

CM098 - Análise Numérica III

Notas de Aula

Saulo P. Oliveira
Departamento de Matemática
Universidade Federal do Paraná

Primeiro Semestre de 2018

Sumário

Aula 1: Métodos Numéricos para PVIs	3
1.1 Métodos de passo simples	3
1.2 Método de Euler explícito	4
Aula 2: Existência e Unicidade de solução de PVIs	6
Aula 3: Polinômio de Taylor e ordem de convergência	9
3.1 Ordem de convergência	11
Aula 4: Método da série de Taylor	12
Aula 5: Análise de erros do método da série de Taylor	14
5.1 Erro local	14
Aula 6: Métodos de Runge-Kutta	16
6.1 Expansão de Taylor 2D	16
6.2 Métodos de Runge-Kutta de primeira ordem (RK1)	17
Aula 7: Métodos de Runge-Kutta de segunda ordem (RK2)	19
Aula 8: Método de Euler modificado	21
Aula 9: Método de Runge-Kutta de quarta ordem	23
Aula 10: Métodos de Runge-Kutta adaptativos	25
Aula 11: Métodos de passos múltiplos	28
11.1 Fórmula de Adams-Bashforth	28
Aula 12: Fórmula de Adams-Moulton	34
12.1 Fórmula de Adams-Moulton de um passo ($k = 1$)	34
12.2 Fórmula de Adams-Moulton de dois passos ($k = 2$)	34
12.3 Fórmula de Adams-Moulton: caso geral	35
Aula 13: Implementação dos métodos de passos múltiplos	37

Aula 14: Consistência, Estabilidade e Convergência	38
14.1 Estabilidade	39
Aula 15: Consistência dos métodos de passos múltiplos	41
15.1 Condição alternativa de consistência	42
Aula 16: Convergência dos métodos de passos múltiplos	44
Aula 17: Sistemas de EDOs	46
Aula 18: Sistemas rígidos de equações	48
Aula 19: Problemas de Valores de Contorno	49
Aula 20: Fórmulas de diferenças finitas	53
Aula 21: Métodos de diferenças finitas	54
Aula 22: Métodos de diferenças finitas (continuação)	57
Aula 23: Métodos de projeção	59
23.1 Método de Galerkin	59
23.2 Método de colocação	60
Aula 24: Representação vetorial dos métodos de projeção	61
Aula 25: Funções polinomiais de Lagrange por partes	64
Aula 26: Interpolação polinomial por partes	66
Aula 27: Funções splines cúbicas	68
Aula 28: Splines cúbicas naturais	70
Aula 29: B-Splines cúbicas	72
Aula 30: Método de Colocação com B-Splines cúbicas	75
Apendice A: Erro de truncamento global	77
Referências	79

Aula 1: Métodos Numéricos para PVIs

O objetivo da primeira parte desta disciplina é aproximar a solução de **problemas de valor inicial (PVI)** na forma

$$\begin{cases} x'(t) = f(t, x(t)) \\ x(t_0) = x_0 \end{cases} \quad (1.1)$$

Em vez de obter a solução $x(t)$ de (1.1) para todo $t \geq t_0$, vamos fixar $T > 0$ e escolher N pontos adicionais $t_1, \dots, t_N$ tais que

$$t_0 < t_1 < \dots < t_N = T.$$

Estes pontos formam uma **partição** do intervalo $[t_0, T]$ (Fig. 1):

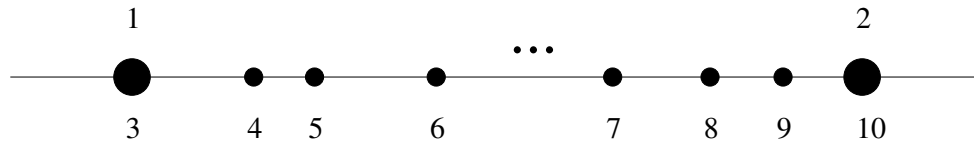


Figura 1: Partição do intervalo $[t_0, T]$.

Por simplicidade, vamos sempre considerar¹ que a partição é uniforme, ou seja,

$$t_{n+1} - t_n = h, \quad 0 \leq n \leq N-1.$$

Neste caso,

$$\begin{aligned} T - t_0 &= t_N - t_0 \\ &= (t_N - t_{N-1}) + t_{N-1} - t_0 = h + t_{N-1} - t_0 \\ &= h + (t_{N-1} - t_{N-2}) + t_{N-2} - t_0 = 2h + t_{N-1} - t_0 \\ &= 2h + (t_{N-2} - t_{N-3}) + t_{N-3} - t_0 = 3h + t_{N-3} - t_0 \\ &= \dots = Nh + t_{N-N} - t_0 = Nh, \end{aligned} \quad \implies \quad h = \frac{T - t_0}{N}.$$

Vamos estudar métodos numéricos que fornecem $X_0, X_1, \dots, X_N$ tais que

$$\begin{cases} X_n \approx x(t_n), & 1 \leq n \leq N \\ X_0 = x(t_0) = x_0 \end{cases}$$

A primeira classe de métodos que veremos para calcular estas aproximações é a dos métodos de passo simples.

1.1 Métodos de passo simples

Nos métodos de passo simples, as aproximações $X_n \approx x(t_n)$ são definidas do seguinte modo:

$$\begin{cases} X_{n+1} = \Phi(X_{n+1}, X_n, h), & n \geq 0 \\ X_0 = x_0 \end{cases} \quad (1.2)$$

Estes métodos podem ser classificados em

- Métodos explícitos, se Φ não depende de X_{n+1} ;
- Métodos implícitos, caso contrário.

¹Exceto quando tratarmos de métodos de Runge-Kutta adaptativos (Aula 9).

1.2 Método de Euler explícito

O método de Euler explícito é o mais básico dos métodos de passo simples, e pode ser deduzido de diversas maneiras. Veremos a seguir duas delas:

1.2.1 Dedução por meio do Teorema Fundamental do Cálculo

Vamos integrar a equação do PVI (1.1) de $t = t_n$ a $t = t_{n+1}$:

$$\int_{t_n}^{t_{n+1}} x'(t) dt = \int_{t_n}^{t_{n+1}} f(t, x(t)) dt.$$

Pelo Teorema Fundamental do Cálculo,

$$x(t_{n+1}) - x(t_n) = \int_{t_n}^{t_{n+1}} f(t, x(t)) dt.$$

Definindo a função composta $g(t) = f(t, x(t))$, podemos escrever a equação acima na forma

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} g(t) dt. \quad (1.3)$$

Se $h = t_{n+1} - t_n$ for suficientemente pequeno, podemos aproximar $g(t)$ em $[t_n, t_{n+1}]$ pela função constante $g_n(t) = g(t_n)$:

$$\int_{t_n}^{t_{n+1}} g(t) dt \approx \int_{t_n}^{t_{n+1}} g_n(t) dt = g(t_n) \int_{t_n}^{t_{n+1}} dt = g(t_n)(t_{n+1} - t_n) = g(t_n)h.$$

Assim, $x(t)$ em (1.3) pode ser aproximado da seguinte maneira:

$$x(t_{n+1}) \approx x(t_n) + g(t_n)h \approx x(t_n) + hf(t_n, x(t_n)). \quad (1.4)$$

Isto motiva propor $\Phi(X_{n+1}, X_n, h) = X_n + hf(t_n, X_n)$ em (1.2), ou seja,

$$\begin{cases} X_{n+1} = X_n + hf(t_n, X_n), & n \geq 0 \\ X_0 = x_0 \end{cases} \quad (1.5)$$

Estas equações definem o **método de Euler explícito**.

1.2.2 Dedução por meio da fórmula de Taylor

Se a solução $x = x(t)$ de (1.1) é uma função m vezes diferenciável, temos pela fórmula de Taylor que

$$x(t+h) = x(t) + hx'(t) + \frac{h^2}{2}x''(t) + \dots + \frac{h^n}{n!} \frac{d^n x}{dx^n}(t) + r_n(h), \quad \lim_{h \rightarrow 0} \frac{r_n(h)}{h^n} = 0.$$

Como $x(t)$ é ao menos uma vez diferenciável (lembre-se que $x'(t) = f(t, x(t))$), temos que

$$x(t+h) = x(t) + hx'(t) + r_1(h), \quad \lim_{h \rightarrow 0} \frac{r_1(h)}{h} = 0.$$

Em particular,

$$\begin{aligned} x(t_{n+1}) &= x(t_n + h) = x(t_n) + hx'(t_n) + r_1(h) \\ &= x(t_n) + hf(t_n, x(t_n)) + r_1(h) \\ &\approx x(t_n) + hf(t_n, x(t_n)), \end{aligned}$$

que é a mesma aproximação obtida em (1.4).

1.2.3 Exemplo

Exemplo 1.1 :: Faça duas iterações do método de Euler explícito para aproximar $x(1/2)$, sendo $x(t)$ a solução do PVI

$$\begin{cases} x'(t) = 2x(t) \\ x(0) = 1. \end{cases}$$

Como $f(t, x(t)) = 2x(t)$, segue que $f(t, x) = 2x$. Além disso, temos que

$$h = \frac{T - t_0}{2} = \frac{1/2 - 0}{2} = \frac{1}{4},$$

logo

$$t_0 = 0, \quad t_1 = h = \frac{1}{4}, \quad t_2 = 2h = \frac{1}{2}.$$

Segue de (1.5) que

$$\begin{aligned} X_0 &= x(0) = 1 \\ X_{n+1} &= X_n + hf(t_n, X_n) = X_n + \frac{1}{4}(2X_n) = X_n + \frac{X_n}{2} = \left(1 + \frac{1}{2}\right) X_n = \frac{3}{2}X_n, \end{aligned}$$

assim,

$$X_1 = \frac{3}{2}X_0 = \frac{3}{2}(1) = 3/2, \quad X_2 = \frac{3}{2}X_1 = \frac{3}{2}\left(\frac{3}{2}\right) = 9/4 = 2.25.$$

Para saber se $X_2 = 2.25$ é uma boa aproximação para $x(t_2) = x(1/2)$, note que a solução exata é $x(t) = e^{2t}$, de modo que

$$x\left(\frac{1}{2}\right) = e = 2.7183...$$

Assim, o erro é dado por $|X_2 - x(t_2)| = e - 2.25 \approx 0.47$.

Em geral, se aproximamos $x(1/2)$ com N passos do método de Euler explícito, teremos

$$\begin{aligned} h &= \frac{1/2 - 0}{N} = \frac{1}{2N} \\ X_0 &= 1 \\ X_{n+1} &= X_n + hf(t_n, X_n) = X_n + \frac{1}{2N}(2X_n) = X_n + \frac{1}{N}X_n = \left(1 + \frac{1}{N}\right) X_n, \end{aligned}$$

de modo que

$$X_N = \left(1 + \frac{1}{N}\right) X_{N-1} = \left(1 + \frac{1}{N}\right) \left[\left(1 + \frac{1}{N}\right) X_{N-2}\right] = \left(1 + \frac{1}{N}\right)^2 X_{N-2} = \dots = \left(1 + \frac{1}{N}\right)^N X_{N-N},$$

ou seja,

$$X_N = \left(1 + \frac{1}{N}\right)^N X_0 = \left(1 + \frac{1}{N}\right)^N (1) = \left(1 + \frac{1}{N}\right)^N.$$

Por exemplo se $N = 100$, então $X_N = 2.7048...$, com erro $|X_N - x(t_N)| \approx 0.013$. Além disso,

$$\lim_{N \rightarrow \infty} X_N = \lim_{N \rightarrow \infty} \left(1 + \frac{1}{N}\right)^N = e = x(1/2).$$

Aula 2: Existência e Unicidade de solução de PVI

Esta aula trata de aspectos teóricos das equações diferenciais ordinárias que influenciam na aproximação numérica destas equações. O exemplo a seguir ilustra a importância destes aspectos.

Exemplo 2.1 : Faça 20 iterações do método de Euler explícito (1.5) para aproximar $x(2)$, sendo $x(t)$ a solução do PVI

$$\begin{cases} x'(t) = 1 + x^2(t) \\ x(0) = 0 \end{cases} \quad (2.1)$$

Como $f(t, x(t)) = 1 + x^2(t)$, $f(t, x) = 1 + x^2$. Além disso, temos que $h = (T - t_0)/20 = (2 - 0)/20 = 1/10$. Assim, $X_0 = x(0) = 0$ e, para $n \geq 0$,

$$X_{n+1} = X_n + hf(t_n, X_n) = X_n + 0.1(1 + X_n^2).$$

Passemos às iterações:

$$X_1 = X_0 + (1/10)(1 + X_0^2) = 0 + 0.1(1 + 0^2) = 0.1$$

$$X_2 = X_1 + (1/10)(1 + X_1^2) = 0.1 + 0.1(1 + 0.1^2) = 0.1 + 0.1(1.01) = 0.1 + 0.101 = 0.201$$

$$X_3 = X_2 + (1/10)(1 + X_2^2) = 0.201 + 0.1(1 + 0.201^2) = 0.201 + 0.1(1.04041) = 0.3050401$$

$\vdots$

$$X_{20} = 925.949..., \quad \text{porém,}$$

$$\int \frac{dx}{1+x^2} = \int dt + C \implies x(t) = \tan(t + c),$$

e de $x(0) = 0$ segue que $x(t) = \tan(t)$, logo $x(2) = \tan(2) = -2.185...$

No exemplo anterior, o aumento do valor de N fez com que X_N se aproximasse mais de $x(t_N)$. Entretanto, se por exemplo dobrarmos o valor de N para aproximar o problema (2.1), encontraríamos

$$X_N = X_{40} \approx 3.203 \times 10^{25}.$$

O que aconteceu ?

Note pela Fig. 2 que há um ponto no intervalo $[t_0, T] = [0, 2]$ em que a solução é indefinida:

$$\nexists \lim_{t \rightarrow \pi/2} x(t).$$

A sequência gerada pelo método de Euler explícito foi incapaz de lidar com este ponto em que não há existência de solução. Especificamente, como $f(t, x) \geq 0$, a sequência X_n é crescente com $X_0 = 0$, de modo que $X_n \geq 0 \quad \forall n$, deixando de prever os valores negativos de $x(t)$ para $x > \pi/2$.

Este exemplo ilustra que não devemos aplicar o método de Euler explícito (e em geral, nenhum método numérico) para aproximar a solução de PVI em intervalos $[t_0, T]$ em que a existência e/ou unicidade de solução não for garantida. Para verificar a existência e unicidade de solução, usamos o seguinte teorema (Teoremas 1 e 2 de [10, Sec. 8.1]):

Teorema 2.1 : Se $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ é uma função contínua no conjunto

$$A = \{(t, x) \in \mathbb{R}^2 \mid |t - t_0| \leq \alpha \text{ e } |x - x_0| \leq \beta\}, \quad (2.2)$$

então a solução do PVI (1.1) existe para todo $|t - t_0| \leq \gamma$, sendo γ dado por

$$\gamma = \min\{\alpha, \beta/M\}, \quad M = \max_{(t,x) \in A} |f(t, x)|. \quad (2.3)$$

Além disso, se $f = f(t, x)$ é de classe C^1 em A , então a solução $x = x(t)$ é **única** no intervalo $|t - t_0| \leq \gamma$.

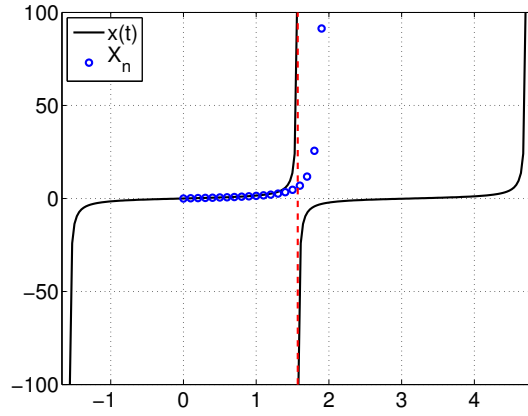


Figura 2: Solução exata e aproximada do Exemplo 2. A linha tracejada (em vermelho) corresponde ao ponto $x = \pi/2$, em que a solução não existe.

OBS: Note que $|t - t_0| \leq \gamma$ implica $t - t_0 \leq \gamma$, ou seja, $t \leq t_0 + \gamma$. Assim, o Teorema 2.1 garante a existência e unicidade em $[0, T]$ no caso em que $T = t_0 + \gamma$.

Exemplo 2.2 : Usando o Teorema 2.1, encontre o maior valor possível de T tal que o problema (2.1) tenha solução única no intervalo $[0, T]$.

