma estimativa do valor de C do Método de Euler para $y(t) \in \mathbb{R}$?*
Solução:

O erro de discretização local para o Método de Euler é dado por

$$\alpha_k = \frac{y(t_{k+1}) - y(t_k)}{h} - f\left(t_k, y(t_k)\right). \tag{2.2.8}$$

Supondo-se que $y(t)$ seja suficientemente diferenciável ao redor de $t = t_k$ e utilizando-se o polinômio de Taylor de primeiro grau com seu respectivo resto de Lagrange, tem-se em $t = t_{k+1}$

$$y(t_{k+1}) = y(t_k) + (t_{k+1} - t_k)y'(t_k) + \frac{(t_{k+1} - t_k)^2}{2!}y''(\xi), \tag{2.2.9}$$

com $\xi \in (t_k, t_{k+1})$. Como $h = t_{k+1} - t_k$, a substituição de (2.2.9) em (2.2.8) resulta em

$$\begin{aligned} \alpha_k &= \frac{y(t_k) + hy'(t_k) + \frac{h^2}{2!}y''(\xi) - y(t_k)}{h} - f\left(t_k, y(t_k)\right) \\ &= y'(t_k) + \frac{h}{2!}y''(\xi) - f\left(t_k, y(t_k)\right) \\ &= \frac{h}{2}y''(\xi). \end{aligned} \tag{2.2.10}$$

Portanto,

$$|\alpha_k| = h\left|\frac{1}{2}y''(\xi)\right| \Rightarrow \max_k |\alpha_k| \leq h \max_{\xi \in I} \frac{|y''(\xi)|}{2}. \tag{2.2.11}$$

Em (2.2.11), supõe-se que $y''(t)$ seja contínua em todo o intervalo de estudo, o que garante, pelo Teorema de Weierstrass, a existência de mínimo e máximo absolutos naquele intervalo. De (2.2.11) e da definição (2.2.7), conclui-se que o Método de Euler tem ordem de consistência 1 (um) com constante $C = \max_{\xi \in I} \frac{|y''(\xi)|}{2}$.

2.2.1 Exercícios

Exercício 2.1. *Verifique que os métodos de Euler Implícito, de Euler Aprimorado e do Trapézio são consistentes com quaisquer problemas de valores iniciais bem postos e têm ordens de consistência um, dois e dois, respectivamente.*

2.3 Erro de discretização global

Definição 2.5 (Erro global)**.** *O erro de discretização global no instante $t = t_k$ é dado por*

$$e(t_k, h) \doteq e_k \doteq y(t_k) - y_k, \tag{2.3.12}$$

onde $y(t)$ é a solução (única) do problema de Cauchy associado e y_k o k-ésimo passo de integração do método numérico para este mesmo problema de Cauchy.

O erro global representa o erro total acumulado cometido até o k-ésimo passo de integração. É importante salientar que y_k não é calculado a partir de valores exatos no instante de tempo anterior, como na definição do erro de discretização local, mas sim a partir de valores obtidos pela aplicação do método numérico (em todos os instantes de tempo anteriores).

2.4 Convergência

Definição 2.6 (Convergência)**.** *Um método numérico é convergente em $t \in [a, b]$ se e somente se*

$$\lim_{h \to 0} e_k = 0, \tag{2.4.13}$$

com $t - t_0 = kh$ fixado. O método numérico é convergente se for convergente para todo t no intervalo de estudo (para qualquer Problema de Cauchy bem posto).

Para determinar quais condições são suficientes para um método de passo único explícito convergir, analisa-se o comportamento do erro de discretização global. O produto do erro de discretização local (2.1.3) pelo passo de integração h fornece

$$y(t_{k+1}) = y(t_k) + h\Phi(t_k, y(t_k), h) + h\,\alpha_k\,, \tag{2.4.14}$$

associada a um método de passo único explícito geral

$$y_{k+1} = y_k + h\,\Phi(t_k, y_k, h). \tag{2.4.15}$$

Efetuando-se a subtração entre (2.4.14) e (2.4.15), tem-se

$$y(t_{k+1}) - y_{k+1} = y(t_k) - y_k + h\,\left[\Phi(t_k, y(t_k), h) - \Phi(t_k, y_k, h)\right] + h\,\alpha_k\,,$$

isto é, a evolução do erro de discretização global é dada por

$$e_{k+1} = e_k + h\,\left[\Phi(t_k, y(t_k), h) - \Phi(t_k, y_k, h)\right] + h\,\alpha_k. \tag{2.4.16}$$

Supondo que a função $\Phi(t, y, h)$ satisfaz a condição de Lipschitz na variável y, ou seja,

$$||\Phi(t, y_1, h) - \Phi(t, y_2, h)|| \leq L||y_1 - y_2||,$$

para quaisquer t e h e para uma constante positiva L, tem-se para (2.4.16) que

$$\begin{aligned} ||e_{k+1}|| &\leq ||e_k|| + h\,||\Phi(t_k, y(t_k), h) - \Phi(t_k, y_k, h)|| + h\,||\alpha_k|| \\ &\leq ||e_k|| + h\,L||y(t_k) - y_k|| + h\,||\alpha_k|| \\ &\leq ||e_k|| + h\,L||e_k|| + h\,||\alpha_k||, \end{aligned}$$

isto é,

$$||e_{k+1}|| \leq (1 + h\,L)||e_k|| + h\,||\alpha_k||\,, \text{ para todo } k > 0. \tag{2.4.17}$$

Supondo-se que o erro de discretização local α_k seja limitado,

$$\max_k ||\alpha_k|| \leq \alpha, \tag{2.4.18}$$

é possível reescrever (2.4.17) como

$$||e_{k+1}|| \leq (1 + h\,L)||e_k|| + h\,\alpha, \quad 0 \leq k \leq n - 1. \tag{2.4.19}$$

Da arbitrariedade do subíndice k em (2.4.19), tem-se

$$\begin{aligned} ||e_1|| &\leq (1 + h\,L)||e_0|| + h\,\alpha \\ ||e_2|| &\leq (1 + h\,L)||e_1|| + h\,\alpha \\ ||e_3|| &\leq (1 + h\,L)||e_2|| + h\,\alpha \\ &\vdots \\ ||e_{k+1}|| &\leq (1 + h\,L)||e_k|| + h\,\alpha. \end{aligned} \tag{2.4.20}$$

Por substituição recursiva de (2.4.20), obtêm-se

$$\begin{aligned} ||e_1|| &\leq (1 + h\,L)||e_0|| + h\,\alpha \\ ||e_2|| &\leq (1 + h\,L)^2||e_0|| + [(1 + h\,L) + 1]\,h\,\alpha \\ ||e_3|| &\leq (1 + h\,L)^3||e_0|| + \left[(1 + h\,L)^2 + (1 + h\,L) + 1\right] h\,\alpha \\ &\vdots \\ ||e_k|| &\leq (1 + h\,L)^k||e_0|| + \\ &+ \left[(1 + h\,L)^{k-1} + \cdots + (1 + h\,L)^2 + (1 + h\,L) + 1\right] h\,\alpha. \end{aligned} \tag{2.4.21}$$

A segunda das parcelas na desigualdade (2.4.21) é a soma dos k termos de uma progressão geométrica de termo inicial 1 (um) e razão $(1 + h\,L)$ [1]. Logo,

$$||e_k|| \leq (1 + h\,L)^k||e_0|| + \frac{(1 + h\,L)^k - 1}{L}\alpha. \tag{2.4.22}$$

[1] A soma S_n dos n termos de uma progressão geométrica é dada por $S_n = a_1\,(q^n - 1)/q - 1$, onde a_1 é o primeiro termo e q é a razão da progressão.

Pela convexidade da função exponencial, pode-se mostrar que $e^t \geq (1+t)$ e, consequentemente,

$$(e^{h\,L})^k \geq (1+h\,L)^\alpha. \tag{2.4.23}$$

Empregando-se a desigualdade (2.4.23) em (2.4.22), constata-se que

$$||e_k|| \leq e^{kh\,L}||e_0|| + \frac{e^{kh\,L}-1}{L}d, \tag{2.4.24}$$

onde d é a constante de limitação do erro de discretização local.

Teorema 2.1 (Delimitação do erro local)**.** *Seja um método de passo único explícito definido por*

$$\left\{ \begin{array}{rcl} y_0 & = & y(t_0) \\ y_{k+1} & = & y_k + h\,\Phi(t_k, y_k, h) \end{array} \right. ,$$

onde $\Phi(t,y,h)$ é uma função contínua em seus argumentos e satisfaz a condição de Lipschitz para a variável y, isto é, existe uma constante $L>0$ tal que

$$||\Phi(t,y_1,h) - \Phi(t,y_2,h)|| \leq L||y_1 - y_2||.$$

Além disso, se o erro de discretização local for limitado, ou seja,

$$\max_k ||\alpha_k|| \leq \alpha,$$

então o erro de discretização global satisfaz a delimitação

$$||e_k|| \leq e^{khL}||e_0|| + \frac{e^{khL}-1}{L}\alpha.$$

Se $||e_0|| = 0$ e se o método numérico for consistente de ordem q, isto é,

$$\max_k ||\alpha_k|| \leq C\,h^q,$$

tem-se

$$0 \leq ||e_k|| \leq \frac{e^{khL}-1}{L}C\,h^q = \frac{e^{(t_k-t_0)L}-1}{L}C\,h^q. \tag{2.4.25}$$

Portanto o método também será convergente de ordem q.

Em (2.4.25), o lado direito da desigualdade tende a zero quando h tende a zero. Dessa forma, pelo *Teorema do Confronto*, o limite do valor absoluto do erro de discretização global tende a zero quando o passo de integração tende a zero. Portanto, um método de passo único consistente é convergente (caso a função Φ seja contínua em seus argumentos e Lipschitziana no segundo argumento). A recíproca contudo não é verdadeira! No segundo exercício resolvido deste capítulo, apresenta-se um método inconsistente e convergente [23]. O teorema a seguir resume as observações anteriores.

Teorema 2.2 (Convergência)**.** *Um método de passo único explícito com*

$$\begin{cases} y_0 &= y(t_0) \\ y_{k+1} &= y_k + h\,\Phi(t_k, y_k, h) \end{cases},$$

onde Φ é Lipschitziana em y e contínua em seus argumentos, que seja ***consistente*** *para qualquer problema de Cauchy bem posto é* ***convergente****. Além disso, a ordem de convergência é pelo menos a mesma ordem da consistência.*

Exemplo 2.3. *O Método de Euler (1.4.23) é consistente com ordem de consistência um. Nesse método, $\Phi(t, y, h) = f(t, y)$. Como $f(t, y)$ é contínua e de Lipschitz (Problema de Cauchy), a função $\Phi(t, y, h)$ também é contínua e de Lipschitz. Portanto, o Método de Euler é convergente, isto é, para um instante de tempo fixado t, as soluções numéricas convergem para $y(t)$.*

2.5 Expansão do erro de discretização global

O comportamento exibido na prática por um determinado método numérico depende fortemente da regularidade da função $f(t, y)$ que define o Problema de Cauchy. O teorema 2.3 [24] fornece os subsídios teóricos necessários para determinar computacionalmente com que ordem o método converge.

Teorema 2.3 (Expansão assintótica do erro de discretização global)**.** *Seja uma função $f(t, y)$ com $N + 2$ derivadas parciais com relação a y, contínuas e limitadas na faixa*

$$\{(t, y); a = t_0 \leq t \leq t_f = b,\ \ y \in \mathbb{R}^n\}.$$

Além disso, seja $\eta(t, h)$ a solução numérica obtida através de um método de passo único

$$\eta_{k+1} = \eta_k + h\Phi(t_k, \eta_k, h)$$

de ordem p para $y(t)$, determinada com o passo de integração h, onde $y(t)$ é a solução do Problema de Cauchy

$$\begin{cases} \dfrac{d}{dt}y(t) &= f(t, y(t)) \\ y(t_0) &= y_0 \end{cases}.$$

Nessas condições, a solução numérica $\eta(t, h)$ admite expansão em potências de h da forma

$$\eta(t, h) = y(t) + h^p e_p(t) + h^{p+1} e_{p+1}(t) + \cdots + h^N e_N(t) + h^{N+1} E_{N+1}(t, h) \quad (2.5.26)$$

com $e_j(t_0) = 0$, $j = p, p + 1, ...$, válida para todo $t \in [a, b]$ e para todo $h = \dfrac{t - t_0}{n}$, $n = 1, 2, ...$

Observe que as funções $e_j(t)$ são independentes de h e o resto $E_{N+1}(t, h)$ é limitado para t fixado e para todo $h = \dfrac{t - t_0}{n}$, $n = 1, 2, ...$ A expansão em série

assintótica (2.5.26) pode ser escrita em função do erro de discretização global no instante t como

$$-e(t,h) = \eta(t,h) - y(t) = \sum_{j=p}^{N} h^j e_j(t) + h^{N+1} E_{N+1}(t,h). \tag{2.5.27}$$

O resultado teórico (2.5.27) pode ser utilizado na prática para estimar o erro de discretização global e a ordem do método, sendo também extremamente útil no processo de depuração do código computacional.

2.5.1 Estimativa do erro de discretização global

Para um dado passo de integração h e um instante de tempo fixado t, calculam-se a soluções numéricas $\eta(t,h)$ e $\eta\left(t,\frac{h}{2}\right)$. Se h for suficientemente pequeno, em primeira aproximação, tem-se de (2.5.27)

$$-e(t,h) = \eta(t,h) - y(t) \approx e_p(t)h^p, \tag{2.5.28}$$

$$-e(t,\frac{h}{2}) = \eta\left(t,\frac{h}{2}\right) - y(t) \approx e_p(t)\left(\frac{h}{2}\right)^p. \tag{2.5.29}$$

Calculando-se a diferença entre (2.5.28) e (2.5.29), chega-se a

$$\eta(t,h) - \eta\left(t,\frac{h}{2}\right) \approx e_p(t)\left[h^p - \left(\frac{h}{2}\right)^p\right] = e_p(t)\left(\frac{h}{2}\right)^p (2^p - 1)$$

e, portanto, a

$$e_p(t)\left(\frac{h}{2}\right)^p \approx \frac{\eta(t,h) - \eta(t,\frac{h}{2})}{2^p - 1}. \tag{2.5.30}$$

Substituindo-se (2.5.30) em (2.5.29), considerando $h > 0$ suficientemente pequeno, obtém-se

$$e(t,\frac{h}{2}) \approx -\frac{\eta(t,h) - \eta(t,\frac{h}{2})}{2^p - 1}, \tag{2.5.31}$$

uma estimativa do erro de discretização global em t, tendo sido calculadas as aproximações $\eta(t,h)$ e $\eta\left(t,\frac{h}{2}\right)$ e sendo conhecida a ordem do método empregado. As considerações anteriores assumem tacitamente que a ordem do método seja menor que a regularidade de f. A ordem de convergência exibida na prática por um método numérico, $\bar{p}$, depende fortemente do quão diferenciável é f e pode exibir um comportamento distinto da ordem p. Tem-se $p = \bar{p}$ apenas se f for suficientemente diferenciável.

2.5.2 Estimativa da ordem de convergência

A expansão (2.5.26) fornece um meio útil para se estimar a ordem com a qual o método numérico em uso converge para a solução de um determinado Problema de Cauchy. Para tanto, calculam-se $\eta(t,2h)$, $\eta(t,h)$ e $\eta\left(t,\frac{h}{2}\right)$, as aproximações numéricas no instante t empregando passos de integração $2h$, h e $\dfrac{h}{2}$, respectivamente, para um passo de integração $h > 0$ suficientemente pequeno. O valor absoluto

do quociente entre as diferenças $\eta(t,2h) - \eta(t,h)$ e $\eta(t,h) - \eta\left(t, \frac{h}{2}\right)$ fornece, em primeira aproximação,

$$\left| \frac{\eta(t,2h) - \eta(t,h)}{\eta(t,h) - \eta\left(t, \frac{h}{2}\right)} \right| \approx \left| \frac{e_{\bar{p}}(t)\,(\,2^{\bar{p}} - 1\,)\,h^{\bar{p}}}{e_{\bar{p}}(t)(1 - 2^{-\bar{p}})h^{\bar{p}}} \right| = 2^{\bar{p}}. \tag{2.5.32}$$

Calculando o logaritmo de base 2 de (2.5.32)

$$\log_2 \left(\left| \frac{\eta(t,2h) - \eta(t,h)}{\eta(t,h) - \eta\left(t, \frac{h}{2}\right)} \right| \right) \approx \log_2 \left(2^{\bar{p}}\right) = \bar{p},$$

tem-se uma estimativa para a ordem exibida pelo método.

Este procedimento deve ser executado para várias triplas de passos sucessivamente menores $\left(2h, h, \frac{h}{2}\right)$, $\left(h, \frac{h}{2}, \frac{h}{4}\right)$, ..., obtendo-se assim uma sequência de aproximações $\bar{p}_1$, $\bar{p}_2$, ..., que converge para a ordem $\bar{p}$ que o método apresenta para o Problema de Cauchy em questão.

2.5.3 Depuração do código computacional

Nas seções anteriores, as estimativas apresentadas assumem que o método numérico já tenha passado por um processo conhecido como *verificação / validação* e que, portanto, ele esteja implementado corretamente e funcionando perfeitamente. Durante a programação, entretanto, é necessário o uso de estratégias de depuração para remover eventuais erros de lógica ou de coeficientes ou parâmetros que tenham sido introduzidos inadvertidamente. Para isto, empregam-se a aproximação do erro de discretização global (2.5.28) e um Problema de Cauchy com solução exata conhecida. Tal estratégia é denominada *verificação por solução manufaturada*.

Para verificar se o código computacional está correto, escolhe-se um Problema de Cauchy com solução suficientemente diferenciável (isto é, com um número de derivadas superior à ordem do método). Em (2.5.28), o erro de discretização global é conhecido uma vez que se tem à mão a solução exata $y(t)$ do Problema de Cauchy. Deseja-se então certificar-se que o método tem, de fato, a ordem prevista na teoria, p. Para tanto, determinam-se $\eta(t,h)$ e $\eta\left(t, \frac{h}{2}\right)$, as aproximações numéricas do problema no instante t, empregando-se passos de integração $h > 0$ e $\dfrac{h}{2}$, respectivamente. O valor absoluto do quociente entre (2.5.28) e (2.5.29) fornece

$$\left| \frac{\eta(\,t,h) \;-\; y(t)}{\eta\left(t, \frac{h}{2}\right) - y(t)} \right| \approx \left| \frac{e_{\bar{p}^\star}(t) h^{\bar{p}^\star}}{e_{\bar{p}^\star}(t)\left(\frac{h}{2}\right)^{\bar{p}^\star}} \right| = 2^{\bar{p}^\star}, \tag{2.5.33}$$

ou seja,

$$\bar{p}^\star \approx \log_2 \left(\left| \frac{\eta(\,t,h) \;-\; y(t)}{\eta\left(t, \frac{h}{2}\right) - y(t)} \right| \right). \tag{2.5.34}$$

Quando esta estratégia é executada para passos de integração progressivamente menores h, $\dfrac{h}{2}$, $\dfrac{h}{4}$, $\dfrac{h}{8}$, ..., obtém-se pelo uso sucessivo de (2.5.34) uma sequência $\bar{p}_1^\star$, $\bar{p}_2^\star$, $\bar{p}_3^\star$, $\bar{p}_4^\star$, ..., que converge à ordem p que a teoria prevê para o método numérico, desde que este tenha sido implementado corretamente e aplicado a um Problema de Cauchy com garantias de solução única suficientemente diferenciável.

O procedimento de verificação por solução manufaturada não pode, por motivo algum, ser negligenciado. Para o bom programador, ele precede o uso regular do método no problema que se deseja solucionar numericamente. Usualmente, o procedimento de verificação deve ser efetuado para várias soluções manufaturadas com complexidade variada.

Exemplo 2.4. *Comprove numericamente a ordem de convergência do Método de Euler aplicado à solução do problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt}y(t) &= -20y(t), \quad t \in [0,1] \\ y(0) &= 1 \end{cases}, \tag{2.5.35}$$

cuja solução exata é $y(t) = e^{-20t}$ *(verifique!).*

$h = n^{-1}$	$\lvert e(t,h)\rvert$	$\left\lvert \dfrac{e(t,h)}{e\left(t,\frac{h}{2}\right)} \right\rvert$	$\log_2 \left\lvert \frac{e(t,h)}{e\left(t,\frac{h}{2}\right)} \right\rvert$
$2{,}000000\times10^{-1}$	2,430000E+02		
$1{,}000000\times10^{-1}$	9,999999E-01	2,430000E+02	7,924813E+00
$5{,}000000\times10^{-2}$	2,061154E-09	4,851652E+08	2,885390E+01
$2{,}500000\times10^{-2}$	2,060244E-09	1,000441E+00	6,367371E-04
$1{,}250000\times10^{-2}$	1,960019E-09	1,051135E+00	7,194787E-02
$6{,}250000\times10^{-3}$	1,534787E-09	1,277063E+00	3,528297E-01
$3{,}125000\times10^{-3}$	9,876378E-10	1,553998E+00	6,359843E-01
$1{,}562500\times10^{-3}$	5,632015E-10	1,753614E+00	8,103308E-01
$7{,}812500\times10^{-4}$	3,010566E-10	1,870749E+00	9,036163E-01
$3{,}906250\times10^{-4}$	1,556782E-10	1,933839E+00	9,514678E-01
$1{,}953125\times10^{-4}$	7,916378E-11	1,966533E+00	9,756547E-01
$9{,}765625\times10^{-5}$	3,991780E-11	1,983170E+00	9,878083E-01

Tabela 2.1: Verificação da ordem de convergência do Método de Euler aplicado ao problema de valor inicial (2.5.35) no instante $t = 1$.

A Tabela (2.1) apresenta as razões entre erros de discretização global para passos de integração progressivamente menores (razão de refinamento dois) em $t = 1$. *Observa-se que a ordem estimada, localizada na última coluna, tende a um, a ordem de convergência prevista pela teoria para o Método de Euler.*

Exemplo 2.5. *Verifique a ordem de convergência numérica do Método de Euler aplicado à solução do problema de valor inicial [5]*

$$\begin{cases} \dfrac{d}{dt}y(t) &= -y(t)\tan(t) - \dfrac{1}{cos(t)} \\ y(0) &= 1 \end{cases} \tag{2.5.36}$$

em $t = 1,292695719373$ *(raiz da equação* $e^t\cos(t) = 1$*), sabendo que a solução exata de (2.5.36) é* $y(t) = \cos(t) - \sin(t)$ *(comprove!).*

A Tabela (2.2) apresenta a razão entre erros de discretização global para passos de integração progressivamente menores (razão de refinamento dois) em $t = 1,292695719373$. *Observa-se que a ordem estimada, localizada na última coluna,*

$h = n^{-1}$	$\lvert e(t,h)\rvert$	$\left\lvert \frac{e(t,h)}{e\left(t,\frac{h}{2}\right)} \right\rvert$	$\log_2 \left\lvert \frac{e(t,h)}{e\left(t,\frac{h}{2}\right)} \right\rvert$
$5{,}000000\times10^{-2}$	1,130400E-03		
$2{,}500000\times10^{-2}$	2,561790E-04	4,412540E+00	2,141609E+00
$1{,}250000\times10^{-2}$	6,115026E-05	4,189336E+00	2,066722E+00
$6{,}250000\times10^{-3}$	1,494962E-05	4,090422E+00	2,032250E+00
$3{,}125000\times10^{-3}$	3,696597E-06	4,044157E+00	2,015839E+00
$1{,}562500\times10^{-3}$	9,191362E-07	4,021816E+00	2,007847E+00
$7{,}812500\times10^{-4}$	2,291629E-07	4,010842E+00	2,003905E+00
$3{,}906250\times10^{-4}$	5,721340E-08	4,005406E+00	2,001949E+00
$1{,}953125\times10^{-4}$	1,429410E-08	4,002588E+00	2,000933E+00
$9{,}765625\times10^{-5}$	3,573300E-09	4,000251E+00	2,000091E+00

Tabela 2.2: Verificação da ordem de convergência do Método de Euler aplicado ao problema de valor inicial (2.5.36) no instante $t = 1,292695719373$ [5].

tende a dois, diferente da ordem de convergência um prevista pela teoria para o Método de Euler. A ordem mais elevada é justificada pelo cancelamento do termo mais significativo na expansão assintótica do erro de discretização global para $t = 1,292695719373$*, raiz da equação* $e^t \cos(t) = 1$ *[5]. Para qualquer outro instante de tempo que não seja raiz dessa equação, a ordem estimada tende a um, a ordem de convergência prevista pela teoria para o Método de Euler.*

Exemplo 2.6. *Verifique a ordem de convergência numérica do Método de Euler aplicado à solução do problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt}y(t) &= -\dfrac{ty(t)}{1-t^2} \\ y(0) &= 1 \end{cases} \tag{2.5.37}$$

em $t = 1$*, sabendo que a solução exata de (2.5.37) é* $y(t) = \sqrt{1-t^2}$ *(comprove!).*

A Tabela (2.3) apresenta a razão entre erros de discretização global para passos de integração progressivamente menores (razão de refinamento dois) em $t = 1$*. Observa-se que a ordem estimada, localizada na última coluna, tende a meio, diferente da ordem de convergência um prevista pela teoria para o Método de Euler. A justificativa para a perda de ordem é que a Condição de Lipschitz não é mantida quando* $t = 1$ *e* $y = 0$ *[5].*

2.5.4 Exercícios

Exercício 2.2. *Estime numericamente a ordem de convergência do Método de Euler Aprimorado (1.4.25) com a estratégia de solução manufaturada empregando o problema modelo (2.5.35) no instante final* $t = 1$*.*

Exercício 2.3. *Estime numericamente a ordem de convergência do Método de Euler (1.4.23) com a estratégia de solução manufaturada empregando o problema modelo (2.5.36) no instante final* $t = \dfrac{\pi}{4}$*.*

$h = n^{-1}$	$\lvert e(t,h)\rvert$	$\left\lvert \frac{e(t,h)}{e\left(t,\frac{h}{2}\right)} \right\rvert$	$\log_2 \left\lvert \frac{e(t,h)}{e\left(t,\frac{h}{2}\right)} \right\rvert$
$1{,}2500000000\times10^{-1}$	3,012018700E-01		
$6{,}2500000000\times10^{-2}$	2,072697687E-01	1,453188E+00	5,392210E-01
$3{,}1250000000\times10^{-2}$	1,441738248E-01	1,437638E+00	5,237004E-01
$1{,}5625000000\times10^{-2}$	1,009724646E-01	1,427853E+00	5,138473E-01
$7{,}8125000000\times10^{-3}$	7,100787890E-02	1,421990E+00	5,079109E-01
$3{,}9062500000\times10^{-3}$	5,005564440E-02	1,418579E+00	5,044464E-01
$1{,}9531250000\times10^{-3}$	3,533418900E-02	1,416635E+00	5,024680E-01
$9{,}7656250000\times10^{-4}$	2,496156840E-02	1,415544E+00	5,013562E-01
$4{,}8828125000\times10^{-4}$	1,764145320E-02	1,414938E+00	5,007392E-01
$2{,}4414062500\times10^{-4}$	1,247093200E-02	1,414606E+00	5,004001E-01
$1{,}2207031250\times10^{-4}$	8,816964600E-03	1,414425E+00	5,002153E-01
$6{,}1035156250\times10^{-5}$	6,234037200E-03	1,414327E+00	5,001153E-01
$3{,}0517578125\times10^{-5}$	4,407942200E-03	1,414274E+00	5,000615E-01

Tabela 2.3: Verificação da ordem de convergência do Método de Euler aplicado ao problema de valor inicial (2.5.37) no instante $t = 1$ [5].

Exercício 2.4. *Estime numericamente a ordem de convergência do Método de Euler (1.4.23) com a estratégia de solução manufaturada empregando o problema modelo (2.5.37) no instante final* $t = \dfrac{1}{2}$.

2.6 Suplemento teórico

Teorema 2.4 (Teorema de Weierstrass)**.** *Seja* $f : A \subset \mathbb{R}^n \to \mathbb{R}$ *uma função contínua no conjunto* A *fechado e limitado. Então, existem pontos de máximo e mímino absoluto de* f *em* A*, isto é, existem* $x_0, x_1 \in A$ *tais que*

$$f(x_0) \leq f(x) \leq f(x_1)$$

para todo $x \in A$.

Teorema 2.5 (Teorema do Confronto)**.** *Se* $f(t) \leq g(t) \leq h(t)$ *quando* t *está próximo de* a *(exceto possivelmente em* a*) e*

$$\lim_{t\to a} f(t) = \lim_{t\to a} h(t) = L,$$

então

$$\lim_{t\to a} g(t) = L.$$

2.7 Exercícios resolvidos

Exercício Resolvido 2.1. *Considere o Método do Trapézio.*

(a) Calcule o erro local de discretização do método;

(b) Verifique que o método é consistente;

(c) Determine um delimitante superior do erro global de discretização.

Solução:

(a)

$$\begin{aligned}
\alpha_k &= \frac{y(t_{k+1})-y(t_k)}{h} - \Phi\left(t_k, y(t_k), y(t_{k+1}), h\right)\\
\alpha_k &= \frac{y(t_{k+1})-y(t_k)}{h} - \frac{f(t_k, y(t_k)+f(t_{k+1}, y(t_{k+1}))}{2}\\
2h\,\alpha_k &= 2\left[y(t_{k+1})-y(t_k)\right] - h\left[f(t_k, y(t_k)) + f(t_{k+1}, y(t_{k+1}))\right]\\
2h\,\alpha_k &= 2\left[y(t_k)+h\,y'(t_k)+\frac{h^2}{2!}y''(t_k)+\frac{h^3}{3!}y'''(\varepsilon_1)-y(t_k)\right]+\\
&- h\left[y'(t_k)+y'(t_{k+1})\right]\\
2h\,\alpha_k &= 2\left[h\,y'(t_k)+\frac{h^2}{2!}y''(t_k)+\frac{h^3}{3!}y'''(\varepsilon_1)\right]+\\
&- h\left[y'(t_k)+y'(t_k)+h\,y''(t_k)+\frac{h^2}{2!}y'''(\varepsilon_2)\right]\\
2h\,\alpha_k &= 2h\,y'(t_k)+h^2y''(t_k)+\frac{h^3}{3}y'''(\varepsilon_1)+\\
&- 2h\,y'(t_k)-h^2y''(t_k)-\frac{h^3}{2}y'''(\varepsilon_2)
\end{aligned}$$

$$\begin{aligned}
2h\,\alpha_k &= \frac{h^3}{3}y'''(\varepsilon_1)-\frac{h^3}{2}y'''(\varepsilon_2)\\
\alpha_k &= \frac{h^2}{6}y'''(\varepsilon_1)-\frac{h^2}{4}y'''(\varepsilon_2)\\
\alpha_k &= h^2\left[\frac{y'''(\varepsilon_1)}{6}-\frac{y'''(\varepsilon_2)}{4}\right]
\end{aligned}$$

Portanto, o erro local de discretização é dado por

$$\boxed{\alpha_k = h^2\left[\frac{y'''(\varepsilon_1)}{6}-\frac{y'''(\varepsilon_2)}{4}\right],\ \varepsilon_1,\varepsilon_2 \in]t_k, t_{k+1}[.}$$

(b)

$$\lim_{h\to 0}\alpha_k = \lim_{h\to 0} h^2\left[\frac{y'''(\varepsilon_1)}{6}-\frac{y'''(\varepsilon_2)}{4}\right] = 0.$$

Logo, o Método do Trapézio é consistente.

(c)

$$\begin{aligned}
&y(t_{k+1}) = y(t_k) + h\,\Phi(t_k, y(t_k), y(t_{k+1}), h) + h\,\alpha_k \qquad A\\
&y_{k+1} = y_k + h\,\Phi(t_k, y_k, y_{k+1}, h) \qquad B
\end{aligned}$$

Calculando-se A-B:

$$\begin{aligned}
y(t_{k+1})-y_{k+1} &= y(t_k)-y_k +\\
&+ h\left[\Phi(t_k, y(t_k), y(t_{k+1}), h) - \Phi(t_k, y_k, y_{k+1}, h)\right] +\\
&+ h\alpha_k;
\end{aligned}$$

$$\begin{aligned}
\|e_{k+1}\| &= \|e_k + \\
&+ h\left[\frac{f(t_k, y(t_k)) + f(t_{k+1}, y(t_{k+1}))}{2} - \frac{f(t_k, y_k) + f(t_{k+1}, y_{k+1})}{2}\right] + \\
&+ h\,\alpha_k\|; \\
\|e_{k+1}\| &= \|e_k + \\
&+ \frac{h}{2}\left[f(t_k, y(t_k)) - f(t_k, y_k) + f(t_{k+1}, y(t_{k+1})) - f(t_{k+1}, y_{k+1})\right] + \\
&+ h\,\alpha_k\|; \\
\|e_{k+1}\| &\leq \|e_k\| + \frac{h}{2}\|f(t_k, y(t_k)) - f(t_k, y_k)\| + \\
&+ \frac{h}{2}\|f(t_{k+1}, y(t_{k+1})) - f(t_{k+1}, y_{k+1})\| + \\
&+ h\,\|\alpha_k\|; \\
\|e_{k+1}\| &\leq \|e_k\| + \frac{L_1 h}{2}\|y(t_k) - y_k\| + \frac{L_2 h}{2}\|y(t_{k+1}) - y_{k+1}\| + h\,\|\alpha_k\|.
\end{aligned}$$

Sendo $L = \max\{L_1, L_2\}$ *e* $d = \max_k \|\alpha_k\|$*:*

$$\begin{aligned}
\|e_{k+1}\| &\leq \|e_k\| + \frac{Lh}{2}\|e_k\| + \frac{Lh}{2}\|e_{k+1}\| + h\,d; \\
\left(1 - \frac{Lh}{2}\right)\|e_{k+1}\| &\leq \left(1 + \frac{Lh}{2}\right)\|e_k\| + h\,d;
\end{aligned}$$

Para h *suficiente pequeno, isto é,* $Lh \ll 2$*,*

$$\|e_{n+1}\| \leq \left(\frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)\|e_n\| + \frac{h\,d}{1 - \frac{Lh}{2}}. \tag{2.7.38}$$

Para $n = 0$ *em (2.7.38):*

$$\|e_1\| \leq \left(\frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)\|e_0\| + \frac{hd}{1 - \frac{Lh}{2}}.$$

Para $n = 1$ *em (2.7.38):*

$$\begin{aligned}
\|e_2\| &\leq \left(\frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)\|e_1\| + \frac{hd}{1 - \frac{Lh}{2}}; \\
\|e_2\| &\leq \left(\frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)\left[\left(\frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)\|e_0\| + \frac{h\,d}{1 - \frac{Lh}{2}}\right] + \frac{h\,d}{1 - \frac{Lh}{2}}; \\
\|e_2\| &\leq \left(\frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)^2\|e_0\| + \left(1 + \frac{1 + \frac{Lh}{2}}{1 - \frac{Lh}{2}}\right)\frac{h\,d}{1 - \frac{Lh}{2}}.
\end{aligned}$$

Para $n = 2$ *em (2.7.38):*

$$\begin{aligned}\|e_3\| &\leq \left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^3 \|e_0\| + \\ &+ \left[1+\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}+\left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^2\right]\frac{h\,d}{1-\frac{Lh}{2}}.\end{aligned}$$

Para $n = k-1$ em (2.7.38):

$$\begin{aligned}\|e_k\| &\leq \left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^k \|e_0\| + \\ &+ \left[1+\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}+\cdots+\left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^{k-1}\right]\frac{h\,d}{1-\frac{Lh}{2}};\\ \|e_k\| &\leq \left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^k \|e_0\| + \left[\frac{\left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^k-1}{\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}-1}\right]\frac{h\,d}{1-\frac{Lh}{2}};\\ \|e_k\| &\leq \left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^k \|e_0\| + \left[\left(\frac{1+\frac{Lh}{2}}{1-\frac{Lh}{2}}\right)^k-1\right]\frac{d}{L},\end{aligned}$$

que define um limitante superior para o erro global.

Exercício Resolvido 2.2. *Considere o Problema de Cauchy*

$$\begin{cases} y' = f(t,y), & t \in [0,1], \\ y(0) = y_0, \end{cases} \tag{2.7.39}$$

e o seguinte método para sua solução,

$$y_{k+1} = y_k + hf(t_k, y_k) + h2^{k-n}, \tag{2.7.40}$$

onde $h = \dfrac{1}{n}$, $t_k = kh$ e $k = 0, 1, 2, ..., n$.

Mostre que o método é inconsistente e convergente [23].

Solução:

O método é de passo único, explícito, com função de iteração

$$\Phi(t_k, y_k, h) = f(t_k, y_k) + \theta(t_k, h), \tag{2.7.41}$$

onde $\theta_k = \theta(t_k, h) = 2^{k-n} = 2^{\frac{t_k-1}{h}}$. Se f for contínua e de Lipschitz na variável y, então Φ também será, pois θ_k não depende de y.

Para constatar a falta de consistência, basta verificar que $\Phi(t, y, 0) \neq f(t, y)$ para algum $t \in [0, 1]$. Como

$$\Phi(t, y, h) = f(t, y) + 2^{\frac{t-1}{h}},$$

para $t = 1$ tem-se que

$$\Phi(1, y, h) = f(1, y) + 1.$$

Logo,

$$\Phi(1, y, 0) = f(1, y) + 1 \neq f(1, y)$$

e, portanto, o método não é consistente com a equação diferencial ordinária.

Pode-se também verificar a falta de consistência com o erro local de truncamento, supondo que a solução é suficientemente diferenciável. Usando a Série de Taylor de $y(t_{k+1})$ ao redor de $y(t_k)$, o erro local de truncamento em t_k será dado por

$$\begin{aligned}
\alpha_k &= \frac{y(t_{k+1}) - y(t_k)}{h} - \Phi(t_k, y(t_k), h) \\
&= \frac{y(t_k) + hy'(t_k) + \frac{h^2}{2}y''(\tilde{t}_k) - y(t_k)}{h} - f(t_k, y(t_k)) - 2^{k-n} \\
&= \frac{h}{2}y''(\tilde{t}_k) - 2^{k-n},
\end{aligned}$$

para algum $\tilde{t}_k \in [0, 1]$. Em $k = n - 1$,

$$|\alpha_{n-1}| = \left|\frac{h}{2}y''(\tilde{t}_{n-1}) - \frac{1}{2}\right| = \frac{1}{2}\left|hy''(\tilde{t}_{n-1}) - 1\right|.$$

Considerando-se n suficientemente grande (h suficientemente pequeno), tal que

$$h < \max_{t\in[0,1]} \frac{1}{2y''(t)},$$

obtém-se

$$\max_k |\alpha_k| \geq \frac{1}{4}.$$

Assim, o método (2.7.40) não é consistente.

Para mostrar que o método (2.7.40) é convergente, emprega-se a Série de Taylor de $y(t_{k+1})$ ao redor de $y(t_k)$. O erro global de discretização no tempo t_{k+1} será tal que

$$\begin{aligned}
|e_{k+1}| &= |y_{k+1} - y(t_{k+1})| \\
&= |y_k + hf(t_k, y_k) + h\theta_k - y(t_k) - hy'(t_k) + O(h^2)| \\
&= |y_k + hf(t_k, y_k) + h\theta_k - y(t_k) - hf(t_k, y(t_k)) + O(h^2)| \\
&= |e_k + h\theta_k + h(f(t_k, y_k) - f(t_k, y(t_k)) + O(h^2)| \\
&\leq |e_k| + h\theta_k + h|f(t_k, y_k) - f(t_k, y(t_k)| + O(h^2) \\
&\leq |e_k| + h\theta_k + hL|e_k| + O(h^2) \\
&\leq (1 + Lh)|e_k| + h\theta_k + O(h^2),
\end{aligned}$$

supondo-se f de Lipschitz em y com constante L.

Observa-se que

$$\begin{aligned}
|e_0| &= 0, \\
|e_1| &\leq h\theta_1 + O(h^2), \\
|e_2| &\leq (1 + Lh)\theta_1 + h\theta_2 + O(h^2),
\end{aligned}$$

e, por fim, que

$$|e_{k+1}| \leq \sum_{j=0}^{k} (1+Lh)^{k-j} h\theta_{j+1} + O(h^2).$$

Portanto, o erro global máximo pode ser majorado pelo erro do final da integração numérica. Assim,

$$e(t_n) = \max_{k=1,\ldots,n} |e_k| \leq |e_n| \leq h\beta_n + O(h^2)$$

com

$$\begin{aligned} \beta_n = \sum_{j=0}^{n} \left(1+\frac{L}{n}\right)^{n-j} 2^{j+1-n} &= \sum_{j=0}^{n} \left(1+\frac{L}{n}\right)^{n-j} \left(\frac{1}{2}\right)^{n-j-1} \\ &= 2\sum_{j=0}^{n} \left(\frac{1+\frac{L}{n}}{2}\right)^{n-j}. \end{aligned} \tag{2.7.42}$$

Logo, β_n é duas vezes a soma de uma progressão geométrica de razão

$$q = \frac{1+\frac{L}{n}}{2} = \frac{1+hL}{2}.$$

A série (2.7.42) tem somente termos positivos e, para $n > 2L$, $q \in (0,1)$. Portanto, existe uma constante positiva C tal que $\beta_n < C$ para todo $n > 2L$. Consequentemente, para $h > 0$ pequeno,

$$e(t_n) \leq h\beta_n + O(h^2) \leq Ch + O(h^2),$$

e o método (2.7.40) é convergente.