Temos que $f(t, x) = 1 + x^2$ é de classe C^1 no conjunto A definido em (2.2) para quaisquer $\alpha, \beta \in \mathbb{R}$. Como o Teorema 2.1 garante a existência e unicidade em $[0, T]$ com $T = t_0 + \gamma = \gamma$, devemos escolher α, β tais que

$$\gamma = \min \left\{ \alpha, \frac{\beta}{\max_{(t,x) \in A} |1 + x^2|} \right\} = \min \left\{ \alpha, \frac{\beta}{\max_{|x| \leq \beta} 1 + x^2} \right\} = \min \left\{ \alpha, \frac{\beta}{1 + \beta^2} \right\}$$

seja o maior possível. Vamos primeiro escolher α . Note que, como $\gamma = \min \{ \alpha, \beta/(1 + \beta^2) \}$,

- Se $\alpha \geq \beta/(1 + \beta^2)$, então $\gamma = \beta/(1 + \beta^2)$;
- Se $\alpha < \beta/(1 + \beta^2)$, então $\gamma = \alpha < \beta/(1 + \beta^2)$,

ou seja (vide Fig. 3),

$$\gamma = \gamma(\alpha) = \begin{cases} \alpha, & \alpha \leq \beta/(1 + \beta^2), \\ \beta/(1 + \beta^2), & \alpha > \beta/(1 + \beta^2). \end{cases}$$

Portanto, devemos ter $\alpha \geq \beta/(1 + \beta^2)$, por exemplo $\alpha = \beta/(1 + \beta^2)$. Assim, o maior valor de γ é da forma

$$\gamma = \frac{\beta}{1 + \beta^2}.$$

Basta achar o valor de β que maximiza γ . Seja $g(x) = x/(1 + x^2)$. Temos que

$$g'(x) = \frac{(1)(1 + x^2) - (x)(2x)}{(1 + x^2)^2} = \frac{1 - x^2}{(1 + x^2)^2},$$

de modo que $g'(x) = 0$ se $x = \pm 1$. Como $\beta \geq 0$, temos que $\beta = 1$ é um candidato a máximo de $\gamma = g(\beta)$. Note que

- Se $\beta > 1$, então $g'(\beta) < 0$, ou seja, g é decrescente. Logo, $g(\beta) < g(1)$ para todo $\beta > 1$;
- Se $0 \leq \beta < 1$, então $g'(\beta) > 0$, ou seja, g é crescente. Logo, $g(1) > g(\beta)$ para todo $0 \leq \beta < 1$;

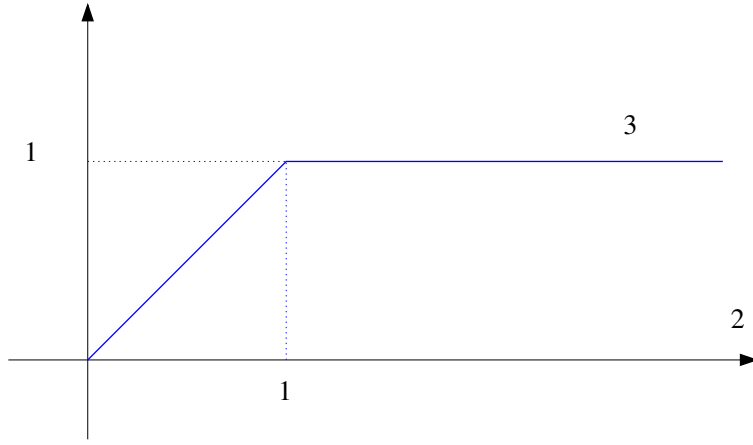


Figura 3: Relação entre os parâmetros γ e α .

Portanto, $\max_{\beta \geq 0} g(\beta) = g(1) = 1/(1 + 1^2) = 1/2$, ou seja, o maior valor possível de γ , segundo o Teorema 2.1, é $\gamma = 1/2$.

OBS: Neste exemplo conhecemos a solução exata ($x(t) = \tan(t)$), que é bem definida em $[0, T]$ para qualquer $T < \pi/2$. Assim, poderíamos por exemplo escolher $\gamma = 3/2 > 1/2$, para o qual a existência e unicidade ainda é válida. Isto enfatiza que o Teorema 2.1 fornece uma condição **suficiente**, mas não necessária.

O Teorema 3 de [10, Sec. 8.1] fornece outra condição suficiente para a existência e unicidade:

Teorema 2.2 : Se $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ é uma função contínua e tal que

$$|f(t, x_1) - f(t, x_2)| \leq L|x_1 - x_2| \quad \forall t \in [a, b], \quad \forall x_1, x_2 \in \mathbb{R}, \quad (2.4)$$

então a solução do PVI (1.1) existe e é única para todo $t \in [a, b]$.

Exemplo 2.3 : Estude, por meio do Teorema 2.2, a existência e unicidade de solução do problema

$$\begin{cases} x'(t) = (t + \sin(x(t)))^2 \\ x(0) = 3 \end{cases} \quad (2.5)$$

Para este problema, temos $f(t, x) = (t + \sin(x))^2$, logo

$$\begin{aligned} |f(t, x_1) - f(t, x_2)| &= |(t + \sin(x_1))^2 - (t + \sin(x_2))^2| \\ &= |(t + \sin(x_1) + t + \sin(x_2))(t + \sin(x_1) - t - \sin(x_2))| \\ &= |(2t + \sin(x_1) + \sin(x_2))(\sin(x_1) - \sin(x_2))| \\ &= |2t + \sin(x_1) + \sin(x_2)| |\cos(\xi)| |x_1 - x_2| \end{aligned}$$

para algum ξ entre x_1 e x_2 (pelo Teorema do Valor Médio). Como $|\cos(\xi)| \leq 1$ e

$$|2t + \sin(x_1) + \sin(x_2)| \leq |2t| + |\sin(x_1)| + |\sin(x_2)| \leq 2|t| + 2,$$

segue que, dado $T > 0$ fixo porém arbitrário,

$$|f(t, x_1) - f(t, x_2)| \leq L|x_1 - x_2| \quad \forall t \in [0, T],$$

sendo $L = 2T + 2$. Portanto, a solução do PVI (2.5) existe e é única para todo $t \in [0, T]$, para qualquer $T > 0$.

Aula 3: Polinômio de Taylor e ordem de convergência

Vamos revisar a aproximação de uma função f por polinômios de Taylor. Por simplicidade, vamos inicialmente assumir que f é infinitamente diferenciável.

Seja $\mathcal{P}_m$ o espaço vetorial dos polinômios de grau $\leq m$. Um caso particular do polinômio de Taylor de grau m é o polinômio de Maclaurin de grau m ,

$$p_m(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m \in \mathcal{P}_m,$$

que satisfaz as seguintes condições:

$$\left\{ \begin{array}{l} p_m(0) = f(0) \\ p'_m(0) = f'(0) \\ p''_m(0) = f''(0) \\ \vdots \\ p_m^{(m)}(0) = f^{(m)}(0), \quad f^{(m)}(x) = \frac{\partial^m f}{\partial x^m}(x). \end{array} \right.$$

Estas $m+1$ condições permitem determinar as $m+1$ incógnitas $a_0, a_1, \dots, a_m$ em $p_m(x)$:

$$\begin{aligned} p_m(x) &= a_0 + a_1x + a_2x^2 + \dots + a_mx^{m-1} \\ p_m(0) &= a_0 + 0 + 0 + \dots + 0 = a_0 \implies a_0 = f(0) \\ p'_m(x) &= (1)a_1 + 2a_2x + 3a_3x^2 + \dots + ma_mx^{m-1} \\ p'_m(0) &= a_1 + 0 + 0 + \dots + 0 = a_1 \implies a_1 = f'(0) \\ p''_m(x) &= (2)(1)a_2 + (3)(2)a_3x + (4)(3)a_4x^2 + \dots + (m)(m-1)a_mx^{m-2} \\ p''_m(0) &= 2!a_2 + 0 + 0 + \dots + 0 = 2!a_2 \implies a_2 = f''(0)/2! \\ p'''_m(x) &= (3)(2)(1)a_3 + (4)(3)(2)a_4x + (5)(4)(3)a_5x^2 + \dots + (m)(m-1)(m-2)a_mx^{m-3} \\ p'''_m(0) &= 3!a_3 + 0 + 0 + \dots + 0 = 3!a_3 \implies a_3 = f'''(0)/3! \\ &\vdots \\ p_m^{(m)}(0) &= m!a_m \implies a_m = f^{(m)}(0)/m! \end{aligned}$$

Portanto, $a_k = f^{(k)}(0)/k!$ ($0 \leq k \leq m$), ou seja,

$$p_m(x) = f(0) + f'(0)x + \frac{1}{2!}f''(0)x^2 + \dots + \frac{1}{m!}f^{(m)}(0)x^m = \sum_{k=0}^m \frac{f^{(k)}(0)}{k!}x^k.$$

O mesmo procedimento poderia ser utilizado para determinar o **polinômio de Taylor** de grau m de f em torno de $x = a$: o objetivo seria determinar $a_0, \dots, a_m$ tais que o polinômio

$$p_{m,a}(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m \in \mathcal{P}_m$$

satisfaça as condições

$$p_{m,a}^{(k)}(a) = f^{(k)}(a), \quad 0 \leq k \leq m.$$

Estas condições resultam no seguinte sistema de equações para os coeficientes do polinômio:

$$\left\{ \begin{array}{l} p_m(a) = a_0 + a_1a + a_2a^2 + \dots + a_ma^{m-1} = f(a) \\ p'_m(a) = (1)a_1 + 2a_2a + 3a_3a^2 + \dots + ma_ma^{m-1} = f'(a) \\ p''_m(a) = 2!a_2 + (3)(2)a_3a + (4)(3)a_4a^2 + \dots + (m)(m-1)a_ma^{m-2} = f''(a) \\ p'''_m(a) = 3!a_3 + (4)(3)(2)a_4a + \dots + (m)(m-1)(m-2)a_ma^{m-3} = f'''(a) \\ \vdots \\ p_m^{(m)}(a) = m!a_m = f^{(m)}(a), \end{array} \right.$$

ou na forma matricial,

$$\begin{bmatrix} 1 & a & a^2 & a^3 & \dots & a^{m-1} \\ 0 & 1! & 2a & 3a_3a^2 & \dots & \frac{m!}{(m-1)!}a^{m-1} \\ 0 & 0 & 2! & \frac{3!}{1!}a_3a & \frac{4!}{2!}a^2 & \dots & \frac{m!}{(m-2)!}a^{m-2} \\ 0 & 0 & 0 & 3! & \frac{4!}{1!}a & \dots & \frac{m!}{(m-3)!}a^{m-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & (m-1)! & \frac{m!}{1!}a \\ 0 & 0 & 0 & 0 & \dots & 0 & m! \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{m-1} \\ a_m \end{bmatrix} = \begin{bmatrix} f(a) \\ f'(a) \\ f''(a) \\ f'''(a) \\ \vdots \\ f^{(m-1)}(a) \\ f^{(m)}(a) \end{bmatrix}$$

Embora seja um sistema triangular (portanto fácil de ser resolvido), as expressões dos coeficientes a_k vão se tornando mais complicadas à medida que k diminui:

$$\begin{aligned} a_m &= f^{(m)}(a)/m! \\ (m-1)!a_{m-1} + m!aa_m &= f^{(m-1)}(a) \implies a_{m-1} = f^{(m-1)}(a)/(m-1)! - af^{(m)}(a)/(m-1)! \\ &\vdots \end{aligned}$$

Para simplificar os coeficientes, vamos escrever $p_{m,a}(x)$ em termos de outra base:

$$p_{m,a}(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + \dots + a_m(x-a)^m.$$

Neste caso,

$$\begin{aligned} p_m(x) &= a_0 + a_1(x-a) + a_2(x-a)^2 + \dots + a_m(x-a)^{m-1} \\ p_m(a) &= a_0 + 0 + 0 + \dots + 0 = a_0 \implies a_0 = f(a) \\ p'_m(x) &= (1)a_1 + 2a_2(x-a) + 3a_3(x-a)^2 + \dots + ma_m(x-a)^{m-1} \\ p'_m(a) &= a_1 + 0 + 0 + \dots + 0 = a_1 \implies a_1 = f'(a) \\ p''_m(x) &= (2)(1)a_2 + (3)(2)a_3(x-a) + (4)(3)a_4(x-a)^2 + \dots + (m)(m-1)a_m(x-a)^{m-2} \\ p''_m(a) &= 2!a_2 + 0 + 0 + \dots + 0 = 2!a_2 \implies a_2 = f''(a)/2! \\ &\vdots \\ p^{(m)}_m(a) &= m!a_m \implies a_m = f^{(m)}(a)/m!, \end{aligned}$$

o que resulta numa expressão análoga à do polinômio de Maclaurin:

$$p_{m,a}(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(m)}(a)}{m!}(x-a)^m = \sum_{k=0}^m \frac{f^{(k)}(a)}{k!}(x-a)^k. \quad (3.1)$$

As condições $p_{m,a}^{(k)}(a) = f^{(k)}(a)$ para todo $0 \leq k \leq m$ fazem com que o polinômio de Taylor $p_{m,a}^{(k)}(x)$ seja o polinômio de grau m que melhor aproxima $f(x)$ na vizinhança de $x = a$ (Fig. 4).

O erro da aproximação de $f(x)$ por $p_{m,a}^{(k)}(x)$ é descrito pela fórmula de Taylor com resto de Lagrange:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(m)}(a)}{m!}(x-a)^m + \frac{f^{(m+1)}(\xi)}{(m+1)!}(x-a)^{m+1}, \quad (3.2)$$

sendo ξ entre a e x . Esta fórmula é válida no caso em que f é uma função de classe C^m e $m+1$ vezes diferenciável. Por outro lado, na situação (menos restritiva) em que f é uma função m vezes diferenciável, vale a seguinte fórmula de Taylor:

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2 + \dots + \frac{1}{m!}f^{(m)}(a)(x-a)^m + r_m(h), \quad \lim_{h \rightarrow 0} \frac{r_m(h)}{h^m} = 0,$$

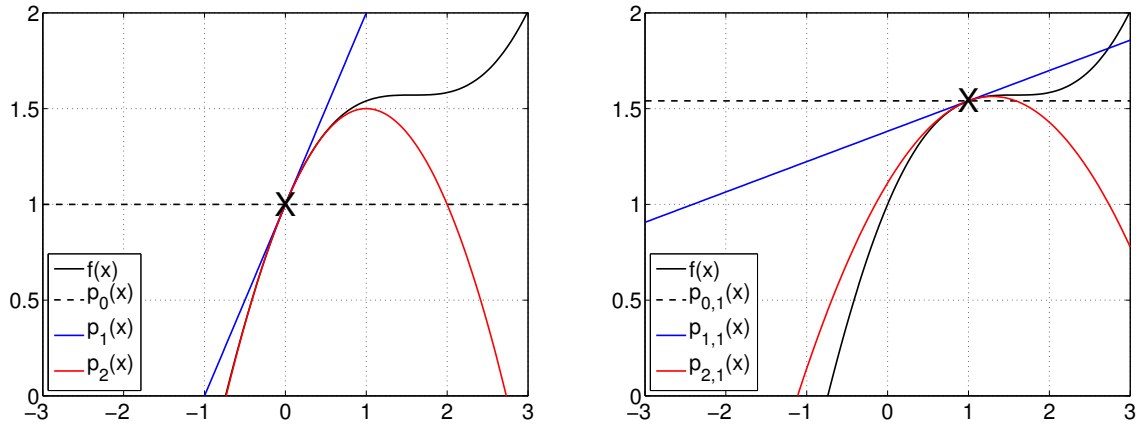


Figura 4: Aproximação de $f(x) = \cos(x) + x$ pelos polinômios de Maclaurin (esquerda) e de Taylor com $a = 1$ (direita).

3.1 Ordem de convergência

Segundo [10, Sec. 1.2], dadas duas funções f e g , dizemos que

$$f(x) = \mathcal{O}(g(x)) \quad (x \rightarrow x^*)$$

se

$$\exists C > 0, \epsilon > 0 \quad | \quad |f(x)| \leq C|g(x)| \quad \forall x, |x - x^*| \leq \epsilon.$$

Daqui em diante, com a expressão $f(x) = \mathcal{O}(g(x))$ (sem referência a x^*) subentende-se que $x^* = 0$, ou seja, $f(x) = \mathcal{O}(g(x))$ se

$$\exists C > 0, \epsilon > 0 \quad | \quad |f(x)| \leq C|g(x)| \quad \forall x, |x| \leq \epsilon. \quad (3.3)$$

Exemplo 3.1 : Mostre que se $f(x) = x^4 + 3x^2 + x + 2x^5$, então $f(x) = \mathcal{O}(x)$.

Devemos encontrar $C > 0$ e $\epsilon > 0$ tais que $|f(x)| \leq C|x| \quad \forall x, |x| \leq \epsilon$. Note que

$$|f(x)| = |x^4 + 3x^2 + x + 2x^5| \leq |x|^4 + 3|x|^2 + |x| + 2|x|^5,$$

logo basta encontrar $\epsilon > 0$ tal que $|x|^n \leq |x|$ para todo $|x| \leq \epsilon$ e $n \geq 1$. Como sabemos das potências de números reais que, dados $n > m$,

$$b^n > b^m > 1, \quad b > 1 \quad \text{e} \quad b^n < b^m < 1, \quad 0 \leq b < 1,$$

podemos escolher $\epsilon = 1$. Neste caso,

$$|f(x)| \leq |x| + 3|x| + |x| + 2|x| = 7|x|$$

Portanto, (3.3) vale com $C = 7$, $\epsilon = 1$, $g(x) = x$, ou seja, $f(x) = \mathcal{O}(x)$. Note que, em geral, $C \geq 7$ e $\epsilon \leq 1$.

OBS: Pela fórmula de Taylor com resto de Lagrange (3.2), o erro de aproximação do polinômio de Maclaurin satisfaz $f(x) - p_m(x) = \mathcal{O}(x^{m+1})$ no caso em que f é de classe C^{m+1} . De fato, segue do teorema de Weierstrass [12] que

$$\exists M > 0 \quad | \quad |f^{(m+1)}(x)| \leq M \quad \forall x \in \mathbb{R},$$

de modo que

$$|f(x) - p_m(x)| = |f(x) - p_{m,0}(x)| = \left| \frac{1}{(m+1)!} f^{(m+1)}(\xi)(x-0)^{m+1} \right| \leq \frac{M}{(m+1)!} |x^{m+1}|.$$

ou seja, (3.3) vale com $C = M/(m+1)!$, $g(x) = x^{m+1}$ e ϵ qualquer.

Aula 4: Método da série de Taylor

O polinômio de Taylor de grau m da solução $x(t)$ do PVI (1.1) em torno do ponto $a = t_n$ é dada por

$$p_{m,t_n}(t) = x(t_n) + x'(t_n)(t - t_n) + \frac{1}{2!}x''(t_n)(t - t_n)^2 + \dots + \frac{1}{m!}x^{(m)}(t_n)(t - t_n)^m.$$

Observando que $x'(t) = f(t, x(t))$ e tomando em particular $t = t_{n+1}$, obtemos

$$p_{m,t_n}(t_{n+1}) = x(t_n) + f(t_n, x(t_n))h + \frac{1}{2!}f'(t_n, x(t_n))h^2 + \dots + \frac{1}{m!}f^{(m-1)}(t_n, x(t_n))h^m.$$

Note que $p_{m,t_n}(t_{n+1}) \approx x(t_{n+1})$. Dada a aproximação $X_n \approx x(t_n)$ no instante t_n , consideramos a seguinte aproximação de $x(t_{n+1})$:

$$\begin{aligned} x(t_{n+1}) &\approx p_{m,t_n}(t_{n+1}) = x(t_n) + f(t_n, x(t_n))h + \frac{f'(t_n, x(t_n))}{2!}h^2 + \dots + \frac{1}{m!}f^{(m-1)}(t_n, x(t_n))h^m \\ &\approx X_n + f(t_n, X_n)h + \frac{1}{2!}f'(t_n, X_n)h^2 + \dots + \frac{1}{m!}f^{(m-1)}(t_n, X_n)h^m. \end{aligned}$$

Esta aproximação define o **método da série de Taylor de ordem m**:

$$\begin{cases} X_{n+1} = X_n + f(t_n, X_n)h + \frac{1}{2!}f'(t_n, X_n)h^2 + \dots + \frac{1}{m!}f^{(m-1)}(t_n, X_n)h^m \\ X_0 = x_0 \end{cases} \quad (4.1)$$

Exemplo 4.1 : Deduza a equação do método da série de Taylor de terceira ordem para a equação

$$x'(t) = \cos(t) - \sin(x(t)) + t^2$$

Observando a regra de derivação de funções compostas, obtemos

$$\begin{aligned} f(t, x(t)) &= \cos(t) - \sin(x(t)) + t^2 \\ f'(t, x(t)) &= -\sin(t) - (\sin(x(t)))' + 2t \\ &= -\sin(t) - \cos(x(t))x'(t) + 2t \\ &= -\sin(t) - \cos(x(t))f(t, x(t)) + 2t \\ f''(t, x(t)) &= -\cos(t) - (\cos(x(t))f(t, x(t)))' + 2 \\ &= -\cos(t) - (\cos(x(t)))'f(t, x(t)) - \cos(x(t))f'(t, x(t)) + 2 \\ &= -\cos(t) + \sin(x(t))x'(t)f(t, x(t)) - \cos(x(t))f'(t, x(t)) + 2 \\ &= -\cos(t) + \sin(x(t))f(t, x(t))^2 - \cos(x(t))f'(t, x(t)) + 2. \end{aligned}$$

Note acima que expressamos $f^{(k)}(t_n, x(t))$ em termos das derivadas de ordem inferior (e em termos de $f(t, x(t))$). Isto permitirá escrever estes termos de maneira recursiva.

O próximo passo é avaliar $(t, x(t))$ em (t_n, X_n) ,

$$\begin{cases} f(t_n, X_n) = \cos(t_n) - \sin(X_n) + t_n^2 \\ f'(t_n, X_n) = -\sin(t_n) - \cos(X_n)f(t_n, X_n) + 2t_n \\ f''(t_n, X_n) = -\cos(t_n) + \sin(X_n)f(t_n, X_n)^2 - \cos(X_n)f'(t_n, X_n) + 2, \end{cases}$$

de modo que a equação do método é dada por

$$\begin{aligned} X_{n+1} &= X_n + f(t_n, X_n)h + \frac{f'(t_n, X_n)}{2!}h^2 + \frac{f''(t_n, X_n)}{3!}h^3 \\ &= X_n + (\cos(t_n) - \sin(X_n) + t_n^2)h + \frac{-\sin(t_n) - \cos(X_n)f(t_n, X_n) + 2t_n}{2!}h^2 \\ &\quad + \frac{-\cos(t_n) + \sin(X_n)f(t_n, X_n)^2 - \cos(X_n)f'(t_n, X_n) + 2}{3!}h^3. \end{aligned}$$

Por clareza, e para utilizar a relação recursiva entre as derivadas, convém escrever esta equação com o auxílio de passos intermediários:

$$\begin{cases} F_1 &= \cos(t_n) - \sin(X_n) + t_n^2 \\ F_2 &= -\sin(t_n) - \cos(X_n)F_1 + 2t_n \\ F_3 &= -\cos(t_n) + \sin(X_n)F_1^2 - \cos(X_n)F_2 + 2 \\ X_{n+1} &= X_n + F_1h + \frac{1}{2!}F_2h^2 + \frac{1}{3!}F_3h^3. \end{cases}$$

Exemplo 4.2 : Deduza a equação do método da série de Taylor de terceira ordem para a equação

$$x'(t) = 1 + x(t)^2$$

Analogamente ao exemplo anterior,

$$\begin{aligned} f(t, x(t)) &= 1 + x(t)^2 \\ f'(t, x(t)) &= 0 + 2x(t)x'(t) = 2x(t)f(t, x(t)) \\ f''(t, x(t)) &= 2x'(t)f(t, x(t)) + 2x(t)f'(t, x(t)) = 2f(t, x(t))^2 + 2x(t)f'(t, x(t)). \end{aligned}$$

Assim,

$$\begin{cases} f(t_n, X_n) &= 1 + X_n^2 \\ f'(t_n, X_n) &= 0 + 2X_nf(t_n, X_n) \\ f'(t_n, X_n)' &= 2f(t_n, X_n)^2 + 2X_nf'(t_n, X_n), \end{cases}$$

de obtemos o conjunto de equações

$$\begin{cases} F_1 &= 1 + X_n^2 \\ F_2 &= 2X_nF_1 \\ F_3 &= 2F_1^2 + 2X_nF_2 \\ X_{n+1} &= X_n + F_1h + \frac{1}{2!}F_2h^2 + \frac{1}{3!}F_3h^3. \end{cases}$$

Exemplo 4.3 : Aplique dois passos do método da série de Taylor de segunda ordem para aproximar $x(0.002)$, sendo $x(t)$ a solução do PVI (2.1).

Temos do exemplo anterior que o método de segunda ordem é dado por

$$\begin{cases} F_1 &= 1 + X_n^2 \\ F_2 &= 2X_nF_1 \\ X_{n+1} &= X_n + F_1h + \frac{1}{2!}F_2h^2. \end{cases}$$

Como $X_0 = 0$, temos no primeiro passo que $F_1 = 1$ e $F_2 = 0$, logo $X_1 = h$. No segundo passo, temos $F_1 = 1 + h^2$ e $F_2 = 2hF_1 = 2h(1 + h^2)$, logo

$$X_2 = h + (1 + h^2)h + \frac{1}{2}2h(1 + h^2)h^2 = 2h + 2h^3 + h^5.$$

Substituindo $h = (0.02 - 0)/2 = 0.01$, encontramos $X_2 = 0.020002001$. Avaliando a solução exata $x(t) = \tan(t)$, obtemos $x(0.02) = \tan(0.02) = 0.020002667093 \dots$

OBS: O método de primeira ordem (Euler explícito) $X_{n+1} = X_n + F_1h$ resultaria em $X_1 = h$ e

$$X_2 = h + (1 + h^2)h = 2h + h^3 = 0.020001.$$

Aula 5: Análise de erros do método da série de Taylor

Vamos estudar o comportamento do erro de aproximação

$$e_n = x(t_n) - X_n.$$

O primeiro passo é decompor e_n na forma

$$\begin{aligned} e_n &= x(t_n) - \tilde{x}(t_n, t_{n-1}, X_{n-1}) + \tilde{x}(t_n, t_{n-1}, X_{n-1}) - X_n \\ e_n &= [x(t_n) - \tilde{x}(t_n)] + [\tilde{x}(t_n) - X_n] = \tilde{e}_n + e_n^{loc}, \end{aligned}$$

sendo $\tilde{x}(t) = \tilde{x}(t, t_{n-1}, X_{n-1})$ a solução do PVI auxiliar

$$\begin{cases} \tilde{x}'(t) = f(t, \tilde{x}(t)) \\ \tilde{x}(t_{n-1}) = X_{n-1}. \end{cases} \quad (5.1)$$

Obtemos no Apêndice A que, se

$$\left| \frac{\partial f}{\partial x}(t, x) \right| \leq M \quad \forall (t, x) \in [t_0, T] \times \mathbb{R}, \quad (5.2)$$

então o erro total e_n satisfaz

$$|e_n| \leq \delta \frac{e^{M(T-t_0)} - 1}{hM}, \quad \delta = \max_{0 \leq j \leq n} |e_n^{loc}|. \quad (5.3)$$

OBS: A condição (5.2) garante a validade do Teorema 2.2, de modo que a solução $x(t)$ do PVI (1.1) existe e é única para qualquer $(t_0, x_0) \in [t_0, T] \times \mathbb{R}$.

Mostraremos a seguir que $e_n^{loc} = \mathcal{O}(h^{m+1})$, de modo que $\delta \leq Ch^{m+1}$ e assim segue de (5.3) que

$$|e_n| \leq Ch^{m+1} \frac{e^{M(T-t_0)} - 1}{hM} = \frac{C}{M} (e^{M(T-t_0)} - 1) h^m = \tilde{C} h^m,$$

ou seja, $e_n = \mathcal{O}(h^m)$. Considere, por exemplo, $m = 3$. Se o espaçamento da malha h decresce de $h = 10^{-1}$ para $h = 10^{-2}$ (uma casa decimal), então o erro, que era inferior a $\tilde{C}(10^{-1})^3 = 10^{-3}\tilde{C}$, torna-se inferior a $10^{-6}\tilde{C}$ (decai três casas decimais). Além disso, como $m \geq 1$, temos em particular que

$$\lim_{h \rightarrow 0} e_n = 0 \quad (0 \leq n \leq N),$$

ou seja, o método é convergente.

5.1 Erro local

O erro e_n^{loc} é relacionado ao erro cometido na passagem de X_{n-1} para X_n assumindo que **não houve erros** nos passos anteriores (ou seja, se $X_j = x(t_j)$ para todo $j < n$). Se isso fosse verdade, teríamos em particular $X_{n-1} = x(t_{n-1})$ e assim $\tilde{x}(t)$ torna-se a solução do PVI

$$\begin{cases} \tilde{x}'(t) = f(t, \tilde{x}(t)) \\ \tilde{x}(t_{n-1}) = x(t_{n-1}). \end{cases} \quad (5.4)$$

Como $x(t)$ trivialmente satisfaz o PVI (5.4) teríamos, pela existência e unicidade de solução que $\tilde{x}(t) = x(t)$, ou seja, $e_n = e_n^{loc}$. Por este motivo a parcela e_n^{loc} do erro e_n recebe a seguinte denominação:

Definição 5.1 : Seja $\tilde{x}(t) = \tilde{x}(t, t_{n-1}, X_{n-1})$ a solução do PVI auxiliar (5.1). Denominamos a expressão

$$e_n^{loc} = \tilde{x}(t_n, t_{n-1}, X_{n-1}) - X_n$$

o **erro local** da aproximação $X_n = x(t_n)$.

Lema 5.1 : A expressão $e_n^{loc} = \tilde{x}(t_n, t_{n-1}, X_{n-1}) - X_n$ satisfaz $e_n^{loc} = \mathcal{O}(h^{m+1})$ para todo n .

Note que, pela equação que define o método,

$$X_n = X_{n-1} + f(t_n, X_{n-1})h + \frac{f'(t_n, X_{n-1})}{2!}h^2 + \dots + \frac{f^{(m-1)}(t_n, X_{n-1})}{m!}h^m. \quad (5.5)$$

Por outro lado, pela fórmula de Taylor com resto de Lagrange para $\tilde{x}(t)$,

$$\tilde{x}(t) = \tilde{x}(t_{n-1}) + \tilde{x}'(t_{n-1})(t - t_{n-1}) + \dots + \frac{\tilde{x}^{(m)}(t_{n-1})}{m!}(t - t_{n-1})^m + \frac{\tilde{x}^{(m+1)}(\bar{t})}{(m+1)!}(t - t_{n-1})^{m+1},$$

com $\bar{t}$ entre t_{n-1} e t . Em particular se $t = t_n$,

$$\tilde{x}(t_n) = \tilde{x}(t_{n-1}) + \tilde{x}'(t_{n-1})h + \dots + \frac{\tilde{x}^{(m)}(t_{n-1})}{m!}h^m + \frac{\tilde{x}^{(m+1)}(\bar{t})}{(m+1)!}(\bar{t})h^{m+1} \quad (t_{n-1} \leq \bar{t} \leq t_n).$$

Usando o fato que $\tilde{x}' = f$,

$$\tilde{x}(t_n) = \tilde{x}(t_{n-1}) + f(t_{n-1}, \tilde{x}(t_{n-1}))h + \dots + \frac{f^{(m-1)}(t_{n-1}, \tilde{x}(t_{n-1}))}{m!}h^m + \frac{f^{(m-1)}(\bar{t}, \tilde{x}(\bar{t}))}{(m+1)!}h^{m+1}.$$

Finalmente, pela condição inicial de (5.1),

$$\tilde{x}(t_n) = X_{n-1} + f(t_{n-1}, X_{n-1})h + \dots + \frac{f^{(m-1)}(t_{n-1}, X_{n-1})}{m!}h^m + \frac{f^{(m-1)}(\bar{t}, \tilde{x}(\bar{t}))}{(m+1)!}h^{m+1}. \quad (5.6)$$

Segue de (5.5) e (5.6) que

$$\tilde{x}(t_n) = X_n + g(h), \quad g(h) = \frac{f^{(m)}(\bar{t}, \tilde{x}(\bar{t}))}{(m+1)!}h^{m+1},$$

ou seja, $e_n^{loc} = g(h)$. Assumindo adicionalmente que

$$\exists L > 0 \quad | | f^{(m)}(t, x) | | \leq L \quad \forall (t, x) \in [t_0, T] \times \mathbb{R}, \quad (5.7)$$

segue que, para todo $n \geq 0$ e $|h| = (T - t_0)/N \leq T - t_0$,

$$|g(h)| \leq Ch^{m+1}, \quad C = \frac{L}{(m+1)!}$$

ou seja, $e_n^{loc} = g(h) = \mathcal{O}(h^{m+1})$. □

Aula 6: Métodos de Runge-Kutta

O defeito do método da série de Taylor é depender das derivadas da função f . Isto dificulta uma implementação computacional universal para o método (ou seja, o algoritmo escrito para um PVI em geral não pode ser aproveitado em outro PVI).