Capítulo 3

Métodos de passo único de altas ordens

Dentre os métodos apresentados até aqui, apenas métodos de passo único de no máximo segunda ordem foram vistos. Neste Capítulo, duas técnicas para se obter métodos de passo único de ordens mais elevadas são apresentadas. Numa primeira abordagem, utiliza-se a Fórmula de Taylor com resto de Lagrange para se obter os *métodos da Série de Taylor*, Seção 3.1. Na prática, tais métodos são pouco utilizados devido ao grande número requerido de derivações de $f(t, y)$, o que imprime a tais métodos complexidade algébrica elevada (impactando, inclusive, a eficiência computacional). Numa segunda abordagem, substituindo-se as derivações por cálculos de $f(t, y)$ em pontos estrategicamente posicionados, obtêm-se os *métodos de Runge-Kutta*, Seção 3.2, bastante difundidos e utilizados, os quais são relativamente simples de serem implementados. No que vem a seguir, supõe-se que a solução do Problema de Cauchy tenha tantas derivadas quantas forem necessárias para que se possa apresentar estas duas abordagens.

3.1 Métodos da Série de Taylor

Da Fórmula de Taylor de grau q com centro em $t = t_k$, tem-se a expressão para $y(t_k + h)$,

$$\begin{aligned} y(t_k + h) &\doteq y(t_{k+1}) = y(t_k) + h\, y'(t_k) + \frac{h^2}{2!} y''(t_k) + \cdots + \frac{h^q}{q!} y^{(q)}(t_k) + \\ &+ \frac{h^{q+1}}{(q+1)!} y^{(q+1)}(\xi), \end{aligned} \tag{3.1.1}$$

na qual $\xi \in (t_k, t_k + h)$. A substituição $y'(t) = f(t, y(t))$ em (3.1.1), fornece

$$\begin{aligned} y(t_{k+1}) &= y(t_k) + h\, f(t_k, y(t_k)) + \frac{h^2}{2!} \frac{d}{dt} f(t_k, y(t_k)) + \cdots + \\ &+ \frac{h^q}{q!} \frac{d^{q-1}}{dt^{q-1}} f(t_k, y(t_k)) + \frac{h^{q+1}}{(q+1)!} \frac{d^q}{dt^q} f(\xi, y(\xi)). \end{aligned} \tag{3.1.2}$$

Como

$$\begin{aligned}
\frac{d}{dt}f(t,y(t)) &= \frac{\partial f}{\partial t}\frac{dt}{dt}+\frac{\partial f}{\partial y}\frac{dy}{dt}= \\
&= \left(\frac{\partial}{\partial t}+f\frac{\partial}{\partial y}\right)[f](t,y(t)), \\
\frac{d^2}{dt^2}f(t,y(t)) &= \left(\frac{\partial^2 f}{\partial t^2}+2f\frac{\partial^2 f}{\partial t\partial y}+f^2\frac{\partial^2 f}{\partial y^2}+\frac{\partial f}{\partial t}\frac{\partial f}{\partial y}+f\left(\frac{\partial f}{\partial y}\right)^2\right)= \\
&= \left(\frac{\partial}{\partial t}+f\frac{\partial}{\partial y}\right)^2[f](t,y(t)), \\
&\vdots \\
\frac{d^j}{dt^j}f(t,y(t)) &= \left(\frac{\partial}{\partial t}+f\frac{\partial}{\partial y}\right)^j[f](t,y(t)),
\end{aligned}$$

de (3.1.2), vem de maneira natural a proposta do método numérico da forma descrita a seguir.

Método 3.1 (Série de Taylor)**.**

$$\left\{\begin{array}{rcl} y_0 &=& y(t_0), \\ y_{k+1} &=& y_k + h\,\Phi(t_k,y_k,h), \end{array}\right. \tag{3.1.3}$$

no qual

$$\Phi(t,y,h) = f(t,y)+\frac{h}{2!}Df(t,y)+\frac{h^2}{3!}D^2f(t,y)+\cdots+\frac{h^{q-1}}{q!}D^{q-1}f(t,y),$$

com $D=\dfrac{\partial}{\partial t}+f\dfrac{\partial}{\partial y}$.

Este será denominado de Método da Série de Taylor de ordem q se (3.1.3) contiver termos de ordem até $q-1$ (isto é, seu erro de discretização local é $O(h^q)$). Verifique!

3.2 Métodos de Runge-Kutta explícitos

Os métodos de Runge-Kutta surgiram como uma extensão do método de Euler para ordens mais altas, usando apenas avaliações de f, sem a necessidade de cálculos de suas derivadas [5]. Neste sentido, um método é dito de Runge-Kutta explícito se ele:

1. for um método de passo único explícito,
2. substituir as derivadas de $f(t,y)$ por cálculos de $f(t,y)$ em pontos "convenientemente escolhidos" e, por fim,
3. concordar com o Método da Série de Taylor (3.1.3) até o termo de q-ésima ordem, para algum $q>0$.

Veja um pouca da história dos métodos de Runge-Kutta em [4].

3.2.1 Métodos de Runge-Kutta explícitos de 2-estágios

O método de Euler necessita de apenas uma avaliação de f (um estágio). Vamos supor um método de passo único explícito que use duas avaliações de f (2 estágios), da seguinte forma:

$$y_{k+1} = y_k + h(b_1 f(t_k, y_k) + b_2 f(\tilde{t}_k, \tilde{y}_k)), \tag{3.2.4}$$

onde b_1 e b_2 são constantes e $\tilde{y}_k$ define uma aproximação para $y(\tilde{t}_k)$, para algum $\tilde{t}_k \in [t_k, t_k + h]$, com

$$\begin{aligned} \tilde{t}_k &= t_k + \theta h, \\ \tilde{y}_k &= y_k + \theta h f(t_k, y_k). \end{aligned} \tag{3.2.5}$$

Seja $f_k = f(t_k, y_k)$ e $Df_k = \dfrac{\partial f}{\partial t}(t_k, y_k) + f(t_k, y_k)\dfrac{\partial f}{\partial y}(t_k, y_k)$. Agora, expandindo $f(\tilde{t}_k, \tilde{y}_k)$ em Série de Taylor bivariada em torno de (t_k, y_k), temos que (verifique!)

$$f(\tilde{t}_k, \tilde{y}_k) = f_k + \theta h D f_k + O(h^2). \tag{3.2.6}$$

Para garantirmos que um método desta forma seja de segunda ordem, concordando portanto com o Método da Série de Taylor até segunda ordem, precisamos que

$$b_1 f_k + b_2(f_k + \theta h D f_k) = f_k + \frac{h}{2} D f_k. \tag{3.2.7}$$

Portanto,

$$b_1 + b_2 = 1 \tag{3.2.8}$$

$$b_2 \theta = \frac{1}{2}, \tag{3.2.9}$$

e então o método terá que ter a seguinte forma

$$y_{k+1} = y_k + h\left(\left[1 - \frac{1}{2\theta}\right] f(t_k, y_k) + \frac{1}{2\theta} f(t_k + \theta h, y_k + \theta h f(t_k, y_k))\right). \tag{3.2.10}$$

Variando θ, restrito a $\theta \in (0, 1)$, obtemos diferentes métodos de Runge-Kutta de 2 estágios, com garantia de ordem 2 de consistência. Tomando $\theta = 1$ temos o Método de Euler Aprimorado (veja a equação (1.4.25)) e com $\theta = 1/2$ temos o Método de Euler Modificado, descrito no exemplo a seguir.

Método 3.2 (Euler Modificado ou do Ponto Médio Explícito)**.**

$$\left\{ \begin{array}{rcl} y_0 &=& y(t_0), \\ y_{k+1} &=& y_k + h\Phi(t_k, y_k, h)\,, \end{array} \right. \tag{3.2.11}$$

com $t_{k+1} = t_k + h$, $h = \dfrac{b-a}{n}$, $0 \le k \le n-1$ *e*

$$\Phi(t_k, y_k, h) = f\left(t_k + \frac{h}{2}, y_k + \frac{h}{2} f(t_k, y_k)\right).$$

Exemplo 3.1. *O Método de Euler Modificado é um Método de Runge-Kutta de 2 estágios de segunda ordem.*

Por inspeção, observa-se que o método (3.2.11) é de passo único e explícito. Além disso, não existem cálculos de derivadas de $f(t,y)$. Basta então verificar se (3.2.11) concorda com o Método da Série de Taylor (3.1.3) para alguma ordem q.

Desenvolvendo-se $f\left(t+\Delta t, y+\Delta y\right)$ em Série de Taylor em torno de (t_k, y_k), onde $\Delta t = \dfrac{h}{2}$ e $\Delta y = \dfrac{h}{2} f(t,y)$, tem-se

$$\begin{aligned} f\left(t_k+\Delta t, y_k+\Delta y\right) &= f(t_k,y_k) + \left[\begin{array}{cc} \dfrac{\partial}{\partial t} f(t_k,y_k) & \dfrac{\partial}{\partial y} f(t_k,y_k) \end{array}\right] \left[\begin{array}{c} \Delta t \\ \Delta y \end{array}\right] + \\ &+ \frac{1}{2!} \left[\begin{array}{cc} \Delta t & \Delta y \end{array}\right] \left[\begin{array}{cc} \frac{\partial^2}{\partial t^2} f(t_k,y_k) & \frac{\partial^2}{\partial t \partial y} f(t_k,y_k) \\ \frac{\partial^2}{\partial t \partial y} f(t_k,y_k) & \frac{\partial^2}{\partial y^2} f(t_k,y_k) \end{array}\right] \left[\begin{array}{c} \Delta t \\ \Delta y \end{array}\right] + \\ &+ O(\Delta t^3, \Delta y^3), \end{aligned}$$

ou ainda,

$$\begin{aligned} \Phi(t_k,y_k,h) &= f_k + \frac{h}{2}\left(\frac{\partial f_k}{\partial t} + f_k \frac{\partial f_k}{\partial y}\right) + \frac{1}{2!}\frac{h^2}{4}\left(\frac{\partial^2 f_k}{\partial t^2} + 2 f_k \frac{\partial^2 f_k}{\partial t \partial y} + f_k^2 \frac{\partial^2 f_k}{\partial y^2}\right) + \\ &+ O(h^3), \end{aligned} \tag{3.2.12}$$

sendo $f_k = f(t_k, y_k)$.

Substituindo-se (3.2.12) em (3.2.11), obtém-se

$$\begin{aligned} y_{k+1} &= y_k + h f_k + \frac{h^2}{2!}\left(\frac{\partial f_k}{\partial t} + f_k \frac{\partial f_k}{\partial y}\right) + \frac{1}{2!}\frac{h^3}{4}\left(\frac{\partial^2 f_k}{\partial t^2} + 2 f_k \frac{\partial^2 f_k}{\partial t \partial y} + f_k^2 \frac{\partial^2 f_k}{\partial y^2}\right) + \\ &+ O(h^4), \end{aligned} \tag{3.2.13}$$

cujos termos concordam com os do Método da Série de Taylor (3.1.3) até a segunda ordem (verifique que a partir dos termos de terceira ordem não há coincidência!). Desta maneira, o Método de Euler Modificado é um Método de Runge-Kutta de ordem 2.

3.2.2 Métodos de Runge-Kutta explícitos de R-estágios

Nos métodos de Runge-Kutta (ou RK), o número de estágios diz respeito ao número de vezes que $f(t,y)$ deve ser calculada. Os métodos RK explícitos com R-estágios clássicos têm a forma descrita a seguir.

Método 3.3 (Rung-Kutta explícito de R-estágios)**.**

$$\left\{\begin{array}{rcl} y_0 &=& y(t_0), \\ y_{k+1} &=& y_k + h\Phi(t_k, y_k, h)\,, \end{array}\right.$$

onde

$$\Phi(t,y,h) = \sum_{r=1}^{R} b_r \kappa_r, \tag{3.2.14}$$

com

$$\begin{aligned}
\kappa_1(t,y) &= f(t,y), \\
\kappa_2(t,y) &= f(t+hc_2, y+ha_{21}\kappa_1), \\
\kappa_3(t,y) &= f(t+hc_3, y+ha_{31}\kappa_1+hb_{32}\kappa_2), \\
&\vdots \\
\kappa_r(t,y) &= f\left(t+hc_r, y+h\sum_{s=1}^{r-1}a_{rs}\kappa_s\right), \quad 2\le r\le R. \qquad (3.2.15)
\end{aligned}$$

Os parâmetros a_{rs}, b_r e c_r satisfazem as relações

$$\text{(i)} \qquad \sum_{r=1}^{R} b_r = 1, \qquad (3.2.16)$$

$$\text{(ii)} \qquad c_r = \sum_{s=1}^{r-1} a_{rs}, \quad 2\le r\le R. \qquad (3.2.17)$$

A primeira condição, (i), é suficiente, e necessária, para obtermos consistência. De fato, quando $h\to 0$, $\kappa_r(t,y)\to f(t,y)$. Portanto $\Phi(t,y,0)=f(t,y)\sum_{r=1}^{R}b_r$ e precisamos da primeira condição para termos $\Phi(t,y,0)=f(t,y)$, sendo esta também suficiente para garantir a consistência.

A segunda condição, (ii), é conveniente para garantirmos que o método concorde com o termo de ordem 2 (referente a $y''=Df$) do Método de Série de Taylor. De tal forma, um método com mais de um estágio ($R>1$) terá condições necessárias para ser, pelo menos, de segunda ordem. A Série de Taylor de κ_r em torno do ponto (t_k,y_k) será da forma

$$\kappa_r(t_k,y_k) = f_k + hc_r\frac{\partial f_k}{\partial t} + h\left(\sum_{s=1}^{r-1}a_{rs}\kappa_s\right)\frac{\partial f_k}{\partial y} + O(h^2). \qquad (3.2.18)$$

Agora, substituímos a Série de Taylor de κ_s, que pode ser escrita, em primeira ordem, como $\kappa_s(t_k,y_k)=f_k+O(h)$, na série de Taylor de κ_r acima. Com isso, notamos que para que o método concorde com a Φ do Método da Série de Taylor até o termo de ordem um (relativo a Df), será necessário termos

$$c_r\frac{\partial f_k}{\partial t} + \left(\sum_{s=1}^{r-1}a_{rs}f_k\right)\frac{\partial f_k}{\partial y} = \theta_r\left(\frac{\partial f_k}{\partial t} + f_k\frac{\partial f_k}{\partial y}\right), \qquad (3.2.19)$$

para algum $\theta_r>0$. Portanto, a condição (ii) em (3.2.17) garante que, se $r>1$, o método resultante tem chances de ter pelo menos ordem 2, sendo esta uma condição necessária para atingir segunda ordem. Ela não é suficiente para obtermos segunda ordem pois, para concordar com o método da Série de Taylor até pelo menos ordem 2, temos que considerar também os valores de b_r.

Os métodos Runge-Kutta podem ser expressos na forma de uma tabela, conhecida por Tabela de Butcher [5]. Ela é composta pelos coeficientes a_{rs}, b_r e c_r da seguinte forma,

$$\begin{array}{c|ccccccc} 0 & & & & & & & \\ c_2 & a_{21} & & & & & & \\ c_3 & a_{31} & a_{32} & & & & & \\ c_4 & a_{41} & a_{42} & a_{43} & & & & \\ \vdots & \vdots & \vdots & \vdots & & & & \\ c_R & a_{R1} & a_{R2} & a_{R3} & \cdots & a_{R,R-1} & & \\ \hline & b_1 & b_2 & b_3 & \cdots & b_{R-1} & b_R \end{array}$$

ou, na forma compacta,

$$\begin{array}{c|c} c & A \\ \hline & b^t \end{array}$$

onde c é vetor de coeficientes c_r, denominados como nós. b^t é a transposta do vetor de coeficientes b_r, denominados como pesos[1]. A matrix A, definida pelos coeficientes a_{rs}, é conhecida como matriz de Runge-Kutta.

Retomando o caso de dois estágios ($R = 2$), de (3.2.14)-(3.2.17) tem-se

$$\begin{aligned} \Phi(t_k, y_k, h) &= b_1\kappa_1 + b_2\kappa_2 \\ &= b_1 f\left(t_k, y_k\right) + b_2 f\left(t_k + hc_2, y_k + ha_{21}f(t_k, y_k)\right), \end{aligned} \quad (3.2.20)$$

com $c_2 = a_{21}$, que é igual ao θ da seção anterior (vide equação (3.2.10)).

Exemplo 3.2. *Tem-se no Método de Euler Modificado (3.2.11)*

$$\Phi(t, y, h) = f\left(t + \frac{h}{2}, y + \frac{h}{2}f(t, y)\right). \quad (3.2.21)$$

Comparando-se (3.2.21) com (3.2.20), constata-se que $b_1 = 0$ *e* $b_2 = 1$ *(logo* $b_1 + b_2 = 1$*) e que* $c_2 = a_{21} = \frac{1}{2}$*. Assim, o Método de Euler Modificado é um Método de Runge-Kutta de segunda ordem com* 2 *estágios.*

Exemplo 3.3. *Tem-se no Método de Euler Aprimorado (1.4.25),*

$$\Phi(t, y, h) = \frac{1}{2}f\left(t, y\right) + \frac{1}{2}f\left(t + h, y + hf(t, y)\right). \quad (3.2.22)$$

Comparando-se (3.2.22) com (3.2.20), constata-se que $b_1 = \frac{1}{2}$ *e* $b_2 = \frac{1}{2}$ *(logo* $b_1 + b_2 = 1$*) e que* $c_2 = a_{21} = 1$*. Assim, o Método de Euler Aprimorado é um Método de Runge-Kutta de segunda ordem com* 2 *estágios.*

3.2.3 Métodos de Runge-Kutta de ordens mais elevadas

Método 3.4 (Runge-Kutta de Terceira Ordem com Três Estágios (RK33))**.**

$$\Phi(t, y, h) = \frac{1}{6}\left(\kappa_1 + 4\kappa_2 + \kappa_3\right), \quad (3.2.23)$$

[1] O termo "pesos" vem dos pesos usados em um método de integração numérica para equação em sua forma integral.

com

$$\begin{cases} \kappa_1 = f\left(t, y\right), \\ \kappa_2 = f\left(t + \dfrac{h}{2}, y + \dfrac{h}{2}\kappa_1\right), \\ \kappa_3 = f\left(t + h, y - h\kappa_1 + 2h\kappa_2\right). \end{cases}$$

Método 3.5 (Runge-Kutta de Quarta Ordem com Quatro Estágios (RK44))**.**

$$\Phi(t, y, h) = \frac{1}{6}\left(\kappa_1 + 2\kappa_2 + 2\kappa_3 + \kappa_4\right), \tag{3.2.24}$$

com

$$\begin{cases} \kappa_1 = f\left(t, y\right), \\ \kappa_2 = f\left(t + \dfrac{h}{2}, y + \dfrac{h}{2}\kappa_1\right), \\ \kappa_3 = f\left(t + \dfrac{h}{2}, y + \dfrac{h}{2}\kappa_2\right), \\ \kappa_4 = f\left(t + h, y + h\kappa_3\right). \end{cases}$$

Método 3.6 (Runge-Kutta de Quarta Ordem com Cinco Estágios (RK45))**.**

$$\Phi(t, y, h) = \frac{25}{216}\kappa_1 + \frac{1408}{2565}\kappa_3 + \frac{2197}{4104}\kappa_4 - \frac{1}{5}\kappa_5, \tag{3.2.25}$$

com

$$\begin{cases} \kappa_1 = f\left(t, y\right), \\ \kappa_2 = f\left(t + \dfrac{h}{4}, y + \dfrac{h}{4}\kappa_1\right), \\ \kappa_3 = f\left(t + \dfrac{3h}{8}, y + \dfrac{3h}{32}\kappa_1 + \dfrac{9h}{32}\kappa_2\right), \\ \kappa_4 = f\left(t + \dfrac{12h}{13}, y + \dfrac{1932h}{2197}\kappa_1 - \dfrac{7200h}{2197}\kappa_2 + \dfrac{7296h}{2197}\kappa_3\right), \\ \kappa_5 = f\left(t + h, y + \dfrac{439h}{216}\kappa_1 - 8h\kappa_2 + \dfrac{3680h}{513}\kappa_3 - \dfrac{845h}{4104}\kappa_4\right). \end{cases}$$

Método 3.7 (Runge-Kutta de Quinta Ordem com Seis Estágios (RK56))**.**

$$\Phi(t, y, h) = \frac{16}{135}\kappa_1 + \frac{6656}{12825}\kappa_3 + \frac{28561}{56430}\kappa_4 - \frac{9}{50}\kappa_5 + \frac{2}{55}\kappa_6, \tag{3.2.26}$$

com

$$\begin{cases} \kappa_1 = f\left(t, y\right), \\ \kappa_2 = f\left(t + \dfrac{h}{4}, y + \dfrac{h}{4}\kappa_1\right), \\ \kappa_3 = f\left(t + \dfrac{3h}{8}, y + \dfrac{3h}{32}\kappa_1 + \dfrac{9h}{32}\kappa_2\right), \\ \kappa_4 = f\left(t + \dfrac{12h}{13}, y + \dfrac{1932h}{2197}\kappa_1 - \dfrac{7200h}{2197}\kappa_2 + \dfrac{7296h}{2197}\kappa_3\right), \\ \kappa_5 = f\left(t + h, y + \dfrac{439h}{216}\kappa_1 - 8h\kappa_2 + \dfrac{3680h}{513}\kappa_3 - \dfrac{845h}{4104}\kappa_4\right), \\ \kappa_6 = f\left(t + \dfrac{h}{2}, y - \dfrac{8h}{27}\kappa_1 + 2h\kappa_2 - \dfrac{3544h}{2565}\kappa_3 + \dfrac{1859h}{4104}\kappa_4 - \dfrac{11h}{40}\kappa_5\right). \end{cases}$$

3.2.4 Exercícios

Exercício 3.1. *Proponha um Método RK de primeira ordem e dois estágios.*

Exercício 3.2. *Mostre que o Método de Euler Aprimorado (1.4.25) concorda com o Método da Série de Taylor apenas até o termo de segunda ordem.*

Exercício 3.3. *Interprete geometricamente o Método de Euler Modificado e o Método de Euler Aprimorado.*

Exercício 3.4. *Mostre que todos os métodos de ordens mais elevadas (maiores que 2), apresentados na seção 3.2.3, são de fato consistentes e de, pelo menos, segunda ordem.*

Exercício 3.5. *Encontre as condições necessárias sobre os parâmetros da tabela de Butcher para que um método de 3 estágios seja de ordem 3.*

Exercício 3.6. *Considere o método com a seguinte tabela de Butcher*

$$\begin{array}{c|cc} 0 & & \\ 2/3 & 2/3 & \\ \hline & 1/4 & 3/4 \end{array},$$

conhecido como método de Ralston. Mostre que o método tem ordem 2 e estime o erro local de discretização. Este método é conhecido por minimizar a constante do erro local de discretização.

3.3 Controle automático do passo de integração

Já sabemos que o erro dos métodos estudados dependem do tamanho do passo de integração h. Discutiremos nesta seção uma forma de escolher h, ao longo da integração, de tal forma a garantir que o erro associado a discretização seja controlado.

A ideia consiste basicamente em usar informações de 2 métodos com ordens diferentes para controlar o erro. Isso é computacionalmente mais eficiente se os métodos coincidirem nas avaliações de f, portanto, essa forma de controle comumente faz uso de métodos de Runge-Kutta que tenham κ_r iguais.

Considere dois métodos explícitos de passo único com ordens $p+1$ e p,

$$y_{k+1}^{p+1} = y_k^{p+1} + h\Phi^{p+1}(t_k, y_k^{p+1}, h), \tag{3.3.27}$$
$$y_{k+1}^{p} = y_k^{p} + h\Phi^{p}(t_k, y_k^{p}, h), \tag{3.3.28}$$

definidos pelas funções de iterações Φ^{p+1} e Φ^p,

$$\Phi^{p+1}(t,y,h) = \sum_{r=1}^{R} b_r^{p+1}\kappa_r(t,y,h) \tag{3.3.29}$$

$$\Phi^{p}(t,y,h) = \sum_{r=1}^{R} b_r^{p}\kappa_r(t,y,h), \tag{3.3.30}$$

onde R é o número de estágios do método de ordem mais alta. Caso o método de ordem mais baixa tenha número de estágios, R_p, menor que R, os últimos estágios devem assumir valor nulo ($b_r^p = 0$ para $r = R_p + 1, R_p + 2, ..., R$).

Subtraindo os erros locais de discretizações (α_k^p e α_k^{p+1}), considerando que os métodos tem avaliações de κ_r iguais (mesmos c_r e a_{rs}), temos que

$$\begin{aligned}\alpha_k^p - \alpha_k^{p+1} &= \Phi^p(t_k, y(t_k), h) - \Phi^{p+1}(t_k, y(t_k), h)\\ &= \sum_{r=1}^{s}(b_r^p - b_r^{p+1})\kappa_r(t_k, y(t_k), h).\end{aligned}$$

Como α_k^{p+1} tem ordem $p+1$, podemos estimar α_k^p como base em

$$\alpha_k^p = \sum_{r=1}^{s}(b_r^p - b_r^{p+1})\kappa_r(t_k, y(t_k), h) + O(h^{p+1}), \tag{3.3.31}$$

usando

$$\alpha_k^p \approx \sum_{r=1}^{s}(b_r^p - b_r^{p+1})\kappa_r(t_k, y_k, h). \tag{3.3.32}$$

Note que, usando as definições de erros locais de discretizações, isso também poderia ser estimado como

$$\alpha_k^p \approx \frac{1}{h}(y_{k+1}^p - y_{k+1}^{p+1}), \tag{3.3.33}$$

ao custo de termos que avaliar ambos métodos. Por outro lado, usando (3.3.32), aproveitamos as contas dos κ_r.

A estimativa de α_k^p pode ser usada para escolhermos um passo de integração h dentro de uma tolerância para o erro local de discretização. Seja $\epsilon > 0$ uma tolerância fixada. Suponha que demos um passo de integração, com tamanho h_k, para o qual, com base na estimativa de α_k, temos

$$|\alpha_k^p(h_k)| \leq \epsilon, \tag{3.3.34}$$

Sabemos que $\alpha_k^p(h) \approx Ch_k^p$, para algum C que pode depender de t_k, mas não depende de h_k. Gostaríamos de achar h_{k+1} de tal forma a garantir que o erro permaneça dentro da tolerância. Assim, precisamos que

$$|\alpha_{k+1}(h_{k+1})| \approx |C|h_{k+1}^p \leq \epsilon. \tag{3.3.35}$$

Portanto, uma primeira aproximação para h_{k+1} pode ser obtida considerando que, tanto no passo k quanto em $k+1$, temos a mesma constante C, de forma a obter que

$$|\alpha_k^p(h_k)|\frac{h_{k+1}^p}{h_k^p} \leq \epsilon. \tag{3.3.36}$$

Podemos então adotar como próximo passo de integração

$$h_{k+1} = h_k \left(\frac{\epsilon}{|\alpha_k(h_k)|}\right)^{1/p}. \tag{3.3.37}$$

Com base no h_{k+1} calculado, estima-se $\alpha_{k+1}(h_{k+1})$. Se este for menor que a tolerância, passa-se para o próximo passo de integração, senão, deve-se diminuir o h_{k+1} e re-calcular $\alpha_{k+1}(h_{k+1})$ até obtermos a precisão desejada.

Implementações mais simples estimam diretamente $\alpha_k(h_k)$ a cada passo de tempo e consideram, de forma iterativa, diferentes valores de h_k até que a tolerância seja atingida ($|\alpha_k(h_k)| < \epsilon$). Uma possibilidade é, por exemplo, multiplicar sucessivamente h_k por $0 < \beta < 1$ até obter a precisão desejada.

Exemplo 3.4. *Ilustraremos o procedimento com os métodos de Euler Modificado (3.2.11) e de Runge-Kutta RK33 (3.2.23). Para estes métodos temos, respectivamente, os seguinte erros de discretização locais são*

$$\alpha_k^{EM} = \frac{y(t_{k+1}) - y(t_k)}{h} - \kappa_2 \tag{3.3.38}$$

e

$$\alpha_k^{RK33} = \frac{y(t_{k+1}) - y(t_k)}{h} - \frac{1}{6}(\kappa_1 + 4\kappa_2 + \kappa_3), \tag{3.3.39}$$

para os quais é possível observar que ambos métodos tem a mesma avaliação de κ_2.

Calculando-se a diferença entre (3.3.38) e (3.3.39), chega-se a

$$\begin{aligned} \alpha_k^{EM} - \alpha_k^{RK33} &= -\kappa_2 + \frac{1}{6}(\kappa_1 + 4\kappa_2 + \kappa_3) = & (3.3.40)\\ &= \frac{1}{6}(\kappa_1 - 2\kappa_2 + \kappa_3). & (3.3.41) \end{aligned}$$

Como $\alpha_k^{RK33} = O(h^3)$, pode-se reescrever (3.3.40) na forma

$$\alpha_k^{EM} = \frac{1}{6}(\kappa_1 - 2\kappa_2 + \kappa_3) + O(h^3),$$

ou aproximadamente

$$\alpha_k^{EM} \approx \frac{1}{6}(\kappa_1 - 2\kappa_2 + \kappa_3). \tag{3.3.42}$$

Empregando-se (3.3.42) como uma estimativa para o erro de discretização local do Método de Euler Modificado, pode-se propor uma estratégia ingênua para controlar o passo de integração no Método de Euler Modificado. Dados (t_k, y_k), $h > 0$ e $\epsilon > 0$, um passo típico de integração dessa estratégia, sem usar (3.3.37), é descrita no Algoritmo (3.3.1).

O algoritmo 3.3.1 possui duas constantes, β e δ, que agem da seguinte forma. β reduz o passo de integração para que a tolerância seja alcançada, e δ aumenta o

1. Calcule κ_i, $i = 1, 2, 3$ e α_k^{EM},

   $\kappa_1 = f(t_k, y_k)$,
   $\kappa_2 = f(t_k + \frac{h}{2}, y_k + \frac{h}{2}\kappa_1)$,
   $\kappa_3 = f(t_k + h, y_k - h\kappa_1 + 2h\kappa_2)$,

   $\alpha_k^{EM} = \frac{1}{6}(\kappa_1 - 2\kappa_2 + \kappa_3)$.

2. Ache um h adequado a tolerância ϵ.

   Enquanto $\alpha_k^{EM} > \epsilon$ faça

   - Reduza o passo de integração, $h \leftarrow \beta h$, $0 < \beta < 1$;
   - Recalcule κ_i, $i = 1, 2, 3$ e $\alpha_k^{EM} = \frac{1}{6}(\kappa_1 - 2\kappa_2 + \kappa_3)$.

3. Avance a solução numérica e atualize o marcador de tempo:

   $y_k \leftarrow y_k + h\Phi_{EM}$,
   $t_k \leftarrow min\{t_k + h, t_{final}\}$.

4. Se $t_k < t_{final}$ então

   - aumente o passo de integração, $h \leftarrow \delta h$, $\delta > 1$,
     e retorne a 1,

   senão FIM

Algoritmo 3.3.1: Controle do passo de integração no Método de Euler Modificado.

passo de integração para que sempre consideremos passos suficientemente grandes e evitarmos darmos passos pequenos demais (correndo o risco de ficarmos quase parados!). O algoritmo pode facilmente incluir também a estimativa para o h do próximo passo com base na equação (3.3.37).

A combinação dos Métodos RK45 (3.2.25) e RK56 (3.2.26) é conhecida como *Método de Runge-Kutta-Fehlberg* [3, 24]. Essa combinação é comumente empregada na prática para controlar automaticamente o passo de integração.

3.3.1 Exercícios

Exercício 3.7. *Escreva um algoritmo que empregue o Método de Euler com controle automático do passo de integração baseado nos erros locais de discretização dos Métodos de Euler e de Euler Modificado.*

Exercício 3.8. *Escreva um algoritmo que empregue o Método de Runge-Kutta* 45 *com controle automático do passo de integração baseado nos erros locais de discretização dos Métodos de Runge-Kutta* 45 *(3.2.25) e de Runge-Kutta* 56 *(3.2.26) (Runge-Kutta-Felhberg). Verifique que a estimativa para o erro local de discretização é dada por*

$$\alpha_k^{RKF} \approx \frac{1}{360}\kappa_1 - \frac{128}{4275}\kappa_3 - \frac{2197}{75240}\kappa_4 + \frac{1}{50}\kappa_5 + \frac{2}{55}\kappa_6.$$

(Sugestão: veja o algoritmo proposto por Burden [3].)

3.4 Suplemento teórico

Teorema 3.1 (Barreira Explícita (Butcher)). *Se um método de Runge-Kutta com R-estágios tem ordem p, então $R \geq p$. Além disso, se $p \geq 5$, então $R > p$.*

Este importante teorema nos diz que são necessários pelo menos q estágios para obter um método de Runge-Kutta de ordem q. A demonstração exige definições que vão além do escopo dessas notas, mas pode ser encontrada em [5].

3.5 Exercícios resolvidos

Exercício Resolvido 3.1. *Determine os parâmetros a_2, c_1 e c_2 do Método de Runge-Kutta com 2 estágios de modo que este tenha ordem máxima. Mostre que a ordem não pode exceder dois e deduza os métodos de Euler Aprimorado*

$$\begin{cases} y_0 = y(t_0), \\ y_{k+1} = y_k + h\Phi(t_k, y_k, h), \quad t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1, \end{cases}$$

$$\Phi(t_k, y_k, h) = \frac{\kappa_1 + \kappa_2}{2} \quad e \quad \begin{array}{l} \kappa_1 = f(t_k, y_k), \\ \kappa_2 = f(t_k + h, y_k + h\kappa_1), \end{array}$$

e de Euler Modificado

$$\begin{cases} y_0 = y(t_0), \\ y_{k+1} = y_k + h\Phi(t_k, y_k, h), \quad t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1, \end{cases}$$

$$\Phi(t_k, y_k, h) = \kappa_2 \quad e \quad \begin{array}{l} \kappa_1 = f(t_k, y_k), \\ \kappa_2 = f(t_k + \dfrac{h}{2}, y_k + \dfrac{h}{2}\kappa_1), \end{array}$$

ambos com $h = \dfrac{b-a}{n}$.

Solução:

O Método da Série de Taylor de ordem 3 é dado por

$$y_{k+1} = y_k + hy^{(1)}(t_k) + \frac{h^2}{2!}y^{(2)}(t_k) + \frac{h^3}{3!}y^{(3)}(t_k) + O(h^4)$$

onde as derivadas são dadas por

$$y^{(1)}(t) = f(t,y), \quad y^{(2)}(t) = f_t(t,y) + f_y(t,y)f(t,y) \quad e \quad y^{(3)}(t) = f_{tt}(t,y) +$$

$$+2f_{ty}(t,y)f(t,y) + f_{yy}(t,y)f^2(t,y) + f_t(t,y)f_y(t,y) + f_y^2(t,y)f(t,y).$$

Em uma única expressão, tem-se

$$\begin{aligned} y_{k+1} &= y_k + hf(t_k,y_k) + \frac{h^2}{2!}\left[f_t(t_k,y_k) + f_y(t_k,y_k)f(t_k,y_k)\right] + \\ &+ \frac{h^3}{3!}\left[f_{tt}(t_k,y_k) + 2f_{ty}(t_k,y_k)f(t_k,y_k) + f_{yy}(t_k,y_k)f^2(t_k,y_k)\right] + \\ &+ \frac{h^3}{3!}\left[f_t(t_k,y_k)f_y(t_k,y_k) + f_y^2(t_k,y_k)f(t_k,y_k)\right] + O(h^4). \end{aligned} \tag{3.5.43}$$

Um Método de Runge-Kutta de 2 estágios se escreve como

$$y_{k+1} = y_k + h\left(c_1\kappa_1 + c_2\kappa_2\right), \tag{3.5.44}$$

com $\kappa_1 = f(t_k, y_k)$ *e* $\kappa_2 = f(t_k + a_2h, y_k + hb_{21}\kappa_1)$. *Como* $a_2 = b_{21}$, *tem-se para* κ_2

$$\kappa_2 \quad = \quad f(t_k + a_2h, y_k + ha_2\kappa_1). \tag{3.5.45}$$

Truncando-se a expansão de (3.5.45) em Série de Taylor ao redor de (t_k, y_k) *nos termos de ordem 2, obtém-se*

$$\begin{aligned}\kappa_2 &= f(t_k, y_k) + a_2h\left[f_t(t_k, y_k) + \kappa_1 f_y(t_k, y_k)\right] + \\ &+ \frac{a_2^2h^2}{2!}\left[f_{tt}(t_k, y_k) + 2\kappa_1 f_{ty}(t_k, y_k) + \kappa_1^2 f_{yy}(t_k, y_k)\right].\end{aligned} \tag{3.5.46}$$

Após reorganização em potências de h, *a substituição de (3.5.46) em (3.5.44) resulta em*

$$\begin{aligned}y_{k+1} &= y_k + (c_1 + c_2)hf(t_k, y_k) + \\ &+ a_2c_2h^2\left[f_t(t_k, y_k) + f(t_k, y_k)f_y(t_k, y_k)\right] + \\ &+ \frac{a_2^2c_2h^3}{2!}\left[f_{tt}(t_k, y_k) + 2f(t_k, y_k)f_{ty}(t_k, y_k) + f^2(t_k, y_k)f_{yy}(t_k, y_k)\right].\end{aligned} \tag{3.5.47}$$

Comparando-se (3.5.43) e (3.5.47), conclui-se que

$$\begin{cases} c_1 + c_2 &= 1, \\ a_2c_2 &= \dfrac{1}{2!}. \end{cases}$$

Assim, se $c_1 = c_2 = \dfrac{1}{2}$ *então* $a_2 = 1$ *e tem-se o Método de Euler Aprimorado. Se* $c_1 = 0$ *e* $c_2 = 1$ *então* $a_2 = \dfrac{1}{2}$ *e tem-se o Método de Euler Modificado. Não é possível estabelecer concordância entre os termos de terceira ordem de (3.5.43) e (3.5.47). Em (3.5.47) há três termos com o fator* h^3, *enquanto que em (3.5.43) há cinco termos com esse fator. Portanto, a ordem não pode exceder dois.*

Exercício Resolvido 3.2. *Obtenha um método de Runge-Kutta de três estágios de terceira ordem para o qual* $c_2 = c_3$ *e* $a_2 = a_3$.
Solução:

$$\begin{aligned}y_{k+1} &= y_k + h\sum_{r=1}^{3} c_r\kappa_r = \\ &= y_k + h\left(c_1\kappa_1 + c_2\kappa_2 + c_3\kappa_3\right).\end{aligned}$$

Como $c_2 = c_3$, *tem-se que*

$$y_{k+1} = y_k + h\left(c_1\kappa_1 + c_2\kappa_2 + c_2\kappa_3\right), \tag{3.5.48}$$

sendo

$$\begin{aligned}\kappa_1 &= f(t, y), \\ \kappa_2 &= f(t + ha_2, y + hb_{21}\kappa_1), \\ \kappa_3 &= f(t + ha_3, y + hb_{31}\kappa_1 + hb_{32}\kappa_2).\end{aligned}$$

De $a_r = \sum_{s=1}^{r-1} b_{rs}$, $0 < r \leq 3$, *e* $a_2 = a_3$, *obtêm-se* $a_2 = b_{21}$, $a_3 = b_{31} + b_{32}$ *e* $a_2 = b_{31} + b_{32}$, *o que implica que* $b_{31} = a_2 - b_{32}$.

Assim,

$$\begin{aligned} \kappa_1 &= f(t,y), \\ \kappa_2 &= f(t + ha_2, y + ha_2\kappa_1), \\ \kappa_3 &= f(t + ha_2, y + h(a_2 - b_{32})\kappa_1 + hb_{32}\kappa_2). \end{aligned}$$

Considere que $f(t,y)$ *tenha derivadas parciais contínuas até a ordem necessária, então expande-se* κ_2 *e* κ_3 *em torno do ponto* (t,y) *até o termo de segunda ordem. Durante a expansão, todos os termos de terceira ordem são desprezados porque na substituição em (3.5.48) esses termos acabam compondo termos de ordem quatro.*

- *Expansão de* κ_2 *(com* $\kappa_1 = f$*):*

$$\begin{aligned} \kappa_2 &= f + ha_2(f_t + ff_y) + \frac{h^2a_2^2}{2}(f_{tt} + 2ff_{ty} + f^2f_{yy}) = \\ &= f + ha_2F + \frac{h^2a_2^2}{2}G, \end{aligned} \tag{3.5.49}$$

onde $F = f_t + ff_y$ *e* $G = f_{tt} + 2ff_{ty} + f^2f_{yy}$.

- *Expansão de* κ_3 *(com* $\kappa_1 = f$*):*

$$\begin{aligned} \kappa_3 &= f + h\left\{a_2f_t + \left[(a_2 - b_{32})f + b_{32}\kappa_2\right]f_y\right\} + \\ &+ \frac{h^2}{2}\{a_2^2f_{tt} + 2a_2\left[(a_2 - b_{32})f + b_{32}\kappa_2\right]f_{ty} + \\ &+ \left[(a_2 - b_{32})f + b_{32}\kappa_2\right]^2 f_{yy}\} = \\ &= f + h\boxed{A} + \frac{h^2}{2}\boxed{B}, \end{aligned}$$

onde $\boxed{A} = \{a_2f_t + \left[(a_2 - b_{32})f + b_{32}\kappa_2\right]f_y\}$ *e*

$\boxed{B} = \{a_2^2f_{tt} + 2a_2\left[(a_2 - b_{32})f + b_{32}\kappa_2\right]f_{ty} + \left[(a_2 - b_{32})f + b_{32}\kappa_2\right]^2 f_{yy}\}$.

Desenvolvimento de $\boxed{A}$*:*

$$\begin{aligned} \boxed{A} &= a_2f_t + \left[(a_2 - b_{32})f + b_{32}\kappa_2\right]f_y \\ &= a_2f_t + a_2ff_y - b_{32}ff_y + b_{32}f_y\left(f + ha_2F + \frac{h^2a_2^2}{2}G\right) \\ &= a_2f_t + a_2ff_y - b_{32}ff_y + b_{32}ff_y + ha_2b_{32}Ff_y + \frac{h^2a_2^2}{2}b_{32}Gf_y \\ &= a_2(f_t + ff_y) - b_{32}ff_y + b_{32}ff_y + ha_2b_{32}Ff_y + \frac{h^2a_2^2}{2}b_{32}Gf_y \\ &= a_2F + ha_2b_{32}Ff_y + \frac{h^2a_2^2b_{32}}{2}Gf_y. \end{aligned}$$

Assim,

$$\boxed{A} = a_2F + ha_2b_{32}Ff_y + \frac{h^2a_2^2b_{32}}{2}Gf_y. \tag{3.5.50}$$

Desenvolvimento de $\boxed{B}$*:*

$$\begin{aligned}\boxed{B} &= a_2^2f_{tt} + 2a_2\left[(a_2-b_{32})f + b_{32}\kappa_2\right]f_{ty} + \\ &+ \left[(a_2-b_{32})f + b_{32}\kappa_2\right]^2 f_{yy} = \\ &= a_2^2f_{tt} + 2a_2^2ff_{ty} - 2a_2b_{32}ff_{ty} + \underbrace{2a_2b_{32}f_{ty}\kappa_2}_{B_1} + \\ &+ \underbrace{\left[(a_2-b_{32})f + b_{32}\kappa_2\right]^2 f_{yy}}_{B_2}.\end{aligned} \tag{3.5.51}$$

Substituindo-se (3.5.49) em B_1 *e* B_2*, chega-se a:*

$$\begin{aligned}B_1 &= 2a_2b_{32}ff_{ty} + 2ha_2^2b_{32}Ff_{ty}, \\ B_2 &= a_2^2f^2f_{yy} + 2ha_2^2b_{32}Fff_{yy}.\end{aligned}$$

Em B_1 *e* B_2*, descartaram-se termos com fator* h^p*,* $p > 1$*. Substituindo-se* B_1 *e* B_2 *em (3.5.51), obtém-se*

$$\boxed{B} = a_2^2G + 2ha_2^2b_{32}Ff_{ty} + 2ha_2^2b_{32}Fff_{yy}. \tag{3.5.52}$$

Substituindo-se (3.5.50) e (3.5.52) em (3.5.50), tem-se

$$\begin{aligned}\kappa_3 &= f + h\left(a_2F + ha_2b_{32}Ff_y + \frac{h^2a_2^2b_{32}}{2}Gf_y\right) \\ &+ \frac{h^2}{2}\left(a_2^2G + 2ha_2^2b_{32}Ff_{ty} + 2ha_2^2b_{32}Fff_{yy}\right) = \\ &= f + ha_2F + \frac{h^2}{2}\left(a_2^2G + 2a_2b_{32}Ff_y\right).\end{aligned} \tag{3.5.53}$$

Em (3.5.53), descartou-se o termo com fator h^3*.*

Substituindo κ_1*,* κ_2 *e* κ_3 *em (3.5.48), tem-se*

$$\begin{aligned}y_{k+1} &= y_k + h\left[c_1f + c_2\left(f + ha_2F + \frac{h^2a_2^2}{2}G\right)\right] + \\ &+ hc_2\left[f + ha_2F + \frac{h^2}{2}\left(a_2^2G + 2a_2b_{32}Ff_y\right)\right] = \\ &= y_k + h(c_1+2c_2)f + 2h^2a_2c_2F + h^3\left(a_2^2c_2 + a_2b_{32}c_2Ff_y\right).\end{aligned} \tag{3.5.54}$$

O Método da Série de Taylor de terceira ordem é dado pela expressão

$$y_{k+1} = y_k + hf + \frac{h^2}{2}F + \frac{h^3}{6}\left(G + Ff_y\right). \tag{3.5.55}$$

Para que (3.5.54) coincida com (3.5.55), deve-se ter

$$\begin{cases} c_1 + 2c_2 = 1 \\ 2a_2c_2 = \dfrac{1}{2} \\ a_2^2c_2 = \dfrac{1}{6} \\ a_2^2b_{32}c_2 = \dfrac{1}{6} \end{cases} \Rightarrow \begin{cases} a_2 = \dfrac{2}{3} \\ b_{32} = \dfrac{2}{3} \\ c_1 = \dfrac{1}{4} \\ c_2 = \dfrac{3}{8} \end{cases}.$$

Com os valores anteriores, tem-se $a_3 = \dfrac{2}{3}$, $c_3 = \dfrac{3}{8}$ *e* $b_{31} = 0$. *Assim, o Método de Runge-Kutta de terceira ordem, com as condições impostas, é dado por*

$$y_{k+1} = y_k + \frac{h}{4}\left(\kappa_1 + \frac{3}{2}\kappa_2 + \frac{3}{2}\kappa_3\right),$$

com $\kappa_1 = f(t,y)$, $\kappa_2 = f\left(t + \dfrac{2h}{3}, y + \dfrac{2h}{3}\kappa_1\right)$ *e* $\kappa_3 = f\left(t + \dfrac{2h}{3}, y + \dfrac{2h}{3}\kappa_2\right)$.

Exercício Resolvido 3.3. *Mostre que o Método de Runge-Kutta 44 clássico,*

$$\begin{cases} y_0 = y(t_0), \\ y_{k+1} = y_k + h\Phi(t_k, y_k, h), \quad t_{k+1} = t_k + h, \quad 0 \le k \le n-1, \end{cases}$$

onde $\Phi(t_k, y_k, h) = \frac{1}{6}(\kappa_1 + 2\kappa_2 + 2\kappa_3 + \kappa_4)$, *com*

$$\begin{aligned} \kappa_1 &= f(t_k, y_k), \\ \kappa_2 &= f\left(t_k + \tfrac{h}{2}, y_k + \tfrac{h}{2}\kappa_1\right), \\ \kappa_3 &= f\left(t_k + \tfrac{h}{2}, y_k + \tfrac{h}{2}\kappa_2\right), \\ \kappa_4 &= f(t_k + h, y_k + h\kappa_3) \end{aligned}$$

e $h = \dfrac{(b-a)}{n}$, *aplicado à equação diferencial* $\dot{y} = py$, p *constante, fornece*

$$\frac{y_{k+1}}{y_k} = e^{ph} + O(ph)^5.$$

Solução:

Calculando-se κ_1, κ_2, κ_3 *e* κ_4 *para a equação* $\dot{y} = py$, *têm-se*

$$\begin{aligned} \kappa_1 &= f(t_k, y_k) = py_k, \\ \kappa_2 &= f\left(t_k + \frac{h}{2}, y_k + \frac{h}{2}\kappa_1\right) = p\left(y_k + \frac{h}{2}py_k\right) = \left(p + \frac{p^2h}{2}\right)y_k, \\ \kappa_3 &= f\left(t_k + \frac{h}{2}, y_k + \frac{h}{2}\kappa_2\right) = p\left[y_k + \frac{h}{2}\left(p + \frac{p^2h}{2}\right)y_k\right] = \\ &= \left(p + \frac{p^2h}{2} + \frac{p^3h^2}{4}\right)y_k, \\ \kappa_4 &= f(t_k + h, y_k + h\kappa_3) = p\left[y_k + h\left(p + \frac{p^2h}{2} + \frac{p^3h^2}{4}\right)y_k\right] = \\ &= \left(p + p^2h + \frac{p^3h^2}{2} + \frac{p^4h^3}{4}\right)y_k. \end{aligned}$$

Aplicando-se os resultados acima no Método de Runge-Kutta de quarta ordem, obtém-se

$$y_{k+1} = \left(1 + ph + \frac{p^2h^2}{2} + \frac{p^3h^3}{6} + \frac{p^4h^4}{24}\right) y_k,$$

donde

$$\frac{y_{k+1}}{y_k} = 1 + ph + \frac{p^2h^2}{2!} + \frac{p^3h^3}{3!} + \frac{p^4h^4}{4!}. \tag{3.5.56}$$

Como

$$e^{ph} = \sum_{n=0}^{\infty} \frac{(ph)^n}{n!} = 1 + ph + \frac{p^2h^2}{2!} + \frac{p^3h^3}{3!} + \frac{p^4h^4}{4!} + O\left((ph)^5\right),$$

pode-se reescrever (3.5.56) como

$$\frac{y_{k+1}}{y_k} = e^{ph} + O(p^5h^5).$$

Exercício Resolvido 3.4. *Considere o problema de valor inicial*

$$\begin{cases} \dot{y} = e^{2t}, \ 0 \leq t \leq 1, \\ y(0) = \frac{1}{2}. \end{cases}$$

(a) Estime h para que o erro de discretização local para o Método de Euler seja menor que 10^{-4}.

(b) A estimativa para h obtida no item (a) pode ser usada como aproximação inicial do passo de integração para o Método de Runge-Kutta-Fehlberg? Justifique.

Solução:

(a) O Método de Euler

$$\begin{cases} y_0 = y(t_0) \\ y_{k+1} = y_k + hf(t_k, y_k) \end{cases}$$

tem erro local de discretização dado por

$$\alpha_k = \frac{y(t_{k+1}) - y(t_k)}{h} - f\left((t_k, y(t_k)\right). \tag{3.5.57}$$

Truncando-se no segundo termo a expansão em série de Taylor de $y(t_{k+1})$ com centro em $t = t_k$, obtém-se

$$y(t_{k+1}) = y(t_k) + (t_{k+1} - t_k)y'(t_k) + \frac{(t_{k+1} - t_k)^2}{2!}y''(\xi), \tag{3.5.58}$$

com $\xi \in (t_k, t_{k+1})$.

Denotando-se $h = t_{k+1} - t_k$ e substituindo-se (3.5.58) em (3.5.57), tem-se que

$$\begin{aligned} \alpha_k &= \frac{y(t_k) + hy'(t_k) + \frac{h^2}{2!}y''(\xi) - y(t_k)}{h} - f\left(t_k, y(t_k)\right) \\ &= y'(t_k) + \frac{h}{2!}y''(\xi) - f\left(t_k, y(t_k)\right) \\ &= \frac{h}{2}y''(\xi). \end{aligned} \tag{3.5.59}$$

Assim,

$$\begin{aligned} |\alpha_k| &= \frac{h}{2}\,|y''(\xi)|\,, \\ \max_k |\alpha_k| &\leq \frac{h}{2}\max_{\xi\in I}|y''(\xi)|, \\ \frac{h}{2}\max_{\xi\in I}|y''(\xi)| &< 10^{-4}, \\ h &< \frac{2.10^{-4}}{\max\limits_{\xi\in I}|y''(\xi)|}. \end{aligned} \tag{3.5.60}$$

Ainda,

$$\begin{aligned} I &= [0,1], \\ y(t) &= \frac{e^{2t}}{2}, \\ y'(t) &= e^{2t}, \\ y''(t) &= 2e^{2t}. \end{aligned} \tag{3.5.61}$$

Como a função (3.5.61) é estritamente crescente no intervalo $[0,1]$*, assumindo o valor máximo* $2e^2$ *em* $t=1$*, pode-se reescrever (3.5.60) como*

$$h < \frac{2.10^{-4}}{2e^2},$$

$$\boxed{h < \frac{10^{-4}}{e^2} \approx 1,35335x10^{-5}.} \tag{3.5.62}$$

(b) Uma aproximação para o erro local de discretização do Método de Runge-Kutta-Felhberg é dada por

$$\alpha_k^{RKF} \approx \frac{1}{360}\kappa_1 - \frac{128}{4275}\kappa_3 - \frac{2197}{75240}\kappa_4 + \frac{1}{50}\kappa_5 + \frac{2}{55}\kappa_6. \tag{3.5.63}$$

Calculando (3.5.63) com $f(t,y(t)) = e^{2t}$*,* $t_0 = 0$ *e* $h = \dfrac{10^{-4}}{e^2}$*, obtém-se*

$$\begin{aligned} \kappa_1 &= (t_0, y_0) = e^{2t_0} = e^0 = 1, \\ \kappa_2 &= f\left(t_0 + \frac{h}{4}, y_0 + \frac{h}{4}\kappa_1\right) \\ &= e^{2\left(t_0+\frac{h}{4}\right)} = e^{\frac{h}{2}} \approx 1,000006767, \\ \kappa_3 &= f\left(t_0 + \frac{3h}{8}, y_0 + \frac{3h}{32}\kappa_1 + \frac{9h}{32}\kappa_2\right) \\ &= e^{2\left(t_0+\frac{3h}{8}\right)} = e^{\frac{3h}{4}} \approx 1,00001015, \end{aligned}$$

$$\begin{aligned}\kappa_4 &= f\left(t_0+\frac{12h}{13}, y_0+\frac{1932h}{2197}\kappa_1-\frac{7200h}{2197}\kappa_2+\frac{7296h}{2197}\kappa_3\right)\\ &= e^{2\left(t_0+\frac{12h}{13}\right)}=e^{\frac{24h}{13}}\approx 1,000024985,\end{aligned}$$

$$\begin{aligned}\kappa_5 &= f\left(t_0+h, y_0+\frac{439h}{216}\kappa_1-8h\kappa_2+\frac{3680h}{513}\kappa_3-\frac{845h}{4104}\kappa_4\right)\\ &= e^{2(t_0+h)}=e^{2h}\approx 1,000027067,\end{aligned}$$

$$\begin{aligned}\kappa_6 &= f\left(t_0+\frac{h}{2}, y_0-\frac{8h}{27}\kappa_1+2h\kappa_2-\frac{3544h}{2565}\kappa_3+\frac{1859h}{4104}\kappa_4-\frac{11h}{40}\kappa_5\right)\\ &= e^{2\left(t_0+\frac{h}{2}\right)}=e^{h}\approx 1,000013534,\end{aligned}$$

$$\alpha_0^{RKF} \approx 1,96784x10^{-11}. \tag{3.5.64}$$

Como no Método de Runge-Kutta Fehlberg tem-se que

$$\left|\alpha_0^{RKF}\right| = 1,96784x10^{-11} < \epsilon = 10^{-4}, \tag{3.5.65}$$

pode-se usar $\dfrac{10^{-4}}{e^2}$ *como valor inicial para o passo de integração* h*. O mesmo se verifica para* $h < \dfrac{10^{-4}}{e^2}$*.*

Capítulo 4

Estabilidade absoluta dos métodos de passo único

Considere o Problema de Cauchy

$$\left\{ \begin{array}{rcl} \dfrac{d}{dt}y(t) & = & -10y(t), \quad t \in [2,6], \\ y(2) & = & 1000, \end{array} \right. \tag{4.0.1}$$

cuja solução exata é $y(t) = 1000e^{-10t+20}$ (verifique!). A obtenção de uma solução numérica para (4.0.1) demanda, na prática, a escolha de um passo de integração $h > 0$ além, é claro, de um método numérico. Por exemplo, considere as aproximações obtidas com o Método de Euler (1.4.23),

$$\begin{array}{rcl} y_{k+1} & = & y_k + hf\left(t_k, y_k\right) = \\ & = & y_k + h\left(-10y_k\right) = \\ & = & (1-10h)y_k, \end{array}$$

com passos de integração $h = 0,125$ e $h = 0,5$. A Tabela 4.1 mostra o erro de discretização global para cada um destes passos de integração.

			h=0,125	h=0,5
k	t_k	$y(t_k)$	$\lvert y(t_k) - y_k \rvert$	$\lvert y(t_k) - y_k \rvert$
0	2,0	1,000E+03	0,000E+00	0,000E+00
1	2,5	6,738E+00	2,832E+00	4,007E+03
2	3,0	4,540E-02	3,014E-02	1,600E+04
3	3,5	3,059E-04	2,463E-04	6,400E+04
4	4,0	2,061E-06	1,828E-06	2,600E+05
5	4,5	1,389E-08	1,298E-08	1,024E+06
6	5,0	9,358E-11	9,003E-11	4,096E+06
7	5,5	6,305E-13	6,166E-13	1,638E+07
8	6,0	4,248E-15	4,194E-15	6,554E+07

Tabela 4.1: Erros de discretização global produzidos pelo Método de Euler na solução numérica de (4.0.1) para dois tamanhos de passos de integração.

Da análise da Tabela 4.1, constata-se que o Método de Euler produz um erro de discretização global aceitável para $h = 0,125$, porém inadmissível para $h = 0,5$.

Qual é então o motivo para o comportamento da solução numérica em ambos os casos? Como escolher, na prática, um passo de integração que permita uma análise confiável dos resultados obtidos? A escolha apropriada do passo de integração $h > 0$ está associada ao conceito de *estabilidade absoluta* ($h > 0$, fixado).

4.1 Estabilidade absoluta

Para compreender a origem dos problemas de estabilidade que se pode ter na escolha de um passo de integração $h > 0$ para um determinado método numérico, considere o Método de Euler

$$\begin{cases} y_0 &= y(t_0), \\ y_{k+1} &= y_k + hf(t_k, y_k), \end{cases}$$

aplicado ao Problema de Cauchy modelo

$$\begin{cases} \dfrac{d}{dt}y(t) &= \lambda y, \\ y(t_0) &= y_0, \end{cases}$$

cuja solução exata é dada por $y(t) = y_0 e^{\lambda(t-t_0)}$ (verifique!). Assim, tem-se

$$\begin{aligned} y_{k+1} &= y_k + hf(t_k, y_k) = \\ &= y_k + \lambda h y_k = \\ &= (1 + \lambda h) y_k. \end{aligned} \tag{4.1.2}$$

Estabelecendo em (4.1.2) uma dependência da condição inicial y_0, chega-se a

$$\begin{aligned} y_1 &= (1 + \lambda h)\, y_0, \\ y_2 &= (1 + \lambda h)^2\, y_0, \\ y_3 &= (1 + \lambda h)^3\, y_0, \\ &\vdots \\ y_k &= (1 + \lambda h)^k\, y_0, \end{aligned} \tag{4.1.3}$$

onde o parâmetro λ pode ser real ou complexo. O fator $(1 + \lambda\, h)$ é denominado *fator de amplificação*.

Há duas situações possíveis para $\lambda \in \mathbb{R}$:

1. Se $\lambda < 0$ então $\lim\limits_{t\to\infty} y(t) = \lim\limits_{t\to\infty} y_0 e^{\lambda(t-t_0)} = 0$.

   A solução numérica y_k terá esse comportamento se, e só se,

   $$|1 + \lambda h| < 1 \quad \Rightarrow \quad -1 < 1 + \lambda h < 1 \Rightarrow -2 < \lambda h < 0.$$

   Assim, a solução numérica tem o mesmo comportamento da solução exata se, e só se, $\lambda h \in (-2, 0)$.

2. Se $\lambda \geq 0$,

   $$\lim_{t\to\infty} y(t) = \lim_{t\to\infty} y_0 e^{\lambda(t-t_0)} = \begin{cases} y_0, & \text{se} \quad \lambda = 0, \\ +\infty, & \text{se} \quad \lambda > 0, \quad y_0 > 0, \\ -\infty, & \text{se} \quad \lambda > 0, \quad y_0 < 0. \end{cases}$$

Neste caso, o *valor absoluto* da solução numérica

$$y_k = (1 + \lambda h)^k \, y_0$$

apresenta sempre o mesmo comportamento da solução exata: tende ao infinito.

Diz-se que o Método de Euler é um método *condicionalmente estável* cujo intervalo de estabilidade absoluta é dado por $\lambda h \in (-2, 0)$, $\lambda < 0$.

É interessante observar que para $\lambda \in \mathbb{C}$ tem-se $\lambda h = z = a + b\,i \in \mathbb{C}$, $a, b \in \mathbb{R}$, e que, portanto,

$$\begin{aligned} z &= \lambda h = a + bi, \;\; a, b \in \mathbb{R} \\ |1 + \lambda h| < 1 &\Rightarrow |z + 1| < 1 \\ &\Leftrightarrow |a + bi + 1| < 1 \\ &\Leftrightarrow |(a + 1) + bi| < 1 \\ &\Leftrightarrow \sqrt{(a+1)^2 + b^2} < 1 \\ &\Leftrightarrow (a + 1)^2 + b^2 < 1 \\ &\Leftrightarrow [a - (-1)]^2 + b^2 < 1. \end{aligned} \tag{4.1.4}$$

Em (4.1.4), tem-se o conjunto dos pontos interiores a um disco centrado no ponto $(-1, 0)$ e de raio 1. O intervalo de estabilidade é definido por $Re\,(\lambda h) \in (-2, 0)$. A parte imaginária, $Im\,(\lambda h)$, é responsável pelo comportamento oscilatório da solução. Representamos a região de estabilidade (no plano complexo) na Figura 4.1.

Definição 4.1 (Estabilidade absoluta)**.** *Seja um método de passo único que, aplicado ao Problema de Cauchy modelo*

$$\begin{cases} \dfrac{d}{dt} y(t) = \lambda y(t), \quad \lambda \in \mathbb{C}, \\ \\ y(t_0) = y_0, \end{cases}$$

conduz a

$$y_{k+1} = \psi\,(\lambda h)\, y_k.$$

O conjunto

$$\Omega = \{\mu \in \mathbb{C}\,| \quad |\psi\,(\mu)| < 1\}$$

é denominado região de estabilidade absoluta ($h > 0$, fixado) e $\psi\,(\lambda h)$ é o fator de amplificação. A intersecção da região Ω com a reta real determina o intervalo de estabilidade absoluta do método de passo único.

No intervalo de estabilidade absoluta, solução exata e numérica apresentam qualitativamente (em algum sentido) o mesmo comportamento. A Tabela 4.2 apresenta os intervalos de estabilidade absoluta para alguns métodos de Runge-Kutta. É possível mostrar que, todos os métodos com R estágios e ordem $p = R$ têm o mesmo intervalo de estabilidade absoluta [16]. Por exemplo, o método do ponto médio explícito e o método do trapézio explícito, ambos métodos de Runge-Kutta de 2 estágios e ordem 2, possuem o mesmo intervalo de estabilidade absoluta, descrito na Tabela 4.2. Verifique na Figura 4.1 as regiões de estabilidade absoluta para métodos de Runge-Kutta de R estágios e ordem R.

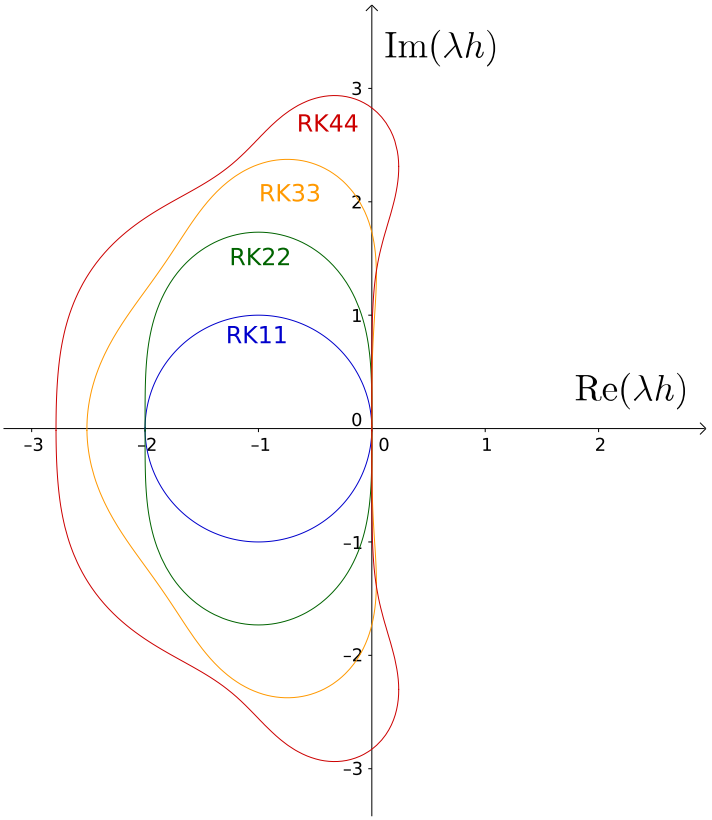


Figura 4.1: Região de estabilidade para métodos de Runge-Kutta. O interior das curvas fechadas indicam λh tais que $|\psi(\lambda h)| < 1$.

Exemplo 4.1 (Euler Implícito)**.** *Considere o método de Euler implícito, para o qual temos iterações da forma*

$$y_{k+1} = y_k + hf(t_{k+1}, y_{k+1}).$$

Substituindo no problema modelo $y' = \lambda y$*, temos*

$$y_{k+1} = y_k + \lambda h y_{k+1}.$$

Re-organizando os termos e isolando y_{k+1} *notamos que*

$$\psi(\lambda h) = \frac{1}{1 - \lambda h}.$$

Para $\lambda < 0$ *e* $h > 0$*,* $|\psi(\lambda h)| < 1$*, portanto o método é incondicionalmente absolutamente estável. No plano complexo, a sua região de estabilidade é o complementar do círculo centrado em* $(1, 0)$ *com raio* 1*, ilustrado na Figura 4.1.*

4.1.1 Exercícios

Exercício 4.1. *Explique o comportamento observado para o Método de Euler aplicado à solução do p.v.i. (4.0.1) com passos de integração* $h = 0,5$ *e* $h = 0,125$*.*

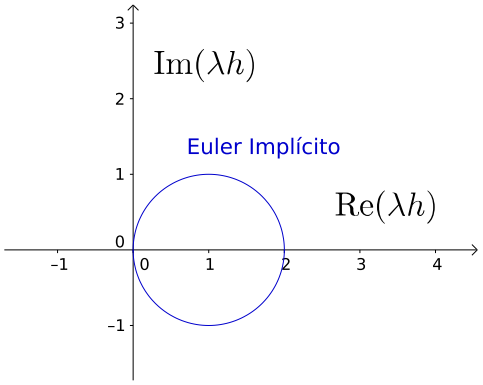


Figura 4.2: Região de estabilidade para o método de Euler Implícito. O interior do círculo indica λh tais que $|\psi(\lambda h)| > 1$, portanto a região que garante estabilidade absoluta é tudo que está fora do círculo.

Exercício 4.2. *A Tabela 4.2 traz o fator de amplificação e o intervalo de estabilidade absoluta para os Métodos de Runge-Kutta de ordem R com R estágios [16]. Verifique que, de fato, os fatores de amplificação estão corretos. Além disso, verifi-*

R	Fator de amplificação	Intervalo Ω
1	$1+\lambda h$	(-2 , 0)
2	$1+\lambda h+\frac{(\lambda h)^2}{2}$	(-2 , 0)
3	$1+\lambda h+\frac{(\lambda h)^2}{2}+\frac{(\lambda h)^3}{6}$	(-2,51 , 0)
4	$1+\lambda h+\frac{(\lambda h)^2}{2}+\frac{(\lambda h)^3}{6}+\frac{(\lambda h)^4}{24}$	(-2,78 , 0)

Tabela 4.2: Intervalos de estabilidade absoluta para Métodos de Runge-Kutta de ordem R com R estágios [16].

que que o método do trapézio explícito (Euler Aprimorado) (1.4.25) *e o método do ponto médio explícito (Euler Modificado)* (3.2.11) *possuem o mesmo intervalo de estabilidade absoluta.*

Exercício 4.3. *Comprove o efeito da estabilidade (ou instabilidade) usando o Método de Runge-Kutta de quarta ordem com quatro estágios (3.2.24) para calcular a solução numérica do problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt}y(t) &= -5ty^2(t)+\dfrac{5}{t}-\dfrac{1}{t^2}, \quad 1\le t\le 4, \\ y(1) &= 1, \end{cases} \tag{4.1.5}$$

com $h=0,2$ e $h=0,4$. Sugestão: Mostre que a solução exata do problema de valor inicial (4.1.5) é $y(t)=\dfrac{1}{t}$. Com a solução exata e a solução numérica, calcule o erro global de discretização.

4.2 Suplemento teórico

4.2.1 Instabilidade inerente

Seja o problema de valor inicial

$$\begin{cases} \dfrac{d}{dt}y(t) = y(t) - t, \quad t \in [0,5] \\ \\ y(0) = 1 \end{cases} . \tag{4.2.6}$$

Em (4.2.6), tem-se uma equação diferencial ordinária linear, de primeira ordem, não homogênea. A solução exata de (4.2.6) é dada por

$$y(t) = t + 1, \tag{4.2.7}$$

que se obtém usando o fator integrante e^{-t}.

Perturbando a condição inicial em (4.2.6) em 1%, isto é, $y(0) = 1 \pm 0,01$, tem-se os problemas de valor inicial

$$\begin{cases} \dfrac{d}{dt}y(t) = y(t) - t, \quad t \in [0,5], \\ \\ y(0) = 0,99, \end{cases} \tag{4.2.8}$$

e

$$\begin{cases} \dfrac{d}{dt}y(t) = y(t) - t, \quad t \in [0,5], \\ \\ y(0) = 1,01. \end{cases} \tag{4.2.9}$$

As soluções exatas de (4.2.8) e (4.2.9) são dadas, respectivamente, por

$$y(t) = -0,01e^t + t + 1 \quad \text{e} \tag{4.2.10}$$

$$y(t) = 0,01e^t + t + 1. \tag{4.2.11}$$

As soluções (4.2.7), (4.2.10) e (4.2.11) em $t = 5$ valem, respectivamente,

$$\begin{aligned} y(5) &= 5 + 1 = 6 \\ y(5) &= -0,01e^5 + 5 + 1 \approx 4,5 \\ y(5) &= 0,01e^5 + 5 + 1 \approx 7,5. \end{aligned}$$

Como pode ser visto, perturbando-se a condição inicial em (4.2.6) em 1%, a solução varia cerca de 25%. No exemplo dado, nenhum método numérico será capaz de produzir um erro inferior a 25% se a condição inicial for perturbada em 1%. Este é um problema de estabilidade intrínseco ao problema de valor inicial e, por este motivo, denominado de *instabilidade inerente*.

4.3 Exercícios resolvidos

Exercício Resolvido 4.1. *Determine a região de estabilidade absoluta para o Método do Trapézio (Implícito).*
Solução:

Tome o problema-modelo

$$\left\{\begin{array}{rcl} \dfrac{d}{dt}y(t) & = & \lambda y, \quad \lambda < 0, \\ y(t_0) & = & y_0, \end{array}\right.$$

cuja solução exata é

$$y(t) = y_0 e^{\lambda(t-t_0)}.$$

Aplicando-se o Método do Trapézio, obtém-se

$$\begin{array}{rcl} y_{k+1} & = & y_k + h\Phi(t_k, y_k, y_{k+1}, h) \\ y_{k+1} & = & y_k + h\dfrac{f(t_k, y_k) + f(t_{k+1}, y_{k+1})}{2} \\ y_{k+1} & = & y_k + h\dfrac{\lambda y_k + \lambda y_{k+1}}{2} \\ \left(1 - \dfrac{\lambda h}{2}\right) y_{k+1} & = & \left(1 + \dfrac{\lambda h}{2}\right) y_k \\ y_{k+1} & = & \left(\dfrac{2+\lambda h}{2-\lambda h}\right) y_k. \end{array} \tag{4.3.12}$$

Estabelecendo-se em (4.3.12) uma dependência da condição inicial y_0,

$$\begin{array}{rcl} y_1 & = & \left(\dfrac{2+\lambda h}{2-\lambda h}\right) y_0, \\ y_2 & = & \left(\dfrac{2+\lambda h}{2-\lambda h}\right) y_1 = \left(\dfrac{2+\lambda h}{2-\lambda h}\right)\left(\dfrac{2+\lambda h}{2-\lambda h}\right) y_0 = \left(\dfrac{2+\lambda h}{2-\lambda h}\right)^2 y_0, \\ y_3 & = & \left(\dfrac{2+\lambda h}{2-\lambda h}\right) y_2 = \left(\dfrac{2+\lambda h}{2-\lambda h}\right)\left(\dfrac{2+\lambda h}{2-\lambda h}\right)^2 y_0 = \left(\dfrac{2+\lambda h}{2-\lambda h}\right)^3 y_0, \\ y_4 & = & \left(\dfrac{2+\lambda h}{2-\lambda h}\right) y_3 = \left(\dfrac{2+\lambda h}{2-\lambda h}\right)\left(\dfrac{2+\lambda h}{2-\lambda h}\right)^3 y_0 = \left(\dfrac{2+\lambda h}{2-\lambda h}\right)^4 y_0, \\ \vdots & & \\ y_k & = & \left(\dfrac{2+\lambda h}{2-\lambda h}\right)^k y_0. \end{array} \tag{4.3.13}$$

O fator de amplificação é $\psi(\lambda h) = \dfrac{2+\lambda h}{2-\lambda h}$.

Para estabelecer o intervalo de estabilidade absoluta do método é preciso analisar as condições para que $|\psi(\lambda h)| < 1$, *isto é,*

$$\left|\frac{2+\lambda h}{2-\lambda h}\right| < 1 \Rightarrow -1 < \frac{2+\lambda h}{2-\lambda h} < 1 \Rightarrow \lambda h < 0 \Rightarrow \lambda h \in (-\infty, 0).$$

O Método do Trapézio é **incondicionalmente estável**, *ou seja, não restrição na escolha do passo de integração.*

Observação:

Quando $\lambda \in \mathbb{C}$, se $z = \lambda h = a + b\,i$, $a, b \in \mathbb{R}$,

$$\begin{aligned}
\left|\frac{2+\lambda h}{2-\lambda h}\right| < 1 \quad &\Rightarrow \quad \left|\frac{2+z}{2-z}\right| < 1 \\
&\Rightarrow \quad \left|\frac{2+a+bi}{2-a-bi}\right| < 1 \\
&\Rightarrow \quad \left|\frac{(a+2)+bi}{(-a+2)-bi}\right| < 1 \\
&\Rightarrow \quad \frac{|(a+2)+bi|}{|(-a+2)-bi|} < 1 \\
&\Rightarrow \quad \frac{\sqrt{(a+2)^2+b^2}}{\sqrt{(-a+2)^2+(-b)^2}} < 1 \\
&\Rightarrow \quad \sqrt{(a+2)^2+b^2} < \sqrt{(-a+2)^2+b^2} \\
&\Rightarrow \quad a^2+4a+4+b^2 < a^2-4a+4+b^2 \\
&\Rightarrow \quad 8a < 0 \\
&\Rightarrow \quad a < 0 \\
&\Rightarrow \quad Re\,(\lambda h) < 0. \qquad (4.3.14)
\end{aligned}$$

Em (4.3.14), tem-se o conjunto dos pontos do semiplano à esquerda da origem. O intervalo de estabilidade é definido por $Re\,(\lambda h) \in (-\infty, 0)$, enquanto que $Im\,(\lambda h)$ é responsável apenas por um comportamento oscilatório da solução (veja Schwarz [21]).

Exercício Resolvido 4.2. *Considere o Método de Runge-Kutta de dois estágios*

$$\left\{\begin{array}{rcl} y_0 &=& y\,(t_0)\,, \\ y_{k+1} &=& y_k + h\,(c_1\kappa_1 + c_2\kappa_2)\,, \end{array}\right. \qquad (4.3.15)$$

com

$$\left\{\begin{array}{rcl} \kappa_1 &=& f(t,y), \\ \kappa_2 &=& f\,(t+ah, y+hb\kappa_1)\,, \end{array}\right.$$

onde a, b, c_1, e c_2 são constantes, aplicado ao problema-modelo

$$\left\{\begin{array}{rcl} \dfrac{d}{dt}y(t) &=& \lambda y,\ \lambda < 0, \\ y\,(t_0) &=& y_0. \end{array}\right. \qquad (4.3.16)$$

Calcule o fator de amplificação e o intervalo de estabilidade absoluta do método (4.3.15) supondo $a = b$ e $c_1 + c_2 = 1$.

Solução:

Na aproximação obtida no ponto t_{k+1}, tem-se

$$\begin{aligned}
y_{k+1} &= y_k + hc_1\kappa_1 + hc_2\kappa_2 \\
&= y_k + hc_1 f\,(t_k, y_k) + hc_2 f\,(t_k + ah, y_k + hbf\,(t_k, y_k)) \\
&= y_k + hc_1\lambda y_k + hc_2\lambda\,(y_k + hb\lambda y_k) \\
&= y_k + hc_1\lambda y_k + hc_2\lambda y_k + c_2 b\,(\lambda h)^2\, y_k \\
&= \left(1 + (c_1+c_2)\,\lambda h + c_2 b\,(\lambda h)^2\right) y_k \\
&= \left(1 + \lambda h + c_2 b\,(\lambda h)^2\right) y_k.
\end{aligned}$$

Assim, recursivamente, obtém-se

$$y_{k+1} = \left(1 + \lambda h + c_2 b \left(\lambda h\right)^2\right)^k y_0.$$

Denotando-se $\alpha = c_2 b$ *e* $\lambda h = x$*, o fator de amplificação é dado por*

$$\psi(x) = 1 + x + \alpha x^2. \tag{4.3.17}$$

Para determinar o intervalo de estabilidade absoluta do método (4.3.15) é necessário analisar as condições para que $|\psi(x)| < 1$*, isto é,*

$$-1 < \psi(x) < 1 \Rightarrow -1 < \alpha x^2 + x + 1 < 1 \Rightarrow -2 < \alpha x^2 + x < 0.$$

Para:

1. $\alpha x^2 + x > -2 \Rightarrow \alpha x^2 + x + 2 > 0$*, há três casos possíveis:*

   (a) *Se* $\alpha < 0$ *então*

   $$\begin{cases} \Delta &= 1 - 8\alpha > 0 \\ x_1 &= \dfrac{-1 - \sqrt{1 - 8\alpha}}{2\alpha} \\ x_2 &= \dfrac{-1 + \sqrt{1 - 8\alpha}}{2\alpha} \end{cases}$$

   e o intervalo é dado por (x_1, x_2)*;*

   (b) *Se* $0 < \alpha \leq \dfrac{1}{8}$ *então*

   $$\begin{cases} \Delta &= 1 - 8\alpha \geq 0 \\ x_1 &= \dfrac{-1 - \sqrt{1 - 8\alpha}}{2\alpha} \\ x_2 &= \dfrac{-1 + \sqrt{1 - 8\alpha}}{2\alpha} \end{cases}$$

   e o intervalo é dado por $(-\infty, x_1) \cup (x_2, \infty)$*;*

   (c) *Se* $\alpha > \dfrac{1}{8}$ *então* $\alpha x^2 + x + 2 > 0$*,* $\forall x \in \mathbb{R}$*.*

2. $\alpha x^2 + x < 0$*, há dois casos possíveis:*

   (a) *Se* $\alpha > 0$ *então*

   $$\alpha x^2 + x < 0 \Rightarrow x\left(\alpha x + 1\right) < 0$$

   e o intervalo é dado por $\left(-\dfrac{1}{\alpha}, 0\right)$*;*

   (b) *Se Se* $\alpha < 0$ *então*

   $$\alpha x^2 + x < 0 \Rightarrow x\left(\alpha x + 1\right) < 0$$

   e o intervalo é dado por $(-\infty, 0) \cup \left(-\dfrac{1}{\alpha}, \infty\right)$*.*

Sendo $\alpha = c_2 b$*, na análise do problema modelo (4.3.16) tem-se os seguintes intervalos de estabilidade:*

- $\alpha < 0$: $\left\{(-\infty, 0) \cup \left(-\frac{1}{\alpha}, \infty\right)\right\} \cap (x_1, x_2)$;
- $0 < \alpha \leq \frac{1}{8}$: $\left(-\frac{1}{\alpha}, 0\right) \cap \{(-\infty, x_1) \cup (x_2, \infty)\}$;
- $\alpha > \frac{1}{8}$: $\left(-\frac{1}{\alpha}, 0\right)$.

Exemplos

1. *Método de Euler Modificado (Método do Ponto Médio)*

$$y_{k+1} = y_k + hf\left(t + \frac{h}{2}, y + \frac{h}{2} f(t, y)\right)$$

$$a = b = \frac{1}{2},\ c_1 = 0,\ c_2 = 1 \Rightarrow \alpha = c_2 b = \frac{1}{2} \Rightarrow z = \lambda h \in (-2, 0)$$

2. *Método de Euler Aprimorado*

$$y_{k+1} = y_k + \frac{h}{2}\left[f(t, y) + f(t + h, y + hf(t, y))\right]$$

$$a = b = 1,\ c_1 = \frac{1}{2},\ c_2 = \frac{1}{2} \Rightarrow \alpha = c_2 b = \frac{1}{2} \Rightarrow z = \lambda h \in (-2, 0)$$

3. *Método de Ralston*

$$y_{k+1} = y_k + \frac{h}{4}\left[f(t, y) + 3f\left(t + \frac{2}{3}h, y + \frac{2}{3}hf(t, y)\right)\right]$$

$$a = b = \frac{2}{3},\ c_1 = \frac{1}{4},\ c_2 = \frac{3}{4} \Rightarrow \alpha = c_2 b = \frac{1}{2} \Rightarrow z = \lambda h \in (-2, 0)$$

Capítulo 5

Métodos de passo múltiplo lineares

Para se obter resultados mais precisos, em geral, aumenta-se a ordem do método numérico ou diminui-se o passo de integração utilizado. Ambas as estratégias resultam num algoritmo mais custoso do ponto de vista computacional pois, de um lado, na primeira abordagem requerem-se mais e mais cálculos da função f que define o Problema de Cauchy e, de outro, na segunda abordagem aumenta-se a quantidade de passos de integração para se alcançar um mesmo instante final de estudo (e, portanto, mais e mais cálculos de f uma vez mais). Se o custo de se calcular f é "baixo" então é possível que qualquer uma das duas estratégias sejam satisfatórias, dependendo da região de estabilidade dos métodos em jogo. Caso contrário, se o custo computacional de se calcular f é alto, uma alternativa interessante é o uso de métodos de passo múltiplo os quais exigem apenas um cálculo de f por passo no tempo (em contrapartida, como será comentado, a um custo do aumento da demando por memória).

5.1 Caracterização dos métodos de passo múltiplo

Definição 5.1 (Método de Passo Múltiplo)**.** *Um método de passo múltiplo linear (ou método de n-passos linear) tem a forma*

$$\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f_{k+j}, \tag{5.1.1}$$

ou seja,

$$\alpha_n y_{k+n} + \cdots + \alpha_1 y_{k+1} + \alpha_0 y_k = h \left[\beta_n f_{k+n} + \cdots + \beta_1 f_{k+1} + \beta_0 f_k\right],$$

onde α_j e β_j são constantes, sendo $\alpha_n \neq 0$ e $\alpha_0^2 + \beta_0^2 \neq 0$ e $f_k = f(t_k, y_k)$.