Os métodos de Runge-Kutta de ordem m (RK m) são métodos de passo simples com erro local de ordem $\mathcal{O}(h^{m+1})$, como o método da série de Taylor de ordem m , mas sem envolver derivadas da função f .

Se $f_1(x) = \mathcal{O}(g(x))$, $f_2(x) = \mathcal{O}(g(x))$ e $f = f_1 \pm f_2$ então $f(x) = \mathcal{O}(g(x))$, ou seja,

$$\mathcal{O}(g(x)) \pm \mathcal{O}(g(x)) = \mathcal{O}(g(x)).$$

Tendo em vista esta propriedade, vamos construir os métodos de RK m na forma

$$X_{n+1} = X_n^{Tm} + \mathcal{O}(h^{m+1}), \quad (6.1)$$

sendo X_n^{Tm} a sequência de aproximações gerada pelo método da série de Taylor de ordem m . Deste modo o erro local dos métodos na forma (6.1) é dado por

$$\begin{aligned} e_n^{loc} &= \tilde{x}(t_n, t_{n-1}, X_{n-1}) - X_n \\ &= \tilde{x}(t_n, t_{n-1}, X_{n-1}) - X_n^{Tm} - \mathcal{O}(h^{m+1}) \\ &= \mathcal{O}(h^{m+1}) - \mathcal{O}(h^{m+1}) = \mathcal{O}(h^{m+1}). \end{aligned}$$

6.1 Expansão de Taylor 2D

Vamos utilizar a expansão de Taylor bidimensional na dedução dos métodos de Runge-Kutta de ordem 1 e 2. Esta expansão, por sua vez, será deduzida a partir das expansões de Taylor unidimensionais de ordem 0 e 1 em torno de $a = 0$:

$$g(x) = g(0) + g'(\bar{x})x, \quad (6.2)$$

$$g(x) = g(0) + g'(0)x + \frac{1}{2}g''(\bar{x})x^2, \quad (6.3)$$

sendo que, para $x > 0$, temos $0 \leq \bar{x} \leq x$ (e o valor de $\bar{x}$ não necessariamente é o mesmo em (6.2) e (6.3)). Para definir a expansão de Taylor de uma função $f = f(t, x)$ em torno de (a, b) , definimos

$$g(s) = f(a + s v_1, b + s v_2), \quad \mathbf{v} = (v_1, v_2) = \frac{1}{s^*}(t - a, x - b), \quad s^* = \|(t - a, x - b)\|. \quad (6.4)$$

Esta função representa um perfil da função f sobre a reta que passa por (a, b) e (t, x) (Fig. 5).

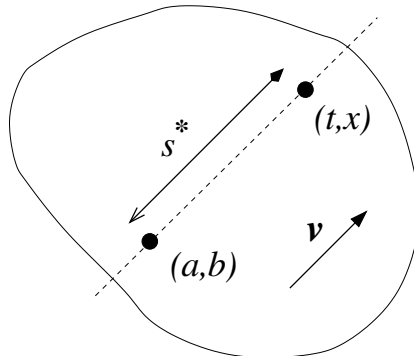


Figura 5: Reta sobre a qual é tomada o perfil de $f(t, x)$ para a dedução da expansão de Taylor 2D.

Segue da definição da função auxiliar $g(s)$ em (6.4) que $g(0) = f(a, b)$ e $g(s^*) = f(t, x)$. Além disso, sua derivada é dada por

$$\begin{aligned} g'(s) &= \frac{\partial f}{\partial t}(a + s v_1, b + s v_2)(a + s v_1)' + \frac{\partial f}{\partial x}(a + s v_1, b + s v_2)(b + s v_2)' \\ &= \frac{\partial f}{\partial t}(a + s v_1, b + s v_2)v_1 + \frac{\partial f}{\partial x}(a + s v_1, b + s v_2)v_2. \end{aligned} \quad (6.5)$$

Tomando $x = s^*$ em (6.2) e observando que $s^*v = (t - a, x - b)$, obtemos

$$\begin{aligned} f(t, x) &= f(a, b) + \frac{\partial f}{\partial t}(a + \bar{s} v_1, b + \bar{s} v_2)v_1 s^* + \frac{\partial f}{\partial x}(a + \bar{s} v_1, b + \bar{s} v_2)v_2 s^* \\ &= f(a, b) + \frac{\partial f}{\partial t}(\bar{a}, \bar{b})(t - a) + \frac{\partial f}{\partial x}(\bar{a}, \bar{b})(x - b), \end{aligned} \quad (6.6)$$

com $(\bar{a}, \bar{b}) = (a + \bar{s} v_1, b + \bar{s} v_2)$ para algum $0 \leq \bar{s} \leq s^*$. Normalizando $\bar{s}$ por meio de $\lambda = \bar{s}/s^*$, temos

$$(\bar{a}, \bar{b}) = (a + \lambda(t - a), b + \lambda(x - b)), \quad 0 \leq \lambda \leq 1.$$

Esta é a expansão de ordem zero com resto de Lagrange. Para obter a expansão de primeira ordem, vamos utilizar a equação (6.3), que por sua vez depende da derivada de (6.5):

$$\begin{aligned} g''(s) &= \left(\frac{\partial^2 f}{\partial t^2}(a + s v_1, b + s v_2)v_1 + \frac{\partial^2 f}{\partial t \partial x}(a + s v_1, b + s v_2)v_2 \right) v_1 \\ &\quad + \left(\frac{\partial^2 f}{\partial x \partial t}(a + s v_1, b + s v_2)v_1 + \frac{\partial^2 f}{\partial x^2}(a + s v_1, b + s v_2)v_2 \right) v_2 \\ &= \frac{\partial^2 f}{\partial t^2}(a + s v_1, b + s v_2)v_1^2 + 2 \frac{\partial^2 f}{\partial t \partial x}(a + s v_1, b + s v_2)v_1 v_2 + \frac{\partial^2 f}{\partial x^2}(a + s v_1, b + s v_2)v_2^2, \end{aligned} \quad (6.7)$$

assumindo-se que f é de classe C^2 . Segue de (6.3), (6.4) com $s = 0$ e (6.5) com $s = \bar{s}$ que

$$\begin{aligned} f(t, x) &= f(a, b) + \frac{\partial f}{\partial t}(a, b)(t - a) + \frac{\partial f}{\partial x}(a, b)(x - b) \\ &\quad + \frac{1}{2} \left(\frac{\partial^2 f}{\partial t^2}(\bar{a}, \bar{b})(t - a)^2 + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{a}, \bar{b})(t - a)(x - b) + \frac{\partial^2 f}{\partial x^2}(\bar{a}, \bar{b})(x - b)^2 \right), \end{aligned} \quad (6.8)$$

em que, novamente, $(\bar{a}, \bar{b}) = (a + \lambda(t - a), b + \lambda(x - b))$ para algum $0 \leq \lambda \leq 1$. As expansões (6.6) e (6.8) serão usadas nos casos particulares em que $(a, b) = (t_n, X_n)$ e $(t, x) = (t_n + h\alpha, X_n + h\beta)$, que resultam em

$$f(t_n + h\alpha, X_n + h\beta) = f(t_n, X_n) + h \left(\frac{\partial f}{\partial t}(\bar{t}, \bar{X})\alpha + \frac{\partial f}{\partial x}(\bar{t}, \bar{X})\beta \right), \quad (6.9)$$

$$\begin{aligned} f(t_n + h\alpha, X_n + h\beta) &= f(t_n, X_n) + h \left(\frac{\partial f}{\partial t}(t_n, X_n)\alpha + \frac{\partial f}{\partial x}(t_n, X_n)\beta \right) \\ &\quad + \frac{h^2}{2} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X})\alpha^2 + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X})\alpha\beta + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X})\beta^2 \right), \end{aligned} \quad (6.10)$$

com $(\bar{t}, \bar{X}) = (t_n + \lambda h\alpha, X_n + \lambda h\beta)$ para algum $0 \leq \lambda \leq 1$.

6.2 Métodos de Runge-Kutta de primeira ordem (RK1)

A princípio, não há motivo para modificar o método da série de Taylor de ordem 1,

$$X_{n+1}^{T1} = X_n + hf(t_n, X_n),$$

que corresponde ao método de Euler explícito, pois este método não envolve derivadas da função f . Assim, a escolha mais natural para (6.1) é dada por

$$X_{n+1} = X_{n+1}^{T1} = X_n + hf(t_n, X_n),$$

para o qual $X_{n+1} = X_{n+1}^{T1} + 0 = X_{n+1}^{T1} + \mathcal{O}(h^{1+1})$. Isso faz com que o método de Euler explícito seja também um caso particular da classe de métodos RK1. Outra possibilidade de escolha para (6.1) é

$$X_{n+1} = X_n + hf(t_{n+1}, X_{n+1}) = X_{n+1}^{T1} + h[f(t_{n+1}, X_{n+1}) - f(t_n, X_n)]. \quad (6.11)$$

Segue que, se $\alpha = 1$ e $\beta = f(t_{n+1}, X_{n+1})$ (ou seja, $t = t_{n+1}$ e $x = X_{n+1}$) em (6.9), então

$$f(t_{n+1}, X_{n+1}) = f(t_n, X_n) + h \left(\frac{\partial f}{\partial t}(\bar{t}, \bar{X}) + \frac{\partial f}{\partial x}(\bar{t}, \bar{X})f(t_{n+1}, X_{n+1}) \right),$$

com $(\bar{t}, \bar{X}) = (t_n + \lambda h, X_n + \lambda hf(t_{n+1}, X_{n+1})) = \lambda(t_n, X_n) + (1 - \lambda)(t_{n+1}, X_{n+1})$ para algum $0 \leq \lambda \leq 1$. Se $f \in C^1$, então segue do Teorema de Weierstrass que

$$\exists M > 0 \quad \text{tal que} \quad \begin{cases} |f(t, x)| \leq M \\ \left| \frac{\partial f}{\partial t}(t, x) \right| \leq M \\ \left| \frac{\partial f}{\partial x}(t, x) \right| \leq M \end{cases} \quad \forall (t, x) \in [t_n, t_{n+1}] \times [X_n, X_{n+1}],$$

logo segue de (6.11) que $X_{n+1} - X_{n+1}^{T1} = h[f(t_{n+1}, X_{n+1}) - f(t_n, X_n)]$ satisfaz

$$|h[f(t_{n+1}, X_{n+1}) - f(t_n, X_n)]| \leq h^2 \left| \frac{\partial f}{\partial t}(\bar{t}, \bar{X}) + \frac{\partial f}{\partial x}(\bar{t}, \bar{X})f(t_{n+1}, X_{n+1}) \right| \leq h^2(M + M^2),$$

ou seja, $X_{n+1} = X_{n+1}^{T1} + \mathcal{O}(h^2)$. O método (6.11) é denominado **método de Euler implícito**.

Aula 7: Métodos de Runge-Kutta de segunda ordem (RK2)

Vamos deduzir um dos métodos RK2 a partir do método da série de Taylor de segunda ordem

$$X_{n+1}^{T2} = X_n + hf(t_n, X_n) + \frac{h^2}{2} f'(t_n, X_n). \quad (7.1)$$

Vamos substituir $f'(t_n, X_n)$ por uma expressão que envolve derivadas parciais da função f . Por exemplo, se $f(t, x) = g(t) + x^2$, então

$$\frac{\partial f}{\partial t}(t, x) = g'(t), \quad \frac{\partial f}{\partial x}(t, x) = 2x. \quad (7.2)$$

Por outro lado,

$$f'(t, x(t)) = (g(t) + x^2(t))' = g'(t) + 2x(t)x'(t) = g'(t) + 2x(t)f(t, x(t)). \quad (7.3)$$

Comparando (7.2) e (7.3), temos que

$$f'(t, x(t)) = \frac{\partial f}{\partial t}(t, x(t)) + \frac{\partial f}{\partial x}(t, x(t)) f(t, x(t)). \quad (7.4)$$

Esta expressão vale para quaisquer funções diferenciáveis f e x , e corresponde à regra da cadeia aplicada à função composta $g(t) = f(t, x(t))$. Portanto, podemos re-escrever (7.1) na forma

$$X_{n+1}^{T2} = X_n + hf(t_n, X_n) + \frac{h^2}{2} \left[\frac{\partial f}{\partial t}(t_n, X_n) + \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right]. \quad (7.5)$$

Falta eliminar as derivadas parciais $\partial f/\partial t$ e $\partial f/\partial x$. Para isto, vamos utilizar (6.10), e para evidenciar a semelhança entre (7.5) e (6.10), vamos escrever (7.5) na forma

$$\begin{aligned} X_{n+1}^{T2} &= X_n + \frac{h}{2} f(t_n, X_n) + \frac{h}{2} f(t_n, X_n) + \frac{h}{2} \left[h \frac{\partial f}{\partial t}(t_n, X_n) + h \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right] \\ &= X_n + \frac{h}{2} f(t_n, X_n) + \frac{h}{2} \left[f(t_n, X_n) + h \left(\frac{\partial f}{\partial t}(t_n, X_n) + \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right) \right] \end{aligned} \quad (7.6)$$

A expressão (7.6) motiva escolher $\alpha = 1$ (logo $t = t_n + h = t_{n+1}$) e $\beta = f(t_n, X_n)$ em (6.10):

$$\begin{aligned} f(t_{n+1}, X_n + hf(t_n, X_n)) &= f(t_n, X_n) + \frac{\partial f}{\partial t}(t_n, X_n)h + \frac{\partial f}{\partial x}(t_n, X_n)hf(t_n, X_n) \\ &+ \frac{h^2}{2} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X}) + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X})f(t_n, X_n) + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X})f(t_n, X_n)^2 \right), \end{aligned} \quad (7.7)$$

sendo que $(\bar{t}, \bar{X}) = (t_n + \lambda h, X_n + \lambda h f(t_{n+1}, X_{n+1})) = (t_n, X_n) + \lambda h(1, f(t_n, X_n))$ para algum $0 \leq \lambda \leq 1$. Assim, a aproximação

$$X_{n+1} = X_n + \frac{h}{2} f(t_n, X_n) + \frac{h}{2} f(t_{n+1}, X_n + hf(t_n, X_n)) \quad (7.8)$$

satisfaz $X_{n+1} = X_{n+1}^{T2} + g(h)$ com

$$\begin{aligned} g(h) &= \frac{h}{2} \left(f(t_{n+1}, X_n + hf(t_n, X_n)) - f(t_n, X_n) - \frac{\partial f}{\partial t}(t_n, X_n)h - \frac{\partial f}{\partial x}(t_n, X_n)hf(t_n, X_n) \right) \\ &= \frac{h^3}{4} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X}) + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X})f(t_n, X_n) + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X})f(t_n, X_n)^2 \right). \end{aligned}$$

Se $f \in C^2$, então segue do Teorema de Weierstrass que

$$\exists M > 0 \quad \text{tal que} \quad \left\{ \begin{array}{l} |f(t, x)| \leq M \\ \left| \frac{\partial^2 f}{\partial t^2}(t, x) \right| \leq M \\ \left| \frac{\partial^2 f}{\partial t \partial x}(t, x) \right| \leq M \\ \left| \frac{\partial^2 f}{\partial x^2}(t, x) \right| \leq M \end{array} \right. \quad \forall (t, x) \in B,$$

sendo B o conjunto fechado

$$B = \{(t, x) \in \mathbb{R}^2 \mid (t, x) = (t_n, X_n) + \lambda h(1, f(t_n, X_n)), \quad 0 \leq \lambda \leq 1\}.$$

Assim,

$$|g(h)| \leq \frac{h^3}{4} (M + 2M^2 + M^3) = \frac{M(M+1)^2}{4} h^3,$$

ou seja, $g(h) = \mathcal{O}(h^3)$. Portanto o método definido por (7.8) satisfaz $X_{n+1} = X_{n+1}^{T2} + \mathcal{O}(h^{2+1})$, ou seja, é um método da classe RK2. Analogamente ao método da série de Taylor, este método pode ser escrito na forma

$$\left\{ \begin{array}{l} F_1 = hf(t_n, X_n) \\ F_2 = hf(t_{n+1}, X_n + F_1) \\ X_{n+1} = X_n + \frac{1}{2}F_1 + \frac{1}{2}F_2. \end{array} \right. \quad (7.9)$$

Este método é conhecido como **método de Heun**.