A relação (5.1.1) é uma equação de diferenças linear cuja solução é uma sequência $\{y_n\}$, $n \in \mathbb{N}$. O método numérico definido por (5.1.1) é implicíto quando $\beta_n \neq 0$ e explícito quando $\beta_n = 0$. Sem perda de generalidade, assume-se que $\alpha_n = 1$.

Exemplo 5.1. *Usando a notação empregada no Capítulo 1, o Método de Adams-Bashforth de 4 passos se escreve como*

$$\begin{cases} y_0 = y(t_0)\ ,\ y_p\ \textit{pré-determinado}\ ,\ 1 \leq p \leq 3, \\ \\ y_{k+1} = y_k + \frac{h}{24}(55f_k - 59f_{k-1} + 37f_{k-2} - 9f_{k-3})\ ,\ 3 \leq k \leq n-1. \end{cases}$$

Na notação utilizada em (5.1.1), ele se escreve como

$$\begin{cases} y_0 = y(t_0)\ ,\ y_p\ \textit{pré-determinado}\ ,\ 1 \leq p \leq 3, \\ \\ y_{k+4} = y_{k+3} + \frac{h}{24}(55f_{k+3} - 59f_{k+2} + 37f_{k+1} - 9f_k)\ ,\ k = 0, 1, ... \end{cases} \tag{5.1.2}$$

Comparando-se (5.1.2) com (5.1.1), constata-se que

$$\alpha_0 = 0,\ \alpha_1 = 0,\ \alpha_2 = 0,\ \alpha_3 = -1,\ \alpha_4 = 1$$

e

$$\beta_0 = \frac{-9}{24},\ \beta_1 = \frac{37}{24},\ \beta_2 = -\frac{59}{24},\ \beta_3 = \frac{55}{24},\ \beta_4 = 0.$$

O método (5.1.2) é explicito ($\beta_4 = 0$), de 4-passos e y_p, para $1 \leq p \leq 3$, é obtido numericamente por intermédio de um método de passo único. Observe que alguma consideração é necessária para a escolha do método de passo único que se utiliza para *inicializar* um método de passo múltiplo pois erros introduzidos inicialmente (e.g. introduzidos por métodos de passo único de ordens demasiadamente baixas) persistem e contaminam a solução numérica final. Na prática, por cautela, é comum escolher-se um método de passo único com mesma ordem do método de passo múltiplo em uso, embora a teoria preveja que é suficiente escolher-se um que tenha sua ordem um a menos que a ordem do método de passo múltiplo.

5.2 Dedução de métodos de passo múltiplo

Dentre inúmeras estratégias disponíveis, para se deduzir tais métodos é comum utilizar-se:

1. a forma diferencial do Problema de Cauchy (1.1.5);
2. a forma integral do Problema de Cauchy (1.2.8) e quadratura numérica (e.g. Regra dos Trapézios ou de Simpson);
3. interpolação polinomial (e.g. interpolação de Newton ou *splines*).

Exemplo 5.2 (Método de Simpson)**.** *Considere a Tabela 5.1*

t	t_k	t_{k+1}	t_{k+2}
$f(t, y(t)) = f$	f_k	f_{k+1}	f_{k+2}

Tabela 5.1: Valores de f em três instantes sucessivos de tempo.

Integrando a forma diferencial do Problema de Cauchy para $t \in [t_k, t_{k+2}]$

$$\int_{t_k}^{t_{k+2}} \frac{d}{ds} y(s) ds = \int_{t_k}^{t_{k+2}} f\left(s, y(s)\right) ds$$

obtém-se

$$y(t_{k+2}) - y(t_k) \quad = \quad \int_{t_k}^{t_{k+2}} f\left(s, y(s)\right) ds, \tag{5.2.3}$$

da qual, aplicando-se a quadratura definida pela Regra de Simpson e utilizando os pontos dados pela Tabela (5.1), chega-se à aproximação

$$y(t_{k+2}) - y(t_k) \approx \frac{h}{3}\left(f_k + 4f_{k+1} + f_{k+2}\right). \tag{5.2.4}$$

A expressão (5.2.4) dá origem ao método de 2 passos implícito

$$y_{k+2} - y_k = \frac{h}{3}\left(f_{k+2} + 4f_{k+1} + f_k\right) \tag{5.2.5}$$

conhecido como Método de Simpson.

Comparando-se (5.2.5) com (5.1.1), constatam-se que

$$\alpha_0 = -1,\ \alpha_1 = 0,\ \alpha_2 = 1$$

e

$$\beta_0 = \frac{1}{3},\ \beta_1 = \frac{4}{3},\ \beta_2 = \frac{1}{3}.$$

5.3 Erro de discretização local

Definição 5.2 (Erro local)**.** *Dado o método numérico de passo múltiplo linear (5.1.1), o erro de discretização local em relação a um problema de valor inicial é definido por*

$$\tau_k \doteq \frac{1}{h}\sum_{j=0}^{n} \alpha_j y(t_k + jh) - \sum_{j=0}^{n} \beta_j f(t_k + jh, y(t_k + jh)), \tag{5.3.6}$$

onde $y(t)$ é a solução (única) de um problema de valor inicial.

Adotamos a notação τ_k para o erro local de discretização, ao invés de α_k, usado nos métodos de passo único, apenas para evitar um conflito de notação com os coeficientes α_j do método.

Seja

$$d_k \doteq h\tau_k = \sum_{j=0}^{n} \alpha_j y(t_k + jh) - h\sum_{j=0}^{n} \beta_j f(t_k + jh, y(t_k + jh)). \tag{5.3.7}$$

Assim como nos métodos de passo único, d_k é o erro produzido pelo método numérico para avançar a solução um passo de integração h partindo de valores exatos $y(t_k + jh)$, $0 \leq j \leq n-1$.

Considere a solução numérica obtida a partir da resolução da equação de diferenças

$$\sum_{j=0}^{n} \alpha_j y_{k+j} - h\sum_{j=0}^{n} \beta_j f(t_{k+j}, y_{k+j}) = 0. \tag{5.3.8}$$

Calculando-se a diferença entre (5.3.7) e (5.3.8), obtém-se

$$\begin{aligned} d_k &= \sum_{j=0}^{n} \alpha_j \left[y(t_k + jh) - y_{k+j} \right] + \\ &- h \sum_{j=0}^{n} \beta_j \left[f(t_k + jh, y(t_k + jh)) - f(t_{k+j}, y_{k+j}) \right]. \end{aligned} \tag{5.3.9}$$

Assumindo valores exatos $y_{k+j} = y(t_k + jh)$, $0 \leq j \leq n-1$, tem-se para $j = n$ em (5.3.9) que

$$d_k = y(t_{k+n}) - y_{k+n} - h\beta_n \left[f(t_k + nh, y(t_k + nh)) - f(t_{k+n}, y_{k+n}) \right]. \tag{5.3.10}$$

Há duas possibilidades em (5.3.10):

1. $\beta_n = 0$ e o método é explícito com erro de discretização local dado por

$$d_k = y(t_{k+n}) - y_{k+n},$$

2. $\beta_n \neq 0$ e o método é implícito com erro de discretização de uma passo de integração dado por

$$\begin{aligned} d_k &= y(t_k + nh) - y_{k+n} - h\beta_n \underbrace{\left[f(t_k + nh, y(t_k + nh)) - f(t_{k+n}, y_{k+n}) \right]}_{\text{Teorema do Valor Médio}} = \\ &= y(t_k + nh) - y_{k+n} - h\beta_n \frac{\partial f}{\partial y}(t_{k+n}, \xi_{k+n}) \left(y(t_k + nh) - y_{k+n} \right) = \\ &= \left(1 - h\beta_n \frac{\partial f}{\partial y}(t_{k+n}, \xi_{k+n}) \right) \left(y(t_{k+n}) - y_{k+n} \right). \end{aligned} \tag{5.3.11}$$

Em métodos explícitos, vê-se que o erro de discretização local é dado diretamente pela diferença entre a solução exata e a aproximação numérica obtida a partir de valores exatos em instantes de tempo anteriores. Em métodos implícitos, tal erro é apenas proporcional a esta diferença.

Expandindo-se (5.3.7)

$$d_k = h\tau_k = \sum_{j=0}^{n} \alpha_j (y(t_k + jh)) - h \sum_{j=0}^{n} \beta_j \underbrace{f(t_k + jh, y(t_k + jh))}_{y^{(1)}(t_k + jh)}$$

em Série de Taylor ao redor de $t = t_k$, tem-se

$$d_k = \sum_{j=0}^{n} \alpha_j \left[y(t_k) + jhy^{(1)}(t_k) + \frac{j^2h^2}{2!} y^{(2)}(t_k) + \cdots + \frac{j^p h^p}{p!} y^{(p)}(t_k) + \cdots \right] +$$

$$- \sum_{j=0}^{n} \beta_j \left[hy^{(1)}(t_k) + jh^2 y^{(2)}(t_k) + \frac{j^2h^3}{2!} y^{(3)}(t_k) + \cdots + \frac{j^{p-1}h^p}{(p-1)!} y^{(p)}(t_k) + \cdots \right],$$

da qual

$$d_k = h\tau_k = C_0 y(t_k) + C_1 hy^{(1)}(t_k) + C_2 h^2 y^{(2)}(t_k) + \cdots + C_p h^p y^{(p)}(t_k) + \cdots \tag{5.3.12}$$

onde os coeficientes C_p são definidos por

$$\begin{aligned}
C_0 &= \sum_{j=0}^{n} \alpha_j, \\
C_1 &= \sum_{j=0}^{n} j\alpha_j - \sum_{j=0}^{n} \beta_j, \\
C_2 &= \sum_{j=0}^{n} \frac{j^2\alpha_j}{2!} - \sum_{j=0}^{n} j\beta_j, \\
C_3 &= \sum_{j=0}^{n} \frac{j^3\alpha_j}{3!} - \sum_{j=0}^{n} \frac{j^2\beta_j}{2!}, \\
&\vdots
\end{aligned}$$

$$\boxed{C_p = \sum_{j=0}^{n} \frac{j^p\alpha_j}{p!} - \sum_{j=0}^{n} \frac{j^{p-1}\beta_j}{(p-1)!}}. \tag{5.3.13}$$

5.4 Consistência

Definição 5.3 (Consistência)**.** *Um método de passo múltiplo linear é consistente com o problema de valor inicial associado se, e somente se,*

$$\lim_{h\to 0} \tau_k(t,h) = 0,$$

considerando o limite com $t - t_0 = kh$ fixado.

Note que o conceito de consistência está sempre associado a um problema de valor inicial bem posto dado. Geralmente, quando falamos de forma geral que o método é *consistente*, queremos dizer que o método é consistente *para qualquer problema de valor inicial bem posto.* A seguir ilustraremos essa questão.

Teorema 5.1. *Um método de passo múltiplo linear é consistente com um problema de valor inicial bem posto associado se*

$$\begin{aligned}
C_0 &= \sum_{j=0}^{n} \alpha_j = 0, \\
C_1 &= \sum_{j=0}^{n} j\alpha_j - \sum_{j=0}^{n} \beta_j = 0.
\end{aligned}$$

O teorema segue como resultado direto da equação (5.3.12), na qual vemos que para que τ_k vá a zero com $h \to 0$ é suficiente que $C_0 = 0$ e $C_1 = 0$. Aqui estamos assumindo suficiente regularidade de f, e que portanto $y^{(p)}$ é limitada para qualquer $p = 0, 1, 2, 3, ...$ no intervalo de integração.

Observe que a volta não é necessariamente verdade, isto é, existem métodos com $C_0 \neq 0$ e/ou $C_1 \neq 0$ que, para certos problemas de valores iniciais com f suave, são consistentes.

Exemplo 5.3 ([23])**.** *Considere o problema de valor inicial $y' = 0$, $y(0) = 0$, com solução exata dada por $y(t) = 0$ e o seguinte método de passo múltiplo de 2-passos para esse problema:*

$$y_{k+2} = y_{k+1} + y_k \tag{5.4.14}$$

com $y_0 = y_1 = 0$. Primeiro note que o método é exato, isto é, $y_k = y(t_k) = 0$ para qualquer $k = 0, 1, 2, \ldots$. O erro local de discretização é dado por

$$\tau_k = \frac{y(t_{k+2}) - y(t_{k+1}) - y(t_k)}{h}, \tag{5.4.15}$$

mas como a solução exata é $y(t) = 0$, temos $\tau_k = 0$. Por outro lado, $C_0 = -1 \neq 0$.

Porém, se exigirmos que o método seja consistente para qualquer problema de valor inicial, então valerá a volta também.

Teorema 5.2. *Dado um método de passo múltiplo linear que seja consistente para qualquer problema de valor inicial bem posto este terá, necessariamente,*

$$\begin{aligned} C_0 &= \sum_{j=0}^{n} \alpha_j = 0, \\ C_1 &= \sum_{j=0}^{n} j\alpha_j - \sum_{j=0}^{n} \beta_j = 0. \end{aligned}$$

Note que o exemplo 5.3 não cabe como contra-exemplo deste último teorema, pois o método não é consistente com o problema de valor inicial $y' = 1$, $y(0) = 1$ (verifique!), e portanto não é consistente para qualquer problema de valor inicial bem posto, logo, pode ter $C_0 \neq 0$.

Teorema 5.3 (Ordem de consistência)**.** *A ordem de consistência de um método de passo múltiplo é "p", com $\tau(t, h) = O(h^p)$, se para todo o instante t se tem*

$$C_0 = C_1 = \cdots = C_p = 0 \quad e \quad C_{p+1} \neq 0. \tag{5.4.16}$$

Da substituição de (5.4.16) em (5.3.12), tem-se

$$h\tau_k = C_{p+1} h^{p+1} y^{(p+1)}(\xi_k) \quad \text{ou} \quad h\tau_k = C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O\left(h^{p+2}\right). \tag{5.4.17}$$

Obtém-se a segunda das formas em (5.4.17) utilizando-se representações polinomiais de Taylor de ordem apropriada e conservando-se o resto na forma de Lagrange.

De (5.3.11) e de (5.4.17) tem-se, respectivamente,

$$\begin{aligned} h\tau_k &= \left(1 - h\beta_n \frac{\partial f}{\partial y}(\xi_{k+n})\right)(y(t_{k+n}) - y_{k+n}) \qquad \text{e} \\ h\tau_k &= C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O\left(h^{p+2}\right) \end{aligned}$$

então, a diferença entre a solução exata e a solução numérica se escreve como

$$y(t_{k+n}) - y_{k+n} = C'_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}). \tag{5.4.18}$$

Em (5.4.18), o termo $C'_{p+1} h^{p+1} y^{(p+1)}(t_k)$ é denominado *erro de discretização local principal*. Para métodos explícitos tem-se

$$C'_{p+1} = C_{p+1}$$

e para métodos implícitos

$$C'_{p+1} = \frac{C_{p+1}}{\left(1 - h\beta_n \frac{\partial f}{\partial y}\right)}.$$

5.5 Inicialização

Para ser utilizado, o método de n-passos linear

$$\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f_{k+j} \tag{5.5.19}$$

requer aproximações da solução previamente calculadas em n instantes de tempo. Por exemplo, para se obter y_n, é necessário que se tenha à disposição aproximações da solução nos instantes t_{n-1}, t_{n-2}, ..., t_0. Sendo assim, (5.5.19) pode apenas ser utilizado para se obter aproximações de y_k para $k \geq n$. Como obter as aproximações requeridas nos instantes anteriores $t = t_k$, $0 \leq k < n$?

Adotando a condição inicial original no Problema de Cauchy, $y_0 = y(t_0)$, as demais aproximações, $y_1, y_2, \cdots, y_{n-1}$, são geralmente obtidas numericamente com um método de passo único. Gostaríamos que o método de inicialização não afetasse a convergência global do método numérico, por isso introduzimos o seguinte conceito de inicialização consistente.

Definição 5.4 (Consistência do método de inicialização)**.** *Considere um método de passo múltiplo linear com n passos inicializado por um método de passo único. Dizemos que o método de inicialização é consistente com o problema de valor inicial* $y' = f(t, y)$, $y(t_0) = y_0$, *se, e somente se,*

$$\lim_{h \to 0} s_k = y_0, \quad k = 0, 1, 2, ..., n-1, \tag{5.5.20}$$

onde s_k *é a sequência de inicialização e estamos considerando o limite com* $t - t_0 = hk$ *fixado.*

O método de inicialização é consistente de ordem p se

$$\max_{l=0,1,...,n-1} \|y(t_l) - s_l\| = O(h^p), \quad \textit{quando} \quad h \to 0, \tag{5.5.21}$$

onde $y(t)$ *é a solução exata do problema de valor inicial. [23]*

Alguns textos, como [5], se referem a uma inicialização consistente apenas como a *pré-consistência* de um método de passo múltiplo.

Nos resta discutir qual é a ordem ideal do método de passo único para inicializarmos um método de n-passos lineares de ordem de consistência p, isto é, com $\tau_k = O(h^p)$. Da equação (5.4.18), considerando $k = 0$, vemos que

$$y(t_n) - y_n = C'_{p+1} h^{p+1} y^{(p+1)}(t_0) + O(h^{p+2}), \tag{5.5.22}$$

isto é, assumindo que os primeiros passos y_0,, y_1, ..., y_{n-1} tenham sido dados exatamente (inicialização exata), o erro do próximo passo será de ordem $O(h^{p+1})$. Portanto, à primeira, deveríamos garantir que a inicialização não afete a ordem desse primeiro passo, o que pode ser obtido com um método inicialização de ordem $p + 1$. Assim, basta termos um método de inicialização de uma ordem a mais que o método de n-passos para garantir que a ordem seja preservada.

Por outro lado, sabemos que os erros do método numérico se acumulam ao avançarmos para $k > n$. Mais adiante vamos mostrar que o erro global do método terá na verdade ordem p se o método tiver consistência de ordem p. Por hora, vamos fazer uma análise simplificada, baseada em [5], de como os erros se acumulam. A consequência de termos o erro global tem ordem p, e não $p + 1$, indica que não precisamos de uma inicialização de ordem $p + 1$, pois basta uma ordem a menos.

Suponha que, conforme a equação (5.4.18), a cada passo de tempo erramos proporcionalmente a $D_1 h^{p+1}$ e que o método de inicialização seja de ordem p com erro proporcional a $D_0 h^p$. Supondo que o método aproxima a solução para qualquer t e que o erro não cresce exponencialmente[1], é razoável supormos que os erros se acumulam linearmente. Portanto, para atingirmos um certo tempo $\bar{t} = t_0 + hk$, $k > n$, precisamos de k passos de integração, sendo os n primeiros passos dados pela inicialização. Assim, para chegarmos em $\bar{t}$ teríamos errado

$$y(\bar{t}) - y_k \approx (k - n) D_1 h^{p+1} + n D_0 h^p.$$

Lembrando que h pode ser escrito como $h = (\bar{t} - t_0)/k$, temos que

$$y(\bar{t}) - y_k \approx k \frac{\bar{t} - t_0}{k} D_1 h^p + n D_1 h^{p+1} + n D_0 h^p,$$

e como n está fixo, pois é definido pelo número de passos do método, temos que $y(\bar{t}) - y_k \approx O(h^p)$ mesmo com uma inicialização de ordem p.

De forma resumida, como temos um número de passos fixados na inicialização, o acumulo desse erro inicial não afeta a ordem do método de inicialização. Porém, o acumulo de erros da integração derruba o erro local d_k de ordem $p+1$ para ordem p. De fato, é por isso que definimos um método com ordem de consistência p aquele que tem $\tau_k = O(h^p)$, e não com $d_k = h\tau_k = O(h^{p+1})$.

Se a escolha do método de passo único que será usado na inicialização for menor que a ordem de consistência do método de passo múltiplo, então esta associação faz com que o método se comporte com uma ordem menor do que aquela prevista na teoria, restrita pela inicialização.

É importante notar que a ordem do método depende também do grau de regularidade da f. Se $f \in C^m$, pode-se mostrar que é suficiente que o método utilizado em sua inicialização para gerar as condições iniciais numéricas tenha ordem $m - 1$, pois de fato a ordem máxima atingível para o método de passo múltiplo será teoricamente também de $m - 1$ (veja a equação (5.3.12)). Se o Problema de Cauchy em estudo não for suficientemente suave, o método poderá apresentar um comportamento diferente daquele previsto pela teoria mesmo que se empreguem condições iniciais numéricas com elevada precisão (ou mesmo *exatas* - se estas estiverem à disposição).

[1] Isto será discutido mais adiante junto com o conceito de estabilidade.

5.6 Convergência

Definição 5.5 (Convergência)**.** *Um método de passo múltiplo linear de n-passos*

$$\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f(t_{k+j}, y_{k+j}), \tag{5.6.23}$$

com método de inicialização consistente é convergente se, e somente se, para todo Problema de Cauchy bem posto com solução $y(t)$ tem-se

$$\lim_{h \to 0} y_k = y(t), \ \forall t \in [a = t_0, t_f = b], \tag{5.6.24}$$

considerando o limite com $kh = t - t_0$ fixado. [16, 23]

Em um método de passo múltiplo linear convergente com inicialização consistente, temos que

$$y_{k+j} \longrightarrow y(t), \text{ para } h \longrightarrow 0, \ j = 0, 1, \cdots, n,$$

para qualquer $kh = t - t_0$ fixado. Equivalentemente,

$$y(t) = y_{k+j} + \theta_{k,j}(h), \ j = 0, 1, \cdots, n,$$

na qual o erro $\theta_{k,j}(h)$ satisfaz $\lim_{h \to 0} \theta_{k,j}(h) = 0$. A consistência da inicialização garante que exista essa função $\theta_{k,j}(h)$ para $k = 0, 1, 2, ..., n$ e a convergência garante a existência para $k \geq n$.

Dessa forma,

$$\sum_{j=0}^{n} \alpha_j y(t) \quad = \quad \sum_{j=0}^{n} \alpha_j y_{k+j} + \sum_{j=0}^{n} \alpha_j \theta_{j,k}(h),$$

a qual, após rearranjo dos termos e usar-se o fato de se ter um Problema de Cauchy, assume a forma

$$y(t) \sum_{j=0}^{n} \alpha_j \quad = \quad h \sum_{j=0}^{n} \beta_j f(t_{k+j}, y_{k+j}) + \sum_{j=0}^{n} \alpha_j \theta_{j,k}(h). \tag{5.6.25}$$

Considerando $h \to 0$ em (5.6.25), conclui-se que

$$y(t) \sum_{j=0}^{n} \alpha_j = 0 \Rightarrow \sum_{j=0}^{n} \alpha_j = C_0 = 0.$$

Supondo-se ainda a convergência do método, temos também que

$$\frac{y_{k+j} - y_k}{jh} \quad \stackrel{h \to 0}{\longrightarrow} \quad \frac{d}{dt} y(t) \ , \ j = 1, 2, \cdots, n, \text{ para } kh = t - t_0 \text{ fixado.}$$

Equivalentemente, espera-se então que

$$\begin{aligned}
\frac{d}{dt}y(t) &= \frac{y_{k+j}-y_k}{jh} + \underbrace{\eta_{j,k}(h)}_{\downarrow 0\ ,\ h\downarrow 0},\\
y_{k+j}-y_k &= jh\frac{d}{dt}y(t) + jh\eta_{j,k}(h),\\
\sum_{j=0}^{n}\alpha_j y_{k+j} - \sum_{j=0}^{n}\alpha_j y_k &= h\sum_{j=0}^{n}j\alpha_j\frac{d}{dt}y(t) + h\sum_{j=0}^{n}j\alpha_j\eta_{j,k}(h),\\
h\sum_{j=0}^{n}\beta_j f_{k+j} - y_k\underbrace{\sum_{j=0}^{n}\alpha_j}_{C_0=0} &= h\frac{d}{dt}y(t)\sum_{j=0}^{n}j\alpha_j + h\sum_{j=0}^{n}j\alpha_j\eta_{j,k}(h),\\
h\sum_{j=0}^{n}\beta_j f_{k+j} &= h\frac{d}{dt}y(t)\sum_{j=0}^{n}j\alpha_j + h\sum_{j=0}^{n}j\alpha_j\eta_{j,k}(h),\\
\sum_{j=0}^{n}\beta_j f_{k+j} &= \frac{d}{dt}y(t)\sum_{j=0}^{n}j\alpha_j + \sum_{j=0}^{n}j\alpha_j\eta_{j,k}(h). \qquad (5.6.26)
\end{aligned}$$

Como $f_{k+j}\to f(t,y(t)) = \dfrac{d}{dt}y(t)$ e $\eta_{j,k}(h)\to 0$ quando $h\to 0$, pode-se concluir de (5.6.26) que

$$\begin{aligned}
\sum_{j=0}^{n}\beta_j f(t,y(t)) &= \frac{d}{dt}y(t)\sum_{j=0}^{n}j\alpha_j,\\
f(t,y(t))\sum_{j=0}^{n}\beta_j &= \frac{d}{dt}y(t)\sum_{j=0}^{n}j\alpha_j,
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}y(t)\sum_{j=0}^{n}\beta_j &= \frac{d}{dt}y(t)\sum_{j=0}^{n}j\alpha_j,\\
\sum_{j=0}^{n}\beta_j &= \sum_{j=0}^{n}j\alpha_j,\\
C_1 &= 0.
\end{aligned}$$

Com $C_0 = C_1 = 0$ o método de passo múltiplo linear é consistente. Das observações anteriores, vê-se que consistência é uma condição necessária à convergência. Seria ela também uma condição suficiente? Vimos que para métodos de passo único, consistência era suficiente para convergência. Mas para métodos de passo múltiplo isso irá depender do conceito de zero-estabilidade, a ser discutido no próximo capítulo.

Teorema 5.4. *Um método linear de passo múltiplo convergente é consistente (para qualquer problema de valor inicial bem posto).*

5.6.1 Exercícios

Exercício 5.1. *Mostre que o método de passo múltiplo linear*

$$y_{k+2} - y_{k+1} = \frac{h}{3}(3f_{k+1} - 2f_k)$$

não é consistente.

Exercício 5.2. *Mostre que o método de passo múltiplo linear (Adams-Bashford de 2 passos)*

$$y_{k+2} - y_{k+1} = \frac{h}{2}(3f_{k+1} - f_k)$$

tem ordem 2 de consistência.

Exercício 5.3. *Mostre que a ordem do método de passo múltiplo linear*

$$y_{k+2} + (b-1)y_{k+1} - by_k = \frac{h}{4}[(b+3)f_{k+2} + (3b+1)f_k]$$

é 3 se $b = -1$ e 2 caso contrário.

5.7 Exercícios resolvidos

Exercício Resolvido 5.1. *Mostre que a ordem do método de passo múltiplo linear*

$$y_{k+2} - y_{k+1} = \frac{h}{12}(4f_{k+2} + 8f_{k+1} - f_k) \tag{5.7.27}$$

é zero. Mostre que o método é divergente usando a solução teórica do Problema de Cauchy

$$\begin{cases} \dot{y}(t) = 1, \ 0 \le t \le 2, \\ y(0) = 0. \end{cases} \tag{5.7.28}$$

Solução:

Para o método dado (5.7.27) tem-se seguintes parâmetros:

$$\alpha_0 = 0, \quad \alpha_1 = -1, \quad \alpha_2 = 1, \quad \beta_0 = -\frac{1}{12}, \quad \beta_1 = \frac{2}{3} \quad e \quad \beta_2 = \frac{1}{3}.$$

De (5.4.16), para se mostrar que a ordem do método é zero, basta mostrar que $C_0 = 0$ e que $C_1 \neq 0$:

$$\begin{aligned} C_0 &= \sum_{j=0}^{2} \alpha_j = \alpha_0 + \alpha_1 + \alpha_2 = 0 - 1 + 1 = 0, \\ C_1 &= \sum_{j=0}^{2} j\alpha_j - \sum_{j=0}^{2} \beta_j = \alpha_1 + 2\alpha_2 - \beta_0 - \beta_1 - \beta_2 = \\ &= -1 + 2 + \frac{1}{12} - \frac{2}{3} - \frac{1}{3} = \frac{1}{12} \neq 0. \end{aligned}$$

Como $C_0 = 0$ e $C_1 \neq 0$, conclui-se que o método tem ordem zero.

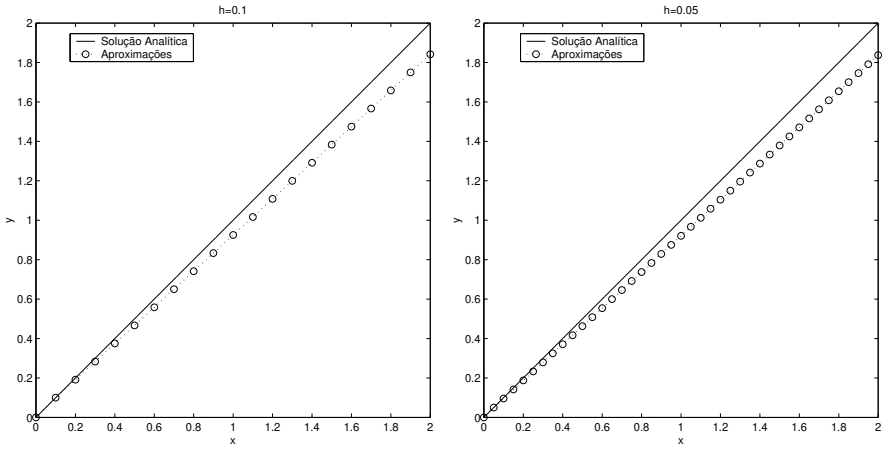
No Problema de Cauchy (5.7.28), tem-se $f(t, y(t)) = 1$ e solução exata $y(t) = t$ (verifique!). Para se mostrar que o método considerado diverge neste caso, considere o instante $t = 1$, fixo. Empregando-se o método (5.7.27) a (5.7.28), tem-se

$$\begin{aligned} y_{k+2} - y_{k+1} &= \frac{h}{12}(4 + 8 - 1) = \frac{11}{12}h, \ \textit{isto é,} \\ y_{k+2} &= y_{k+1} + \frac{11}{12}h. \end{aligned}$$

Logo,

$$\begin{aligned} y_0 &= 0, \\ y_1 &= h \ (y_1 = y_0 + h = 0 + h), \\ y_2 &= y_1 + \frac{11}{12}h = h + \frac{11}{12}h = \frac{23}{12}h, \\ y_3 &= y_2 + \frac{11}{12}h = \frac{34}{12}h, \\ y_4 &= y_3 + \frac{11}{12}h = \frac{45}{12}h, \\ &\vdots \\ y_k &= h + (k-1)\,\frac{11}{12}. \end{aligned} \tag{5.7.29}$$

Assim, para $kh = 2$ fixado, quando $h \to 0$ tem-se $y_k \to 0$. Porém, $y(2) = 2$, isto é, o método diverge para $t = 2$. Graficamente, as figuras (5.1) e (5.2) exibem este fato. Nelas, traçam-se simultaneamente a solução exata e as aproximações numéricas geradas pelo método. Os passos de integração considerados foram, respectivamente, $h = 0,1$, $h = 0,05$ e $h = 0,025$.



(a) (b)

Figura 5.1: Comparação entre as soluções exata e numérica do Problema de Cauchy (5.7.28). A solução numérica foi obtida aplicando-se o método (5.7.27) com (a) $h = 0,1$ e (b) $h = 0,05$.

Observa-se nas Figuras (5.1) e (5.2) o comportamento divergente do método de passo múltiplo linear (5.7.27). O refinamento do passo de integração não diminui a diferença entre os valores exato e numérico. Pelo fato da ordem do método ser zero, haverá para todo passo h uma diferença quase que constante entre as soluções analítica $y(t_k)$ e numérica y_k em todos os pontos t_k. Fixando-se o ponto $t = 2,0$, o comportamento do erro nas aproximações de $y(2)$ com o passo h sendo reduzido pela metade sucessivamente, por vinte vezes, a partir de $h = 0,5$, pode ser observado na Figura (5.3), apresentada em escala logarítmica. Nessa figura, nota-se que o erro quase se estabiliza para um passo de integração h menor do que aproximadamente 10^{-2}.

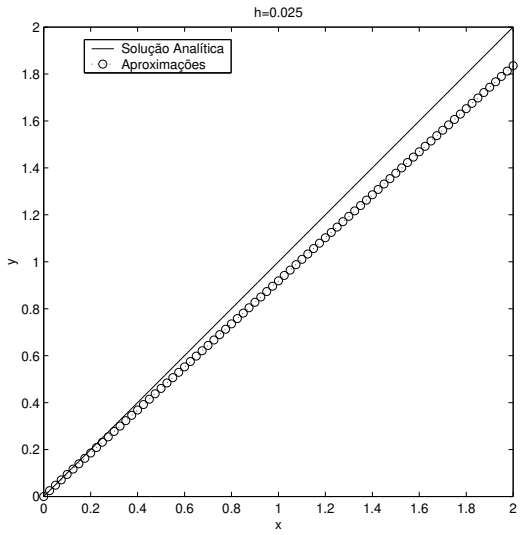


Figura 5.2: Comparação entre as soluções exata e numérica do Problema de Cauchy (5.7.28). A solução numérica foi obtida aplicando-se o método (5.7.27) com $h = 0,025$.

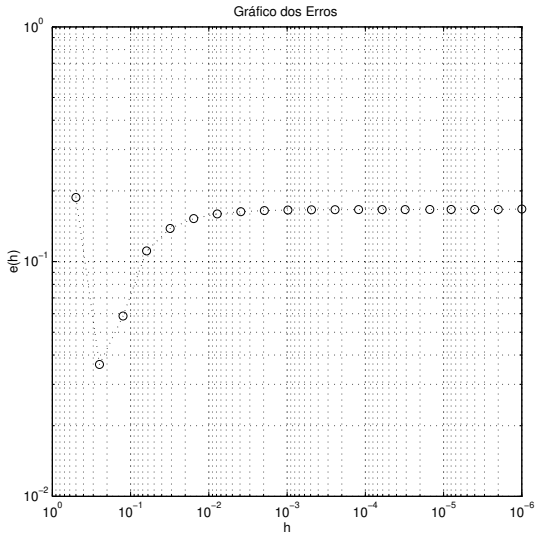


Figura 5.3: Comportamento do erro na solução numérica do Problema de Cauchy (5.7.28) obtida através do método (5.7.27) com refinamento do passo de integração h.

Exercício Resolvido 5.2. *Construa um método de dois passos implícito de ordem máxima contendo um parâmetro livre, $\alpha_0 = a$. Determine a ordem do método obtido.*
Solução:

Método de dois passos implícito:

$$y_{k+2} + \alpha_1 y_{k+1} + a y_k = h\left[\beta_2 f_{k+2} + \beta_1 f_{k+1} + \beta_0 f_k\right], \tag{5.7.30}$$

com $\beta_2 \neq 0$, $\alpha_2 = 1$ *e* $\alpha_0 = a$.

Em (5.7.30) tem-se quatro incógnitas. Logo, são necessárias quatro equações. Para tanto, considere-se o método (5.7.30) consistente de ordem três, o que implica $C_0 = C_1 = C_2 = C_3 = 0$ *e* $C_4 \neq 0$.

$$C_0 = 0 \Rightarrow \sum_{j=0}^{2} \alpha_j = 0 \Rightarrow 1 + \alpha_1 + a = 0 \Rightarrow \alpha_1 = -(a+1) \tag{5.7.31}$$

$$\begin{aligned} C_1 = 0 \quad &\Rightarrow \quad \sum_{j=0}^{2} j\alpha_j - \sum_{j=0}^{2} \beta_j = 0 \Rightarrow \alpha_1 + 2 - \beta_0 - \beta_1 - \beta_2 = 0 \\ &\Rightarrow \quad 1 - a = \beta_0 + \beta_1 + \beta_2, \end{aligned} \tag{5.7.32}$$

$$\begin{aligned} C_2 = 0 \quad &\Rightarrow \quad \sum_{j=0}^{2} \frac{j^2}{2!}\alpha_j - \sum_{j=0}^{2} j\beta_j = 0 \Rightarrow \frac{1}{2}\alpha_1 + 2 - \beta_1 - 2\beta_2 = 0 \\ &\Rightarrow \quad \frac{1}{2}\underbrace{\left[-(a+1)\right]}_{(5.7.31)} + 2 = \beta_1 + 2\beta_2 \\ &\Rightarrow \quad 3 - a = 2\beta_1 + 4\beta_2, \end{aligned} \tag{5.7.33}$$

$$\begin{aligned} C_3 = 0 \quad &\Rightarrow \quad \sum_{j=0}^{2} \frac{j^3}{3!}\alpha_j - \sum_{j=0}^{2} \frac{j^2}{2!}\beta_j = 0 \Rightarrow \frac{1}{6}\alpha_1 + \frac{8}{6} - \frac{1}{2}\beta_1 - 2\beta_2 = 0 \\ &\Rightarrow \quad \frac{1}{6}\underbrace{\left[-(a+1)\right]}_{(5.7.31)} + \frac{8}{6} = \frac{1}{2}\beta_1 + 2\beta_2 \\ &\Rightarrow \quad 7 - a = 3\beta_1 + 12\beta_2. \end{aligned} \tag{5.7.34}$$

Subtraindo-se três vezes (5.7.33) de (5.7.34), tem-se

$$2a - 2 = -3\beta_1 \Rightarrow \beta_1 = \frac{2-2a}{3}. \tag{5.7.35}$$

Substituindo (5.7.35) em (5.7.33), obtém-se

$$4\beta_2 = 3 - a - 2\frac{2-2a}{3} \Rightarrow \beta_2 = \frac{a+5}{12}. \tag{5.7.36}$$

Substituindo (5.7.35) e (5.7.36) em (5.7.32), chega-se a

$$\beta_0 = 1 - a - \frac{2-2a}{3} - \frac{a+5}{12} \Rightarrow \beta_0 = -\frac{5a+1}{12}.$$

Calculando C_4, obtém-se

$$\begin{aligned} C_4 &= \sum_{j=0}^{2} \frac{j^4}{4!}\alpha_j - \sum_{j=0}^{2} \frac{j^3}{3!}\beta_j = \frac{1}{24}\alpha_1 + \frac{16}{24} - \frac{1}{6}\beta_1 - \frac{8}{6}\beta_2 \\ &= \frac{1}{24}\underbrace{[-(a+1)]}_{(5.7.31)} + \frac{16}{24} - \frac{1}{6}\underbrace{\frac{2-2a}{3}}_{(5.7.35)} - \frac{8}{6}\underbrace{\frac{a+5}{12}}_{(5.7.36)} \\ &= \frac{45-3a+8a-8-8a-40}{72} = -\frac{a+1}{24}. \end{aligned} \tag{5.7.37}$$

Em (5.7.37), se $a \neq -1$, então o método (5.7.30) tem ordem 3 ($C_0 = C_1 = C_2 = C_3 = 0$ e $C_4 \neq 0$); se $a = -1$, então $\alpha_1 = 0$, $\alpha_0 = -1$, $\beta_2 = \frac{1}{3}$, $\beta_1 = \frac{4}{3}$, $\beta_0 = \frac{1}{3}$ e C_5 é igual a

$$\begin{aligned} C_5 &= \sum_{j=0}^{2} \frac{j^5}{5!}\alpha_j - \sum_{j=0}^{2} \frac{j^4}{4!}\beta_j = \frac{1}{120}\alpha_1 + \frac{32}{120} - \frac{1}{24}\beta_1 - \frac{16}{24}\beta_2 \\ &= \frac{32}{120} - \frac{1}{24}\frac{4}{3} - \frac{16}{24}\frac{1}{3} = -\frac{1}{90} \neq 0, \end{aligned} \tag{5.7.38}$$

e o método tem ordem quatro.

Portanto, o método numérico (5.7.30) tem a forma

$$y_{k+2} - (a+1)y_{k+1} + ay_k = h\left[\frac{a+5}{12}f_{k+2} - \frac{2a-2}{3}f_{k+1} - \frac{5a+1}{12}f_k\right]. \tag{5.7.39}$$

Em (5.7.39), se $a \neq -1$ então o método tem ordem 3. Se $a = -1$ então o método é o Método de Simpson,

$$y_{k+2} = y_k + \frac{h}{3}\left[f_{k+2} + 4f_{k+1} + f_k\right],$$

e tem ordem 4. Se $a \neq -5$ então o método é implícito.

Capítulo 6

Zero-estabilidade e convergência dos métodos de passo múltiplo lineares

Apresentam-se neste Capítulo, condições necessárias e suficientes à convergência dos métodos de passo múltiplo lineares. Conhecimentos sobre equações de diferenças são necessários à apresentação e, por conveniência, são incluídos em seu início. Maior detalhamento do conteúdo aqui exposto e demonstrações de alguns dos resultados apresentados podem ser encontrados, por exemplo, em [24, 16, 17, 18].

6.1 Equações de diferenças lineares

Uma equação de diferenças linear é uma expressão do tipo

$$\gamma_n y_{k+n} + \gamma_{n-1} y_{k+n-1} + \cdots + \gamma_0 y_k = \sum_{j=0}^{n} \gamma_j y_{k+j} = \phi_k \ , \ k = 0, 1, \ldots , \qquad (6.1.1)$$

onde γ_j, $j = 0, 1, ..., n$, são constantes independentes de k, com $\gamma_0 \neq 0$ e $\gamma_n \neq 0$. As soluções desse tipo de equação são sequências numéricas $y = \{y_k\}_{k\in\mathbb{N}}$ para as quais quaisquer $n+1$ elementos sucessivos se relacionem como em (6.1.1). A essa equação associamos *uma equação de diferenças homogênea* dada por

$$\gamma_n y_{k+n} + \gamma_{n-1} y_{k+n-1} + \cdots + \gamma_0 y_k = \sum_{j=0}^{n} \gamma_j y_{k+j} = 0 \ , \ k = 0, 1, \ldots . \qquad (6.1.2)$$

Se $w = \{w_k\}_{k\in\mathbb{N}}$ e $z = \{z_k\}_{k\in\mathbb{N}}$ são duas soluções de (6.1.1), então a sequência dada pela diferença $\widehat{y} = \{\widehat{y}_k\}_{k\in\mathbb{N}} = \{w_k - z_k\}_{k\in\mathbb{N}}$ é solução da equação de diferenças homogênea associada, isto é,

$$\begin{aligned} \sum_{j=0}^{n} \gamma_j \widehat{y}_{k+j} &= \sum_{j=0}^{n} \gamma_j (w_{k+j} - z_{k+j}) = \\ = \sum_{j=0}^{n} \gamma_j w_{k+j} - \sum_{j=0}^{n} \gamma_j z_{k+j} &= \phi_k - \phi_k = 0, \quad \forall k = 0, 1, \ldots \end{aligned} \qquad (6.1.3)$$

De (6.1.3), vê-se que (qualquer) uma solução $\{w_k\}_{k\in\mathbb{N}}$ de (6.1.1) decompõe-se como uma soma da solução da equação homogênea com uma *solução particular* $\{z_k\}_{k\in\mathbb{N}}$ a qual satisfaz

$$\sum_{j=0}^{n} \gamma_j z_{k+j} = \phi_k, \quad k = 0, 1, 2, \ldots$$

Em outras palavras, das observações anteriores, conclui-se que a solução geral de uma equação de diferenças linear se escreve como a soma da solução de sua equação homogênea associada com uma solução particular, isto é,

$$w_k = \widehat{y}_k + z_k, \quad j = 0, 1, 2, \ldots, \tag{6.1.4}$$

onde $\widehat{y}$ é solução da equação homogênea e z é uma solução particular da equação não homogênea.

Portanto, percebe-se que a equação de diferenças (6.1.1) não possui solução única; de fato, possui infinitas soluções! Para garantirmos unicidade, a equação de n passos deve vir acompanhada de n condições iniciais. Assim, dados $y_0, y_1, \ldots, y_{n-1}$, podemos calcular de forma única y_n, e consequente de forma única os valores y_k, $k > n$, usando a relação

$$y_{k+n} = \phi_k - \sum_{j=0}^{n-1} \gamma_j y_{k+j}, \quad k = 0, 1, 2, ... \tag{6.1.5}$$

Um caminho para calcularmos a solução única do problema não homogêneo é:

(i) achar um conjunto de soluções possíveis para o problema homogêneo;

(ii) achar uma solução particular da equação não homogênea;

(iii) definir a solução geral do problema como sendo combinação linear das soluções possíveis do problema homogêneo com a solução particular;

(iv) buscar dentro do conjunto de soluções gerais a única solução do problema não homogêneo que satisfaça as condições iniciais.

Definição 6.1 (Independência linear)**.** *Dizemos que m sequências* $y^{(1)} = \{y_k^{(1)}\}_{k\in\mathbb{N}}$, $y^{(2)} = \{y_k^{(2)}\}_{k\in\mathbb{N}}$, ..., $y^{(m)} = \{y_k^{(m)}\}_{k\in\mathbb{N}}$ *são* linearmente independentes *quando a combinação linear*

$$a_1 y_k^{(1)} + a_2 y_k^{(2)} + ...a_m y_k^{(m)} = 0, \quad k \in \mathbb{N} \tag{6.1.6}$$

implicar em $a_l = 0$, $l = 1, 2, ..., m$.

Teorema 6.1 (Sistema de soluções fundamentais)**.** *Sejam* $y^{(1)} = \{y_k^{(1)}\}_{k\in\mathbb{N}}$, $y^{(2)} = \{y_k^{(2)}\}_{k\in\mathbb{N}}$, ..., $y^{(n)} = \{y_k^{(n)}\}_{k\in\mathbb{N}}$ *sequências linearmente independentes, soluções da equação de diferenças homogênea com n passos (6.1.2). Esse conjunto de n sequências linearmente independentes forma o que chamamos de um* sistema fundamental de soluções*. Toda solução de (6.1.2) é dada como combinação linear de sequencias desse conjunto. Portanto, w é solução de (6.1.2) se, e somente se,*

$$w = \sum_{l=1}^{n} a_l y^{(l)},$$

com $a_l \in \mathbb{R}$, $l = 0, 1, ..., n$.

A demonstração desse teorema vem de alguns resultados de Algebra Linear. Em particular, o espaço de soluções de equações de diferenças homogêneas com n passos tem dimensão n. As n sequências $y^{(l)}$, $l = 1, 2, ..., n$ são linearmente independentes, portanto formam uma base para o espaço. Logo, a demonstração desse teorema decorre de que simplesmente qualquer solução pode ser escrita nessa base.

O primeiro passo na busca de soluções de uma equação de diferenças linear homogênea com n passos é achar as soluções fundamentais, isto é, n sequências linearmente independentes, todas soluções da equação, que juntas formam uma base para definir qualquer solução da equação homogênea.

Uma forma de achar uma solução para o problema homogêneo é supor que a sequência tenha a forma $y_k = r^k$, com k potência de r, para algum $r \in \mathbb{R}$. Substituindo essa proposta de solução na equação homogênea (6.1.2) temos

$$\sum_{j=0}^{n} \gamma_j r^{k+j} = r^k \sum_{j=0}^{n} \gamma_j r^j = 0 \tag{6.1.7}$$

$$\Rightarrow$$

$$\begin{cases} r^k = 0 \Rightarrow r = 0 : \text{ solução trivial,} \\ \\ \displaystyle\sum_{j=0}^{n} \gamma_j r^j = 0 \Rightarrow r \text{ é uma raiz do polinômio} \\ \\ \qquad\qquad \rho_n(r) = \displaystyle\sum_{j=0}^{n} \gamma_j r^j. \end{cases}$$

Definição 6.2 (Polinômio característico)**.** *Definimos como polinômio característico associado à equação de diferenças homogênea de n passos* (6.1.2) *o polinômio de grau n dado por*

$$\rho_n(r) = \sum_{j=0}^{n} \gamma_j r^j. \tag{6.1.8}$$

Se todas as raízes do polinômio característico $\rho_n(r)$ forem distintas, então acabamos de achar n sequências que resolvem o problema homogêneo. Sejam $r_1, r_2, ..., r_n$ as n raízes de ρ_n e considere as sequências $y^{(l)} = \{y_k^{(l)}\}_{k\in\mathbb{N}} = \{r_l^k\}_{k\in\mathbb{N}}$, $l = 1, 2, ..., n$. Observe a troca de índices por potências para cada termo da sequência. $y_k^{(l)} = r_l^k$ significa que o k-ésimo termo da l-ésima sequência tem valor r_l^k, isto é, a l-ésima raiz de ρ_n (r_l) elevada à potência k. Da equação (6.1.7), vemos que as sequências $y^{(l)}$ satisfazem a equação de diferenças homogênea (6.1.2).

Para que essas sequências formem um sistema de soluções fundamentais (base) para o problema homogêneo é necessário que elas sejam linearmente independentes. Isso pode ser analisado considerando o seguinte sistema linear,

$$\begin{cases} a_1 & + & a_2 & + & \cdots & + & a_n & = & 0 \\ a_1 r_1 & + & a_2 r_2 & + & \cdots & + & a_n r_n & = & 0 \\ a_1 r_1^2 & + & a_2 r_2^2 & + & \cdots & + & a_n r_n^2 & = & 0 \\ \vdots & + & \vdots & + & \vdots & + & \vdots & = & \vdots \\ a_1 r_1^n & + & a_2 r_2^n & + & \cdots & + & a_n r_n^n & = & 0 \end{cases} . \tag{6.1.9}$$

Se for possível acharmos $a_1, a_2, ..., a_n$, não todos nulos, tal que o sistema seja satisfeito, então as sequências são linearmente *dependentes*, pois pode-se escrever uma em função das demais. Caso contrário, as sequências são linearmente *independentes*. A matriz que forma o sistema linear é uma matriz conhecida como matriz de Vandermonde, cujo determinante é dado por

$$D = \prod_{1 \leq i < j \leq n} (r_i - r_j). \tag{6.1.10}$$

Como as raízes são distintas, $D \neq 0$, e portanto o sistema tem solução única. Como a solução trivial (nula) resolve o sistema, a única solução do sistema será com $a_l = 0$, $l = 1, 2, ..., n$. Logo, as sequências são linearmente independentes.

Concluímos portanto que, caso o polinômio característico tenha raízes distintas, qualquer solução do problema homogêneo pode ser escrita como

$$w = \sum_{l=1}^{n} a_l y^{(l)}, \tag{6.1.11}$$

ou em notação de sequências

$$\{w_k\}_{k \in \mathbb{N}} = \sum_{l=1}^{n} a_l \{r_l^k\}_{k \in \mathbb{N}}. \tag{6.1.12}$$

Se conhecermos uma sequência que seja solução particular do problema não homogêneo, $\psi = \{\psi_k\}_{k \in \mathbb{N}}$, então a solução geral da equação de diferenças linear não homogênea será dada por

$$w_k = \sum_{l=1}^{n} a_l r_l^k + \psi_k.$$

Se o problema homogêneo estiver acompanhado de n condições iniciais (y_0, y_1, y_2, ...,y_{n-1}), então podemos encontrar os coeficientes a_l, de forma a obter a solução única da equação de diferenças (6.1.1), resolvendo o sistema (não singular)

$$\begin{cases} a_1 + a_2 + \cdots + a_n = y_0 - \psi_0 \\ a_1 r_1 + a_2 r_2 + \cdots + a_n r_n = y_1 - \psi_1 \\ a_1 r_1^2 + a_2 r_2^2 + \cdots + a_n r_n^2 = y_2 - \psi_2 \\ \vdots + \vdots + \vdots + \vdots = \vdots \\ a_1 r_1^n + a_2 r_2^n + \cdots + a_n r_n^n = y_{n-1} - \psi_{n-1} \end{cases}. \tag{6.1.13}$$

Um caso particular de interesse para a teoria a ser exposta a seguir ocorre quando $\sum_{j=0}^{n} \gamma_j \neq 0$ e se tem o lado direito não homogêneo de (6.1.1) dado por uma sequência *constante* $\{\phi_k\}_{k \in \mathbb{N}}$ para a qual se tem $\phi_k = \phi$, $k = 0, 1, 2, \ldots$. Nestas condições, não é difícil de se verificar que uma solução particular do problema é $\psi = \{\psi_k\}_{k \in \mathbb{N}}$ com

$$\psi_k = \frac{\phi}{\sum_{j=0}^{n} \gamma_j}, \quad k = 0, 1, \cdots,$$

desde que $\sum_{l=0}^{n} \gamma_l \neq 0$.

No caso de raízes múltiplas, alguns ajustes são necessários, e usaremos o seguinte teorema para nos auxiliar.

Teorema 6.2 (Solução de equações de diferenças lineares homogêneas)**.** *Considere a seguinte equação de diferenças linear homogênea*

$$\gamma_n y_{k+n} + \gamma_{n-1} y_{k+n-1} + \cdots + \gamma_0 y_k = \sum_{j=0}^{n} \gamma_j y_{k+j} = 0 \ , \ k = 0, 1, \ldots , \qquad (6.1.14)$$

com $\gamma_n \neq 0$, $\gamma_0 \neq 0$, $\gamma_j \in \mathbb{R}$, $j = 0, 1, ..., n$. *Sejam* r_l, $l = 1, 2, .., n_r$ *raízes do polinômio característico associado,*

$$\rho_n(r) = \sum_{l=0}^{n} \gamma_l r^l,$$

com respectivas multiplicidades m_1, m_2, ..., m_{n_r}, *satisfazendo* $\sum_{l=1}^{n_r} m_l = n$.

Se uma sequência $\{y_k\}$ *de números complexos satisfaz a equação de diferenças homogênea, então*

$$y_k = \sum_{l=1}^{n_r} p_l(k) r_l^k, \qquad (6.1.15)$$

onde $p_l(k)$ *é um polinômio em* k *de grau* $m_l - 1$. *Se todas as raízes forem simples* ($m_l = 1$)*, então* p_l *são constantes. [9]*

O caso de raízes simples já foi demonstrado anteriormente. Ressaltamos que caso hajam raízes complexas, é possível que observemos soluções também complexas. Porém, dadas condições iniciais reais, a sequência única a ser encontrada terá necessariamente apenas valores reais, tendo em vista que os próximos termos são obtidos a partir dos anteriores e os coeficiente γ_l são reais.

Comentaremos brevemente o caso no qual as raízes tem multiplicidades maiores ou iguais a 1.

Quando as raízes eram simples, tomávamos para cada raiz r_l uma sequência da forma $y^{(l)} = \{r_l^k\}_{k\in\mathbb{N}}$. Se r_l tem agora multiplicidade m_l, precisamos de m_l sequências relacionadas a esta raiz. Podemos definir tais sequências da seguinte forma:

$$\begin{aligned}
y^{(l,1)} &= \{r_l^k\}_{k\in\mathbb{N}} \\
y^{(l,2)} &= \{k r_l^k\}_{k\in\mathbb{N}} \\
y^{(l,3)} &= \{k(k-1) r_l^k\}_{k\in\mathbb{N}} \\
\vdots &= \vdots \\
y^{(l,m_l)} &= \{k(k-1)\ldots(k-m_l+2) r_l^k\}_{k\in\mathbb{N}}
\end{aligned}$$

É fácil mostrar que as sequências $y^{(l,k)}$, $k = 1, 2, ..., m_l$, satisfazem a equação homogênea, pois sabemos, da multiplicidade da raiz, que $\rho_n(r_l) = \rho_n'(r_l) = \rho_n''(r_l) = \ldots = \rho_n^{(m_l-1)}(r_l) = 0$. Resta mostrar que as sequências são linearmente independentes. O determinante do sistema associado a analise de independência linear tem

como determinante algo parecido com o caso de raízes simples, sendo este dependente de um produto de diferenças entre as raízes (com potências relacionadas às multiplicidades). Portanto, as sequências são linearmente independentes.

Assim, as soluções do problema homogêneo no caso de termos raízes com multiplicidade maior que um são da forma

$$w = \sum_{l=1}^{n_r} \sum_{j=1}^{m_l} a_{lj} y^{(l,j)}, \tag{6.1.16}$$

onde n_r é o número de raízes do polinômio característico, m_l são as respectivas multiplicidades e a_{lj} definem $n = \sum_{l=1}^{n_r} m_l$ coeficientes a serem encontrados com base nas condições iniciais.

Por exemplo, se r_1 for uma raiz de multiplicidade dois, e as demais raízes multiplicidade 1, tem-se como solução geral da equação não homogênea

$$y_k = a_{1,1} r_1^k + a_{1,2} k r_1^k + \sum_{l=2}^{n} a_l r_l^k + \psi_k, \tag{6.1.17}$$

onde ψ_k vem de uma solução particular da equação.

Explicitamente, se temos n_r raízes, com cada raiz r_l tendo multiplicidade m_l, a solução geral será

$$\begin{aligned} y_k &= \left[a_{1,1} + a_{1,2}k + \cdots a_{1,m_1} k(k-1)(k-2)...(k-m_1+2)\right] r_1^k \\ &+ \left[a_{2,1} + a_{2,2}k + \cdots + a_{2,m_2} k(k-1)(k-2)...(k-m_2+2)\right] r_2^k \\ &\vdots \\ &+ \left[a_{n_r,1} + a_{n_r,2}k + \cdots + a_{n_r,m_{n_r}} k(k-1)(k-2)...(k-m_{n_r}+2)\right] r_{n_r}^k \\ &+ \psi_k. \end{aligned} \tag{6.1.18}$$

Exemplo 6.1. *Considere a equação de diferenças*

$$y_{k+2} + y_{k+1} - 2y_k = 0, \tag{6.1.19}$$

com $y_0 = 0$, $y_1 = 1$. Encontre a solução (única) deste problema.
Solução: O polinômio característico,

$$\rho_2(r) = r^2 + r - 2 = (r-1)(r+2), \tag{6.1.20}$$

tem raízes $r_1 = 1$, $r_2 = -2$. Portanto, toda solução da equação pode ser escrita como

$$y_k = a_1(1)^k + a_2(-2)^k. \tag{6.1.21}$$

Impondo as condições iniciais temos,

$$a_1(1)^0 + a_2(-2)^0 = a_1 + a_2 = y_0 = 0, \tag{6.1.22}$$

$$a_1(1)^1 + a_2(-2)^1 = a_1 - 2a_2 = y_1 = 1. \tag{6.1.23}$$

$$\tag{6.1.24}$$

Portanto, $a_1 = 1/3$ e $a_2 = -1/3$ e a solução única resultante é

$$y_k = \frac{1}{3} - \frac{1}{3}(-2)^k. \tag{6.1.25}$$

Exemplo 6.2 ([16])**.** *Encontre a solução da equação de diferenças linear*

$$y_{k+4} - 4y_{k+3} + 5y_{k+2} - 4y_{k+1} + 4y_k = 4 \tag{6.1.26}$$

satisfazendo $y_0 = 5$, $y_1 = 0$, $y_2 = -4$ *e* $y_3 = -12$.
Solução:

$$\begin{aligned} p_4(r) &= \sum_{j=0}^{4} \gamma_j r^j = \gamma_4 r^4 + \gamma_3 r^3 + \gamma_2 r^2 + \gamma_1 r + \gamma_0 \\ p_4(r) &= r^4 - 4r^3 + 5r^2 - 4r + 4 \\ &= (r^2+1)(r-2)^2 \\ &= (r-i)(r+i)(r-2)^2 \\ &\textit{Raízes:}\quad r_1 = 2 \text{ ; } r_2 = -i \text{ e } r_3 = i. \end{aligned} \tag{6.1.27}$$

Solução particular:

$$\psi_k = \frac{\phi}{\sum_{j=0}^{n} \gamma_j} = \frac{4}{1-4+5-4+4} = 2 \ , \ k \in \mathbb{N}.$$

Como $r_1 = 2$ *é uma raiz de multiplicidade dois e a solução particular é* $\psi_k = 2$, *a solução geral da equação não homogênea é dada por*

$$y_k = a_{1,1} r_1^k + a_{1,2} k r_1^k + a_2 r_2^k + a_3 r_3^k + 2. \tag{6.1.28}$$

Das condições iniciais

$$\begin{aligned} y_0 &= a_{1,1} + a_2 + a_3 + 2 = 5, \\ y_1 &= a_{1,1} 2 + a_{1,2} 1 \cdot 2 + a_2(-i) + a_3(i) + 2 = 0, \\ y_2 &= a_{1,1} 2^2 + a_{1,2} 2 \cdot 2^2 + a_2(-i)^2 + a_3(i)^2 + 2 = -4, \\ y_3 &= a_{1,1} 2^3 + a_{1,2} 3 \cdot 2^3 + a_2(-i)^3 + a_3(i)^3 + 2 = -12, \end{aligned}$$

obtêm-se um sistema linear de quarta ordem cujas incógnitas são $a_{1,1}$, $a_{1,2}$, a_2 *e* a_3. *A solução desse sistema é*

$$a_{1,1} = 1 \ , \ a_{1,2} = -1 \ , \ a_2 = 1 - i \ , \ a_3 = 1 + i. \tag{6.1.29}$$

Substituindo as raízes (6.1.27) e os coeficientes (6.1.29) na equação (6.1.28), tem-se

$$\begin{aligned} y_k &= 1 \cdot 2^k + (-1) k 2^k + (1-i)(-i)^k + (1+i)(i)^k + 2 \\ &= (1-k)2^k + (1-i)(-i)^k + (1+i)(i)^k + 2 \ , \ k = 0, 1, \ldots . \end{aligned} \tag{6.1.30}$$

(Verifique que de fato essa sequência resolve a equação de diferenças!)

6.1.1 Exercícios

Exercício 6.1. *Calcule os termos* y_4, y_5 *e* y_6 *da solução (6.1.30) empregando*

(a) a própria solução (6.1.30);

(b) a equação de diferenças linear não homogenêa (6.1.26).

Exercício 6.2. *Como os coeficientes de (6.1.26) são reais e as condições iniciais também, y_k é necessariamente real. Mostre que $(1-i)(-i)^k + (1+i)(i)^k$ é um número real para qualquer natural k ou, equivalentemente, mostre que*

$$y_k = 2\left[cos\left(\frac{k\pi}{2}\right) - sen\left(\frac{k\pi}{2}\right)\right] + (1-k)2^k + 2$$

lembrando que $i = e^{i\frac{\pi}{2}} = cos\left(\frac{\pi}{2}\right) + isen\left(\frac{\pi}{2}\right)$.

Exercício 6.3. *A sequência de Fibonacci é uma sequência de números inteiros onde $y_0 = 0$, $y_1 = 1$ e cada um dos demais termos é dado pela soma dos dois termos precedentes.*

(a) Escreva a equação de diferenças linear que define a sequência de Fibonacci.

(b) Determine a solução empregando a teoria estudada.

(c) Determine os dez primeiros termos da sequência de Fibonacci.

(d) Determine y_{100}, usando a solução analítica calculada.

(e) Determine y_{101}/y_{100} e analise o limite de y_{k+1}/y_k quando $k \to \infty$.

6.2 Exemplo de divergência

Discutiremos aqui um exemplo de um método de terceira ordem de consistência que não converge para a solução esperada do problema de Cauchy. Partes deste exemplo estão descritas em detalhes em [24].

6.2.1 Consistência e divergência

Considere o método de 2-passos explícito

$$y_{k+2} + \alpha_1 y_{k+1} + \alpha_0 y_k = h\left[\beta_1 f_{k+1} + \beta_0 f_k\right]. \tag{6.2.31}$$

Calculando-se os coeficientes do erro de discretização local obtêm-se

$$\begin{aligned}
C_0 &= \alpha_0 + \alpha_1 + 1 = 0,\\
C_1 &= 0\cdot\alpha_0 + 1\cdot\alpha_1 + 2\cdot 1 - \beta_0 - \beta_1 = 0,\\
C_2 &= \frac{0^2\cdot\alpha_0 + 1^2\cdot\alpha_1 + 2^2\cdot 1}{2!} - \frac{0\cdot\beta_0 + 1\cdot\beta_1}{1!} = 0,\\
C_3 &= \frac{0^3\cdot\alpha_0 + 1^3\cdot\alpha_1 + 2^3\cdot 1}{3!} - \frac{0^2\cdot\beta_0 + 1^2\cdot\beta_1}{2!} = 0.
\end{aligned} \tag{6.2.32}$$

A solução do sistema linear (6.2.32) é $\alpha_0 = -5$, $\alpha_1 = 4$, $\beta_0 = 2$ e $\beta_1 = 4$. Com os coeficientes determinados, o método resultante,

$$y_{k+2} + 4y_{k+1} - 5y_k = h\left[4f_{k+1} + 2f_k\right], \tag{6.2.33}$$

assume a maior ordem possível. Portanto, obtivemos um método de terceira ordem de consistência, mas que possui apenas 2 passos. Para ser útil, este método deveria funcionar para qualquer problema de valor inicial que tenha f suficientemente suave.

Exemplo 6.3. *Solucione o Problema de Cauchy*

$$\begin{cases} \dfrac{d}{dt}y(t) = 0, \ \ t \in [0,1], \\ y(0) = 0, \end{cases} \tag{6.2.34}$$

utilizando o método de terceira ordem (6.2.33) com inicialização consistente dada por $y_1 = h$.
Solução: O problema de valor inicial está bem posto e tem como solução única a soluçã $y(t) = 0$. *O método pode ser escrito para este problema como*

$$y_{k+2} + 4y_{k+1} - 5y_k = 0, \tag{6.2.35}$$

que é uma equação de diferenças linear homogênea. O polinômio característico tem raízes $r_1 = 1$ *e* $r_2 = -5$, *portanto as solução devem ser da forma*

$$y_k = a_1(1)^k + a_2(-5)^k. \tag{6.2.36}$$

Imponto as condições iniciais temos que

$$\begin{aligned} a_1 + a_2 \quad &= 0 \\ a_1 - 5a_2 \quad &= h, \end{aligned}$$

que definem $a_1 = -h/6$ *e* $a_2 = h/6$. *Assumindo um* $t \in [0,1]$ *fixado, e portanto que* $t - 0 = kh$ *é constante, temos que a solução única será dada por*

$$y_k = -\frac{t}{6k} + \frac{t}{6k}(-5)^k. \tag{6.2.37}$$

Agora note que, por exemplo em $t = 1$, *temos*

$$\lim_{h\to 0} y_k = \lim_{k\to\infty}(-\frac{1}{6k} + \frac{(-5)^k}{6k}) = \lim_{k\to\infty}\frac{(-5)^k}{6k} \neq 0, \tag{6.2.38}$$

portanto o método, apesar de ser consistente com o problema, não converge para solução analítica do mesmo com $h \to 0$. *O pequeno erro cometido na inicialização, ainda que esta tenha sido feita de forma consistente, é exponencialmente amplificado pelo termo* $(-5)^k$ *da solução, que também fará com que a solução numérica oscile em torno do zero. Note também que isso vale para qualquer* $t > 0$ *fixado.*

O exemplo anterior mostra que o método apresentado aqui não será útil mesmo para um problema muito simples (este será instável) se tivermos pequenos erros na inicialização. Se a condição inicial tivesse sido imposta de forma exata no exemplo anterior, com $y_1 = 0$, então o método seria exato, sem erro algum, com $y_k = 0$, $\forall k > 0$. Para compreender melhor esse fenômeno, vamos dar um exemplo menos trivial com inicialização sem erro.

Exemplo 6.4. *Solucione o Problema de Cauchy*

$$\begin{cases} \dfrac{d}{dt}y(t) = -y(t), \ \ t \in [0,1], \\ y(0) = 1, \end{cases} \tag{6.2.39}$$

utilizando o método de terceira ordem (6.2.33) com inicialização $y_0 = 1$, $y_1 = e^{-h}$ *e* $h = 0,01$.
Solução:

A solução exata do problema é $y(t) = e^{-t}$, portanto a inicialização está sendo feita de forma analítica. A discretização de (6.2.39) por (6.2.33) é dada por

$$\begin{aligned} y_{k+2} &= -4y_{k+1} + 5y_k + h\,[4f_{k+1} + 2f_k] = \\ &= -4y_{k+1} + 5y_k + h\,[-4y_{k+1} - 2y_k] = \\ &= -(4+4h)y_{k+1} + (5-2h)y_k, \end{aligned}$$

ou, reescrevendo-a, por

$$y_{k+2} + (4+4h)y_{k+1} + (-5+2h)y_k = 0. \tag{6.2.40}$$

A equação (6.2.40) é uma equação de diferenças linear homogênea com $\gamma_2 = 1$, $\gamma_1 = 4+4h$ e $\gamma_0 = -5+2h$, cuja solução (exata) se escreve como

$$y_k = a_1 r_1^k + a_2 r_2^k, \tag{6.2.41}$$

onde r_j são as raízes do polinômio

$$p_2(r) = \sum_{j=0}^{2} \gamma_j r^j = \gamma_2 r^2 + \gamma_1 r^1 + \gamma_0 = r^2 + (4+4h)r + (-5+2h)$$

e as contantes a_j, $j = 1, 2$, são calculadas a partir das condições iniciais da equação de diferenças

$$y(t_0) = y_0 = 1 \quad e \quad y_1 = y(t_0 + h) = y(0 + h) = e^{-h}.$$

Calculando-se as raízes de $p_2(r)$, obtêm-se

$$r_1 = -2 - 2h + 3\sqrt{1 + \frac{2}{3}h + \frac{4}{9}h^2}$$

e

$$r_2 = -2 - 2h - 3\sqrt{1 + \frac{2}{3}h + \frac{4}{9}h^2},$$

e calculando-se a_1 e a_2 chega-se a

$$\begin{cases} 1 = y_0 = a_1 + a_2 \\ e^{-h} = y_1 = a_1 r_1 + a_2 r_2 \end{cases} \Rightarrow \quad a_1 = \frac{r_2 - e^{-h}}{r_2 - r_1} \quad e \quad a_2 = \frac{e^{-h} - r_1}{r_2 - r_1} \ .$$

Assim,

$$y_k = \frac{r_2 - e^{-h}}{r_2 - r_1} r_1^k + \frac{e^{-h} - r_1}{r_2 - r_1} r_2^k \tag{6.2.42}$$

é a solução exata da equação de diferenças (6.2.40), resultante da aplicação do esquema numérico (6.2.33) ao Problema de Cauchy (6.2.39).

As diferenças entre a solução numérica (6.2.42) e a solução exata, para vários valores de instantes de tempo, encontram-se na Tabela 6.1. Observa-se que o método de passo múltiplo linear consistente (6.2.33) diverge da solução do problema (6.2.39).

É importante destacar que o resultado exibido pela Tabela 6.1 não melhora fazendo-se com que o passo de integração h diminua. No limite para h tendendo a zero, r_1 tende a um e r_2 tende a -5. O caráter ilimitado do erro de discretização global é devido à raiz com módulo maior que um, associada ao polinômio característico da equação de diferenças (6.2.40) quando $h = 0$. Como conclusão, vê-se que, mesmo consistentes e com ordem máxima, métodos de passo múltiplo lineares podem ser divergentes. O que falta para a convergência é o estudo da zero-estabilidade de tais esquemas numéricos.

k	$y_k - e^{t_k}$
2	$-1,4 \times 10^{-9}$
3	$+5,01 \times 10^{-9}$
4	$-3,00 \times 10^{-8}$
5	$+1,44 \times 10^{-7}$
$\vdots$	$\vdots$
98	$-2,57 \times 10^{58}$
99	$+1,29 \times 10^{59}$
100	$-6,52 \times 10^{59}$

Tabela 6.1: Erro de discretização global obtido na solução do problema de valor inicial (6.2.39) com o método (6.2.33) e $h = 0,01$.[24]

6.2.2 Exercícios

Exercício 6.4. *Calcule $y_2, \cdots, y_{100}$ usando a solução exata da equação de diferenças, (6.2.42), e usando a própria equação de diferenças, (6.2.33).*

6.2.3 Relação com as raízes do polinômio característico

Os resultados presentes na Tabela 6.1 mostram que consistência e alta ordem não bastam à convergência. Usando a solução (6.2.41)-(6.2.42) da equação de diferenças, pode-se explicar as causas desse comportamento.

Ao se expandir as raízes r_1 e r_2, assim como das constantes a_1 e a_2, em Série de Taylor como função de h em torno de $h = 0$, tem-se

$$\begin{aligned} r_1 &= -2 - 2h + 3\sqrt{1 + \frac{2}{3}h + \frac{4}{9}h^2} = 1 - h + \frac{1}{2}h^2 - \frac{1}{6}h^3 + \frac{1}{72}h^4 + O(h^5) \\ &\text{e} \\ r_2 &= -2 - 2h - 3\sqrt{1 + \frac{2}{3}h + \frac{4}{9}h^2} = -5 - 3h - \frac{1}{2}h^2 + \frac{1}{6}h^3 - \frac{1}{72}h^4 + O(h^5) \end{aligned}$$

pois

$$\sqrt{1 + \frac{2}{3}h + \frac{4}{9}h^2} = 1 + \frac{1}{3}h + \frac{1}{6}h^2 - \frac{1}{18}h^3 + \frac{1}{216}h^4 + O(h^5).$$

Lembrando-se de que a função exponencial tem desenvolvimento em série de Taylor dado por

$$e^{-h} = \sum_{n=0}^{\infty} \frac{(-h)^n}{n!} = 1 - h + \frac{1}{2}h^2 - \frac{1}{6}h^3 + \frac{1}{24}h^4 + O(h^5),$$

para as constantes a_1 e a_2, obtêm-se

$$\begin{aligned} a_1 = \frac{r_2 - e^{-h}}{r_2 - r_1} &= 1 + O(h^2), \\ a_2 = \frac{e^{-h} - r_1}{r_2 - r_1} &= -\frac{1}{216}h^4 + O(h^5). \end{aligned}$$

Dessa forma, das observações anteriores, têm-se a solução para o tempo $t = hk$

$$\begin{aligned} y_k &= a_1 r_1^k + a_2 r_2^k \\ &= \left[1 + O\left(\left(\frac{t}{k}\right)^2\right)\right]\left[1 - \left(\frac{t}{k}\right) + O\left(\left(\frac{t}{k}\right)^2\right)\right]^k + \\ &- \frac{1}{216}\left(\frac{t}{k}\right)^4\left[1 - O\left(\frac{t}{k}\right)\right]\left[-5 - 3\left(\frac{t}{k}\right) + O\left(\left(\frac{t}{k}\right)^2\right)\right]^k = \\ &= \left[1 + O\left(\left(\frac{t}{k}\right)^2\right)\right]\left[1 - \left(\frac{t}{k}\right) + O\left(\left(\frac{t}{k}\right)^2\right)\right]^k + \\ &- \frac{1}{216}\left(\frac{t}{k}\right)^4 (-5)^k\left[1 - O\left(\frac{t}{k}\right)\right]\left[1 + \frac{3}{5}\left(\frac{t}{k}\right) + O\left(\left(\frac{t}{k}\right)^2\right)\right]^k, \end{aligned} \tag{6.2.43}$$

onde y_k representa a solução numérica, ou seja, a solução da equação de diferenças linear.

Lembrando-se que $kh = t - 0$ permanece fixado, à medida que k tende a infinito, observa-se que a primeira parcela do lado direito de (6.2.43) tende a e^{-t}, a solução exata do Problema de Cauchy (6.2.39). Entretanto, a segunda parcela comporta-se como

$$-\frac{t^4}{216}\frac{(-5)^k}{k^4}e^{\frac{3t}{5}}. \tag{6.2.44}$$

Nas considerações anteriores, empregam-se os limites fundamentais

$$\lim_{k\to\infty}\left(1 + \frac{1}{k}\right)^k = e,$$

$$\lim_{k\to 0}(1+k)^{\frac{1}{k}} = e$$

e

$$\lim_{k\to\infty}\left(1 + \frac{a}{b}\frac{1}{k}\right)^k = e^{\frac{a}{b}}.$$

Em (6.2.44),

$$\lim_{k\to\infty}\left|\frac{(-5)^k}{k^4}\right| = \infty, \tag{6.2.45}$$

uma vez que $5^k >> k^4$ quando $k \to \infty$.

Portanto, o limite (6.2.45) explica o comportamento oscilatório e ilimitado da solução y_k quando $k \to \infty$ ($k \to \infty \iff h \to 0$).

De (6.2.43), conclui-se então que o comportamento da solução da equação de diferenças é controlado pela maior raiz do polinômio

$$p_2(r) = r^2 + (4 + 4h)r + (-5 + 2h),$$

ou seja, $r_1 = -2 - 2h + 3\sqrt{1 + \frac{2}{3}h + \frac{4}{9}h^2}$. Com h tendendo a zero, tem-se

$$p_2(r) \to r^2 + 4r - 5,$$

cujas raízes são $r_1 = 1$ e $r_2 = -5$.

6.3 Primeiro e segundo polinômios característicos

Considere o método de passo múltiplo linear

$$y_{k+n} + \sum_{j=0}^{n-1} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f_{k+j}. \tag{6.3.46}$$

O primeiro e o segundo polinômios característicos associados ao método são definidos, respectivamente, por ρ e σ,

$$\rho(r) = \sum_{j=0}^{n} \alpha_j r^j \tag{6.3.47}$$

$$\sigma(r) = \sum_{j=0}^{n} \beta_j r^j, \tag{6.3.48}$$

onde α_n é tomado como sendo um.

Observe que, em relação aos polinômios característicos (6.3.47) e (6.3.48), um método de passo múltiplo linear (6.3.46) é consistente para qualquer problema de valor inicial bem posto se, e somente se,

$$\rho(1) = 0$$

e

$$\rho'(1) - \sigma(1) = 0.$$

Veja

$$\begin{aligned} \rho(r) &= \sum_{j=0}^{n} \alpha_j r^j = \alpha_0 r^0 + \alpha_1 r^1 + \alpha_2 r^2 + \alpha_3 r^3 + \cdots + \alpha_n r^n, \\ \rho(1) &= \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_n = \sum_{j=0}^{n} \alpha_j = C_0. \end{aligned}$$

Portanto,

$$\rho(1) = 0 \Rightarrow C_0 = 0.$$

Além disso,

$$\begin{aligned} \rho'(r) &= \alpha_1 + 2\alpha_2 r^1 + 3\alpha_3 r^2 + 4\alpha_4 r^3 + \cdots + n\alpha_n r^{n-1}, \\ \rho'(1) &= \alpha_1 + 2\alpha_2 + 3\alpha_3 + 4\alpha_4 + \cdots + n\alpha_n = \sum_{j=0}^{n} j\alpha_j \sigma(r), \end{aligned}$$

$$\begin{aligned} \sigma(r) &= \sum_{j=0}^{n} \beta_j r^j = \beta_0 r^0 + \beta_1 r^1 + \beta_2 r^2 + \beta_3 r^3 + \cdots + \beta_n r^n, \\ \sigma(1) &= \beta_0 + \beta_1 + \beta_2 + \beta_3 + \cdots + \beta_n = \sum_{j=0}^{n} \beta_j \end{aligned}$$

$$\rho'(1) - \sigma(1) = \sum_{j=0}^{n} j\alpha_j - \sum_{j=0}^{n} \beta_j = C_1$$

portanto,

$$\rho'(1) - \sigma(1) = 0 \Rightarrow C_1 = 0.$$

Em um método de passo múltiplo linear, o primeiro polinômio característico $\rho(r)$ tem sempre uma raiz igual a 1. Esta raiz é denominada *raiz principal* e geralmente denotada por r_1. Assim, a consistência de um método de passo múltiplo linear depende apenas da raiz principal $r_1 = 1$.

6.4 Zero-estabilidade

A zero-estabilidade de um método de passo múltiplo linear diz respeito ao estudo da estabilidade para no limite para h tendendo a zero. Esta pode ser analisado com base no seguinte problema de valor inicial,

$$\begin{cases} \dfrac{d}{dt}y(t) = 0 \quad t \in [t_0, T] \\ \\ y(t_0) = 0 \end{cases}, \tag{6.4.49}$$

cuja solução exata é dada por $y(t) = 0$. Ao se aplicar o método de passo múltiplo linear

$$\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f_{k+j} \tag{6.4.50}$$

para resolver (6.4.49), tem-se

$$\alpha_n y_{k+n} + \alpha_{n-1} y_{k+n-1} + \cdots + \alpha_0 y_k = 0, \tag{6.4.51}$$

uma vez que $f(t, y(t)) = 0$. A relação (6.4.51) é uma equação de diferenças linear homogênea.

Se o método (6.4.50) for convergente então

$$\lim_{h \to 0} y_k = 0, \quad kh = t - t_0 \text{ fixado}, \quad \forall t \in [t_0, T].$$

Para simplificar a análise num primeiro momento, suponha que as raízes do polinômio característico da equação de diferenças (6.4.51)

$$p_n(r) = \sum_{j=0}^{n} \alpha_j r^j, \tag{6.4.52}$$

o qual coincide com o primeiro polinômio característico do método de passo múltiplo (6.4.50), sejam reais e distintas, denotadas por r_l, $l = 1, 2, \ldots, n$. Neste contexto, a solução de (6.4.51) pode ser escrita como

$$y_k = a_1 r_1^k + a_2 r_2^k + a_3 r_3^k + \cdots + a_n r_n^k, \tag{6.4.53}$$

onde as constantes a_l, $l = 1, 2, \ldots, n$, são obtidas a partir das n primeiras aproximações da solução exata $y_0, y_1, y_2, \ldots, y_{n-1}$, nos instantes $t_0, t_1, t_2, \ldots, t_{n-1}$. É natural que se utilize $y_0 = y(t_0) = 0$ (a condição inicial dada para o problema estudado). As aproximações nos outros $n-1$ instantes de tempo devem ser obtidas

numericamente, por exemplo e para fixar ideias, por um método de passo único de primeira ordem, isto é,

$$y_k = 0 + \mathrm{O}(h), \quad k = 1, 2, \ldots, n-1. \tag{6.4.54}$$

Buscamos soluções da forma (6.4.53) que respeitem as condições iniciais (6.4.54), portanto

$$a_1 r_1^k + a_2 r_2^k + a_3 r_3^k + \cdots + a_n r_n^k = O(h), \quad , k = 1, 2, \ldots, n-1. \tag{6.4.55}$$

É simples ver que as solução desse problema são da forma $a_l = h d_l$, para que a igualdade seja verificada com $h \to 0$. Logo, pode-se escrever a solução do problema homogêneo como y_k, $k \geq n$, da forma

$$y_k = h\left(d_1 r_1^k + d_2 r_2^k + \cdots + d_n r_n^k\right).$$

Considerando a convergência do método para a solução exata do problema (no caso, a solução trivial), tem-se

$$\lim_{\substack{h\to 0 \\ kh = t - t_0}} y_k = 0$$

e, portanto, para cada parcela que compõe a solução numérica y_k devemos ter

$$\lim_{\substack{h\to 0 \\ kh = t - t_0}} h r_j^k = \lim_{k\to\infty} \frac{t - t_0}{k} r_j^k = t - t_0 \lim_{k\to\infty} \frac{r_j^k}{k} = 0. \tag{6.4.56}$$

Note que o método jamais poderá ser convergente se o módulo de qualquer uma das raízes r_j for maior do que 1. Em outras palavras, o limite em (6.4.56) será zero se e somente se $|r_j| \leq 1$, $j = 1, 2, \ldots, n$. Isso nos leva ao seguinte teorema.

Teorema 6.3 (Critério das raízes simples reais)**.** *Um método de passos múltiplos linear de n passos da forma (6.4.50), com inicialização consistente, que tenha o primeiro polinômio característico com raízes simples, distintas e reais é convergente somente se*

$$|r_j| \leq 1, \quad j = 1, \cdots, n \ ,$$

sendo r_j as raízes do seu primeiro polinômio característico.