Aula 8: Método de Euler modificado

Vimos que o método da série de Taylor de segunda ordem pode ser escrito na forma (7.5), e para deduzir o método de Heun, dividimos a segunda parcela do lado direito de (7.5) em duas parcelas iguais, e adicionamos uma delas à terceira parcela. Agora vamos adicionar (sem dividir em duas) a segunda parcela na terceira parcela, e vamos colocar o fator $h/2$ para dentro dela:

$$\begin{aligned} X_{n+1}^{T2} &= X_n + h \left[f(t_n, X_n) + \frac{h}{2} \frac{\partial f}{\partial t}(t_n, X_n) + \frac{h}{2} \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right] \\ &= X_n + h \left[f(t_n, X_n) + h \left(\frac{\partial f}{\partial t}(t_n, X_n) \frac{1}{2} + \frac{\partial f}{\partial x}(t_n, X_n) \frac{f(t_n, X_n)}{2} \right) \right]. \end{aligned} \quad (8.1)$$

Comparando esta expressão com (6.10), convém escolher $\alpha = 1/2$ e $\beta = f(t_n, X_n)/2$:

$$\begin{aligned} f \left(t_n + \frac{h}{2}, X_n + \frac{h}{2} f(t_n, X_n) \right) &= f(t_n, X_n) + h \left(\frac{\partial f}{\partial t}(t_n, X_n) \frac{1}{2} + \frac{\partial f}{\partial x}(t_n, X_n) \frac{f(t_n, X_n)}{2} \right) \\ &\quad + \frac{h^2}{2} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X}) \frac{1}{4} + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X}) \frac{f(t_n, X_n)}{4} + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X}) \frac{f(t_n, X_n)^2}{4} \right), \end{aligned} \quad (8.2)$$

sendo $(\bar{t}, \bar{X}) = (t_n + \lambda h/2, X_n + \lambda h f(t_n, X_n)/2)$, com $0 \leq \lambda \leq 1$. Segue de (8.1) e (8.2) que

$$\begin{aligned} X_{n+1}^{T2} &= X_n + h f \left(t_n + \frac{h}{2}, X_n + \frac{h}{2} f(t_n, X_n) \right) - g(h), \\ g(h) &= \frac{h^3}{8} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X}) + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X}) f(t_n, X_n) + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X}) f(t_n, X_n)^2 \right). \end{aligned}$$

Se $f \in C^2$, podemos seguir o mesmo procedimento da seção anterior para verificar que $g(h) = \mathcal{O}(h^3)$. Assim, o método definido pela equação

$$X_{n+1} = X_n + h f \left(t_n + \frac{h}{2}, X_n + \frac{h}{2} f(t_n, X_n) \right) \quad (8.3)$$

satisfaz $X_{n+1} = X_{n+1}^{T2} + \mathcal{O}(h^3)$, ou seja, é um método da classe RK2. Este método é conhecido como **método de Euler modificado**. Este método foi proposto por Runge em 1895 [4].

Outros métodos da classe RK2 podem ser obtidos considerando um passo fracionário $t = "t_{n+\alpha}" = t_n + \alpha h$ ($0 < \alpha \leq 1$), o que corresponde a escolher $\beta = \alpha f(t_n, X_n)$ na expansão (6.10):

$$\begin{aligned} f(t_n + \alpha h, X_n + \alpha h f(t_n, X_n)) &= f(t_n, X_n) + \alpha h \left(\frac{\partial f}{\partial t}(t_n, X_n) + \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right) \\ &\quad + \frac{\alpha^2 h^2}{2} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X}) + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X}) f(t_n, X_n) + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X}) f(t_n, X_n)^2 \right), \end{aligned} \quad (8.4)$$

com $(\bar{t}, \bar{X}) = (t_n + \lambda \alpha h, X_n + \lambda \alpha h f(t_n, X_n))$, com $0 \leq \lambda \leq 1$. Esta nova combinação de termos na fórmula de Taylor sugere reescrever a equação (8.1) da seguinte forma:

$$\begin{aligned} X_{n+1}^{T2} &= X_n + h \left[f(t_n, X_n) + \frac{h}{2} \frac{\partial f}{\partial t}(t_n, X_n) + \frac{h}{2} \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right] \\ &= X_n + \frac{h}{2} \left[2f(t_n, X_n) + \frac{\partial f}{\partial t}(t_n, X_n) h + \frac{\partial f}{\partial x}(t_n, X_n) h f(t_n, X_n) \right] \\ &= X_n + \frac{h}{2\alpha} \left[2\alpha f(t_n, X_n) + \frac{\partial f}{\partial t}(t_n, X_n) \alpha h + \frac{\partial f}{\partial x}(t_n, X_n) \alpha h f(t_n, X_n) \right] \\ &= X_n + \frac{h}{2\alpha} \left[(2\alpha - 1)f(t_n, X_n) + f(t_n, X_n) + \frac{\partial f}{\partial t}(t_n, X_n) \alpha h + \frac{\partial f}{\partial x}(t_n, X_n) \alpha h f(t_n, X_n) \right] \\ &= X_n + h \left(1 - \frac{1}{2\alpha} \right) f(t_n, X_n) + \frac{h}{2\alpha} \left[f(t_n, X_n) + \alpha h \left(\frac{\partial f}{\partial t}(t_n, X_n) + \frac{\partial f}{\partial x}(t_n, X_n) f(t_n, X_n) \right) \right]. \end{aligned} \quad (8.5)$$

Segue de (8.5) e (8.4) que

$$\begin{aligned} X_{n+1}^{T2} &= X_n + \left(h - \frac{h}{2\alpha}\right) f(t_n, X_n) + \frac{h}{2\alpha} f(t_n + \alpha h, X_n + \alpha h f(t_n, X_n)) - g(h), \\ g(h) &= \frac{\alpha h^3}{4} \left(\frac{\partial^2 f}{\partial t^2}(\bar{t}, \bar{X}) + 2 \frac{\partial^2 f}{\partial t \partial x}(\bar{t}, \bar{X}) f(t_n, X_n) + \frac{\partial^2 f}{\partial x^2}(\bar{t}, \bar{X}) f(t_n, X_n)^2 \right). \end{aligned}$$

Analogamente ao método de Euler modificado, o método

$$X_{n+1} = X_n + h \left(1 - \frac{1}{2\alpha}\right) f(t_n, X_n) + h \frac{1}{2\alpha} f(t_n + \alpha h, X_n + \alpha h f(t_n, X_n))$$

satisfaz $X_{n+1} = X_{n+1}^{T2} + \mathcal{O}(h^3)$, ou seja, também é um método da classe RK2. Ele é usualmente escrito por meio dos parâmetros auxiliares $\beta = \alpha$, $\omega_2 = 1/2\alpha$ e $\omega_1 = 1 - \omega_2$:

$$\begin{cases} F_1 = h f(t_n, X_n) \\ F_2 = h f(t_n + \alpha h, X_n + \beta F_1) \\ X_{n+1} = X_n + \omega_1 F_1 + \omega_2 F_2, \end{cases} \quad \begin{cases} \omega_1 + \omega_2 = 1 \\ \alpha = \beta = \frac{1}{2\omega_2}. \end{cases} \quad (8.6)$$

As equações (8.6) constituem a forma geral dos métodos de Runge-Kutta de segunda ordem. Note que $\omega_1 = 0$ e $\omega_1 = 1/2$ correspondem aos métodos de Euler modificado e de Heun, respectivamente.

Aula 9: Método de Runge-Kutta de quarta ordem

A dedução de métodos de Runge-Kutta RK_m a partir dos métodos da série de Taylor de ordem m torna-se complexa quando $m \geq 3$. Para estas classes de métodos, convém mostrar diretamente (sem recorrer a X_{n+1}^{Tm} , apesar de recorrer à expansão de Taylor da solução exata) que o erro local tem ordem $\mathcal{O}(h^{m+1})$. Butcher [2] provou que se o método

$$\begin{cases} F_1 = hf(t_n, X_n) \\ F_2 = hf(t_n + c_2h, X_n + a_{21}F_1) \\ F_3 = hf(t_n + c_3h, X_n + a_{31}F_1 + a_{32}F_2) \\ F_4 = hf(t_n + c_4h, X_n + a_{41}F_1 + a_{42}F_2 + a_{43}F_3) \\ X_{n+1} = X_n + b_1F_1 + b_2F_2 + b_3F_3 + b_4F_4 \end{cases} \quad (9.1)$$

tem um erro local de ordem $\mathcal{O}(h^{4+1})$, então os coeficientes $C_1, \dots, b_4$ satisfazem as seguintes equações:

$$\begin{cases} b_1 + b_2 + b_3 + b_4 = 1 \\ b_2c_2 + b_3c_4 + b_4c_4 = \frac{1}{2} \\ b_2c_2^2 + b_3c_4^2 + b_4c_4^2 = \frac{1}{3} \\ b_2c_2^3 + b_3c_4^3 + b_4c_4^3 = \frac{1}{4} \end{cases} \quad \begin{cases} b_3a_{32}c_2 + b_4(a_{42}c_2 + a_{43}c_3) = \frac{1}{6} \\ b_3c_3a_{32}c_2 + b_4c_4(a_{42}c_2 + a_{43}c_3) = \frac{1}{8} \\ b_3a_{32}c_2^2 + b_4(a_{42}c_2^2 + a_{43}c_3^2) = \frac{1}{12} \\ b_4a_{43}a_{32}c_2 = \frac{1}{24} \end{cases} \quad \begin{cases} c_2 = a_{21} \\ c_3 = a_{31} + a_{32} \\ c_4 = a_{41} + a_{42} + a_{43} \end{cases}$$

Uma das soluções deste sistema é

$$\begin{aligned} c_2 &= 1/2 & c_3 &= 1/2 & c_4 &= 1 \\ a_{21} &= 1/2 \\ a_{31} &= 0 & a_{32} &= 1/2 \\ a_{41} &= 0 & a_{42} &= 0 & a_{43} &= 1 \\ b_1 &= 1/6 & b_2 &= 2/6 & b_3 &= 2/6 & b_4 &= 1/6, \end{aligned}$$

que resulta no método de Runge-Kutta clássico de quarta ordem (RK4), proposto por Kutta em 1901 [4]:

$$\begin{cases} F_1 = hf(t_n, X_n) \\ F_2 = hf(t_n + h/2, X_n + F_1/2) \\ F_3 = hf(t_n + h/2, X_n + F_2/2) \\ F_4 = hf(t_n + h, X_n + F_3) \\ X_{n+1} = X_n + (F_1 + 2F_2 + 2F_3 + F_4)/6, \end{cases} \quad (9.2)$$

Exemplo 9.1 : Use um passo do método RK4 para aproximar $x(0.02)$, sendo $x(t)$ a solução do PVI

$$\begin{cases} x'(t) = 1 + x(t)^2 \\ x(0) = 0. \end{cases}$$

Temos $X_0 = 0$, $t_0 = 0$, $T = 0.02$ e $h = (T - t_0)/1 = 0.02$. Assim,

$$\begin{aligned} F_1 &= hf(t_n, X_n) = (0.02)(1 + 0^2) = 0.02 \\ F_2 &= hf(t_n + h/2, X_n + F_1/2) = 0.02(1 + (0.02/2)^2) = 0.020002 \\ F_3 &= hf(t_n + h/2, X_n + F_2/2) = 0.02(1 + (0.020002/2)^2) = 0.02000200040002 \\ F_4 &= hf(t_n + h, X_n + F_3) = 0.02(1 + 0.02000200040002^2) = 0.02000800160040004716 \\ X_{n+1} &= X_n + h(F_1 + 2F_2 + 2F_3 + F_4)/6 = 0.02000266706674000972 \end{aligned}$$

Note que $\tan(0.02) = \mathbf{0.0200026670934024}$, ou seja, o método RK4 acertou 10 casas decimais. Em geral, os métodos de Runge-Kutta explícitos com k termos são escritos na forma

$$\left\{ \begin{array}{l} F_i = hf \left(t_n + c_i h, X_n + \sum_{j=1}^{i-1} a_{ij} F_j \right) \quad (1 \leq i \leq k) \\ X_{n+1} = X_n + \sum_{i=1}^k b_i F_i, \end{array} \right. \quad (9.3)$$

com $c_1 = 0$. Os coeficientes acima são organizados de acordo com a seguinte tabela (conhecida como *tabela de Butcher*):

c_1					
c_2	a_{21}				
c_3	a_{31}	a_{32}			
$\vdots$	$\vdots$		$\ddots$		
c_k	a_{k1}	a_{k2}	$\dots$	a_{kk-1}	
	b_1	b_2	$\dots$	b_{k-1}	b_k

Por exemplo, a tabela de Butcher correspondente ao método RK4 é dada por

0				
1/2	1/2			
1/2	0	1/2		
1	0	0	1	
	1/6	2/6	2/6	1/6

Butcher [3] provou que o erro local de um método de Runge-Kutta explícito com k termos não necessariamente é da ordem $\mathcal{O}(h^{k+1})$:

k	4	5	6	7	8	9+
ordem	$\mathcal{O}(h^{4+1})$	$\mathcal{O}(h^{4+1})$	$\mathcal{O}(h^{5+1})$	$\mathcal{O}(h^{6+1})$	$\mathcal{O}(h^{6+1})$	$\mathcal{O}(h^{k-1})$

Aula 10: Métodos de Runge-Kutta adaptativos

Nesta aula vamos considerar métodos de passo simples em que o comprimento de passo h é variável, ou seja, $t_{n+1} = t_n + h_n$. Nestes métodos devemos estabelecer um critério para a escolha de cada h_n .

Um exemplo de situação em que vale a pena usar comprimento de passo variável é quando a derivada da solução é elevada num ponto $t^* \in [t_0, T]$ (Fig. 6).

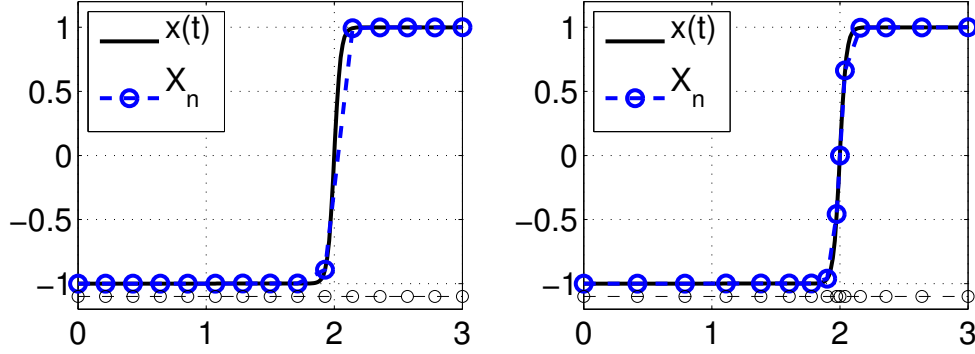


Figura 6: Comparação da função $x(t) = \tanh(20(x - 2))$ com a aproximação X_n (supondo que $X_n = x(t_n)$) para uma partição uniforme (esquerda) e uma partição não-uniforme concentrada em torno de $t^* = 2$ (direita). As partições são exibidas abaixo dos gráficos. Note que a partição não-uniforme usa menos pontos onde a função varia lentamente e mais pontos onde ela varia rapidamente.

Quando não conhecemos de antemão a localização do(s) ponto(s) t^* , uma alternativa é selecionar, na passagem de X_n para X_{n+1} , o valor de h_n tal que

$$|e_{n+1}| = |x(t_{n+1}) - X_{n+1}| = |x(t_n + h_n) - X_{n+1}| \leq \delta,$$

sendo δ uma tolerância máxima escolhida para o erro de aproximação. Os métodos de passo simples **adaptativos** calculam iterativamente o valor de h_n , partindo de uma estimativa inicial h_0 e ajustando-o até que a tolerância máxima seja atingida. Dada a seguinte generalização do método de passo simples (1.2),

$$\begin{cases} X_{n+1} = \Phi(X_{n+1}, X_n, t_{n+1}, t_n), & n \geq 0 \\ X_0 = x_0, \end{cases} \quad (10.1)$$

uma versão adaptativa deste método é dada pelo seguinte algoritmo:

- 1: Dados de entrada: t_0, x_0, T, δ, h_0 .
- 2: $X_0 \leftarrow x_0$;
- 3: $n \leftarrow 0$;
- 4: **while** $t_n < T$ **do**
- 5: $h \leftarrow h_0$;
- 6: $\tilde{X} \leftarrow \Phi(\tilde{X}, X_n, t_n + h, t_n)$;
- 7: $\tilde{e} \leftarrow x(t_n + h) - \tilde{X}$;
- 8: **while** $|\tilde{e}| > \delta$ **do**
- 9: $h \leftarrow h/2$;
- 10: $\tilde{X} \leftarrow \Phi(\tilde{X}, X_n, t_n + h, t_n)$;
- 11: $\tilde{e} \leftarrow x(t_n + h) - \tilde{X}$;
- 12: **end while**
- 13: $X_{n+1} \leftarrow \tilde{X}$;
- 14: $t_{n+1} \leftarrow t_n + h$;
- 15: $n \leftarrow n + 1$;
- 16: **end while**

O defeito deste algoritmo é que ele depende da solução exata (ver linhas 7 e 11). Para contornar este problema, devemos substituir o erro exato $\tilde{e}$ por uma aproximação do erro $\tilde{\varepsilon}$ que não dependa de $x(t_n + h)$. No jargão de métodos adaptativos, a função ou algoritmo que fornece $\tilde{\varepsilon}$ é denominado **estimador do erro**, enquanto a técnica de obtenção de estimadores é denominada **análise a posteriori**.

Seja (10.1) um método de Runge-Kutta de ordem $m - 1$. Uma forma de definir um estimador de erro é aproximar $x(t_n + h)$ por um método **melhor** que o método (10.1). Em particular, podemos considerar

$$\tilde{\varepsilon} = \tilde{Y} - \tilde{X}, \quad \tilde{Y} = \tilde{\Phi}(\tilde{X}, X_n, t_n + h, t_n), \quad (10.2)$$

sendo

$$\begin{cases} X_{n+1} = \tilde{\Phi}(X_{n+1}, X_n, t_{n+1}, t_n), & n \geq 0 \\ X_0 = x_0, \end{cases} \quad (10.3)$$

algum método de Runge-Kutta de ordem m . O estimador (10.2) resolve o problema da dependência da solução exata, mas cria outro problema: aumentamos o custo computacional ao calcular a solução aproximada duas vezes.

Para amenizar este custo, escolhemos para (10.1) e (10.3) dois métodos de Runge-Kutta que possuam cálculos intermediários compartilhados. Mais precisamente,

$$\begin{cases} X_{n+1} = \Phi(X_{n+1}, X_n, t_{n+1}, t_n) = X_n + \sum_{i=1}^m b_i F_i, \\ X_{n+1} = \tilde{\Phi}(X_{n+1}, X_n, t_{n+1}, t_n) = X_n + \sum_{i=1}^m \tilde{b}_i F_i, \end{cases} \quad F_i = hf \left(t_n + c_i h, \sum_{j=1}^{i-1} a_{ij} F_j \right) \quad (1 \leq i \leq m).$$

Os coeficientes são escolhidos de modo que os dois métodos sejam idênticos a menos dos coeficientes b_i e $\tilde{b}_i$, e que $e_n^{loc} = \mathcal{O}(h^m)$ no primeiro método, enquanto $e_n^{loc} = \mathcal{O}(h^{m+1})$ no segundo método. Este par de métodos é usualmente representado pela seguinte tabela de Butcher:

c_1					
c_2	a_{21}				
c_3	a_{31}	a_{32}			
$\vdots$	$\vdots$		$\ddots$		
c_k	a_{k1}	a_{k2}	$\dots$	a_{kk-1}	
	b_1	b_2	$\dots$	b_{k-1}	b_k
	$\tilde{b}_1$	$\tilde{b}_2$	$\dots$	$\tilde{b}_{k-1}$	$\tilde{b}_k$

em que $k \geq m$ (geralmente $k > m$). Note que o erro estimado é dado por

$$\tilde{\varepsilon} = \sum_{i=1}^m (\tilde{b}_i - b_i) F_i,$$

e que uma vez encontrado h_n tal que $|\tilde{\varepsilon}| \leq \delta$, podemos escolher $X_{n+1} = \tilde{Y}$ em vez de $X_{n+1} = \tilde{X}$, dado que $\tilde{Y}$ é uma aproximação resultante de um método de ordem mais alta.

OBS: Kincaid e Cheney [10] indicam outra forma de atualizar o tamanho de passo h (linhas 5 e 9 do algoritmo):

$$h = 0.9h \left(\frac{\delta}{\tilde{\varepsilon}} \right)^{1/m+1}.$$

O método de Runge-Kutta adaptativo clássico é o método de Runge-Kutta-Fehlberg [7], que utiliza $m = 4$ e é definido pela seguinte tabela:

0						
1/4	1/4					
3/8	3/32	9/32				
12/13	1932/2197	-7200/2197	7296/2197			
1	439/216	-8	3680/513	845/4104		
1/2	-8/27	2	-3544/2565	1859/4104	-11/40	
	16/135	0	6656/12825	28561/56430	-9/50	2/55
	25/216	0	1408/2565	2197/4104	-1/5	0

Outros métodos populares são o método de Dormand-Prince [6]

0							
1/5	1/5						
3/10	3/40	9/40					
4/5	44/45	-56/15	32/9				
8/9	19372/6561	-25360/2187	64448/6561	-212/729			
1	9017/3168	-355/33	46732/5247	49/176	-5103/18656		
1	35/384	0	500/1113	125/192	-2187/6784	11/84	
	35/384	0	500/1113	125/192	-2187/6784	11/84	0
	5179/57600	0	7571/16695	393/640	-92097/339200	187/2100	1/40

e o método de Bogacki-Shampine [1]:

0				
1/2	1/2			
3/4	0	3/4		
1	2/9	1/3	4/9	
	2/9	1/3	4/9	0
	7/24	1/4	1/3	1/8

Os dois últimos métodos estão disponíveis no matlab/octave por meio das funções **ode45** **ode23**, respectivamente. Note que, nestes métodos, $\tilde{X} = X_n + F_{k-1}$, de modo que $F_k = hf(t_n + c_k h, \tilde{X})$ e $\tilde{Y} = X_n + \tilde{b}_1 F_1 + \dots \tilde{b}_k F_k$. Portanto, a solução $\tilde{X}$ é um passo intermediário do cálculo da solução $\tilde{Y}$. Métodos com esta propriedade são denominados métodos de Runge-Kutta encaixados (*embedded Runge-Kutta methods*) [13]

Aula 11: Métodos de passos múltiplos

Os métodos de passos múltiplos com k passos são dados por

$$a_0 X_{n+1} + a_1 X_n + a_2 X_{n-1} + \dots + a_k X_{n-(k-1)} = h [b_0 f(t_{n+1}, X_{n+1}) + b_1 f(t_n, X_n) + \dots + b_k f(t_{n-(k-1)}, X_{n-(k-1)})]$$

Daqui em diante vamos empregar a notação $f_k = f(t_k, X_k)$, de modo que

$$\begin{aligned} a_0 X_{n+1} + a_1 X_n + a_2 X_{n-1} + \dots + a_k X_{n-(k-1)} &= h [b_0 f_{n+1} + b_1 f_n + b_2 f_{n-1} + \dots + b_k f_{n-(k-1)}] \\ a_0 X_{n+1} + a_1 X_{n+1-1} + a_2 X_{n+1-2} + \dots + a_k X_{n+1-k} &= h [b_0 f_{n+1} + b_1 f_{n+1-1} + b_2 f_{n+1-2} + \dots + b_k f_{n+1-k}], \end{aligned}$$

ou seja,

$$\sum_{j=0}^k a_j X_{n+1-j} = h \sum_{j=0}^k b_j f_{n+1-j}. \quad (11.1)$$

O método de passos múltiplos (11.1) é dito **explícito** se $b_0 = 0$ e **implícito** caso contrário. Se $k \geq 2$, precisamos calcular previamente as aproximações $X_{k-1}, X_{k-2}, \dots, X_1$, pois estes valores não podem ser obtidos pela equação (11.1). No caso $k = 2$ por exemplo, se tentássemos calcular X_1 tomando $n = 0$ em (11.1), teríamos

$$a_0 X_1 + a_1 X_0 + a_2 X_{-1} = h [b_0 f_1 + b_1 f_0 + b_2 f_{-1}],$$

ou seja, X_1 dependeria de X_{-1} , que é indefinido. Esta dificuldade, entretanto, pode ser superada no momento da implementação computacional, conforme será discutido posteriormente (Aula 12).

11.1 Fórmula de Adams-Bashforth

Considere a versão explícita do método (11.1) ($b_0 = 0$) com $a_0 = 1$, $a_1 = -1$ e $a_j = 0$ para todo $j \geq 2$, ou seja,

$$\begin{aligned} X_{n+1} - X_n &= h \sum_{j=1}^k b_j f_{n+1-j} \\ X_{n+1} &= X_n + h \sum_{j=1}^k b_j f_{n+1-j} \end{aligned} \quad (11.2)$$

Note que as únicas incógnitas em (11.2) são os valores $b_1, \dots, b_k$. A fórmula de Adams-Bashforth com k passos corresponde à equação (11.2) em que os k coeficientes $b_1, \dots, b_k$ são escolhidos de modo que, se x é a solução do PVI

$$\begin{cases} x'(t) = p(t) \\ x(t_0) = x_0 \end{cases} \quad \left(\text{ou seja, } x(t) = x_0 + \int_{t_0}^t p(t) dt \right), \quad (11.3)$$

com $p \in \mathcal{P}_{k-1}$, e além disso, se

$$\begin{aligned} X_0 &= x_0 \\ X_n &= x(t_n) \quad 1 \leq n \leq k-1, \end{aligned} \quad (11.4)$$

então

$$X_n = x(t_n), \quad n \geq k. \quad (11.5)$$

Veremos mais adiante que a condição (11.5) para toda solução $x(t)$ de (11.3) garantirá um erro de truncamento local de ordem $\mathcal{O}(h^{k+1})$.

11.1.1 Fórmula de Adams-Bashforth de um passo ($k = 1$)

Se $p \in \mathcal{P}_{1-1} = \mathcal{P}_0$, então p é uma função constante da forma $p(t) = c$. Se $x(t)$ é a solução do PVI (11.3), então $x(t) = x_0 + c(t - t_0)$. Além disso, para que $X_n = x(t_n)$ para $n \leq k - 1 = 0$, basta que $X_0 = x(t_0) = x_0$. Vamos obter os coeficientes do método (11.2) com $k = 1$ e $X_0 = x_0$, ou seja,

$$\begin{cases} X_{n+1} = X_n + hb_1 f_n, & n \geq 0 \\ X_0 = x_0, \end{cases}$$

tais que $X_n = x(t_n)$ para todo $n \geq 1$. Note que, como $f(t, x(t)) = p(t) = c$, $X_{n+1} = X_n + hb_1 c$. Esta expressão nos permite obter uma fórmula para X_n em termos de X_0 :

$$X_n = X_{n-1} + hb_1 c = (X_{n-2} + hb_1 c) + hb_1 c = X_{n-2} + 2hb_1 c = \dots = X_{n-n} + nhb_1 c.$$

Como $X_0 = x_0$, segue que

$$X_n = x_0 + nhb_1 c \quad (11.6)$$

Por outro lado, como $x(t) = x_0 + c(t - t_0)$, temos que

$$x(t_n) = x_0 + c(t_n - t_0) = x_0 + c(nh). \quad (11.7)$$

Segue de (11.6) e (11.7) que devemos ter $b_1 = 1$. Portanto a fórmula de Adams-Bashforth de um passo (AB1) é dada por

$$\begin{cases} X_{n+1} = X_n + hf_n = X_n + hf(t_n, X_n), & n \geq 0 \\ X_0 = x_0, \end{cases}$$

ou seja, ela coincide com o método de Euler explícito.

11.1.2 Fórmula de Adams-Bashforth de dois passos ($k = 2$)

Se $p \in \mathcal{P}_{2-1} = \mathcal{P}_1$, então existem $c_0, c_1 > 0$ tais que $p(t) = c_0 + c_1 t$. Teremos a seguir manipulações algébricas extremamente trabalhosas, de modo que será importante expressar $p(t)$ de outra maneira.

Se $x(t)$ é a solução do PVI (11.3), então

$$x(t) = x_0 + \int_{t_0}^t (c_0 + c_1 t) dt = x_0 + c_0(t - t_0) + c_1 \frac{t^2 - t_0^2}{2}.$$

Assim, para que $X_n = x(t_n)$ para $n \leq k - 1 = 1$, devemos ter

$$\begin{aligned} X_0 &= x(t_0) = x_0 \\ X_1 &= x(t_1) = x_0 + c_0 h + c_1(t_1^2 - t_0^2)/2 = x_0 + c_0 h + c_1(h^2 + 2ht_0)/2 \end{aligned} \quad (11.8)$$

Assim, devemos obter os coeficientes b_1 e b_2 em

$$\begin{cases} X_{n+1} = X_n + h[b_1 f_n + b_2 f_{n-1}], & n \geq 1 \\ X_0 = x_0 \\ X_1 = x_0 + c_0 h + c_1(h^2 + 2ht_0)/2 \end{cases}$$

Como $f(t, x(t)) = p(t) = c_0 + c_1 t$, temos que, para $n \geq 1$,

$$\begin{aligned} X_{n+1} &= X_n + h[b_1(c_0 + c_1 t_n) + b_2(c_0 + c_1 t_{n-1})] \\ &= X_n + h[c_0(b_1 + b_2) + c_1(b_1 t_n + b_2 t_{n-1})] \\ &= X_n + h[c_0(b_1 + b_2) + c_1(b_1 t_n + b_2 t_n - b_2 h)] \\ &= X_n + h[c_0(b_1 + b_2) - hc_1 b_2 + c_1(b_1 + b_2)t_n] = X_n + h\beta_1 + h\beta_2 t_n, \end{aligned}$$

com $\beta_1 = c_0(b_1 + b_2) - hc_1b_2$ e $\beta_2 = c_1(b_1 + b_2)$. Deste modo,

$$\begin{aligned}
X_n &= X_{n-1} + h\beta_1 + h\beta_2 t_{n-1} \\
&= (X_{n-2} + h\beta_1 + h\beta_2 t_{n-2}) + h\beta_1 + h\beta_2 t_{n-1} = X_{n-2} + 2h\beta_1 + h\beta_2(t_{n-1} + t_{n-2}) \\
&= \dots = X_{n-(n-1)} + (n-1)h\beta_1 + h\beta_2(t_{n-1} + t_{n-2} + \dots + t_{n-(n-1)}) \\
&= X_1 + (n-1)h\beta_1 + h\beta_2 \sum_{j=1}^{n-1} t_j \\
&= X_1 + (n-1)h\beta_1 + h\beta_2 \sum_{j=1}^{n-1} (jh + t_0) \\
&= X_1 + (t_{n-1} - t_0)\beta_1 + h^2\beta_2 \sum_{j=1}^{n-1} j + h\beta_2 \sum_{j=1}^{n-1} t_0 \\
&= X_1 + (t_{n-1} - t_0)\beta_1 + h^2\beta_2 \frac{n(n-1)}{2} + (n-1)h\beta_2 t_0 \\
&= X_1 + (t_{n-1} - t_0)\beta_1 + \beta_2 h(n-1) \left(n\frac{h}{2} + t_0 \right) \\
&= X_1 + (t_{n-1} - t_0)\beta_1 + \beta_2(t_{n-1} - t_0) \frac{nh + 2t_0}{2} \\
&= X_1 + (t_{n-1} - t_0)\beta_1 + \beta_2 \frac{(t_{n-1} - t_0)(t_n + t_0)}{2}.
\end{aligned}$$

Tendo em vista que

$$x(t_n) = x_0 + c_0(t_n - t_0) + c_1 \frac{t_n^2 - t_0^2}{2},$$

vamos escrever a expressão obtida para X_n em termos de $t_n - t_0$:

$$\begin{aligned}
X_n &= X_1 + (t_n - \textcolor{red}{h} - t_0)\beta_1 + \beta_2 \frac{(t_n - \textcolor{red}{h} - t_0)(t_n + t_0)}{2} \\
&= X_1 + (t_n - t_0)\beta_1 - h\beta_1 + \beta_2 \frac{(t_n - t_0)(t_n + t_0)}{2} - h\beta_2 \frac{t_n + t_0}{2} \\
&= X_1 + (t_n - t_0)\beta_1 - h\beta_1 + \beta_2 \frac{t_n^2 - t_0^2}{2} - h\beta_2 \frac{t_n - t_0 + \textcolor{red}{2}t_0}{2} \\
&= X_1 + (t_n - t_0)\beta_1 - h\beta_1 + \beta_2 \frac{t_n^2 - t_0^2}{2} - h\beta_2 \frac{t_n - t_0}{2} - h\beta_2 t_0 \\
&= X_1 + \left(\beta_1 - \frac{h\beta_2}{2} \right) (t_n - t_0) + \beta_2 \frac{t_n^2 - t_0^2}{2} - \beta_1 h - h\beta_2 t_0 \\
&= X_1 + \left(\beta_1 - \frac{h\beta_2}{2} \right) (t_n - t_0) + \beta_2 \frac{t_n^2 - t_0^2}{2} - h(\beta_1 + \beta_2 t_0)
\end{aligned}$$