Se o polinômio (6.4.52) tiver uma raiz com multiplicidade maior que um, por exemplo r_j com multiplicidade $m > 1$, então tal raiz contribuirá para a solução da equação de diferenças com uma parcela da forma

$$\left[a_{j,1} + a_{j,2} k + \cdots + a_{j,m} k(k-1)\cdots(k-m+2)\right] r_j^k.$$

Se $|r_j| \geq 1$, o termo dominante quando $k \to \infty$ ($h \to 0$) é o termo de maior grau em k, referente a $a_{j,m}$. Analisando a forma geral da solução do problema, descrita na equação (6.1.18), vemos que quando houver uma multiplicidade $m > 1$, essa raiz sempre contribuirá com uma parcela da forma $h k^{m-1} r_j^k$, $m > 1$. Para tais parcelas, tem-se, para um instante de tempo t fixado, o limite

$$\lim_{\substack{h\to 0 \\ kh = t - t_0}} h k^{m-1} r_j^k = \lim_{k\to\infty} \frac{t - t_0}{k} k^{m-1} r_j^k = t - t_0 \lim_{k\to\infty} k^{m-2} r_j^k. \tag{6.4.57}$$

Se $m = 2$, então esta equação (6.4.57), e de fato a equação geral (6.4.56), será da forma

$$\lim_{k\to\infty} k\frac{r_j^k}{k}. \tag{6.4.58}$$

Neste caso, como $|r_j| \geq 1$, o limite não vai a zero com $k \to \infty$. Se $|r_j| = 1$, o limite será ± 1, e se $|r_j| > 1$, o limite será $\pm\infty$. Caso $m > 2$, então vemos que a equação (6.4.57) irá para $\pm\infty$ quando $k \to \infty$.

Concluímos, portanto, que se $m > 1$ não há convergência se $|r_j| \geq 1$, mesmo que $|r_j| = 1$. Logo, o método poderá convergir para a solução exata do problema-modelo proposto somente se o limite em (6.4.57) for zero, que só ocorre quando $|r_j| < 1$.

Teorema 6.4 (Critério das raízes)**.** *Um método de passos múltiplos linear de n passos da forma (6.4.50) que tenha primeiro polinômio característico com raízes r_j, $j = 1, 2, ..., n_r$, é convergente somente se*

$$\begin{aligned} &|r_j| \leq 1, && \text{se } r_j \text{ for raiz simples,} \\ &|r_j| < 1, && \text{se } r_j \text{ for raiz com multiplicidade maior que 1.} \end{aligned}$$

É importante destacar que o critério das raízes é necessário para convergência, mas não suficiente, pois mostramos a sua necessidade para apenas um único problema de valor inicial. Para garantia de convergência para qualquer problema de valor inicial bem posto vamos precisar de algo além desse crierio.

Definição 6.3 (Zero-estabilidade)**.** *Um método de passo múltiplo linear é zero-estável se nenhuma raiz do primeiro polinômio característico*

$$\rho(r) = \sum_{j=0}^{n} \alpha_j r^j \tag{6.4.59}$$

tiver módulo maior que 1 e toda raiz com módulo 1 for simples (multiplicidade um).

Note que aqui estamos também interessados em raízes complexas. Portanto, dizemos que o método é zero-estável se as raízes do seu primeiro polinômio característico estiverem no interior do disco unitário fechado do plano complexo, sendo que aquelas que estiveram no bordo (tenham módulo igual a 1) devem ter multiplicidade 1.

Se analisarmos o caso dos métodos de passo único explícitos (2.0.1), vemos que estes sempre tem primeiro polinômio característico dado por

$$\rho(r) = r - 1, \tag{6.4.60}$$

com raiz simples e igual a 1. Se o método for consistente com qualquer problema de Cauchy bem posto, de forma a garantir que a Φ do lado direito de (2.0.1) seja de pelo menos $O(h^p)$, $p \geq 0$, e portanto respeitando o problema modelo $y' = 0$ quando $h \to 0$, temos o seguinte resultado.

Teorema 6.5. *Todo método de passo único explícito consistente é zero-estável.*

O principal resultado deste capítulo relaciona zero-estabilidade com convergência, e é descrito a seguir.

Teorema 6.6 (Teorema de Equivalência de Dahlquist)**.** *Um método de n passos linear* ***consistente*** *com a equação diferencial $y' = f(t, y)$, sendo f Lipschitz, com condições iniciais consistentes, isto é, $y_j = \theta_j(h)$, $j = 0, 1, ..., n-1$ são tais que $\theta_j(h) \to 0$ quando $h \to 0$, é* ***convergente*** *se e somente se é* ***zero-estável****. Além disso, se f for suficientemente diferenciável, e o erro local de discretização e inicialização tiverem ambas ordem p, então o erro global também terá ordem p.*

A demonstração da necessidade de um método ser zero-estável para obtermos convergência segue do critério das raízes. Porém, a volta, que garante que um método consistente zero-estável é sempre convergente, exige outros resultados que estão além do escopo destas notas. A prova pode ser encontrada em [13, 24], mas foi primeiro discutida por Dalhquilst [6].

6.5 Ordem de convergência

O teorema de equivalência de Dahlquist nos garante que para buscar métodos de uma certa ordem de convergência, que sejam zero-estáveis, basta analisar a ordem consistência. Com base no Teorema 5.3, se os coeficientes C_p, deduzidos em (5.3.13), forem nulos até o coeficiente de ordem p, isto é $C_j = 0$, $j = 0, 1, ..., p$, então o método é de ordem $O(h^p)$ de consistência. Se for zero-estável, então também terá ordem $O(h^p)$ de convergência.

Suponha um método de n passos linear. Temos $2(n+1)$ coeficientes $(\alpha_j,\ \beta_j)$ a serem determinados. Como podemos, sem perda de generalidade, multiplicar a equação do método por uma constante sem afetar a solução, temos na prática apenas $2n+1$ coeficientes a serem determinados. Portanto, à princípio, poderíamos tentar buscar métodos de até ordem $p = 2n$.

Isso significa, por exemplo, que poderíamos ter um método de 2-passos com ordem 4. Vimos no exemplo (6.2.31) que isso pode ser perigoso. Neste exemplo apresentamos um método de 2 passos explícito com ordem 3, que foi definido unicamente, ou seja, é o único método de 2 passos e ordem 3 explícito possível, portanto seria um método ótimo. Na prática, verificamos que o método era instável.

As restrições impostas pela zero-estabilidade reduzem as possibilidade de termos métodos de n passos com ordem $2n$, de fato, temos o seguinte importante resultado que estabelece uma barreira para a ordem.

Teorema 6.7 (Barreira de Dahlquist)**.** *Nenhum método de passo múltiplo linear de n passos ($n \geq 2$) zero-estável tem ordem maior que $n+1$ se n for ímpar, ou ordem maior que $n+2$ se n for par. [13, 23]*

Além disso, é possível mostrar que apenas métodos implícitos atingem a ordem máxima definida pela barreira de Dahlquist.

Exemplo 6.5. *Quando $n = 1$, temos o método o método do Trapézio implícito dado por*

$$y_{k+1} - y_k = \frac{h}{2}(f_{k+1} + f_k). \tag{6.5.61}$$

Mostre que este método é consistente com qualquer problema de valor inicial bem posto com ordem 2, zero-estável e convergente. Justifique porque o método não pode ter ordem de convergência maior que 2.
Solução:

O método é um método de passo único linear implícito. É simples ver que, para este método temos $C_0 = 0$, $C_1 = 0$, $C_2 = 0$, portanto tem consistência de ordem 2 para qualquer problema de valor inicial bem posto. Como o método é de passo único, pelo Teorema 6.5 ele é zero-estável. Se inicializarmos o método com a condição inicial dado do problema, então a inicialização será consistente. Portanto, pelo teorema da equivalência de Dahlquist, o método é convergente de ordem 2. Concluímos ainda que como $n = 1$ é ímpar, a maior ordem possível de convergência é 2, pelo Teorema da Barreira de Dahlquist.

6.5.1 Exercícios

Exercício 6.5. *Vimos no Teorema 6.5 que todo método de passo único consistente é zero-estável. Por outro lado, vimos no Exercício Resolvido 2.2 que um método de passo único pode ser convergente e ao mesmo tempo inconsistente. Justifique por que não há contradição destas observações com o Teorema de Equivalência de Dahlquist 6.6.*

Exercício 6.6. *Verifique se o método de passo múltiplo linear*

$$y_{k+2} + 4y_{k+1} - 5y_k = h\left[4f_{k+1} + 2f_k\right]$$

é zero-estável.

Exercício 6.7. *Analise a zero-estabilidade do método de passo múltiplo linear*

$$y_{k+2} - (1+a)y_{k+1} + ay_k = \frac{h}{2}\left[(3-a)f_{k+1} - (1+a)f_k\right]$$

para

1. $a = 0$;
2. $a = -5$.

Exercício 6.8. *Mostre que o método de passo múltiplo linear*

$$y_{k+2} + (b-1)y_{k+1} - by_k = \frac{h}{4}\left[(b+3)f_{k+2} + (3b+1)f_k\right]$$

é zero-instável se $b = -1$.

Exercício 6.9. *Empregue o método de passo múltiplo linear*

$$y_{k+2} + (b-1)y_{k+1} - by_k = \frac{h}{4}\left[(b+3)f_{k+2} + (3b+1)f_k\right],$$

com $b = 0$ e $b = -1$, para solucionar o p.v.i.

$$\begin{cases} \dfrac{d}{dt}y(t) = y(t), \quad t \in [0,1] \\ \\ y_0 = y_1 = 0 \end{cases}.$$

Comente os resultados obtidos.

Exercício 6.10. *Considere o seguinte método de 2 passos linear implícito*

$$y_{k+2} - y_k = h\left(\frac{1}{3}f_{k+2} + \frac{4}{3}f_{k+1} + \frac{1}{3}f_k\right).$$

Mostre que o método é zero-estável e tem ordem de consistência 4. Observe que esta é a maior ordem possível para um método de 2 passos.

Exercício 6.11. *Considere o seguinte método de 2 passos linear explícito da forma*

$$y_{k+2} + \alpha_1 y_{k+1} + \alpha_0 y_k = h\left(\beta_1 f_{k+1} + \beta_0 f_k\right).$$

Mostre que não é possível acharmos coeficientes α_1, α_0, β_1, β_0 tais que o método seja zero-estável e consistente de ordem 4, que é a maior ordem possível para métodos de 2 passos pela Barreira de Dahlquist.

6.6 Erro de discretização global

Para o método de passo múltiplo linear

$$\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f_{k+j}, \tag{6.6.62}$$

o erro de discretização local multiplicado pelo passo de integração h e a solução numérica são dados, respectivamente, por

$$h\tau_k = \sum_{j=0}^{n} \alpha_j y(t_{k+j}) - h \sum_{j=0}^{n} \beta_j f\left(t_{k+j}, y(t_{k+j})\right), \tag{6.6.63}$$

$$R_k = \sum_{j=0}^{n} \alpha_j y_{k+j} - h \sum_{j=0}^{n} \beta_j f(t_{k+j}, y_{k+j}). \tag{6.6.64}$$

Observe que aqui estamos supondo que a equação de diferenças numérica não é resolvida exatamente, pois esta pode acumular, por exemplo, erros de arredondamento.

Efetuando-se a subtração (6.6.63) - (6.6.64), tem-se

$$\begin{aligned} h\tau_k - R_k &= \sum_{j=0}^{n} \alpha_j \left(y(t_{k+j}) - y_{k+j}\right) + \\ &- h \sum_{j=0}^{n} \beta_j \left(f\left(t_{k+j}, y(t_{k+j})\right) - f(t_{k+j}, y_{k+j})\right). \end{aligned} \tag{6.6.65}$$

Considerando-se em (6.6.65) $\phi_k = h\tau_k - R_k$ e empregando-se o Teorema do Valor Médio, chega-se a

$$\begin{aligned} \phi_k &= \sum_{j=0}^{n} \alpha_j \left(y(t_{k+j}) - y_{k+j}\right) + \\ &- h \sum_{j=0}^{n} \beta_j \frac{\partial}{\partial y} f\left(t_{k+j}, \xi_{k+j}\right)\left(y(t_{k+j}) - y_{k+j}\right), \end{aligned} \tag{6.6.66}$$

para ξ_{k+j} entre $y(t_{k+j})$ e y_{k+j}.

Lembrando-se que o erro de discretização global, no instante de tempo $t = t_{k+j}$, é definido como sendo

$$e_{k+j} = y(t_{k+j}) - y_{k+j},$$

pode-se reescrever (6.6.66) como

$$\phi_k \quad = \quad \sum_{j=0}^{n} \alpha_j e_{k+j} - h \sum_{j=0}^{n} \beta_j \frac{\partial}{\partial y} f\left(t_{k+j}, \xi_{k+j}\right) e_{k+j}. \tag{6.6.67}$$

Supondo-se que para todo $k > 0$ se tenha

$$\begin{aligned} \phi_k &\approx \bar{\phi} = \text{constante}, \\ \frac{\partial}{\partial y} f\left(t_{k+j}, \xi_{k+j}\right) &\approx \lambda = \text{constante}, \end{aligned}$$

de (6.6.67) obtém-se a equação de diferenças linear que descreve aproximadamente a dinâmica do comportamento do erro de discretização global ao longo do tempo,

$$\sum_{j=0}^{n} \left(\alpha_j - h\lambda\beta_j\right) e_{k+j} = \bar{\phi}. \tag{6.6.68}$$

A solução de (6.6.68) é determinada pelas raízes de seu polinômio característico,

$$\pi(r) = \sum_{j=0}^{n} \left(\alpha_j - h\lambda\beta_j\right) r^j, \tag{6.6.69}$$

e pode ser escrita como,

$$e_k = \sum_{j=1}^{n} a_j r_j^k + \frac{\bar{\phi}}{\sum_{j=0}^{n} \left(\alpha_j - h\lambda\beta_j\right)}, \tag{6.6.70}$$

onde r_j são as raízes de $\pi(r)$, supostas simples momentaneamente apenas para simplificar a exposição da teoria. Aqui, $\sum_{j=1}^{n} a_j r_j^k$ é a solução geral da equação de diferenças linear homogênea e

$$\frac{\bar{\phi}}{\sum_{j=0}^{n} \left(\alpha_j - h\lambda\beta_j\right)}$$

é uma solução particular de (6.6.68).

Sendo (6.6.62) consistente, tem-se

$$C_0 = \sum_{j=0}^{n} \alpha_j = 0,$$

e denotando-se por $\bar{h} = h\lambda$, (6.6.70) reescreve-se como

$$e_k = \sum_{j=1}^{n} d_j r_j^k + \frac{\bar{\phi}}{-\bar{h} \sum_{j=0}^{n} \beta_j}, \tag{6.6.71}$$

solução da equação de diferenças linear para o erro global de discretização (6.6.68).

É interessante notar que o polinômio característico da equação de diferenças que descreve aproximadamente a dinâmica do comportamento do erro de discretização global, (6.6.68). Este pode ser reescrito em termos dos polinômios característicos de (6.6.62), sendo

$$\begin{aligned} \pi(r) &= \sum_{j=0}^{n} \left(\alpha_j - \bar{h}\beta_j\right) r^j \\ &= \sum_{j=0}^{n} \alpha_j r^j - \bar{h} \sum_{j=0}^{n} \beta_j r^j \\ &= \rho(r) - \bar{h}\,\sigma(r). \end{aligned} \tag{6.6.72}$$

De (6.6.72), vê-se que as raízes de $\pi(r)$ tendem às raízes do primeiro polinômio característico de (6.6.62), $\rho(r) = \sum_{j=0}^{n} \alpha_j\, r^j$, quando o passo de integração h tende a zero. Como consequência, conclui-se que o comportamento do erro de discretização global, no limite para h tendendo a zero, relaciona-se à zero-estabilidade do método considerado.

No que segue, apresenta-se o conceito de *estabilidade absoluta* também conhecida como *estabilidade fraca*. Enquanto que o conceito de zero-estabilidade relaciona-se à convergência do método numérico, o conceito de estabilidade absoluta relaciona-se à escolha de um passo de integração $h > 0$ o qual, na prática, possa ser utilizado na obtenção de uma solução numérica que qualitativamente preserve as características da solução exata do problema, permitindo assim uma análise e interpretação corretas do fenômeno objeto de estudo.

6.7 Estabilidade absoluta

Vamos considerar o problema de valor inicial adotado na parte de estabilidade absoluta dos métodos de passo únicos,

$$\begin{cases} y' = \lambda y, & t > 0, \\ y(0) = 1, \end{cases} \tag{6.7.73}$$

com $\lambda \in \mathbb{C}$. Vamos supor que a parte real de λ seja negativa ($\mathrm{Re}(\lambda) < 0$) e que portanto $y(t) \to 0$ quando $t \to \infty$. Substituindo esse problema na forma usual dos métodos de passo múltiplos temos

$$\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j \lambda y_{k+j}, \tag{6.7.74}$$

e portanto

$$\sum_{j=0}^{n} \left(\alpha_j - \lambda h \beta_j\right) y_{k+j} = 0 \tag{6.7.75}$$

que define a equação homogênea associada à equação (6.6.68) no estudo do erro global. Portanto, o λ a ser considerado no problema modelo diz respeito a variação de f com relação a y (relativo a $\frac{\partial f}{\partial y}$), que é uma primeira aproximação (linear) de f. Essa aproximação (linearização) é razoável localmente, e esperamos que um método seja capaz de reproduzir o comportamento de convergência para zero do problema (6.7.73) com $t \to \infty$.

Definição 6.4 (Polinômio de estabilidade absoluta)**.** *Definimos como o polinômio de estabilidade absoluta associado a um problema de n-passos múltiplos linear o polinômio,*

$$\pi(r) = \rho(r) - \bar{h}\,\sigma(r), \tag{6.7.76}$$

onde ρ e σ são respectivamente o primeiro e o segundo polinômios característico do método e $\bar{h} \in \mathbb{C}$.

Definição 6.5 (Estabilidade Absoluta)**.** *Um método de passo múltiplo linear é absolutamente estável se e só se, para $\bar{h} = h\lambda$ dado ($\bar{h}$ fixado), todas as raízes, r_j, do polinômio de estabilidade absoluta satisfizerem*

$$|r_j| < 1, \;\; j = 1, 2, \cdots, n. \tag{6.7.77}$$

A região de estabilidade absoluta é definida pelo conjunto de valores de $\bar{h}$ no plano complexo tal que o método é absolutamente estável.

Definição 6.6 (A-estabilidade)**.** *Dizemos que um método é A-estável se, e somente se, a região de estabilidade absoluta contiver o todos os pontos complexos z tais que $Re(z) < 0$.*

Exemplo 6.6. *Considere o método do Trapézio Implícito,*

$$y_{k+1} - y_k = \frac{h}{2}\left(f_k + f_{k+1}\right). \tag{6.7.78}$$

Sabemos que o método é consistente de segunda ordem, pois $C_0 = C_1 = C_2 = 0$, e que é zero-estável pois é de passo único. O polinômio de estabilidade absoluta para este método é

$$\pi(r) = r - 1 - \frac{\bar{h}}{2}(r+1) = \left(1 - \frac{\bar{h}}{2}\right) r - \left(1 + \frac{\bar{h}}{2}\right), \tag{6.7.79}$$

cujas raiz é

$$r = \frac{1 + \frac{\bar{h}}{2}}{1 - \frac{\bar{h}}{2}}. \tag{6.7.80}$$

Vemos que o método é absolutamente estável para todo $\bar{h} = h\lambda$ tal que $hRe(\lambda) < 0$, ou seja, método sempre respeita o comportamento de convergência a zero do problema modelo (6.7.73)*. Este método é portanto A-estável.*

Na prática, podemos estar interessados apenas em valores reais de λ. Neste caso, o intervalo (a, b) da reta real é denominado *intervalo de estabilidade absoluta* se o método de passo múltiplo linear for absolutamente estável para todo $\bar{h} \in (a, b)$. Pode-se determinar o intervalo de estabilidade absoluta da seguinte forma:

1. Calculam-se as raízes de $\pi(r) = \rho(r) - \bar{h}\,\sigma(r)$, $r_j(\bar{h})$, para um conjunto de valores de $\bar{h}$ numa vizinhança da origem;

2. Representam-se graficamente as funções $\left|r_j(\bar{h})\right|$;

3. Observam-se os intervalos para os quais $\left|r_j(\bar{h})\right| < 1$.

Exemplo 6.7. *Considere o seguinte esquema de 2-passos lineares, conhecido como Leap-Frog,*

$$y_{k+2} - y_k = 2hf_{k+1}. \tag{6.7.81}$$

Este método é consistente com segunda ordem, pois $C_0 = C_1 = C_2 = 0$ (verifique!). Ele também é zero-estável, pois o seu primeiro polinômio característico tem raízes $r_1 = 1$, $r_2 = -1$, distintas, simples e com módulos iguais a 1. O seu polinômio de estabilidade absoluta é

$$\pi(r) = r^2 - 1 - \bar{h}2r, \tag{6.7.82}$$

com raízes $r_\pm = \bar{h} \pm \sqrt{\bar{h}^2 + 1}$. Vejamos o caso em que $\lambda \in \mathbb{R}$ e $\lambda < 0$. Neste caso, $|r_\pm| = 1$ somente se $\bar{h} = 0$. Para $\bar{h} \neq 0$, conforme podemos ver na Figura 6.7, temos que

$$\begin{aligned} 0 < r_+ < 1 \qquad & \forall \bar{h} < 0 \\ r_- < -1, \qquad & \forall \bar{h} < 0. \end{aligned}$$

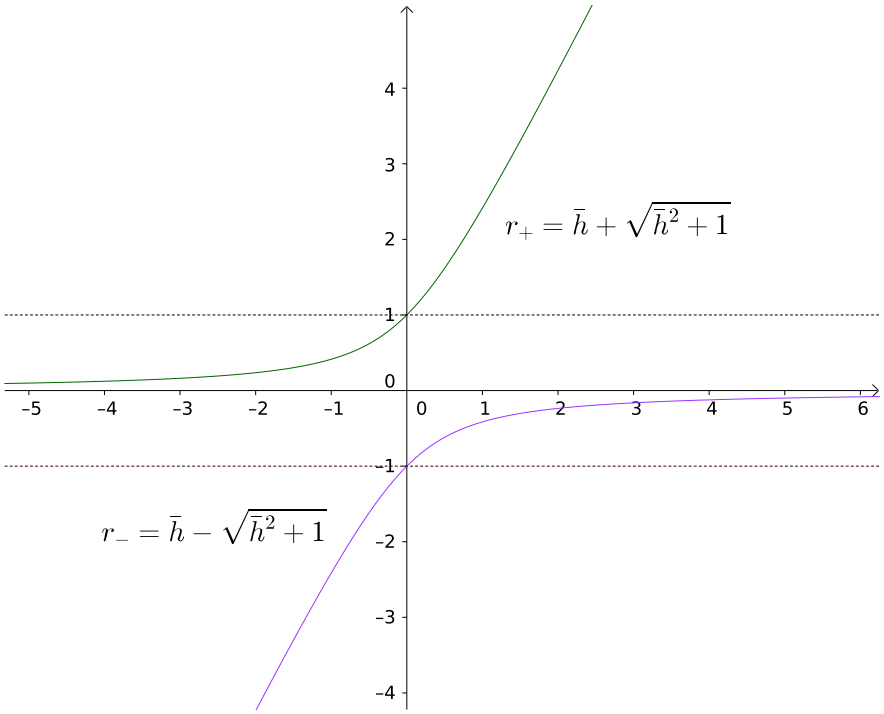


Figura 6.1: Exemplo de construção do intervalo de estabilidade absoluta para o método descrito em (6.7.81)

Então a raiz r_+ não afeta a estabilidade absoluta para $\lambda < 0$. Porém, a raiz r_- nunca terá módulo menor que 1 para $\lambda < 0$. Portanto, o método não é absolutamente estável não importa o quão pequeno for h. Portando, como está, o método não tem utilidade prática. Curiosamente, este método é bastante utilizado na prática, dado seu baixíssimo custo computacional (apenas uma avaliação de f), mas é necessário usar um filtro para obtermos soluções estáveis.

Métodos A-estáveis são na verdade raros. O seguinte resultado mostra o quão difícil pode ser de se obter um método A-estável.

Teorema 6.8 (Segunda Barreira de Dahlquist [7] [9])**.** *Relações entre A-estabilidade e métodos de passos múltiplos lineares:*

(i) Nenhum método linear de passos múltiplos explícito é A-estável.

(ii) Nenhum método A-estável linear de múltiplos passos tem ordem maior que 2.

(iii) O método de segunda ordem A-estável linear de passos múltiplos com menor constante de erro é o método do Trapézio Implícito.

Exemplo 6.8. *Considere o método de 2-passos implícito conhecido como BDF2 (Backwards Differentiation Formulae),*

$$3y_{k+2} - 4y_{k+1} + y_k = 2hf_{k+2}. \tag{6.7.83}$$

Este método é consistente com ordem 2, pois $C_0 = C_1 = C_2 = 0$*. As raízes do primeiro polinômio característico são* 1 *e* $1/3$*, portanto o método é zero-estável. O polinômio de estabilidade absoluta é*

$$\pi(r) = (3 - 2\bar{h})r^2 - 4r + 1,$$

com raízes

$$r_{\pm} = \frac{2 \pm \sqrt{1 + 2\bar{h}}}{3 - 2\bar{h}}.$$

Mostramos na Figura 6.8 que a região de estabilidade contempla todo plano complexo negativo (esquerdo), e que portanto o método é A-estável. O método BDF2 só não é estável para alguns valores de λ *positivos com* $\bar{h}$ *próximo de* 2*.*

6.7.1 Exercícios

Exercício 6.12. *Reveja as noções de estabilidade absoluta e de zero-estabilidade e suas relações com a escolha, na prática, do passo de integração para a análise do problema e com a convergência do método numérico.*

Exercício 6.13. *Considere o método de passo múltiplo linear*

$$y_{k+2} - (1+a)y_{k+1} + ay_k = \frac{h}{12}\left[(5+a)f_{k+2} + 8(1-a)f_{k+1} - (1+5a)f_k\right], \tag{6.7.84}$$

onde $-1 \leq a < 1$*.*

1. *Mostre que o intervalo de estabilidade absoluta do método (6.7.84) é*

$$\left(\frac{6(a+1)}{a-1},\ 0\right).$$

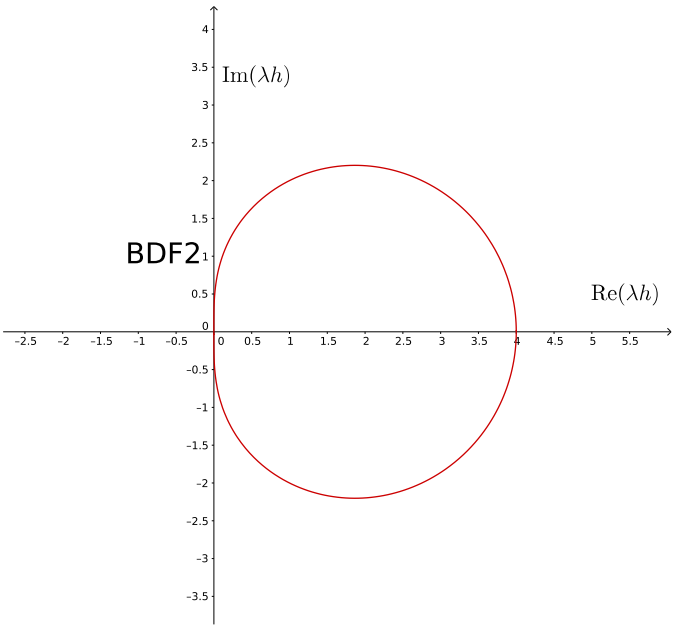


Figura 6.2: Exemplo de região de estabilidade absoluta para o método BDF2 descrito em (6.7.86). O interior da curva indica região com módulo maior que 1, e toda a região exterior à curva define a região onde o método é absolutamente estável.

2. *Para $a = -0,9$, use o método (6.7.84) para solucionar o problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt}y(t) = -20y(t), \quad t \in [0,1], \\ \\ y(0) = 1, \end{cases} \tag{6.7.85}$$

utilizando $h = 0,01$, $h = 0,02$ e $h = 0,04$.

Exercício 6.14. *Considere o método de 3-passos implícito conhecido como BDF3 (Backwards Differentiation Formulae),*

$$11y_{k+3} - 18y_{k+2} + 9y_{k+1} - 2y_k = 6hf_{k+3}. \tag{6.7.86}$$

Mostre que o método tem ordem 3 de consistência e que é zero-estável. Construa a região de estabilidade absoluta do método, verificando que ele não é A-estável (por bem pouco!).

6.8 Exercícios resolvidos

Exercício Resolvido 6.1. *Use o método de passo múltiplo linear*

$$y_{k+2} + 4y_{k+1} + y_k = \frac{h}{2}\left[8f_{k+1} + 4f_k\right] \tag{6.8.87}$$

para calcular a solução numérica do problema de valor inicial

$$\begin{cases} \dot{y} = 4t\sqrt{y}, \\ y(0) = 1, \end{cases} \qquad 0 \le t \le 2, \tag{6.8.88}$$

com $h = 0,1$, $h = 0,05$ e $h = 0,025$. Comente os resultados obtidos.
Solução:

Cálculo da solução exata do problema de valor inicial:

$$\begin{aligned}
\frac{dy}{dt} &= 4t\sqrt{y} \\
\frac{1}{\sqrt{y}}\frac{dy}{dt} &= 4t \\
\int \frac{1}{\sqrt{y}}\frac{dy}{dt}dt &= \int 4tdt \\
\frac{y^{\frac{1}{2}}}{\frac{1}{2}} &= \frac{4t^2}{2} + C \\
y^{\frac{1}{2}} &= t^2 + \frac{C}{2} \\
y &= \left(t^2 + \frac{C}{2}\right)^2 \\
y(0) = 1 &\Rightarrow C = 2 \\
y(t) &= \left(t^2 + 1\right)^2 . \qquad (6.8.89)
\end{aligned}$$

O método (6.8.87) foi implementado utilizando a solução exata (6.8.89) para obter y_1, ou seja,

$$y_1 = y\left(t_0 + h\right) = y(h) = \left(h^2 + 1\right)^2 .$$

Este dado é necessário para iniciar o cálculo das aproximações em um método de dois passos (alternativamente, pode-se empregar um método de passo único para gerar y_1).

O método (6.8.87) não é consistente nem zero-estável (verifique!). Logo, divergente. A redução no passo de integração h aumenta o erro de discretização global. O método (6.8.87) gera aproximações negativas a partir de certos valores de tempo para as quais $f(t, y)$ é um número complexo.

Considerando-se o instante $t = 0.3$, pode-se perceber que a redução do passo de integração h provoca aumento no erro. As tabelas 6.2, 6.3 e 6.4 apresentam os valores gerados até o ponto em que a solução numérica torna-se negativa.

$\mathbf{h = 0,1}$			
t_k	$y(t_k)$	y_k	$\lvert y(t_k) - y_k \rvert$
0,000	1,000000	1,000000	0,000000
0,100	1,020100	1,020100	0,000000
0,200	1,081600	1,081200	0,000400
0,300	**1,188100**	**1,189238**	**0,001138**
0,400	1,345600	1,338866	0,006734
0,500	1,562500	1,592994	0,030494
0,600	1,849600	1,702337	0,147263
0,700	2,220100	2,913023	0,692923
0,800	2,689600	-0,602567	3,292167

Tabela 6.2: Comparação das soluções exata e numérica do problema de valor inicial (6.8.88), onde a solução numérica foi obtida através do método (6.8.87) com $h = 0,1$.

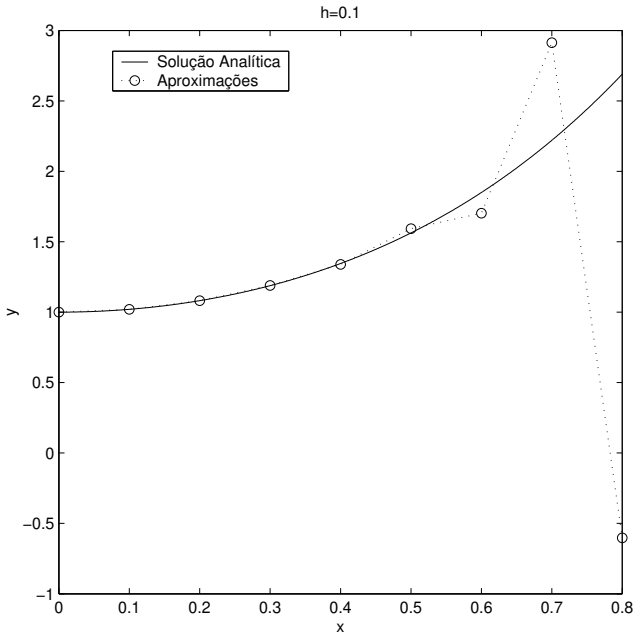


Figura 6.3: Comparação das soluções exata e numérica do problema de valor inicial (6.8.88), onde a solução numérica foi obtida através do método (6.8.87) com $h = 0,1$.

$\mathbf{h = 0,05}$			
t_k	$y(t_k)$	y_k	$\lvert y(t_k) - y_k \rvert$
0,000	1,000000	1,000000	0,000000
0,050	1,005006	1,005006	0,000000
0,100	1,020100	1,020075	0,000025
0,150	1,045506	1,045580	0,000074
0,200	1,081600	1,081158	0,000442
0,250	1,128906	1,130988	0,002082
0,300	**1,188100**	**1,177715**	**0,010385**
0,350	1,260006	1,310883	0,050877
0,400	1,345600	1,095852	0,249748
0,450	1,446006	2,666284	1,220278
0,500	1,562500	-4,430548	5,993048

Tabela 6.3: Comparação das soluções exata e numérica do problema de valor inicial (6.8.88), onde a solução numérica foi obtida através do método (6.8.87) com $h = 0,05$.

Exercício Resolvido 6.2. *Verifique o efeito da falta de estabilidade (instabilidade) do método de passo múltiplo linear*

$$y_{k+2} - y_{k+1} = \frac{h}{3}\left[3f_{k+1} - 2f_k\right] \tag{6.8.90}$$

empregando-o para solucionar numericamente o Problema de Cauchy

$$\begin{cases} \dot{y} = 4t\sqrt{y}, \\ y(0) = 1, \end{cases} \qquad 0 \leq t \leq 2, \tag{6.8.91}$$

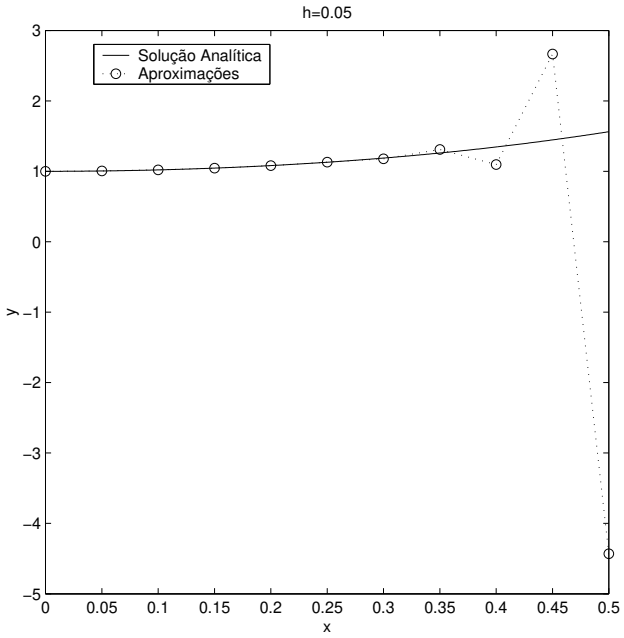


Figura 6.4: Comparação das soluções exata e numérica do problema de valor inicial (6.8.88), onde a solução numérica foi obtida através do método (6.8.87) com $h = 0,05$.

$\mathbf{h = 0,025}$			
t_k	$y(t_k)$	y_k	$\lvert y(t_k) - y_k \rvert$
0,000	1,000000	1,000000	0,000000
0,025	1,001250	1,001250	0,000000
0,050	1,005006	1,005005	0,000002
0,075	1,011282	1,011286	0,000005
0,100	1,020100	1,020072	0,000028
0,125	1,031494	1,031627	0,000133
0,150	1,045506	1,044835	0,000672
0,175	1,062188	1,065521	0,003334
0,200	1,081600	1,065009	0,016591
0,225	1,103813	1,186258	0,082445
0,250	1,128906	0,719318	0,409588
0,275	1,156969	3,187841	2,030872
0,300	**1,188100**	**-8,915969**	**10,104069**

Tabela 6.4: Comparação das soluções exata e numérica do problema de valor inicial (6.8.88), onde a solução numérica foi obtida através do método (6.8.87) com $h = 0,025$.

cuja solução exata é $y(t) = \left(t^2 + 1\right)^2$.

Solução:

O método (6.8.90) é zero-estável porém inconsistente (verifique!). Logo, divergente. As Figuras 6.6 e 6.7a-b foram geradas com passos de integração $h = 0,1$,

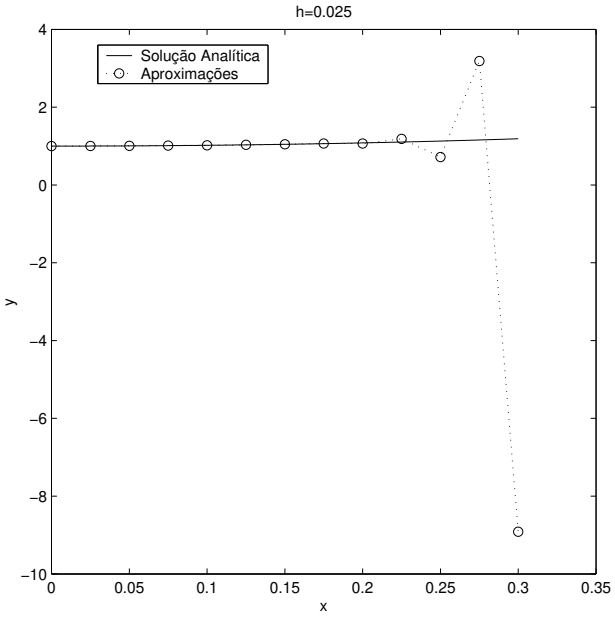


Figura 6.5: Comparação das soluções exata e numérica do problema de valor inicial (6.8.88), onde a solução numérica foi obtida através do método (6.8.87) com $h = 0,025$.

$h = 0,05$ *e* $h = 0,025$. *Note-se que a diferenç entre as aproximações e a solução exata aumenta à medida que o passo de integração é reduzido, fato que caracteriza a divergência do método numérico (6.8.90).*

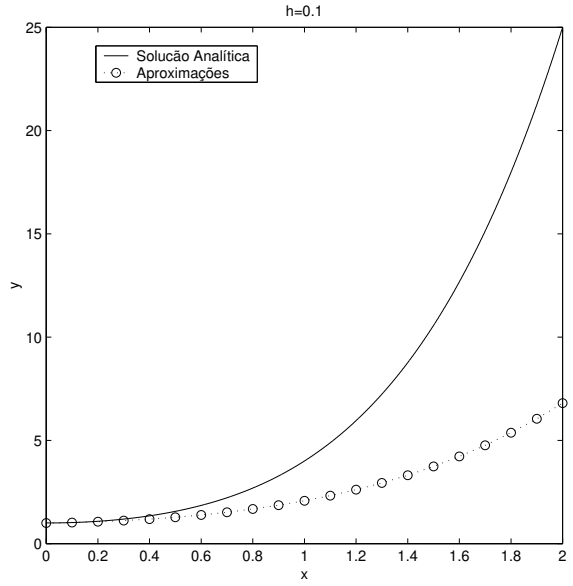
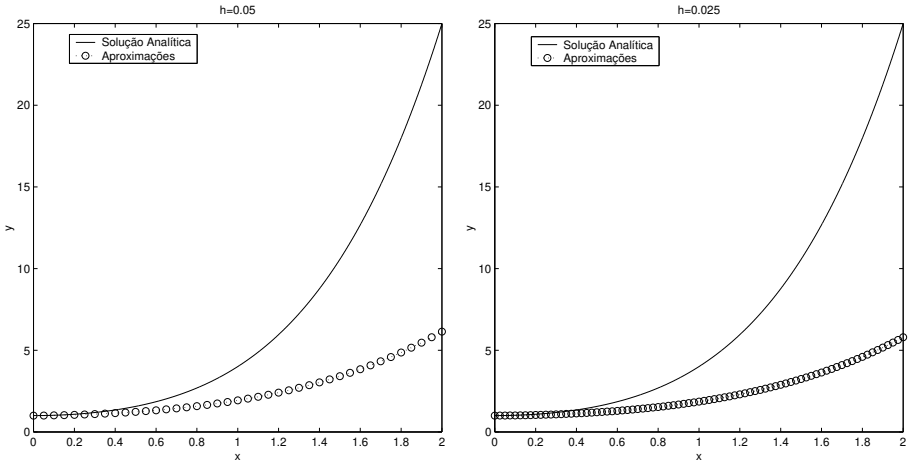


Figura 6.6: Comparação das soluções exata e numérica do problema de valor inicial (6.8.91), onde a solução numérica foi obtida através do método (6.8.90) com $h = 0,1$.



(a) (b)

Figura 6.7: Comparação das soluções exata e numérica do problema de valor inicial (6.8.91), onde a solução numérica foi obtida através do método (6.8.90) com (a) $h = 0,05$ e (b) $h = 0,025$.