Além disso, segue de (11.8) que

$$\begin{aligned}
X_n &= x_0 + c_0 h + c_1 \frac{h^2 + 2ht_0}{2} + \left(\beta_1 - \frac{h\beta_2}{2} \right) (t_n - t_0) + \beta_2 \frac{t_n^2 - t_0^2}{2} - h(\beta_1 + \beta_2 t_0) \\
&= x_0 + \left(\beta_1 - \frac{h\beta_2}{2} \right) (t_n - t_0) + \beta_2 \frac{t_n^2 - t_0^2}{2} - h \left(\beta_1 + \beta_2 t_0 - c_0 - c_1 \frac{h + 2t_0}{2} \right) \\
&= x_0 + \left(\beta_1 - \frac{h\beta_2}{2} \right) (t_n - t_0) + \beta_2 \frac{t_n^2 - t_0^2}{2} - h \left(\beta_1 + \beta_2 t_0 - c_0 - c_1 \frac{h}{2} - c_1 t_0 \right) \\
&= x_0 + \left(\beta_1 - \frac{h\beta_2}{2} \right) (t_n - t_0) + \beta_2 \frac{t_n^2 - t_0^2}{2} - h \left(\beta_1 - \frac{hc_1}{2} - c_0 \right) - ht_0(\beta_2 - c_1)
\end{aligned}$$

Portanto, para que $X_n = x(t_n)$, basta que

$$\beta_1 - \frac{h\beta_2}{2} = c_0, \quad \beta_2 = c_1.$$

Note que esta escolha faz com a última parcela de X_n acima se anule automaticamente. Agora vamos substituir as expressões de β_1 e β_2 de modo a obter condições para b_1 e b_2 . Como $\beta_2 = c_1$,

$$c_1(b_1 + b_2) = c_1 \implies b_1 + b_2 = 1.$$

Da primeira condição, temos que

$$\begin{aligned} \beta_1 - \frac{h\beta_2}{2} &= c_0 \\ c_0(b_1 + b_2) - hc_1b_2 - \frac{hc_1}{2} &= c_0 \\ c_0 - hc_1b_2 - \frac{hc_1}{2} &= c_0 \\ -hc_1b_2 - \frac{hc_1}{2} &= 0 \\ (hc_1)b_2 &= -\frac{hc_1}{2} \implies b_2 = -\frac{1}{2}. \end{aligned}$$

Como $b_1 + b_2 = 1$, temos que $b_1 = 3/2$, e portanto a fórmula de Adams-Bashforth de dois passos (AB2) é dada por

$$\begin{cases} X_{n+1} = X_n + h \left[\frac{3}{2}f_n - \frac{1}{2}f_{n-1} \right], & n \geq 1 \\ X_0 = x_0 \end{cases} \quad (11.9)$$

Não inserimos em (11.9) a fórmula para X_1 dada por (11.8) porque esta somente vale para $f(t, x(t)) = c_0 + c_1t$. A escolha de X_1 será discutida posteriormente.

11.1.3 Fórmula de Adams-Bashforth de dois passos: dedução alternativa

Visando simplificar os cálculos realizados anteriormente, vamos reescrever $p(t)$ na forma

$$p(t) = \tilde{c}_0 + \tilde{c}_1 \frac{t - t_0}{h}. \quad (11.10)$$

Neste caso, $x(t)$ passa a ser escrito na forma

$$x(t) = x_0 + \tilde{c}_0(t - t_0) + \tilde{c}_1 \frac{(t - t_0)^2}{2h},$$

e em particular,

$$x(t_n) = x_0 + \tilde{c}_0(t_n - t_0) + \tilde{c}_1 \frac{(t_n - t_0)^2}{2h} = x_0 + \tilde{c}_0nh + \tilde{c}_1 \frac{n^2h}{2} \quad (11.11)$$

$$x(t_1) = x_0 + \tilde{c}_0h + \tilde{c}_1 \frac{h}{2}. \quad (11.12)$$

A fórmula para X_{n+1} se torna

$$\begin{aligned} X_{n+1} &= X_n + h \left[b_1 \left(\tilde{c}_0 + \tilde{c}_1 \frac{t_n - t_0}{h} \right) + b_2 \left(\tilde{c}_0 + \tilde{c}_1 \frac{t_{n-1} - t_0}{h} \right) \right] \\ &= X_n + h \left[b_1 \left(\tilde{c}_0 + \tilde{c}_1 \frac{nh}{h} \right) + b_2 \left(\tilde{c}_0 + \tilde{c}_1 \frac{(n-1)h}{h} \right) \right] \\ &= X_n + h [b_1\tilde{c}_0 + b_1\tilde{c}_1n + b_2\tilde{c}_0 + b_2\tilde{c}_1(n-1)] \\ &= X_n + h [b_1\tilde{c}_0 + b_2\tilde{c}_0 + b_1\tilde{c}_1n + b_2\tilde{c}_1n - b_2\tilde{c}_1] \\ &= X_n + h [\tilde{c}_0(b_1 + b_2) - \tilde{c}_1b_2 + \tilde{c}_1(b_1 + b_2)n] = X_n + h\beta_1 + h\beta_2n, \end{aligned}$$

com $\beta_1 = \tilde{c}_0(b_1+b_2) - \tilde{c}_1b_2$ e $\beta_2 = \tilde{c}_1(b_1+b_2)$. Repetindo o procedimento da seção anterior, encontramos

$$\begin{aligned}
X_n &= X_{n-1} + h\beta_1 + h\beta_2(n-1) \\
&= [X_{n-2} + h\beta_1 + h\beta_2(n-2)] + h\beta_1 + h\beta_2(n-1) = X_{n-2} + 2h\beta_1 + h\beta_2[(n-1) + (n-2)] \\
&= \dots = X_1 + (n-1)h\beta_1 + h\beta_2 \sum_{j=1}^{n-1} j \\
&= X_1 + (n-1)h\beta_1 + h\beta_2 \frac{n(n-1)}{2} \\
&= X_1 + \beta_1 nh - \beta_1 h + \beta_2 \frac{n^2 h}{2} - \beta_2 \frac{nh}{2} \\
&= X_1 + \left(\beta_1 - \frac{\beta_2}{2} \right) nh + \beta_2 \frac{n^2 h}{2} - \beta_1 h
\end{aligned}$$

Segue de (11.12) que

$$\begin{aligned}
X_n &= x_0 + \tilde{c}_0 h + \tilde{c}_1 \frac{h}{2} + \left(\beta_1 - \frac{\beta_2}{2} \right) nh + \beta_2 \frac{n^2 h}{2} - \beta_1 h \\
&= x_0 + \left(\beta_1 - \frac{\beta_2}{2} \right) nh + \beta_2 \frac{n^2 h}{2} + h \left(\tilde{c}_0 + \frac{\tilde{c}_1}{2} - \beta_1 \right).
\end{aligned}$$

Para que $X_n = x(t_n)$, basta que

$$\beta_1 - \frac{\beta_2}{2} = \tilde{c}_0, \quad \beta_2 = \tilde{c}_1.$$

Analogamente à dedução anterior, $\beta_2 = \tilde{c}_1$ implica que $b_1 + b_2 = 1$. Logo $\beta_1 = \tilde{c}_0 - \tilde{c}_1 b_2$, e além disso,

$$\begin{aligned}
\beta_1 - \frac{\beta_2}{2} &= \tilde{c}_0 \\
\tilde{c}_0 - \tilde{c}_1 b_2 - \frac{\beta_2}{2} &= \tilde{c}_0 \\
-\tilde{c}_1 b_2 - \frac{\tilde{c}_1}{2} &= 0,
\end{aligned}$$

ou seja, $b_2 = -1/2$ (e de $b_1 + b_2 = 1$, $b_2 = 3/2$), conforme deduzido anteriormente.

11.1.4 Segunda dedução alternativa

Vamos simplificar mais ainda os cálculos acima. Considere $n > 0$ fixo, porém arbitrário, e considere a seguinte representação do polinômio p :

$$p(t) = c_0 + c_1 \frac{t - t_n}{h}. \quad (11.13)$$

Vamos escrever $x(t_{n+1})$ em termos de $x(t_n)$ por meio do Teorema Fundamental do Cálculo:

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} \left(c_0 + c_1 \frac{t - t_n}{h} \right) dt = x(t_n) + c_0 h + c_1 \frac{h^2 - 0}{2h} = x(t_n) + c_0 h + c_1 \frac{h}{2}.$$

Portanto, a sequência $Y_n = x(t_n)$ satisfaz

$$\begin{cases} Y_{n+1} &= Y_n + h \left(c_0 + \frac{c_1}{2} \right) \\ Y_0 &= x(t_0) \\ Y_1 &= x(t_1) \end{cases} \quad (11.14)$$

Queremos provar que $X_i = Y_i$ para todo $0 \leq i \leq n+1$. Temos de (11.13) que

$$\begin{aligned}
X_{n+1} &= X_n + h [b_1 p(t_n) + b_2 p(t_{n-1})] \\
&= X_n + h \left[b_1 \left(c_0 + c_1 \frac{t_n - t_n}{h} \right) + b_2 \left(c_0 + c_1 \frac{t_{n-1} - t_n}{h} \right) \right] \\
&= X_n + h \left[b_1 c_0 + b_2 \left(c_0 - c_1 \frac{h}{h} \right) \right] \\
&= X_n + h [c_0(b_1 + b_2) - c_1 b_2],
\end{aligned}$$

logo segue que

$$\begin{cases} X_{n+1} = X_n + h [c_0(b_1 + b_2) + c_1(-b_2)], \\ X_0 = x(t_0) \\ X_1 = x(t_1). \end{cases} \quad (11.15)$$

Para que as **equações de diferenças** (11.15) e (11.14) tenham a mesma solução, basta que os coeficientes das equações de Y_{n+1} e X_{n+1} sejam iguais para quaisquer c_0 e c_1 , isto é,

$$\begin{cases} b_1 + b_2 = 1 \\ -b_2 = 1/2, \end{cases}$$

do qual obtemos novamente $b_2 = -1/2$ e $b_1 = 3/2$.

Em geral, para deduzir a fórmula de Adams-Bashforth de k passos, convém escrever $p(t)$ na forma

$$p(t) = c_0 + c_1 \frac{t - t_n}{h} + c_2 \frac{(t - t_n)^2}{h^2} + \dots + c_{n-1} \frac{(t - t_n)^{k-1}}{h^{k-1}}. \quad (11.16)$$

Aula 12: Fórmula de Adams-Moulton

A fórmula de Adams-Moulton corresponde à versão **implícita** do método (11.1), ou seja, temos $b_0 \neq 0$. Além disso, $a_0 = 1$, $a_1 = -1$ e $a_j = 0$ para todo $j \geq 2$, ou seja,

$$\begin{aligned} X_{n+1} - X_n &= h \sum_{j=0}^{k-1} b_j f_{n+1-j} \\ X_{n+1} &= X_n + h \sum_{j=0}^{k-1} b_j f_{n+1-j}, \end{aligned} \quad (12.1)$$

em que os k coeficientes $b_0, \dots, b_{k-1}$ são escolhidos de modo que, se x é a solução do PVI (11.3) com $p \in \mathcal{P}_{k-1}$, e

$$\begin{aligned} X_0 &= x_0 \\ X_n &= x(t_n) \quad 1 \leq n \leq k-2, \end{aligned} \quad (12.2)$$

então

$$X_n = x(t_n), \quad n \geq k. \quad (12.3)$$

Note que a contagem dos passos inicia em $j = 0$ e termina em $j = k-1$, totalizando k passos. Além disso, ao contrário da fórmula de Adams-Bashforth, a expressão (12.1) não depende de X_{n+1-k} , logo ela permite calcular X_n para $n+1-k \geq 0$, ou seja, $n \geq k-1$. Portanto devemos fornecer externamente $X_0, \dots, X_{k-2}$, o justifica a restrição (12.2) ser ligeiramente diferente de (11.4)

12.1 Fórmula de Adams-Moulton de um passo ($k = 1$)

Vimos que, se $p \in \mathcal{P}_{1-1} = \mathcal{P}_0$, ou seja, $p(t) = c$, então a solução do PVI (11.3) é $x(t) = x_0 + c(t - t_0)$ ou, usando o Teorema Fundamental do Cálculo, $x(t_{n+1}) = x(t_n) + ch$. Assim, a sequência $Y_n = x(t_n)$ satisfaz

$$\begin{cases} Y_{n+1} &= Y_n + hc \\ Y_0 &= x_0 \end{cases} \quad (12.4)$$

Por outro lado, como $f_{n+1} = p(t_{n+1}) = c$, temos de (12.1) com $k = 1$ que

$$\begin{cases} X_{n+1} &= X_n + hb_0 f_{n+1} = X_n + hcb_0 \\ X_0 &= x(t_0) = x_0, \end{cases} \quad (12.5)$$

logo para que $X_n = x(t_n) = Y_n$ as equações de diferenças (12.4) e (12.5) devem ser iguais. Para isto, basta que $b_0 = 1$. Portanto a fórmula de Adams-Moulton de um passo (AM1) é dada por

$$\begin{cases} X_{n+1} &= X_n + hf_{n+1} = X_n + hf(t_{n+1}, X_{n+1}), \quad n \geq 0 \\ X_0 &= x_0. \end{cases} \quad (12.6)$$

Este método também é conhecido como **método de Euler implícito**.

12.2 Fórmula de Adams-Moulton de dois passos ($k = 2$)

Vimos que, para $f(t, x(t)) = p(t)$ na forma

$$p(t) = c_0 + c_1 \frac{t - t_n}{h},$$

a solução $x(t_n)$ do PVI (11.1) satisfaz $x(t_{n+1}) = x(t_n) + c_0 h + c_1 \frac{h}{2}$ e a sequência $Y_n = x(t_n)$ satisfaz

$$\begin{cases} Y_{n+1} &= Y_n + h \left(c_0 + \frac{c_1}{2} \right) \\ Y_0 &= x(t_0) \end{cases} \quad (12.7)$$

Além disso,

$$\begin{aligned}
X_{n+1} &= X_n + h [b_0 p(t_{n+1}) + b_1 p(t_n)] \\
&= X_n + h \left[b_0 \left(c_0 + c_1 \frac{t_{n+1} - t_n}{h} \right) + b_1 \left(c_0 + c_1 \frac{t_n - t_n}{h} \right) \right] \\
&= X_n + h [b_0(c_0 + c_1) + b_1 c_0] = X_n + h [c_0(b_0 + b_1) + c_1 b_0],
\end{aligned}$$

ou seja, a solução gerada pela fórmula de Adams-Moulton de dois passos satisfaz a equação de diferenças

$$\begin{cases} X_{n+1} = X_n + h [c_0(b_0 + b_1) + c_1 b_0], \\ X_0 = x(t_0). \end{cases} \quad (12.8)$$

Para que as equações de diferenças (12.7) e (12.8) tenham a mesma solução, basta que os coeficientes das equações de Y_{n+1} e X_{n+1} sejam iguais para quaisquer c_0 e c_1 , isto é,

$$\begin{cases} b_0 + b_1 = 1 \\ b_0 = 1/2, \end{cases} \quad (12.9)$$

do qual obtemos $b_0 = b_1 = 1/2$. Portanto a fórmula de Adams-Moulton de dois passos (AM2) é dada por

$$\begin{cases} X_{n+1} = X_n + h(f_{n+1} + f_n)/2, \quad n \geq 0 \\ X_0 = x_0. \end{cases} \quad (12.10)$$

Este método também é conhecido como **método trapezoidal**.