Exercício Resolvido 6.3. *Determine a ordem de consistência e o erro de discretização local principal do Método de Quade*

$$y_{k+4} - \frac{8}{19}(y_{k+3} - y_{k+1}) - y_k = \frac{6h}{19}(f_{k+4} + 4f_{k+3} + 4f_{k+1} + f_k). \qquad (6.8.92)$$

Solução:

$$\alpha_4 = 1,\ \alpha_3 = -\frac{8}{19},\ \alpha_2 = 0,\ \alpha_1 = \frac{8}{19},\ \alpha_0 = -1$$

$$\beta_4 = \frac{6}{19},\ \beta_3 = \frac{24}{19},\ \beta_2 = 0,\ \beta_1 = \frac{24}{19},\ \beta_0 = \frac{6}{19}$$

$$
\begin{aligned}
C_0 &= \sum_{j=0}^{4} \alpha_j = -1 + \frac{8}{19} - \frac{8}{19} + 1 = 0 \\
C_1 &= \sum_{j=0}^{4} j\alpha_j - \sum_{j=0}^{4} \beta_j \\
&= \left(\frac{8}{19} - \frac{24}{19} + 4\right) - \left(\frac{12}{19} + \frac{48}{19}\right) \\
&= \frac{60}{19} - \frac{60}{19} = 0 \\
C_2 &= \sum_{j=0}^{4} \frac{j^2}{2!}\alpha_j - \sum_{j=0}^{4} j\beta_j \\
&= \frac{1}{2}\left(\frac{8}{19} - \frac{72}{19} + 16\right) - \left(\frac{24}{19} + \frac{72}{19} + \frac{24}{19}\right) \\
&= \frac{120}{19} - \frac{120}{19} = 0 \\
C_3 &= \sum_{j=0}^{4} \frac{j^3}{3!}\alpha_j - \sum_{j=0}^{4} \frac{j^2}{2!}\beta_j \\
&= \frac{1}{6}\left(\frac{8}{19} - \frac{216}{19} + 64\right) - \frac{1}{2}\left(\frac{24}{19} + \frac{216}{19} + \frac{96}{19}\right) \\
&= \frac{168}{19} - \frac{168}{19} = 0 \\
C_4 &= \sum_{j=0}^{4} \frac{j^4}{4!}\alpha_j - \sum_{j=0}^{4} \frac{j^3}{3!}\beta_j \\
&= \frac{1}{24}\left(\frac{8}{19} - \frac{648}{19} + 256\right) - \frac{1}{6}\left(\frac{24}{19} + \frac{648}{19} + \frac{384}{19}\right) \\
&= \frac{176}{19} - \frac{176}{19} = 0
\end{aligned}
$$

$$\begin{aligned}
C_5 &= \sum_{j=0}^{4}\frac{j^5}{5!}\alpha_j - \sum_{j=0}^{4}\frac{j^4}{4!}\beta_j = \\
&= \frac{1}{120}\left(\frac{8}{19} - \frac{1944}{19} + 1024\right) - \frac{1}{24}\left(\frac{24}{19} + \frac{1944}{19} + \frac{1536}{19}\right) \\
&= \frac{146}{19} - \frac{146}{19} = 0 \\
C_6 &= \sum_{j=0}^{4}\frac{j^6}{6!}\alpha_j - \sum_{j=0}^{4}\frac{j^5}{5!}\beta_j = \\
&= \frac{1}{720}\left(\frac{8}{19} - \frac{5832}{19} + 4096\right) - \frac{1}{120}\left(\frac{24}{19} + \frac{5832}{19} + \frac{6144}{19}\right) \\
&= \frac{100}{19} - \frac{100}{19} = 0 \\
C_7 &= \sum_{j=0}^{4}\frac{j^7}{7!}\alpha_j - \sum_{j=0}^{4}\frac{j^6}{6!}\beta_j = \\
&= \frac{1}{5040}\left(\frac{8}{19} - \frac{17496}{19} + 16384\right) + \\
&- \frac{1}{720}\left(\frac{24}{19} + \frac{17496}{19} + \frac{24576}{19}\right) \\
&= \frac{6121}{1995} - \frac{877}{285} = -\frac{6}{665}
\end{aligned}$$

O Método de Quade (6.8.92) é consistente $(C_0 = C_1 = 0)$ *de ordem* 6. *O erro de discretização local principal do método é*

$$\alpha = C_7 h^6 y^{(7)}(\xi)$$

com constante $C_7 = -\dfrac{6}{665}$.

Exercício Resolvido 6.4. *Verifique que (6.8.92) é convergente.*
Solução:

Primeiro polinômio característico:

$$\begin{aligned}
r^4 - \frac{8}{19}r^3 + \frac{8}{19}r - 1 &= r^4 - 1 - \frac{8}{19}r\left(r^2 - 1\right) \\
&= \left(r^2 - 1\right)\left(r^2 + 1\right) - \frac{8}{19}r\left(r^2 - 1\right) \\
&= \left(r^2 - 1\right)\left(r^2 - \frac{8}{19}r + 1\right) \\
&= (r+1)(r-1)\left(r^2 - \frac{8}{19}r + 1\right).
\end{aligned}$$

Raízes do polinômio (6.8.93):

$$\begin{aligned}
r + 1 = 0 &\Rightarrow r_1 = -1; \\
r - 1 = 0 &\Rightarrow r_2 = 1; \\
r^2 - \frac{8}{19}r + 1 = 0 &\Rightarrow r_3 = \frac{4}{19} + \frac{i}{19}\sqrt{345}; \\
&\qquad r_4 = \frac{4}{19} - \frac{i}{19}\sqrt{345}.
\end{aligned}$$

Módulo das raízes r_3 e r_4:

$$|r_3| = |r_4| = \left[\left(\frac{4}{19}\right)^2 + \left(\frac{\sqrt{345}}{19}\right)^2\right]^{\frac{1}{2}} = \left(\frac{16}{361} + \frac{345}{361}\right)^{\frac{1}{2}} = 1.$$

Como as quatro raízes do polinômio (6.8.93) são distintas e têm módulo 1*, o método é zero-estável.*

O Método de Quade (6.8.92) é convergente uma vez que é consistente e zero-estável.

Exercício Resolvido 6.5. *Considere o método de passo múltiplo linear*

$$y_{k+3} + \alpha\left(y_{k+2} - y_{k+1}\right) - y_k = \frac{h}{2}\left(3 + \alpha\right)\left(f_{k+2} + f_{k+1}\right). \tag{6.8.93}$$

1. *Determine para quais valores de α o método (6.8.93) é zero-estável.*

   Solução:

$$\alpha_3 = 1,\ \alpha_2 = \alpha,\ \alpha_1 = -\alpha,\ \alpha_0 = -1$$

Primeiro polinômio característico:

$$\begin{aligned} r^3 + \alpha r^2 - \alpha r - 1 &= \left(r^3 - 1\right) + \alpha\left(r^2 - r\right) \\ &= \left(r - 1\right)\left(r^2 + r + 1\right) + \alpha r\left(r - 1\right) \\ &= \left(r - 1\right)\left[r^2 + \left(1 + \alpha\right) r + 1\right]. \end{aligned} \tag{6.8.94}$$

Uma das raízes do polinômio (6.8.94) é $r_1 = 1$. As outras duas raízes são dadas por

$$r_{2,3} = \frac{-\left(1 + \alpha\right) \pm \sqrt{\left(1 + \alpha\right)^2 - 4}}{2}. \tag{6.8.95}$$

   (a) *Considerando $\alpha = 1$ ou $\alpha = -3$ em (6.8.95):*

$$\begin{aligned} \alpha = 1 &\Rightarrow \left(1 + \alpha\right)^2 - 4 = 0 \Rightarrow r_2 = r_3 = -1; \\ \alpha = -3 &\Rightarrow \left(1 + \alpha\right)^2 - 4 = 0 \Rightarrow r_2 = r_3 = 1. \end{aligned}$$

   Como em ambos os casos o polinômio tem raiz de multiplicidade 2 *e módulo* 1*, o método não é zero-estável.*

   (b) *Considerando $\alpha > 1$ ou $\alpha < -3$ em (6.8.95):*

$$\begin{aligned} \alpha > 1 &\Rightarrow 1 + \alpha > 2 \Rightarrow -\left(1 + \alpha\right) < -2 \\ &\Rightarrow -\frac{1 + \alpha}{2} < -1 \\ &\Rightarrow -\frac{1 + \alpha}{2} - \frac{\sqrt{\left(1 + \alpha\right)^2 - 4}}{2} < -1 \\ &\Rightarrow r_3 \text{ é menor que } -1; \end{aligned}$$

$$\begin{aligned}
\alpha < -3 \quad &\Rightarrow \quad 1+\alpha < -2 \Rightarrow -(1+\alpha) > 2 \\
&\Rightarrow \quad -\frac{1+\alpha}{2} > 1 \\
&\Rightarrow \quad -\frac{1+\alpha}{2} + \frac{\sqrt{(1+\alpha)^2 - 4}}{2} > 1 \\
&\Rightarrow \quad r_2 \text{ é maior que } 1.
\end{aligned}$$

Como em ambos os casos o polinômio tem raiz de módulo maior que 1*, o método não é zero-estável.*

(c) Considerando $-3 < \alpha < 1$ *em (6.8.95):*

$$\begin{aligned}
-3 < \alpha < 1 \quad &\Rightarrow \quad -2 < 1+\alpha < 2 \Rightarrow (1+\alpha)^2 < 4 \\
&\Rightarrow \quad r_{2,3} = -\frac{1+\alpha}{2} \pm \frac{i}{2}\sqrt{4-(1+\alpha)^2}; \\
|r_2| = |r_3| \quad &= \quad \left[\frac{(1+\alpha)^2}{4} + \frac{4-(1+\alpha)^2}{4}\right]^{\frac{1}{2}} = 1.
\end{aligned}$$

Para $\alpha \in (-3, 1)$ *o método (6.8.93) é zero-estável, uma vez que as três raízes do primeiro polinômio característico (6.8.94) são distintas de módulo* 1*.*

2. Mostre que existe um valor de α *para o qual o método (6.8.93) tem ordem* 4*, mas para que ele seja zero-estável sua ordem não pode exceder* 2*.*

Solução:

$$\alpha_3 = 1,\ \alpha_2 = \alpha,\ \alpha_1 = -\alpha,\ \alpha_0 = -1$$

$$\beta_3 = 0,\ \beta_2 = \frac{3+\alpha}{2},\ \beta_1 = \frac{3+\alpha}{2},\ \beta_0 = 0$$

$$\begin{aligned}
C_0 \quad &= \quad \sum_{j=0}^{3} \alpha_j = -1 - \alpha + \alpha + 1 = 0 \\
C_1 \quad &= \quad \sum_{j=0}^{3} j\alpha_j - \sum_{j=0}^{3} \beta_j \\
&= \quad -\alpha + 2\alpha + 3 - \left(\frac{3+\alpha}{2} + \frac{3+\alpha}{2}\right) \\
&= \quad \alpha + 3 - 3 - \alpha = 0 \\
C_2 \quad &= \quad \sum_{j=0}^{3} \frac{j^2}{2!}\alpha_j - \sum_{j=0}^{3} j\beta_j \\
&= \quad \frac{1}{2}(-\alpha + 4\alpha + 9) - \left(\frac{3+\alpha}{2} + 3 + \alpha\right) \\
&= \quad \frac{1}{2}(9+3\alpha) - \frac{1}{2}(9+3\alpha) = 0
\end{aligned}$$

$$\begin{aligned} C_3 &= \sum_{j=0}^{3} \frac{j^3}{3!}\alpha_j - \sum_{j=0}^{3} \frac{j^2}{2!}\beta_j \\ &= \frac{1}{6}\left(-\alpha + 8\alpha + 27\right) - \frac{1}{2}\left(\frac{3+\alpha}{2} + 4\frac{3+\alpha}{2}\right) \\ &= \frac{1}{6}\left(27 + 7\alpha\right) - \frac{1}{4}\left(15 + 5\alpha\right) \\ &= \frac{18 - 2\alpha}{24} = \frac{9-\alpha}{12} \end{aligned}$$

$$C_3 \neq 0 \quad \Rightarrow \quad \alpha \neq 9$$

Para que o método (6.8.93) seja zero-estável é necessário que a ordem do mesmo seja 2*, uma vez que* $\alpha = 9 \notin (-3, 1)$*.*

$$\begin{aligned} C_4 &= \sum_{j=0}^{3} \frac{j^4}{4!}\alpha_j - \sum_{j=0}^{3} \frac{j^3}{3!}\beta_j \\ &= \frac{1}{24}\left(-\alpha + 16\alpha + 81\right) - \frac{1}{6}\left(\frac{3+\alpha}{2} + 8\frac{3+\alpha}{2}\right) \\ &= \frac{1}{24}\left(81 + 15\alpha\right) - \frac{1}{12}\left(27 + 9\alpha\right) \\ &= \frac{27 - 3\alpha}{24} = \frac{9-\alpha}{8} \end{aligned}$$

$$\alpha = 9 \quad \Rightarrow \quad C_4 = 0$$

$$\begin{aligned} C_5 &= \sum_{j=0}^{3} \frac{j^5}{5!}\alpha_j - \sum_{j=0}^{3} \frac{j^4}{4!}\beta_j \\ &= \frac{1}{120}\left(-\alpha + 32\alpha + 243\right) - \frac{1}{24}\left(\frac{3+\alpha}{2} + 16\frac{3+\alpha}{2}\right) \\ &= \frac{1}{120}\left(243 + 31\alpha\right) - \frac{1}{48}\left(51 + 17\alpha\right) \\ &= \frac{231 - 23\alpha}{240} \end{aligned}$$

$$\alpha = 9 \quad \Rightarrow \quad C_5 = \frac{1}{10}$$

Para $\alpha = 9$*, o método (6.8.93) tem ordem* 4*, porém não é zero-estável.*

Capítulo 7

Métodos preditores-corretores

Ao se propor um método de passo múltiplo linear

$$y_{k+n} + \sum_{j=0}^{n-1} \alpha_j y_{k+j} \quad = \quad h \sum_{j=0}^{n} \beta_j f_{k+j}, \tag{7.0.1}$$

onde $f_{k+j} = f(t_{k+j}, y_{k+j})$, têm-se como opções escolher $\beta_n = 0$, para um método explícito, ou $\beta_n \neq 0$ para um método implícito. No primeiro caso, têm-se $2n$ coeficientes a se determinar ("graus de liberdade"), α_j e β_j, $j = 1, 2, \ldots, n-1$, e no segundo caso tem-se, além destes, β_n a se escolher perfazendo um total de $2n + 1$ coeficientes a se determinar. Imposições sobre os coeficientes (**??**), C_j, $j = 0, 1, \ldots,$ p, constituem *vínculos* com os quais controla-se a ordem objetivada.

Com tal diferença no número de graus de liberdade, não é de se surpreender que métodos explícitos e implícitos com mesmo número de passos (e, portanto, com mesma demanda por armazenamento de informação em memória) possam ter propriedades como ordem e intervalo de estabilidade absoluta diferentes. As tabelas 7.1 e 7.2 mostram ordens, intervalos de estabilidade absoluta e o coeficiente principal de erro para os métodos explícitos de Adams-Bashforth e implícitos de Adams-Moulton, respectivamente.

número de passos	1	2	3	4
ordem (p)	1	2	3	4
coeficiente C_{p+1}	1/2	5/12	3/8	251/720
estabilidade absoluta	$(-2, 0)$	$(-1, 0)$	$(-6/11, 0)$	$(-3/10, 0)$

Tabela 7.1: Propriedades de alguns métodos de Adams-Bashforth (explícitos).

número de passos	1	2	3	4
ordem p	2	3	4	5
coeficiente C_{p+1}	$-1/12$	$-1/24$	$-19/720$	$-3/160$
estabilidade absoluta	$(-\infty, 0)$	$(-6, 0)$	$(-3, 0)$	$(-90/49, 0)$

Tabela 7.2: Propriedades de alguns métodos de Adams-Moulton (implícitos).

Comparando-se as tabelas 7.1 e 7.2, percebem-se que são muitas as vantagens dos métodos implícitos sobre os métodos explícitos (de mesmo número de passos). Dentre elas, podem ser destacadas menores valores absolutos para os coeficientes principais de erro C_{p+1} e maiores amplitudes dos intervalos de estabilidade absoluta. Claro, como principal desvantagem, pode-se mencionar o fato dos métodos implícitos demandarem a resolução de uma equação algébrica em todo o passo de integração no tempo para se determinar y_{k+n} (e.g. via método da Dicotomia, do Ponto-fixo, de Newton, etc.).

Considere o método de passo múltiplo linear implícito

$$y_{k+n} + \sum_{j=0}^{n-1} \alpha_j y_{k+j} = h\beta_n f(t_{k+n}, y_{k+n}) + h \sum_{j=0}^{n-1} \beta_j f_{k+j}. \tag{7.0.2}$$

Ao se calcular y_{k+n} em (7.0.2) usando o Método do Ponto Fixo, tem-se

$$y_{k+n}^{[s+1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j} = h\beta_n f(t_{k+n}, y_{k+n}^{[s]}) + h \sum_{j=0}^{n-1} \beta_j f_{k+j} \quad s = 0, 1, 2, \ldots, \tag{7.0.3}$$

onde $y_{k+n}^{[0]}$ é uma aproximação inicial de y_{k+n} e s é o índice de iteração. Uma condição suficiente para a convergência das iterações em (7.0.3) para y_{k+n} é que se tenha o passo de integração h

$$h < \frac{1}{L|\beta_n|}, \tag{7.0.4}$$

onde L é a constante de Lipschitz da função f do Problema de Cauchy em estudo.

A questão que deve ser considerada neste ponto é: *"É possível beneficiar-se das boas propriedades de ambos os tipos de métodos simultaneamente?"*. Como é visto adiante, a resposta é afirmativa.

7.0.1 Exercícios

Exercício 7.1. *Mostre que se* $h < \frac{1}{L|\beta_n|}$ *então o processo iterativo (7.0.3) converge.*

Há duas opções para se solucionar as iterações advindas da aplicação do Método do Ponto Fixo ao esquema implícito (7.0.2):

1ª iterar até que $|y_{k+n}^{[s+1]} - y_{k+n}^{[s]}| < \epsilon$, onde $\epsilon > 0$ é uma precisão pré-fixada

2ª iterar m vezes, com m fixo (isto é, $s = 0, 1, 2, \ldots, m-1$).

Para diminuir o número total de iterações na primeira opção, é necessário utilizar boas aproximações iniciais $y_{k+n}^{[0]}$. Observe que, dependendo de $y_{k+n}^{[0]}$ e ϵ, muitas avaliações (cálculos) da função f podem ser necessárias.

Uma forma conveniente de se obter $y_{k+n}^{[0]}$ é usar um método de passo múltiplo linear explícito de ordem p^*, denominado *preditor*, e depois calcular $y_{k+n}^{[1]}$ através de um método de passo múltiplo linear implícito de ordem p, denominado *corretor*. Pode-se denotar a estratégia da seguinte maneira [16]:

P : uma aplicação do método explícito preditor, o qual gera uma boa aproximação para $y_{k+n}^{[0]}$;

E : um cálculo de f (*evaluation*);

C : uma aplicação do método implícito corretor, o qual gera a solução numérica para y_{k+n} via Método do Ponto Fixo.

Em geral, a forma de aplicação do Método Preditor-Corretor é descrita como

$$P(EC)^m \tag{7.0.5}$$

ou

$$P(EC)^m E, \tag{7.0.6}$$

sendo

$$\begin{aligned} P &= y_{k+n}^{[0]}, \\ E &= f\left(t_{k+n}, y_{k+n}^{[0]}\right), \\ C &= y_{k+n}^{[1]}, \\ E &= f\left(t_{k+n}, y_{k+n}^{[1]}\right), \end{aligned}$$

e m o número fixo de vezes que y_{k+n} $\left(y_{k+n}^{[m]}\right)$ será calculado. O modo (7.0.6) difere do modo (7.0.5) por uma avaliação a mais da função f. Obviamente, optar pelo modo (7.0.5) ou (7.0.6) afeta o próximo passo de integração.

O corretor, para m geral, é dado por

$$y_{k+n}^{[s+1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[m]} = h\beta_n f(t_{k+n}, y_{k+n}^{[s]}) + h\sum_{j=0}^{n-1} \beta_j f(t_{k+j}, y_{k+j}^{[m-t]}) \quad t = 0, 1. \tag{7.0.7}$$

Para $s = 0$ em (7.0.7), tem-se que

$$y_{k+n}^{[1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[m]} = h\beta_n f(t_{k+n}, y_{k+n}^{[0]}) + h\sum_{j=0}^{n-1} \beta_j f(t_{k+j}, y_{k+j}^{[m-t]}),$$

onde se $t = 1$ tem-se $P(EC)^m$ e se $t = 0$ tem-se $P(EC)^m E$.

Para $s = 1$ em (7.0.7), tem-se que

$$y_{k+n}^{[2]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[m]} = h\beta_n f(t_{k+n}, y_{k+n}^{[1]}) + h\sum_{j=0}^{n-1} \beta_j f(t_{k+j}, y_{k+j}^{[m-t]}),$$

onde se $t = 1$ tem-se $P(EC)^m$ e se $t = 0$ tem-se $P(EC)^m E$.

Definição 7.1. *Sejam os métodos de passo múltiplo lineares usados como preditor e corretor, respectivamente, definidos pelos polinômios característicos*

$$\rho^*(r) = \sum_{j=0}^{n} \alpha_j^* r^j, \;\; \alpha_n^* = 1, \;\; \sigma^*(r) = \sum_{j=0}^{n} \beta_j^* r^j$$

e

$$\rho(r) = \sum_{j=0}^{n} \alpha_j r^j, \;\; \alpha_n = 1, \;\; \sigma(r) = \sum_{j=0}^{n} \beta_j r^j.$$

Os modos $P(EC)^m E$ e $P(EC)^m$ são definidos como:

$P(EC)^mE$**:**

$$\begin{aligned}
y_{k+n}^{[0]} \quad & + \sum_{j=0}^{n-1} \alpha_j^* y_{k+j}^{[m]} = h \sum_{j=0}^{n-1} \beta_j^* f_{k+j}^{[m]} \quad \textit{(prediz)} \\
Para \quad & s = 0, 1, \ldots, m-1 \\
& f_{k+n}^{[s]} = f(t_{k+n}, y_{k+n}^{[s]}) \quad \textit{(avalia)} \\
& y_{k+n}^{[s+1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[m]} = h\beta_n f_{k+n}^{[s]} + h \sum_{j=0}^{n-1} \beta_j f_{k+j}^{[m]} \quad \textit{(corrige)} \\
f_{k+n}^{[m]} \quad & = f(t_{k+n}, y_{k+n}^{[m]}) \quad \textit{(avalia)};
\end{aligned}$$

$P(EC)^m$**:**

$$\begin{aligned}
y_{k+n}^{[0]} \quad & + \sum_{j=0}^{n-1} \alpha_j^* y_{k+j}^{[m]} = h \sum_{j=0}^{n-1} \beta_j^* f_{k+j}^{[m-1]} \quad \textit{(prediz)} \\
Para \quad & s = 0, 1, \ldots, m-1 \\
& f_{k+n}^{[s]} = f(t_{k+n}, y_{k+n}^{[s]}) \quad \textit{(avalia)} \\
& y_{k+n}^{[s+1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[m]} = h\beta_n f_{k+n}^{[s]} + h \sum_{j=0}^{n-1} \beta_j f_{k+j}^{[m-1]} \quad \textit{(corrige)}.
\end{aligned}$$

É interessante observar que à medida que o número total de iterações m cresce tendendo ao infinito as características do método preditor-corretor aproximam-se cada vez mais àquelas do método usado como corretor (e.g. ordem de consistência, intervalo de estabilidade absoluta e coeficiente principal de erro). Iterar até a convergência das iterações não é, a princípio, a primeira estratégia de implementação que se usa pois nunca se sabe, *a priori*, o custo computacional envolvido por passo de integração no tempo. Prefere-se, ao invés, fixar-se um valor para m (tipicamente $m = 2$) e se resolver o problema a um custo computacional conhecido.

7.1 Erro de discretização local

Sejam P o método preditor e C o método corretor, definidos por

$$P: \quad y_{k+n}^{[s]} + \sum_{j=0}^{n-1} \alpha_j^* y_{k+j}^{[m]} = h \sum_{j=0}^{n-1} \beta_j^* f_{k+j}^{[m-t]}, \tag{7.1.8}$$

$$C: \quad y_{k+n}^{[s+1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[m]} = h\beta_n f(t_{k+n}, y_{k+n}^{[s]}) + h \sum_{j=0}^{n-1} \beta_j f_{k+j}^{[m-t]}, \tag{7.1.9}$$

onde $s = 0, 1, \cdots, m-1$, $t = 0 \Rightarrow P(EC)^mE$ e $t = 1 \Rightarrow P(EC)^m$.

O erro local de discretização principal dos métodos preditor e corretor é dado, respectivamente, por

$$P: \quad y(t_{k+n}) - y_{k+n}^{[s]} = C_{p^*+1}^* h^{p^*+1} y^{(p^*+1)}(t_k) + O(h^{p^*+2}) = h\alpha^*, \tag{7.1.10}$$

$$C: \quad y(t_{k+n}) - y_{k+n}^{[s+1]} = C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}) = h\alpha. \tag{7.1.11}$$

O produto do erro local de discretização pelo passo de integração h (suficientemente pequeno) para um método de passo múltiplo linear é dado por

$$\begin{aligned} h\alpha &= y(t_{k+n}) + \sum_{j=0}^{n-1} \alpha_j y(t_{k+j}) - h\beta_n f(t_{k+n}, y(t_{k+n})) + \\ &- h\sum_{j=0}^{n-1} \beta_j f(t_{k+j}, y(t_{k+j})). \end{aligned} \tag{7.1.12}$$

Como até $j = n$ não se cometeu qualquer erro (porquê?), a subtração (7.1.12) - (7.1.9) resulta em

$$h\alpha = y(t_{k+n}) - y_{k+n}^{[s+1]} - h\beta_n \left[f\left(t_{k+n}, y\left(t_{k+n}\right)\right) - f\left(t_{k+n}, y_{k+n}^{[s]}\right)\right]. \tag{7.1.13}$$

Aplicando o Teorema do Valor Médio a (7.1.13), chega-se a

$$\begin{aligned} h\alpha &= y(t_{k+n}) - y_{k+n}^{[s+1]} - h\beta_n \frac{\partial f}{\partial y}\left(t_{k+n}, \eta_{k+n,s}\right)\left(y(t_{k+n}) - y_{k+n}^{[s]}\right), \\ y(t_{k+n}) - y_{k+n}^{[s+1]} &= h\beta_n \frac{\partial f}{\partial y}\left(t_{k+n}, \eta_{k+n,s}\right)\left(y(t_{k+n}) - y_{k+n}^{[s]}\right) + h\alpha, \end{aligned} \tag{7.1.14}$$

onde $\eta_{k+n,s}$ pertence ao intervalo de extremos $y_{k+n}^{[s]}$ e $y(t_{k+n})$.

(a) Caso $p^* \geq p$:

Substituindo (7.1.10) e (7.1.11) em (7.1.14) e considerando $s = 0$, tem-se

$$\begin{aligned} y(t_{k+n}) - y_{k+n}^{[1]} = h\beta_n \frac{\partial f}{\partial y}\left(t_{k+n}, \eta_{k+n,s}\right)\left[C_{p^*+1}^* h^{p^*+1} y^{(p^*+1)}(t_k) + O(h^{p^*+2})\right] \\ + C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}), \end{aligned}$$

onde

$$y(t_{k+n}) - y_{k+n}^{[1]} = C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}). \tag{7.1.15}$$

Substituindo-se (7.1.11) e (7.1.15) em (7.1.14) e tomando-se sucessivamente $s = 1, 2, 3, \ldots, m-1$, conclui-se que

$$y(t_{k+n}) - y_{k+n}^{[m]} = C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}).$$

Logo, o erro de discretização local principal do método preditor-corretor é o mesmo do corretor quando $p^* \geq p \ \forall m \geq 1$.

(b) Caso $p^* = p - 1$:

Substituindo-se (7.1.10) e (7.1.11) em (7.1.14) e considerando $s = 0$, tem-se

$$\begin{aligned} y(t_{k+n}) - y_{k+n}^{[1]} &= h\beta_n \frac{\partial f}{\partial y}\left(t_{k+n}, \eta_{k+n,s}\right)\left(C_p^* h^p y^{(p)}(t_k) + O(h^{p+1})\right) + \\ &+ C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}), \\ y(t_{k+n}) - y_{k+n}^{[1]} &= \beta_n \frac{\partial f}{\partial y}\left(t_{k+n}, \eta_{k+n,s}\right)\left(C_p^* h^{p+1} y^{(p)}(t_k) + O(h^{p+2})\right) + \\ &+ C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}), \\ y(t_{k+n}) - y_{k+n}^{[1]} &= \left[\beta_n \frac{\partial f}{\partial y} C_p^* y^{(p)}(t_k) + C_{p+1} y^{(p+1)}(t_k)\right] h^{p+1} + O(h^{p+2}). \end{aligned}$$

Para $m = 1$, o erro de discretização local do método preditor-corretor é da mesma ordem do corretor, porém não idêntico. Entretanto, com sucessivas substituições em (7.1.14), tem-se para $m \geq 2$

$$y(t_{k+n}) - y_{k+n}^{[m]} = C_{p+1}h^{p+1}y^{(p+1)}(t_k) + O(h^{p+2}),$$

ou seja, o erro de discretização local principal do método preditor-corretor torna-se igual ao do corretor.

(c) Caso $p^* = p - 2$:

Substituindo-se (7.1.10) e (7.1.11) em (7.1.14) e considerando $s = 0$, tem-se

$$\begin{aligned}
y(t_{k+n}) - y_{k+n}^{[1]} &= h\beta_n \frac{\partial f}{\partial y}(t_{k+n}, \eta_{k+n,s}) \left(C_{p-1}^* h^{p-1} y^{(p-1)}(t_k) + O(h^p)\right) + \\
&\quad + C_{p+1}h^{p+1}y^{(p+1)}(t_k) + O(h^{p+2}), \\
y(t_{k+n}) - y_{k+n}^{[1]} &= \beta_n \frac{\partial f}{\partial y}(t_{k+n}, \eta_{k+n,s}) \left(C_{p-1}^* h^{p} y^{(p-1)}(t_k) + O(h^{p+1})\right) + \\
&\quad + C_{p+1}h^{p+1}y^{(p+1)}(t_k) + O(h^{p+2}), \\
y(t_{k+n}) - y_{k+n}^{[1]} &= \left[\beta_n \frac{\partial f}{\partial y} C_{p-1}^* y^{(p-1)}(t_k) + C_{p+1} h y^{(p+1)}(t_k)\right] h^p + O(h^{p+1}).
\end{aligned}$$

Para $m = 1$, o erro de discretização local principal do método preditor-corretor tem ordem um a menos do que a ordem do corretor. Susbtituindo-se a expressão anterior e (7.1.11) em (7.1.14) com $s = 1$, conclui-se que

$$\begin{aligned}
y(t_{k+n}) - y_{k+n}^{[2]} &= \left(\beta_n \frac{\partial f}{\partial y}\right)^2 C_{p-1}^* y^{(p-1)}(t_k)\, h^{p+1} + \\
&+ \left(1 + h^2 \beta_n \frac{\partial f}{\partial y}\right) C_{p+1} y^{(p+1)}(t_k)\, h^{p+1} + O(h^{p+2}).
\end{aligned} \tag{7.1.16}$$

Para $m = 2$, o erro de discretização local principal do método preditor-corretor é da mesma ordem do corretor, porém não idêntico. Entretanto, com sucessivas substituições em (7.1.14), tem-se para $m \geq 3$

$$y(t_{k+n}) - y_{k+n}^{[m]} = C_{p+1}h^{p+1}y^{(p+1)}(t_k) + O(h^{p+2}),$$

ou seja, o erro de discretização local principal do método preditor-corretor torna-se o mesmo do corretor.

É possível demonstrar que para o caso geral, $p^* = p - q, \quad 0 < q \leq p$, vale o teorema [**?**]

Teorema 7.1. *um método preditor-corretor para o qual o preditor tem ordem p^* e o corretor tem ordem p, aplicado no modo $P(EC)^m$ ou $P(EC)^mE$, com $p^*, p, m \in \mathbb{N}$ tais que $p^* \geq 0$, $p \geq 1$ e $m \geq 1$, satisfaz*

- *se $p^* \geq p$ então o erro discretização local principal do método preditor-corretor é igual ao do corretor;*
- *se $p^* = p - q$, $0 < q \leq p$, então o erro de discretização local principal do método preditor-corretor*

(a) é igual ao do corretor quando $m \geq q+1$;

(b) é da mesma ordem que a do corretor, porém não idêntico, quando $m = q$;

(c) é da forma $Kh^{p-q+m+1} + O(h^{p-q+m+2})$ quando $m \leq q-1$.

7.2 Estratégia de Milne

A "Estratégia" de Milne[1] tem por finalidade fornecer uma estimativa para o erro de discretização local de um método preditor-corretor. Considere preditor e corretor dados por

$$P: \quad C^*_{p^*+1} h^{p^*+1} y^{(p^*+1)}(t_k) + O(h^{p^*+2}) = y(t_{k+n}) - y^{[0]}_{k+n}, \tag{7.2.17}$$

$$C: \quad C_{p+1} h^{p+1} y^{(p+1)}(t_k) + O(h^{p+2}) = y(t_{k+n}) - y^{[m]}_{k+n}, \tag{7.2.18}$$

e tome $p^* = p$. Subtraindo-se (7.2.17) de (7.2.18), tem-se em primeira aproximação

$$\left(C^*_{p+1} - C_{p+1}\right) h^{p+1} y^{(p+1)}(t_k) = y^{[m]}_{k+n} - y^{[0]}_{k+n}. \tag{7.2.19}$$

Multiplicando-se (7.2.19) por C_{p+1} e por C^*_{p+1}, obtêm-se

$$C_{p+1} h^{p+1} y^{(p+1)}(t_k) = \frac{C_{p+1}}{C^*_{p+1} - C_{p+1}} \left(y^{[m]}_{k+n} - y^{[0]}_{k+n}\right), \tag{7.2.20}$$

$$C^*_{p+1} h^{p+1} y^{(p+1)}(t_k) = \frac{C^*_{p+1}}{C^*_{p+1} - C_{p+1}} \left(y^{[m]}_{k+n} - y^{[0]}_{k+n}\right), \tag{7.2.21}$$

estimativas para os erros de discretização locais principais do método preditor-corretor e do preditor, respectivamente.

Pode-se melhorar $y^{[m]}_{k+n}$ utilizando-se a estimativa (7.2.20):

$$y(t_{k+n}) - y^{[m]}_{k+n} = C_{p+1} h^{p+1} y^{p+1}(t_k) + O(h^{p+2}),$$

$$y(t_{k+n}) - \underbrace{\left(y^{[m]}_{k+n} + C_{p+1} h^{p+1} y^{p+1}(t_k)\right)}_{\hat{y}^{[m]}_{k+n}} = O(h^{p+2}),$$

$$\hat{y}^{[m]}_{k+n} = y^{[m]}_{k+n} + C_{p+1} h^{p+1} y^{p+1}(t_k). \tag{7.2.22}$$

Substituindo-se (7.2.20) em (7.2.22), obtém-se

$$\hat{y}^{[m]}_{k+n} = y^{[m]}_{k+n} + \frac{C_{p+1}}{C^*_{p+1} - C_{p+1}} \left(y^{[m]}_{k+n} - y^{[0]}_{k+n}\right). \tag{7.2.23}$$

Em (7.2.23), $\hat{y}^{[m]}_{k+n}$ é o *Modificador M* do corretor.

[1] Tradução livre do inglês *Milne's Device* com possíveis variações dadas por *Algoritmo* e *Esquema* de Milne.

A Estratégia de Milne também pode ser empregada para aprimorar a solução do preditor. A expressão (7.2.21) não pode ser empregada pois, no momento de aplicar o preditor $y_{k+n}^{[m]}$ não é conhecido. Nesse caso,

$$C_{p+1}^* h^{p+1} y^{(p+1)}(t_k) = \underbrace{C_{p+1}^* h^{p+1} y^{(p+1)}(t_{k-1})}_{\hat{y}_{k+n}^{[0]} - y_{k+n}^{[0]}} + O(h^{p+2}),$$

$$\hat{y}_{k+n}^{[0]} - y_{k+n}^{[0]} = C_{p+1}^* h^{p+1} y^{(p+1)}(t_{k-1}). \tag{7.2.24}$$

Substituindo-se (7.2.21) em (7.2.24), tem-se

$$\hat{y}_{k+n}^{[0]} = y_{k+n}^{[0]} + \frac{C_{p+1}^*}{C_{p+1}^* - C_{p+1}} \left(y_{k+n-1}^{[m]} - y_{k+n-1}^{[0]} \right). \tag{7.2.25}$$

Em (7.2.25), $\hat{y}_{k+n}^{[0]}$ é o *Modificador M* do preditor.

Os modos $P(EC)^m E$ e $P(EC)^m$ podem ser agora reescritos com os modificadores do preditor e do corretor como

$$PM(ECM)^m E$$

e

$$PM(ECM)^m.$$

Para $p^* = p$, a Estratégia de Milne pode ser utilizada ou para aprimorar a solução $y_{k+n}^{[m]}$ ou para o controle automático do passo de integração, mas não para ambos simultaneamente. Caso se queira valores aproximados mais precisos, é preferível aumentar a ordem do método e usar a Estratégia de Milne para controlar o passo de integração.

7.3 Estabilidade absoluta

Sabe-se que o polinômio cujas raízes controlam as propriedades de estabilidade absoluta de um método de passo múltiplo linear (ou seja, a escolha do passo de integração numa simulação computacional) é o polinômio

$$\Pi(r) = \rho(r) - \bar{h}\sigma(r),$$

onde $\rho(r) = \sum_{j=0}^{n} \alpha_j r^j$ e $\sigma(r) = \sum_{j=0}^{n} \beta_j r^j$ são, respectivamente, o primeiro e o segundo polinômios característicos do método de passo múltiplo e $\bar{h} = h\lambda = h\dfrac{\partial f}{\partial y}$ é uma constante.

Em métodos preditores-corretores, o polinômio que determina as propriedades de estabilidade absoluta $\Pi(r)$ depende de ambos os polinômios, o do preditor e o do corretor. Tal fato pode ser ilustrado, por exemplo, para um método preditor-

corretor aplicado no modo PECE ($m = 1$), explicitamente dado por

$$P: \quad y_{k+n}^{[0]} + \sum_{j=0}^{n-1} \alpha_j^* y_{k+j}^{[1]} = h \sum_{j=0}^{n-1} \beta_j^* f(t_{k+j}, y_{k+j}^{[1]}), \tag{7.3.26}$$

$$C: \quad y_{k+n}^{[1]} + \sum_{j=0}^{n-1} \alpha_j y_{k+j}^{[1]} = h\beta_n f(t_{k+n}, y_{k+n}^{[0]}) + h \sum_{j=0}^{n-1} \beta_j f(t_{k+j}, y_{k+j}^{[1]}), \tag{7.3.27}$$

$$P: \quad y(t_{k+n}) + \sum_{j=0}^{n-1} \alpha_j^* y(t_{k+j}) = h \sum_{j=0}^{n-1} \beta_j^* f(t_{k+j}, y(t_{k+j})) + \alpha^* h, \tag{7.3.28}$$

$$C: \quad y(t_{k+n}) + \sum_{j=0}^{n-1} \alpha_j y(t_{k+j}) = h\beta_n f(t_{k+n}, y(t_{k+n})) + h \sum_{j=0}^{n-1} \beta_j f(t_{k+j}, y(t_{k+j})) + \\ + \alpha h. \tag{7.3.29}$$

Subtraindo-se (7.3.26) de (7.3.28), obtém-se

$$\begin{aligned} y(t_{k+n}) - y_{k+n}^{[0]} \quad &+ \quad \sum_{j=0}^{n-1} \alpha_j^* \left[y(t_{k+j}) - y_{k+j}^{[1]} \right] = \\ &= \quad h \sum_{j=0}^{n-1} \beta_j^* \left[f(t_{k+j}, y(t_{k+j})) - f(t_{k+j}, y_{k+j}^{[1]}) \right] + \alpha^* h, \\ e_{k+n}^{[0]} + \sum_{j=0}^{n-1} \alpha_j^* e_{k+j}^{[1]} \quad &= \quad h \sum_{j=0}^{n-1} \beta_j^* [f(t_{k+j}, y(t_{k+j})) - f(t_{k+j}, y_{k+j}^{[1]})] + \\ &+ \quad \alpha^* h. \end{aligned} \tag{7.3.30}$$

Do mesmo modo, subtraindo-se (7.3.27) de (7.3.29), chega-se a

$$\begin{aligned} y(t_{k+n}) - y_{k+n}^{[1]} \quad &+ \quad \sum_{j=0}^{n-1} \alpha_j \left[y(t_{k+j}) - y_{k+j}^{[1]} \right] = \\ &= \quad h\beta_n \left[f(t_{k+n}, y(t_{k+n})) - f(t_{k+n}, y_{k+n}^{[0]}) \right] + \\ &+ \quad \sum_{j=0}^{n-1} \beta_j \left[f(t_{k+j}, y(t_{k+j})) - f(t_{k+j}, y_{k+j}^{[1]}) \right] + \alpha h, \\ e_{k+n}^{[1]} + \sum_{j=0}^{n-1} \alpha_j e_{k+j}^{[1]} \quad &= \quad h\beta_n \left[f(t_{k+n}, y(t_{k+n})) - f(t_{k+n}, y_{k+n}^{[0]}) \right] + \\ &+ \quad h \sum_{j=0}^{n-1} \beta_j [f(t_{k+j}, y(t_{k+j})) - f(t_{k+j}, y_{k+j}^{[1]})] + \alpha h. \end{aligned} \tag{7.3.31}$$

Utilizando-se o Teorema do Valor Médio nas equações de diferenças lineares (7.3.30) e (7.3.31), obtêm-se

$$e_{k+n}^{[0]} + \sum_{j=0}^{n-1} \alpha_j^* e_{k+j}^{[1]} = h \sum_{j=0}^{n-1} \beta_j^* \frac{\partial f}{\partial y} (t_{k+j}, \eta_{k+j})\, e_{k+j}^{[1]} + \alpha^* h \tag{7.3.32}$$

e

$$e_{k+n}^{[1]} + \sum_{j=0}^{n-1} \alpha_j e_{k+j}^{[1]} = h\beta_n \frac{\partial f}{\partial y}\left(t_{k+n}, \eta_{k+n}\right) e_{k+n}^{[0]} + \tag{7.3.33}$$

$$+ \quad h \sum_{j=0}^{n-1} \beta_j \frac{\partial f}{\partial y}(t_{k+j}, \eta_{k+j}) e_{k+j}^{[1]} + \alpha h.$$

Supondo-se que seja aproximadamente constante $\dfrac{\partial f}{\partial y} \approx \lambda$, denotando $h\lambda = \bar{h}$ e substituindo-se (7.3.32) em (7.3.33), tem-se

$$e_{k+n}^{[1]} + \sum_{j=0}^{n-1} \alpha_j e_{k+j}^{[1]} = \bar{h}\beta_n \left[-\sum_{j=0}^{n-1} \alpha_j^* e_{k+j}^{[1]} + \bar{h} \sum_{j=0}^{n-1} \beta_j^* e_{k+j}^{[1]} + \alpha^* h \right] +$$

$$+ \quad \bar{h} \sum_{j=0}^{n-1} \beta_j e_{k+j}^{[1]} + \alpha h,$$

$$e_{k+n}^{[1]} + \sum_{j=0}^{n-1} \alpha_j e_{k+j}^{[1]} - \bar{h} \sum_{j=0}^{n-1} \beta_j e_{k+j}^{[1]} = \bar{h}\beta_n \left[-\sum_{j=0}^{n-1} \alpha_j^* e_{k+j}^{[1]} + \bar{h} \sum_{j=0}^{n-1} \beta_j^* e_{k+j}^{[1]} \right] +$$

$$+ \quad \bar{h}\beta_n \alpha^* h + \alpha h,$$

$$\sum_{j=0}^{n} \alpha_j e_{k+j}^{[1]} - \bar{h} \sum_{j=0}^{n-1} \beta_j e_{k+j}^{[1]} = -\bar{h}\beta_n \sum_{j=0}^{n-1} (\alpha_j^* - \bar{h}\beta_j^*) e_{k+j}^{[1]} + \phi, \tag{7.3.34}$$

onde $\bar{h} = h\lambda = h\frac{\partial f}{\partial y}$ e $\phi = \bar{h}\beta_n \alpha^* h + \alpha h$.

Adicionado-se $-\bar{h}\beta_n e_{k+n}^{[1]}$ a ambos os lados de (7.3.34) e considerando-se $\alpha_n^* = 1$ e $\beta_n^* = 0$, obtêm-se

$$\sum_{j=0}^{n} (\alpha_j - \bar{h}\beta_j) e_{k+j}^{[1]} = -\bar{h}\beta_n \sum_{j=0}^{n} (\alpha_j^* - \bar{h}\beta_j^*) e_{k+j}^{[1]} + \phi,$$

$$\sum_{j=0}^{n} (\alpha_j - \bar{h}\beta_j) e_{k+j}^{[1]} + \bar{h}\beta_n \sum_{j=0}^{n} (\alpha_j^* - \bar{h}\beta_j^*) e_{k+j}^{[1]} = \phi,$$

$$\sum_{j=0}^{n} \left[\left(\alpha_j - \bar{h}\beta_j\right) + \bar{h}\beta_n \left(\alpha_j^* - \bar{h}\beta_j^*\right) \right] e_{k+j}^{[1]} = \phi. \tag{7.3.35}$$

Em (7.3.35), tem-se uma equação de diferenças linear cuja solução é dada por

$$e_k^{[1]} = \sum_{j=1}^{n} d_j r_j^k + \frac{\phi}{\displaystyle\sum_{j=0}^{n} \left[\left(\alpha_j - \bar{h}\beta_j\right) + \bar{h}\beta_n \left(\alpha_j^* - \bar{h}\beta_j^*\right) \right]},$$

onde r_j são as raízes do polinômio

$$\pi(r) = \sum_{j=0}^{n} \left[(\alpha_j - \bar{h}\beta_j) + \bar{h}\beta_n(\alpha_j^* - \bar{h}\beta_j^*)\right] r^j. \tag{7.3.36}$$

Pode-se reescrever o polinômio (7.3.36) em função dos polinômios característicos do preditor e do corretor. Assim,

$$\pi(r) = \rho(r) - \bar{h}\sigma(r) + \bar{h}\beta_n \left[\rho^*(r) - \bar{h}\sigma^*(r)\right], \tag{7.3.37}$$

onde $\rho^*(r) = \sum_{j=0}^{n} \alpha_j^* r^j$, $\rho(r) = \sum_{j=0}^{n} \alpha_j r^j$, $\sigma^*(r) = \sum_{j=0}^{n} \beta_j^* r^j$ e $\sigma(r) = \sum_{j=0}^{n} \beta_j r^j$.

Caso as raízes r_j do polinômio (7.3.37) satisfaçam a condição

$$|r_j| < 1, \;\; j = 1, 2, \cdots, n,$$

o método preditor-corretor em estudo será absolutamente estável. O intervalo de estabilidade absoluta do Método Preditor-Corretor será (α, β) $\alpha, \beta \in \mathbb{R}$ tal que $\bar{h} = h\lambda \in (\alpha, \beta)$.

7.3.1 Exercícios

7.4 Controle do passo de integração

Para controlar automaticamente o passo de integração $h > 0$ em um método preditor-corretor, utilizam-se simultaneamente três critérios de seleção:

1. a Estratégia de Milne (se $p^* = p$) para estimar o erro de discretização local principal;

2. $h\dfrac{\partial f}{\partial y} = h\lambda = \bar{h}$ deve pertencer ao intervalo de estabilidade absoluta;

3. a condição de convergência do Método do Ponto Fixo

$$h < \frac{1}{L|\beta_n|}.$$

No segundo critério, faz-se necessário estimar $\dfrac{\partial f}{\partial y}$. Uma das estimativas possíveis (Lambert [16]) é dada por

$$\bar{h} = h\lambda = h\frac{\partial f}{\partial y} \approx h\frac{f\left(t_{k+n}, y_{k+n}^{[1]}\right) - f\left(t_{k+n}, y_{k+n}^{[0]}\right)}{y_{k+n}^{[1]} - y_{k+n}^{[0]}}$$

quando $m = 1$ e

$$\bar{h} = h\lambda = h\frac{\partial f}{\partial y} \approx \frac{y_{k+n}^{[m]} - y_{k+n}^{[m-1]}}{\beta_n\left(y_{k+n}^{[m-1]} - y_{k+n}^{[m-2]}\right)}$$

quando $m \geq 2$.

7.4.1 Exercícios

Exercício 7.2. *Deduza as fórmulas que estimam a parte principal do erro de discretização local dos métodos que usam como preditor Adams-Bashforth e como corretor Adams-Moulton, ambos de ordem 4.*

Exercício 7.3. *Defina formalmente os algoritmos de 4 passos que usam o par preditor-corretor de Adams-Bashforth-Moulton de quarta ordem nos modos*

1. *PEC;*
2. *PECE;*
3. *PMEC;*
4. *PMECE.*

Exercício 7.4. *Aplique o Método Preditor-Corretor definido por*

$$\begin{array}{l} y_{n+4} - y_{n+3} = \dfrac{h}{24}(55f_{n+3} - 59f_{n+2} + 37f_{n+1} - 9f_n) \\ \\ y_{n+4} - y_{n+3} = \dfrac{h}{24}(9f_{n+4} + 19f_{n+3} - 5f_{n+2} + f_{n+1}) \end{array},$$

no modo $PECE$, ao problema de valor inicial

$$\begin{cases} \dfrac{d}{dt}y(t) = -20y(t) \qquad 0 \leq t \leq 1 \\ \\ y(0) = 1 \end{cases}.$$

Ilustre a estabilidade usando

1. $h = 0,1$;
2. $h = 0,01$.

Exercício 7.5. *Considere o problema de valor inicial*

$$\begin{cases} \dfrac{d}{dt}y(t) = \dfrac{2t}{y^2}, \qquad 0 \leq t \leq 3, \\ \\ y(0) = 1. \end{cases} \tag{7.4.38}$$

1. *Solucione (7.4.38) usando $h = 0,1$ e o método preditor-corretor*

$$\begin{array}{l} y_{n+1} - y_n = \frac{h}{12}\left(23f_n - 16f_{n-1} + 5f_{n-2}\right) \\ \\ y_{n+1} - y_n = \frac{h}{24}(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2}) \end{array}.$$

2. *Solucione (7.4.38) usando $h = 0,1$ e o método preditor-corretor*

$$\begin{array}{l} y_{n+1} - y_n = \frac{h}{24}\left(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}\right) \\ \\ y_{n+1} - y_n = \frac{h}{24}(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2}) \end{array}.$$

3. *Compare os dois métodos usando y_1, y_2 e y_3 obtidos por um $RK44$.*

P	$\rho^*(\zeta) = \zeta^2 - 1 \quad \sigma^*(\zeta) = (3/2)(\zeta - 1)$
C	$\rho(\zeta) = \zeta^2 - 1 \quad \sigma(\zeta) = (1/2)\left(\zeta^2 + \zeta\right)$

Tabela 7.3: Polinômios característicos que definem o método preditor-corretor.

Exercício 7.6. *Considere o preditor P e o corretor C definidos pelos polinômios característicos presentes na Tabela 7.3.*

1. *Determine a ordem do método.*
2. *Calcule o coeficiente da parte principal do erro de discretização local .*
3. *Mostre que o coeficiente determinado no item anterior é igual ao do corretor.*

7.5 Suplemento teórico

Um número p é um ponto fixo de uma função $g(t)$ se

$$g(p) = p,$$

ou equivalentemente, se

$$g(p) - p = 0.$$

Teorema 7.2 (Teorema do Ponto Fixo)**.** *Seja $g(t) \in C^{(1)}[a,b]$ tal que $g(t) \in [a,b]$ $\forall t \in [a,b]$. Suponha ainda que $g'(t)$ exista em (a,b) com*

$$|g'(t)| \leq k \ \forall t \in (a,b),$$

onde $k < 1$ é uma constante positiva. Então, para um número qualquer $p_0 \in [a,b]$, a sequência definida por

$$p_{n+1} = g\left(p_n\right), \ \ n \geq 0$$

converge para o único ponto fixo $p \in [a,b]$.

Observação: O Teorema do Ponto Fixo permite determina a raiz p de $f(t) = 0$, ou seja, $f(p) = 0$. Para empregá-lo, é necessário reescrever $f(t) = 0$ como $t = g(t)$.

7.6 Exercícios resolvidos

Exercício Resolvido 7.1. *Considere o preditor P e dois corretores $C^{(1)}$ e $C^{(2)}$, definidos pelos polinômios característicos presentes na Tabela (7.4).*

(a) Deduza uma estimativa do erro local de discretização

(i) para um Método Preditor-Corretor que use P e $C^{(1)}$;

(ii) para um Método Preditor-Corretor que use P e $C^{(2)}$.

(b) Escreva o algoritmo que utiliza P e $C^{(1)}$ no modo $PECE$.

P	$\rho^*(\zeta) = \zeta^4 - 1$	$\sigma^*(\zeta) = (4/3)\left(2\zeta^3 - \zeta^2 + 2\zeta\right)$
$C^{(1)}$	$\rho_1(\zeta) = \zeta^2 - 1$	$\sigma_1(\zeta) = (1/3)\left(\zeta^2 + 4\zeta + 1\right)$
$C^{(2)}$	$\rho_2(\zeta) = \zeta^3 - (9/8)\zeta^2 + (1/8)$	$\sigma_2(\zeta) = (3/8)\left(\zeta^3 + 2\zeta^2 - \zeta\right)$

Tabela 7.4: Polinômios característicos que definem o Método Preditor-Corretor.

(c) Escreva o algoritmo que utiliza P e $C^{(2)}$ no modo $PMECME$.

Solução:

Como $\rho^*(r) = \sum_{j=0}^{n} \alpha_j^* r^j$, $\rho(r) = \sum_{j=0}^{n} \alpha_j r^j$, $\sigma^*(r) = \sum_{j=0}^{n} \beta_j^* r^j$, $\sigma(r) = \sum_{j=0}^{n} \beta_j r^j$ *e* $\sum_{j=0}^{n} \alpha_j y_{k+j} = h \sum_{j=0}^{n} \beta_j f_{k+j}$, *os polinômios característicos da Tabela (7.4) definem os seguintes métodos de passo múltiplo lineares:*

$$P: \quad y_{k+4} - y_k = \frac{4}{3} h \left[2f_{k+3} - f_{k+2} + 2f_{k+1}\right]; \tag{7.6.39}$$

$$C^{(1)}: \quad y_{k+2} - y_k = \frac{1}{3} h \left[f_{k+2} + 4f_{k+1} + f_k\right]; \tag{7.6.40}$$

$$C^{(2)}: \quad y_{k+3} - \frac{9}{8} y_{k+2} + \frac{1}{8} y_k = \frac{3}{8} h \left[f_{k+3} + 2f_{k+2} - f_{k+1}\right]. \tag{7.6.41}$$

Pode-se mostrar que os métodos P, $C^{(1)}$ e $C^{(2)}$ têm ordem de consistência 4 (verifique que $C_0 = C_1 = C_2 = C_3 = C_4 = 0$ e $C_5 \neq 0$), o que possibilita o uso do dispositivo de Milne. Ao calcular o coeficiente C_{p+1} do erro, obtêm-se

$$C_5^* = \frac{14}{45}, \tag{7.6.42}$$

$$C_5^{(1)} = -\frac{1}{90}, \tag{7.6.43}$$

$$C_5^{(2)} = -\frac{1}{40}. \tag{7.6.44}$$

Substituindo (7.6.42), (7.6.43) e (7.6.44) em

$$C_{p+1} h^{p+1} y^{(p+1)}(t_k) = \frac{C_{p+1}}{C_{p+1}^* - C_{p+1}} \left(y_{k+n}^{[m]} - y_{k+n}^{[0]}\right),$$

conclui-se que:

$1(a)(i)$

$$\begin{aligned} C_5^{(1)} h^5 y^{(5)}(t_k) &= \frac{-\frac{1}{90}}{\frac{14}{45} + \frac{1}{90}} \left(y_{k+n}^{[m]} - y_{k+n}^{[0]}\right) \\ &= -\frac{1}{29} \left(y_{k+n}^{[m]} - y_{k+n}^{[0]}\right); \end{aligned}$$

$1(a)(ii)$

$$\begin{aligned} C_5^{(2)} h^5 y^{(5)}(t_k) &= \frac{-\frac{1}{40}}{\frac{14}{45} + \frac{1}{40}} \left(y_{k+n}^{[m]} - y_{k+n}^{[0]}\right) \\ &= -\frac{9}{121} \left(y_{k+n}^{[m]} - y_{k+n}^{[0]}\right). \end{aligned}$$

Como o preditor é um método de 4 passos, os corretores $C^{(1)}$ e $C^{(2)}$ podem ser reescritos na forma

$$C^{(1)}: \quad y_{k+4} - y_{k+2} = \frac{1}{3}h\left[f_{k+4} + 4f_{k+3} + f_{k+2}\right], \tag{7.6.45}$$

$$C^{(2)}: \quad y_{k+4} - \frac{9}{8}y_{k+3} + \frac{1}{8}y_{k+1} = \frac{3}{8}h\left[f_{k+4} + 2f_{k+3} - f_{k+2}\right]. \tag{7.6.46}$$

Assim:

1(b)

$$P: \quad y^{[0]}_{k+4} - y^{[1]}_{k} = \frac{4}{3}h\left[2f^{[1]}_{k+3} - f^{[1]}_{k+2} + 2f^{[1]}_{k+1}\right]$$

$$E: \quad f^{[0]}_{k+4} = f\left(t_{k+4}, y^{[0]}_{k+4}\right)$$

$$C: \quad y^{[1]}_{k+4} - y^{[1]}_{k+2} = \frac{1}{3}h\left[f^{[0]}_{k+4} + 4f^{[1]}_{k+3} + f^{[1]}_{k+2}\right]$$

$$E: \quad f^{[1]}_{k+4} = f\left(t_{k+4}, y^{[1]}_{k+4}\right);$$

1(c)

$$P: \quad y^{[0]}_{k+4} - \hat{y}^{[1]}_{k} = \frac{4}{3}h\left[2\hat{f}^{[1]}_{k+3} - \hat{f}^{[1]}_{k+2} + 2\hat{f}^{[1]}_{k+1}\right]$$

$$M: \quad \hat{y}^{[0]}_{k+4} = y^{[0]}_{k+4} + \frac{112}{121}\left(y^{[1]}_{k+3} - y^{[0]}_{k+3}\right)$$

$$E: \quad \hat{f}^{[0]}_{k+4} = f\left(t_{k+4}, \hat{y}^{[0]}_{k+4}\right)$$

$$C: \quad y^{[1]}_{k+4} - \frac{9}{8}\hat{y}^{[1]}_{k+3} + \frac{1}{8}\hat{y}^{[1]}_{k+1} = \frac{3}{8}h\left[\hat{f}^{[0]}_{k+4} + 2\hat{f}^{[1]}_{k+3} - \hat{f}^{[1]}_{k+2}\right]$$

$$M: \quad \hat{y}^{[1]}_{k+4} = y^{[1]}_{k+4} - \frac{9}{121}\left(y^{[1]}_{k+4} - y^{[0]}_{k+4}\right)$$

$$E: \quad \hat{f}^{[1]}_{k+4} = f\left(t_{k+4}, \hat{y}^{[1]}_{k+4}\right).$$

<u>*Observação:*</u>

$$\begin{aligned}
\hat{y}^{[0]}_{k+n} &= y^{[0]}_{k+n} + \frac{C^*_{p+1}}{C^*_{p+1} - C_{p+1}}\left(y^{[m]}_{k+n-1} - y^{[0]}_{k+n-1}\right) \\
\hat{y}^{[0]}_{k+4} &= y^{[0]}_{k+4} + \frac{\frac{14}{45}}{\frac{14}{45} + \frac{1}{40}}\left(y^{[1]}_{k+3} - y^{[0]}_{k+3}\right) \\
\hat{y}^{[0]}_{k+4} &= y^{[0]}_{k+4} + \frac{112}{121}\left(y^{[1]}_{k+3} - y^{[0]}_{k+3}\right)
\end{aligned}$$

Apêndice A

Exercícios complementares

1 Seja o método de passo único descrito abaixo.

Método A.1 (de Runge-Kutta de $2^{\underline{a}}$ ordem com dois estágios (Ralston))**.**

$$\begin{cases} y_0 = y(t_0) \\ y_{k+1} = y_k + h\Phi(t_k, y_k, h) \end{cases}, \tag{A.0.1}$$

com $t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1, \quad h = \dfrac{b-a}{n}$ *e*

$$\Phi(t_k, y_k, h) = \frac{1}{4}\left(\kappa_1 + 3\kappa_2\right),$$

sendo

$$\begin{cases} \kappa_1 = f(t_k, y_k) \\ \\ \kappa_2 = f\left(t_k + \dfrac{2}{3}h, y_k + \dfrac{2}{3}h\kappa_1\right) \end{cases}.$$

1.1 O método de passo simples (A.0.1) é consistente e convergente? Justifique.

1.2 Comprove a ordem de consistência dois do método de passo simples (A.0.1) verificando que ele concorda com o Método da Série de Taylor até o termo de segunda ordem.

1.3 Determine a região de estabilidade absoluta do método de passo simples (A.0.1). Faça a análise para $\lambda \in \mathbb{R}$.

1.4 Use o método de passo simples (A.0.1) para solucionar o Problema de Cauchy

$$\begin{cases} \dfrac{d}{dt}y(t) = te^{3t} - 2y(t), \quad t \in [0,1] \\ \\ y(0) = 0 \end{cases}, \tag{A.0.2}$$

com $h = 0,5$. Calcule o erro global de discretização no instante $t = 1$ e comente o resultado obtido.

1.5 Ao solucionar o Problema de Cauchy numericamente, é mais vantajoso computacionalmente usar o método de passo simples (A.0.1) ou o Método de Euler Aprimorado? Justifique.

2 Considere o método descrito a seguir.

Método A.2 (de passo único)**.**

$$\begin{cases} y_0 = y(t_0) \\ y_{k+1} = y_k + h\Phi(t_k, y_k, h) \end{cases}, \tag{A.0.3}$$

com $t_{k+1} = t_k + h, \quad 0 \leq k \leq n-1, \quad h = \dfrac{b-a}{n}$ *e*

$$\Phi(t_k, y_k, h) = \frac{1}{5}\left(2\kappa_1 + 3\kappa_2\right),$$

sendo

$$\begin{cases} \kappa_1 = f(t_k, y_k) \\ \kappa_2 = f\left(t_k + \dfrac{5}{6}h, y_k + \dfrac{5}{6}h\kappa_1\right) \end{cases}.$$

2.1 O método (A.0.3) é um Método de Runge-Kutta? Justifique.

2.2 O método (A.0.3) é consistente e convergente? Justifique.

2.3 Seja o Problema de Cauchy

$$\begin{cases} \dfrac{d}{dt}y(t) = -50y(t), \quad t \in [0, 10] \\ \\ y(0) = 5 \end{cases}. \tag{A.0.4}$$

a Discretize o problema de valor inicial (A.0.4) usando o Método de Runge-Kutta 44 clássico.

b Analise a estabilidade absoluta do Método de Runge-Kutta 44 clássico, aplicado ao problema de valor inicial (A.0.4), para os seguintes passos de integração: $h = \dfrac{1}{4}$, $h = \dfrac{1}{10}$, $h = \dfrac{3}{50}$, $h = \dfrac{1}{20}$ e $h = \dfrac{1}{100}$.

2.4 Seja o Problema de Cauchy

$$\begin{cases} \dfrac{d}{dt}y(t) = 1 + \dfrac{1}{t}y(t), \quad t \in [1, 2] \\ \\ y(1) = 2 \end{cases}. \tag{A.0.5}$$

a Calcule a solução exata do problema de valor inicial (A.0.5).

b Discretize o problema de valor inicial (A.0.5) empregando o Método de Euler Implícito.

c Determine o erro global de discretização cometido no instante $t = 2$ ao solucionar numericamente o problema de valor inicial (A.0.5) utilizando o Método de Euler Implícito com passo de integração $h = \dfrac{1}{4}$. Comente o resultado obtido.

2.5 Seja o Problema de Cauchy

$$\begin{cases} \dfrac{d}{dt}y(t) = -5y(t) + 5t^2 + 2t, \quad t \in [0,1] \\ \\ y(0) = \dfrac{1}{3} \end{cases} . \tag{A.0.6}$$

a Calcule a solução exata do problema de valor inicial (A.0.6).

b Estime o passo de integração h para que o erro local de discretização para o Método de Euler, aplicado à solução numérica do problema de valor inicial (A.0.6), seja menor que 10^{-3}.

3 Solucione a equação de diferenças linear

$$y_{k+2} + 4y_k = 15$$

com condições iniciais $y_0 = 1$ e $y_1 = 2$.

4 Seja a sequência de Fibonacci

$$y_{k+2} = y_{k+1} + y_k, \ k \geq 0, \tag{A.0.7}$$

onde

$$y_0 = y_1 = 1. \tag{A.0.8}$$

Solucione a equação de diferenças linear (A.0.7), sujeita às condições iniciais (A.0.8).

5 Calcule φ, δ e γ para que o método de passo múltiplo linear

$$y_{k+4} - y_k + \varphi\left(y_{k+3} - y_{k+1}\right) = h\left[\delta\left(f_{k+3} - f_{k+1}\right) + \gamma f_{k+2}\right] \tag{A.0.9}$$

seja consistente de ordem 3.

6 Os métodos de passo múltiplo lineares

$$y_{k+3} = 3y_{k+1} - \frac{1}{2}\left(3y_{k+2} + y_k\right) + 3hf_{k+2} \tag{A.0.10}$$

e

$$y_{k+3} = \frac{1}{8}\left(9y_{k+2} - y_k\right) + \frac{3h}{8}\left(f_{k+3} + 2f_{k+2} - f_{k+1}\right) \tag{A.0.11}$$

são convergentes? Justifique.

7 Sejam o método de passo múltiplo linear

$$y_{k+2} - y_{k+1} = \frac{h}{3}\left(f_{k+2} + 4f_{k+1} + f_k\right) \tag{A.0.12}$$

e o problema de valor inicial

$$\begin{cases} \dfrac{d}{dt}y(t) = -20\left[y(t) - t^2\right] + 2t, \quad t \in [0,10] \\ \\ y(0) = \dfrac{1}{3} \end{cases} . \tag{A.0.13}$$

7.1 O método de passo múltiplo linear (A.0.12) é *absolutamente estável* considerando o p.v.i. (A.0.13) com $h = 0,1$? Justifique.

7.2 Na sua opinião, o método de passo múltiplo linear (A.0.12) gera bons resultados quando aplicado à solução do p.v.i. (A.0.13) com $h = 0,1$? Justifique.

8 Seja o problema de valor inicial

$$\begin{cases} \dfrac{d}{dt}y(t) = \dfrac{1}{t}y(t) - \dfrac{1}{t^2}y^2(t), \quad t \in [1,2] \\ \\ y(1) = 1 \end{cases} . \tag{A.0.14}$$

8.1 Calcule a solução exata do p.v.i. (A.0.14).

8.2 Discretize o p.v.i. (A.0.14) empregando o método de passo múltiplo linear

$$y_{k+4} = y_k + \frac{4h}{3}\left(2f_{k+3} - f_{k+2} + 2f_{k+1}\right). \tag{A.0.15}$$

8.3 Determine o erro global de discretização cometido no instante $t = 2$ ao solucionar o p.v.i. (A.0.14) com o método de passo múltiplo linear (A.0.15) e $h = \dfrac{1}{4}$.

9 Considere os Métodos Preditor P e Corretor C definidos por

$$\begin{aligned} y_{k+2} + 4y_{k+1} - 5y_k &= h\left(4f_{k+1} + 2f_k\right), \\ y_{k+2} - y_k &= \frac{h}{3}\left(f_{k+2} + 4f_{k+1} + f_k\right). \end{aligned}$$

9.1 Determine o erro local de discretização principal do Método Preditor-Corretor.

9.2 Escreva o polinômio

$$\pi(r) = \rho(r) - \bar{h}\sigma(r) + \bar{h}\beta_n\left[\rho^*(r) - \bar{h}\sigma^*(r)\right] \tag{A.0.16}$$

associado à equação de diferenças linear para o erro global de discretização do Método Preditor-Corretor no modo $PECE$. Qual é a aplicabilidade do polinômio (A.0.16)?

9.3 Escreva o algoritmo que utiliza o Método Preditor-Corretor no modo $PECE$.

10 Seja o Método Preditor-Corretor Adams-Bashforth-Moulton definido por

$$\begin{aligned} y_{k+2} - y_{k+1} &= \frac{h}{2}\left(3f_{k+1} - f_k\right), \\ y_{k+2} - y_{k+1} &= \frac{h}{12}\left(5f_{k+2} + 8f_{k+1} - f_k\right). \end{aligned}$$

10.1 Determine o erro local de discretização principal do Método Preditor-Corretor.

10.2 Escreva o algoritmo que utiliza o Método Preditor-Corretor no modo $PECE$.

Apêndice B

Exercícios computacionais

1 Considere o p.v.i.

$$\begin{cases} \dfrac{d}{dt}y(t) = -20y(t), \quad t \in [0,1] \\ \\ y(0) = 1 \end{cases} . \tag{B.0.1}$$

1.1 Solucione numericamente o p.v.i. (B.0.1) usando os Métodos de Euler e de Runge-Kutta de $4^{\underline{a}}$ ordem com 4 estágios e os seguintes passos de integração:

A $h = 0,05$;

B $h = 0,1$;

C $h = 0,2$.

Implemente os métodos numéricos em linguagem C ou Fortran.

1.2 Analise e justifique os resultados obtidos, comparando os dois métodos numéricos implementados em relação a:

A convergência;

B estabilidade;

C consistência.

2 Seja o Método de Runge-Kutta-Fehlberg, definido por

$$\alpha_k^{RKF} \approx \frac{1}{360}\kappa_1 - \frac{128}{4275}\kappa_3 - \frac{2197}{75240}\kappa_4 + \frac{1}{50}\kappa_5 + \frac{2}{55}\kappa_6.$$

2.1 Valide o Método de Runge-Kutta-Fehlberg, em linguagem C ou Fortran, empregando o p.v.i.

$$\begin{cases} \dfrac{d}{dt}y(t) = \lambda y(t), \quad t \in [a,b] \\ \\ y(a) = y(t_0) = y_0 \end{cases} , \tag{B.0.2}$$

onde λ é uma constante.

2.2 Considere o p.v.i.

$$\begin{cases} \dfrac{d}{dt}y(t) = -2t - y(t), \quad t \in [0,10] \\ \\ y(0) = -1 \end{cases} . \tag{B.0.3}$$

A Solucione o p.v.i. (B.0.3) usando o Método de Runge-Kutta-Fehlberg, em linguagem C ou Fortran, com $h = 0,2$ e

(a) $\epsilon = 10^{-6}$;

(b) $\epsilon = 10^{-3}$.

B Compare o erro local de discretização e o passo de integração. Organize os dados em uma tabela.

C Calcule o erro global de discretização no instante $t = 10$ e comente os resultados obtidos.

3 Modelagem da propagação de doenças contagiosas

3.1 Problema

Na teoria de propagação de doenças contagiosas, uma equação diferencial ordinária não linear de primeira ordem homogênea pode ser usada para prever o número de indivíduos infectados em um tempo qualquer (em dias), desde que simplificações adequadas sejam adotadas. Em particular, considera-se que todos os indivíduos de uma população fixa podem ser contaminados e que, uma vez infectados, permanecem nessa condição.

Sejam $x(t)$ o número de indivíduos suscetíveis à infecção e $y(t)$ o número de indivíduos infectados. É razoável supor que a taxa de variação temporal do número de infectados seja proporcional ao produto do número de indivíduos suscetíveis pelo número de indivíduos infectados. Assim, tem-se que

$$\frac{d}{dt}y(t) = k\,x(t)y(t), \tag{B.0.4}$$

onde k é uma constante e

$$x(t) + y(t) = m, \tag{B.0.5}$$

sendo m o tamanho da população. Como $x(t) = m - y(t)$, a equação (B.0.4) pode ser reescrita como

$$\frac{d}{dt}y(t) = k[m - y(t)]y(t). \tag{B.0.6}$$

A equação (B.0.6) é chamada *Equação de Bernoulli*. Esta equação pode, pela substituição

$$u(t) = \frac{1}{y(t)}, \tag{B.0.7}$$

ser reescrita como a equação diferencial ordinária linear de primeira ordem não homogênea (verifique!)

$$\frac{d}{dt}u(t) = k[1 - m\,u(t)]. \tag{B.0.8}$$

3.2 Questões

A Calcule a solução exata (família de soluções) da equação (B.0.6) e determine $\lim_{t\to\infty} y(t)$. Esse limite é aceitável? Justifique.

B Implemente computacionalmente, em linguagem C ou Fortran, os seguintes métodos de aproximação da solução do Problema de Cauchy

$$\begin{cases} \dot{y}(t) = \dfrac{d}{dt}y(t) = f(t, y(t)), \quad t \in [a, b] \\ \\ y(t_0) = y(a) = y_0 \end{cases} :$$

- Euler;
- Euler Aprimorado;
- Trapézio;
- Runge-Kutta-Fehlberg;
- Adams-Bashforth de 4 passos;
- Adams-Moulton de 4 passos.

C Considere na equação (B.0.6):

$$m = 10^5;$$
$$k = 2\text{x}10^{-6};$$
$$y(0) = 10^3;$$
$$t \in [0, 30].$$

(a) Use os métodos implementados no item B para aproximar a solução da equação (B.0.11) no instante de tempo $t = 30$ *dias*. Otimize o passo temporal em cada método e calcule o erro global de discretização.

(b) Compare os métodos empregados. Use tabelas e gráficos (plote o gráfico da solução exata e da solução numérica empregando aplicativos como o winplot, o octave, o maple, o matlab, o mathematica, etc.).

4 Modelagem do crescimento populacional

4.1 Problema

Seja $P(t)$ o número de integrantes de uma população em um determinado instante de tempo t, medido em anos. Se a taxa média de nascimentos b é constante e a taxa média de mortes d é proporcional ao tamanho da população, então a taxa de crescimento da população é dada pela equação logística

$$\frac{d}{dt}P(t) = bP(t) - dP(t), \tag{B.0.9}$$

sendo d dada por

$$d = kP(t). \tag{B.0.10}$$

Substituindo (B.0.10) na equação (B.0.9), obtém-se a equação diferencial ordinária não linear, de primeira ordem, homogênea

$$\frac{d}{dt}P(t) = bP(t) - k[P(t)]^2. \tag{B.0.11}$$

A equação (B.0.11) pode, pela substituição

$$P(t) = \frac{1}{u(t)}, \tag{B.0.12}$$

ser reescrita como a equação diferencial ordinária linear, de primeira ordem, não homogênea (verifique!)

$$\frac{d}{dt}u(t) + bu(t) = k. \tag{B.0.13}$$

4.2 Questões

A Calcule a solução exata (família de soluções) da equação (B.0.11) e determine

$$\lim_{t\to\infty} P(t).$$

Este limite é aceitável? Justifique.

B Implemente computacionalmente, em linguagem C ou Fortran, os seguintes métodos de aproximação da solução do Problema de Cauchy

$$\begin{cases} \dot{y}(t) = \dfrac{d}{dt}y(t) = f(t, y(t)), \quad t \in [a, b] \\ \\ y(t_0) = y(a) = y_0 \end{cases} :$$

- Euler;
- Euler Aprimorado;
- Trapézio;
- Runge-Kutta-Fehlberg;
- Adams-Bashforth de 4 passos;
- Adams-Moulton de 4 passos;
- um método preditor-corretor.

C Considere na equação (B.0.11):

$$b = 2,9\text{x}10^{-2};$$

$$k = 1,4\text{x}10^{-7};$$

$$P(0) = 50.976;$$

$$t \in [0, 30].$$

(a) Use os métodos implementados na questão 2 para aproximar a solução das equações (B.0.11) e (B.0.10) no instante de tempo $t = 30$ *anos*. Otimize o passo temporal em cada método e calcule o erro global de discretização.

(b) Compare os métodos empregados. Use tabelas e gráficos (plote o gráfico da solução exata e da solução numérica empregando aplicativos como o winplot, o octave, o maple, o matlab, o mathematica, etc.).

Bibliografia

[1] BERGMAN, R. N. The minimal model of glucose regulation: a biography. *Mathematical Modeling in Nutrition and Health*, 2001.

[2] BOYCE, W. E.; DIPRIMA, R. C. *Equações diferenciais elementares e problemas de valores de contorno*. 10. ed. [S.l.]: LTC, 2010.

[3] BURDEN, R. L.; FAIRES, J. D. *Numerical analysis*. [S.l.]: Brooks/Cole Publishing Company, 1997.

[4] BUTCHER, J. C. A history of runge-kutta methods. *Applied numerical mathematics*, North-Holland, v. 20, n. 3, p. 247–260, 1996.

[5] BUTCHER, J. C. *Numerical methods for ordinary differential equations*. [S.l.]: John Wiley & Sons, 2003.

[6] DAHLQUIST, G. Convergence and stability in the numerical integration of ordinary differential equations. *Mathematica Scandinavica*, JSTOR, p. 33–53, 1956.

[7] DAHLQUIST, G. G. A special stability problem for linear multistep methods. *BIT Numerical Mathematics*, Springer, v. 3, n. 1, p. 27–43, 1963.

[8] DOERING, C. I.; LOPES, A. O. *Equações diferenciais ordinárias*. Rio de Janeiro: IMPA, 2010.

[9] ENDRE, S.; MAYERS, D. An introduction to numerical analysis. *Cambridge, UK*, 2003.

[10] FIGUEIREDO, D. G.; NEVES, A. F. *Equações diferenciais aplicadas*. Rio de Janeiro: IMPA, 1997.

[11] GERALD, C. F.; WHEATLEY, P. O. *Applied numerical analysis*. [S.l.]: Addison-Wesley Publishing Company, 1994.

[12] GUIDORIZZI, H. L. *Um curso de Cálculo-Volume 1*. [S.l.: s.n.], 2001.

[13] HENRICI, P. Discrete variable methods in ordinary differential equations. Wiley, 1962.

[14] KAPLAN, W. *Cálculo avançado*. [S.l.]: Edgard Blücher, 2006.

[15] KERNIGHAN, B. W.; RITCHIE, D. M. *C - A linguagem de programação*. [S.l.]: Campus, 1986.

[16] LAMBERT, J. D. *Computational methods in ordinary differential equations*. [S.l.]: John Wiley & Sons, 1973.

[17] LIMA, E. L. *Álgebra linear*. Rio de Janeiro: IMPA, 2008.

[18] LIMA, E. L. *Álgebra linear - Exercícios e soluções*. Rio de Janeiro: IMPA, 2009.

[19] MAKROGLOU, A.; LI, J.; KUANG, Y. Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Applied Numerical Mathematics*, v. 56, p. 559–573, 2006.

[20] SCHILDT, H. *C completo e total*. [S.l.]: Makron Books, 1997.

[21] SCHWARZ, H. R. *Numerical analysis - A comprehensive introduction*. [S.l.]: John Wiley & Sons, 1989.

[22] SOTOMAYOR, J. *Lições de equações diferenciais ordinárias*. Rio de Janeiro: IMPA, 1979.

[23] STETTER, H. J. *Analysis of discretization methods for ordinary differential equations*. [S.l.]: Springer Tracts in Natural Philosophy, 1973.

[24] STOER, J.; BULIRSCH, R. *Introduction to numerical analysis*. [S.l.]: Springer, 1992.

[25] VETTERLING, W. T. et al. *Numerical recipes in C - The art of scientific computing*. [S.l.]: Cambridge University Press, 1992.

[26] VETTERLING, W. T. et al. *Numerical recipes in Fortran - The art of scientific computing*. [S.l.]: Cambridge University Press, 1992.

Índice

caracterização dos métodos de passo múltiplo lineares, 71
Condição de Lipschitz, 6, 16
Conjunto convexo, 15

Discretização
  do Problema de Cauchy, 9
  método explícito, 10, 13, 19, 23, 71
  método implícito, 12, 13, 23, 71
  métodos de passo único, 10, 19, 23, 61
  métodos de passo múltiplo lineares, 13, 71, 72
  passo de integração, 9

Equação de diferenças linear, 71, 87
  homogênea, 87
  polinômio característico, 89
  sistema fundamental de soluções, 88
  solução particular, 88
  soluções fundamentais, 88
  soluções linearmente independentes, 88

Métodos
  de Ralston, 70
Método de Adams-Bashforth de 4 passos, 72
Método de inicialização
  consistência, 77
Método de Simpson, 72
Métodos
  condicionalmente estáveis, 63
  da Série de Taylor, 41
  de Adams-Bashforth, 13
  de Adams-Moulton, 13
  de Euler, 18, 19, 26, 61, 62
    fator de amplificação, 62
  de Euler Aprimorado, 11, 14, 25, 27, 46
  de Euler Implícito, 10, 27
  de Euler Modificado, 43, 46
  de Heun, 11
  de Ralston, 48
  de Runge-Kutta, 42
    de 2 estágios, 43
    de ordens superiores clássicos, 46
    de quarta ordem com 4 estágios, 47
    de quarta ordem com 5 estágios, 47
    de quinta ordem com 6 estágios, 47
    de R estágios, 44
    de terceira ordem com 3 estágios, 46
    Fehlberg, 51
  de Simpson, 13, 14, 21, 73
  do Ponto Médio, 43
  do Trapézio, 14, 20, 27, 36
  do Trapézio Explícito, 11
  do Trapézio Implícito, 11
  incondicionalmente estáveis, 67
Métodos de passo único, 10, 19, 23, 61
  consistência, 25
    ordem, 26
  convergência, 27, 30
    estimativa da ordem, 31
  depuração, 32
  erro de discretização global, 27
    estimativa, 31
    expansão assintótica, 30
  erro de discretização global
    delimitação, 29
  erro de discretização local, 23
  estabilidade absoluta, 62
  instabilidade inerente, 66
Métodos de passo múltiplo lineares, 13
  consistência
    ordem, 76
  consistência com a equação diferencial, 75
  convergência, 79, 87

dedução, 72
erro de discretização global, 105
erro de discretização local, 73
principal, 77
erro global de discretização
polinômio característico, 106
estabilidade
estabilidade absoluta, 107
estabilidade fraca, 107
zero-estabilidade, 100
exemplo de divergência, 94
polinômios característicos, 99
raiz principal, 100
zero-estabilidade, 87
Métodos Preditores-Corretores, 123
corretor, 124
erro de discretização local, 126
estabilidade absoluta, 130
métodos de Adams-Moulton, 123
modificador M, 129
preditor, 124
Métodos preditores-corretores
controle do passo de integração, 133
estratégia de Milne, 129
métodos de Adams-Bashforth, 123

Polinômio de Taylor
de uma função de duas variáveis, 18
de uma função de uma variável, 17
polinômio de Taylor com resto de Lagrange, 9
Polinômio interpolador de Lagrange, 13, 20
Ponto Fixo, 135
Problema de Cauchy, 5, 14, 23, 62
forma integral, 6
Problema de valor inicial, 16

Quadratura numérica, 20, 21
Quadratura numérica, 13

Série de Taylor, 13
Solução manufaturada, 32

Tabela
de Butcher, 45
Teorema
de existência e unicidade de solução do Problema de Cauchy, 6
de Picard, 17
do Confronto, 29
do Ponto Fixo, 135
do Valor Médio, 15
do Valor Médio, 6, 17, 74, 105, 127, 131
Fundamental do Cálculo, 17
Fundamental do Cálculo, 7

Unicidade de solução
condições iniciais, 5