12.3 Fórmula de Adams-Moulton: caso geral

Para deduzir a fórmula de Adams-Moulton de k passos, vamos escrever $p \in \mathcal{P}_{k-1}$ na forma

$$p(t) = c_0 + c_1 \frac{t - t_n}{h} + c_2 \frac{(t - t_n)^2}{h^2} + \dots + c_{k-1} \frac{(t - t_n)^{k-1}}{h^{k-1}} = c_0 + \sum_{i=1}^{k-1} c_i \frac{(t - t_n)^i}{h^i}. \quad (12.11)$$

O termo c_0 foi deixado de fora da somatória, pois caso contrário teríamos uma indeterminação do tipo 0^0 em $p(t_n)$. Pelo Teorema Fundamental do Cálculo,

$$x(t_{n+1}) = x(t_n) + \int_{t_n}^{t_{n+1}} p(t) dt = c_0 h + \sum_{i=1}^{k-1} c_i \int_{t_n}^{t_{n+1}} \frac{(t - t_n)^i}{h^i} dt = c_0 h + \sum_{i=1}^{k-1} c_i \frac{h}{i+1}.$$

Assim, a sequência $Y_n = x(t_n)$ é a solução da equação de diferenças

$$\begin{cases} Y_{n+1} = Y_n + h c_0 + h \sum_{i=1}^{k-1} c_i \frac{1}{i+1} \\ Y_j = x(t_j) \quad (j = 0, 1 \leq j \leq k-2). \end{cases} \quad (12.12)$$

Vamos agora tratar da sequência X_n . Temos de (12.11) que

$$\begin{aligned}
f_j &= p(t_j) = c_0 + \sum_{i=1}^{k-1} c_i \frac{(t_j - t_n)^i}{h^i} \\
&= c_0 + \sum_{i=1}^{k-1} c_i \frac{[(jh - t_0) - (nh - t_0)]^i}{h^i} \\
&= c_0 + \sum_{i=1}^{k-1} c_i \frac{(j - n)^i h^i}{h^i} = c_0 + \sum_{i=1}^{k-1} c_i (j - n)^i.
\end{aligned}$$

Assim, segue de (12.1) que

$$\begin{aligned} X_{n+1} &= X_n + h \sum_{j=0}^{k-1} b_j f_{n+1-j} = X_n + h \sum_{j=0}^{k-1} b_j \left(c_0 + \sum_{i=1}^{k-1} c_i (\textcolor{red}{n} + \textcolor{red}{1} - \textcolor{red}{j} - n)^i \right) \\ &= X_n + h \sum_{j=0}^{k-1} b_j \left(c_0 + \sum_{i=1}^{k-1} c_i (1-j)^i \right). \end{aligned}$$

Vamos inverter a ordem das somatórias acima:

$$X_{n+1} = X_n + h c_0 \sum_{j=0}^{k-1} b_j + h \sum_{i=1}^{k-1} c_i \left(\sum_{j=0}^{k-1} (1-j)^i b_j \right).$$

Lembrando da restrição (12.2), temos que a sequência X_n é a solução de

$$\begin{cases} X_{n+1} = X_n + h \sum_{i=0}^{k-1} c_i \left(\sum_{j=0}^{k-1} (1-j)^i b_j \right) \\ X_j = x(t_j) \quad (j = 0, 1 \leq j \leq k-2). \end{cases} \quad (12.13)$$

Comparando (12.12) com (12.13), temos que $b_0, \dots, b_{k-1}$ devem satisfazer o sistema linear

$$\sum_{j=0}^{k-1} b_j = 1, \quad (12.14)$$

$$\sum_{j=0}^{k-1} (1-j)^i b_j = \frac{1}{i+1}, \quad 1 \leq i \leq k-1. \quad (12.15)$$

Por exemplo, se $k = 2$, temos o sistema linear

$$\begin{cases} \sum_{j=0}^1 b_j = b_0 + b_1 = 1 \\ \sum_{j=0}^1 (1-j)^1 b_j = (1-0)b_0 + (1-1)b_1 = \frac{1}{1+1} = \frac{1}{2}, \end{cases}$$

que resulta no mesmo sistema linear que (12.9).

Aula 13: Implementação dos métodos de passos múltiplos

Há duas dificuldades na implementação computacional dos métodos de passos múltiplos:

1. Precisamos calcular previamente as aproximações $X_{k-1}, X_{k-2}, \dots, X_1$;
2. Não é trivial lidar com o termo $b_0 f_{n+1} = b_0 f(t_{n+1}, X_{n+1})$, no caso de métodos implícitos.

Para lidar com a primeira dificuldade, calculamos $X_{k-1}, X_{k-2}, \dots, X_1$ com um método de passo simples. Porém, como devemos garantir que $X_n = x(t_n)$ para $0 \leq n \leq k-1$ no caso em que $f(t, x(t)) = p(t)$ com $p \in \mathcal{P}_{k-1}$, é necessário que este método de passo simples tenha ordem $k+1$.

Exemplo 13.1 : Aproxime X_1 na fórmula de Adams-Bashforth de dois passos,

$$X_{n+1} = X_n + h \left[\frac{3}{2}f_n - \frac{1}{2}f_{n-1} \right], \quad n \geq 1.$$

Podemos usar algum método de Runge-Kutta de segunda ordem, por exemplo o método de Euler modificado, para calcular X_1 :

$$X_1 = X_0 + hf \left(t_0 + \frac{h}{2}, X_0 + \frac{h}{2}f(t_0, X_0) \right) = x_0 + hf \left(t_0 + \frac{h}{2}, x_0 + \frac{h}{2}f(t_0, x_0) \right). \quad (13.1)$$

Deste modo, temos

$$\begin{aligned} X_1 &= x_0 + hf \left(t_0 + \frac{h}{2}, x_0 + \frac{h}{2}f(t_0, x_0) \right) \\ X_2 &= X_1 + h \left[\frac{3}{2}f(t_1, X_1) - \frac{1}{2}f(t_0, x_0) \right] \\ X_3 &= X_2 + h \left[\frac{3}{2}f(t_2, X_2) - \frac{1}{2}f(t_1, X_1) \right] \\ &\dots \end{aligned}$$

Note que, se $f(t, x(t)) = c_0 + c_1(t - t_0)/h$, de modo que $x(t_1) = x_0 + c_0h + c_1h/2$, então segue de (13.1) que

$$X_1 = x_0 + h \left[c_0 + c_1 \left(t_0 + \frac{h}{2} - t_0 \right) / h \right] = x_0 + h[c_0 + c_1/2] = x(t_1),$$

portanto a condição $X_n = x(t_n)$ para $0 \leq n \leq k-1$ é assegurada.

Para lidar com a segunda dificuldade, aproximamos X_{n+1} em $f_{n+1} = f(t_{n+1}, X_{n+1})$ pela solução $\tilde{X}_{n+1}$ gerada por um método explícito com a mesma ordem.

Exemplo 13.2 : Escreva uma aproximação na forma de um método explícito par a fórmula de Adams-Moulton de dois passos,

$$X_{n+1} = X_n + h \left[\frac{1}{2}f_{n+1} + \frac{1}{2}f_n \right], \quad n \geq 1.$$

Uma possibilidade é aproximar X_{n+1} em f_{n+1} pela fórmula de Adams-Bashforth de dois passos. Obtemos assim,

$$\begin{cases} \tilde{X}_{n+1} &= X_n + h \left[\frac{3}{2}f(t_n, X_n) - \frac{1}{2}f(t_{n-1}, X_{n-1}) \right] \\ X_{n+1} &= X_n + h \left[\frac{1}{2}f(t_{n+1}, \tilde{X}_{n+1}) + \frac{1}{2}f(t_n, X_n) \right] \end{cases}$$

O método resultante da estratégia acima de usar um método explícito para inicializar um método implícito aproximado é denominado **método preditor-corretor**. O passo que aproxima (“prediz”) X_{n+1} por $\tilde{X}_{n+1}$ é chamado **passo preditor**, enquanto o passo que calcula (“corrige”) X_{n+1} é chamado **passo corretor**.

Aula 14: Consistência, Estabilidade e Convergência

Sejam f , t_0 e x_0 tais que o PVI

$$\begin{cases} x'(t) = f(t, x(t)) \\ x(t_0) = x_0 \end{cases} \quad (14.1)$$

tenha solução única no intervalo $[t_0, T]$. Seja

$$t_0 < t_1 < \dots < t_N = T$$

uma partição uniforme do intervalo $[t_0, T]$ com $h = (T - t_0)/N$, e seja

$$\begin{cases} X_{n+1} = \Phi([X_0, X_1, \dots, X_n, X_{n+1}], h) \\ X_0 = x_0 \end{cases} \quad (14.2)$$

um método de passo simples ou de passos múltiplos.

Definição 14.1 : O *erro de truncamento local* da aproximação X_n no instante t_{n+1} é definido por

$$\tau_n = x(t_n) - \tilde{X}_n, \quad (14.3)$$

sendo $\{\tilde{X}_n\}$ uma sequência tal que

$$\begin{cases} \tilde{X}_n = \Phi([x(t_0), x(t_1), \dots, x(t_{n-1}), x(t_n)], h), \\ \tilde{X}_0 = x_0. \end{cases} \quad (14.4)$$

Def: Dizemos que o método (14.2) é *consistente* se

$$\lim_{h \rightarrow 0} \frac{1}{h} \tau_n = 0.$$

Seja $x(h, t)$ a função contínua e linear por partes tal que $x(h, t_n) = X_n$ para todo $n = 0, \dots, N$. Vamos adotar a seguinte definição de convergência:

Definição 14.2 : Dizemos que a sequência finita $\{X_0, X_1, \dots, X_N\}$ gerada pelo método (14.2) *converge* para a solução $x(t)$ do PVI (14.1), ou seja, o método (14.2) é *convergente*, se

$$\lim_{h \rightarrow 0} x(h, t) = x(t) \quad \forall t \in [t_0, T].$$

Para relacionar os conceitos de consistência e estabilidade, suponha por simplicidade que a equação principal do método (14.2) possa ser escrita na forma

$$X_{n+1} = X_n + hF([X_0, \dots, X_{n+1}], h). \quad (14.5)$$

Segue que

$$\tau_n = x(t_n) - x(t_{n-1}) - hF = h \left(\frac{x(t_n) - x(t_{n-1})}{h} - F \right),$$

com $F = F([x(t_0), \dots, x(t_n)], h)$. Assim o método é consistente se

$$\lim_{h \rightarrow 0} \frac{x(t_n) - x(t_{n-1})}{h} = \lim_{h \rightarrow 0} F, \quad \text{ou seja,} \quad f(t_n, x(t_n)) = x'(t_n) = \lim_{h \rightarrow 0} F.$$

Assim, segundo Dahlquist e Björck [5, Sec. 8.5.4], “consistência normalmente significa que a equação de diferenças $[\frac{x(t_n) - x(t_{n-1})}{h} - F]$ converge para a equação diferencial $[x' = f]$, enquanto

convergência significa que a **solução** da equação de diferenças converge para a **solução** da equação diferencial.”

Vamos fazer um paralelo com os métodos iterativos para sistemas lineares $A\mathbf{x} = \mathbf{b}$. Seja $\mathbf{e}^{(k)} = \mathbf{x} - \mathbf{x}^{(k)}$ o erro da aproximação $\mathbf{x}^{(k)}$ e seja $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$ o seu resíduo. Temos que

$$\|\mathbf{e}^{(k)}\| = \|(A^{-1}A)(\mathbf{x} - \mathbf{x}^{(k)})\| = \|A^{-1}(A\mathbf{x} - A\mathbf{x}^{(k)})\| \leq \|A^{-1}\| \|\mathbf{b} - A\mathbf{x}^{(k)}\| = \|A^{-1}\| \|\mathbf{r}^{(k)}\|$$

Em particular, o resíduo relativo $\mathbf{r}^{(k)}/\|\mathbf{b}\|$ e o erro relativo $\mathbf{e}^{(k)}/\|\mathbf{x}\|$ satisfazem a desigualdade

$$\begin{aligned} \frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{x}\|} &\leq \frac{\|A^{-1}\| \|\mathbf{r}^{(k)}\|}{\|\mathbf{x}\|} = \frac{\|A^{-1}\| \|\mathbf{r}^{(k)}\| \|\mathbf{b}\|}{\|\mathbf{x}\| \|\mathbf{b}\|} = \|A^{-1}\| \frac{\|\mathbf{b}\|}{\|\mathbf{x}\|} \frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{b}\|} = \|A^{-1}\| \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{b}\|} \\ \frac{\|\mathbf{e}^{(k)}\|}{\|\mathbf{x}\|} &\leq \|A\| \|A^{-1}\| \frac{\|\mathbf{r}^{(k)}\|}{\|\mathbf{b}\|}, \end{aligned}$$

ou seja, $\|\mathbf{e}^{(k)}/\|\mathbf{x}\| \leq \text{cond}(A) \|\mathbf{r}^{(k)}/\|\mathbf{b}\|$, sendo $\text{cond}(A) = \|A\| \|A^{-1}\|$ o *número de condição* da matriz A . Como $\text{cond}(A) \geq 1$, a condição de erro relativo pequeno é mais restritiva que a condição de resíduo relativo pequeno. Analogamente, a condição de convergência é mais restritiva que a condição de consistência. Falta identificar o conceito correspondente ao de número de condição.

OBS: A expressão (14.5) corresponde à equação (11.1) apenas no caso particular em que $a_0 + a_1 = 0$ e $a_2 = \dots = a_k = 0$.

14.1 Estabilidade

Vamos considerar o PVI

$$\begin{cases} x'(t) = \lambda x(t) \\ x(0) = x_0, \end{cases} \quad \lambda < 0, \quad x_0 \neq 0, \quad (14.6)$$

cuja solução é $x^\lambda(t) = x_0 e^{\lambda t}$. Temos que

$$|x^\lambda(t)| = |x_0| |e^{\lambda t}| \leq |x_0| \quad \forall t \geq 0.$$

Assim, esperamos que a sequência de soluções aproximadas $\{X_0^\lambda, X_1^\lambda, \dots\}$ para o PVI (14.6) também satisfaça $|X_n^\lambda| \leq |x_0| = |X_0^\lambda|$ para todo $n \geq 0$.

Vejamos o que acontece com o método de Euler explícito:

$$X_{n+1}^\lambda = X_n^\lambda + hf(t_n, X_n^\lambda) = X_n^\lambda + h\lambda X_n^\lambda = (1 + h\lambda)X_n^\lambda.$$

Assim,

$$X_n^\lambda = (1 + h\lambda)X_{n-1}^\lambda = (1 + h\lambda)^2 X_{n-2}^\lambda = \dots = (1 + h\lambda)^n X_{n-n}^\lambda = (1 + h\lambda)^n x_0,$$

logo $|X_n^\lambda| = |1 + h\lambda|^n |x_0|$. Para que $|X_n^\lambda| \leq |x_0|$, devemos ter $|1 + h\lambda| \leq 1$, ou seja,

$$-2 \leq h\lambda \leq 0.$$

Como $\lambda < 0$, $h\lambda \leq 0$ segue trivialmente. Portanto, basta impor que $h\lambda \geq -2$, ou seja, $h \leq 2/|\lambda|$. Assim, para qualquer $\lambda \leq 0$, sempre podemos escolher h de modo que $|X_n^\lambda| \leq |x_0|$.

Definição 14.3 : A *região de estabilidade absoluta* do método (14.2) é o conjunto

$$R = \left\{ z \in \mathbb{C} \mid z = h\lambda, \text{ com } h \text{ tal que } |X_n^\lambda| \leq |x_0| \quad \forall n \geq 0 \right\}.$$

Definição 14.4 : Dizemos que o método (14.2) é *0-estável* (ou *estável*) se $0 \in R$.

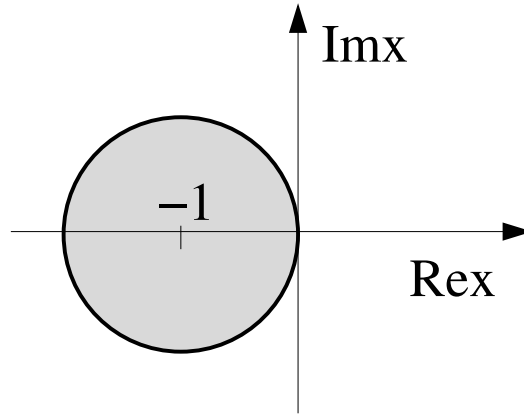


Figura 7: Região de estabilidade absoluta do método de Euler explícito.

Teorema 14.1 : O método (14.2) é convergente se, e somente se, é consistente e estável (vide [10, Teor. 1, Sec. 8.5]).

A região de estabilidade absoluta do método de Euler explícito é $R = \{z \in \mathbb{C} \mid |1 + z| \leq 1\}$, ou seja, o círculo aberto de raio 1 e centro em $z_0 = -1 + 0i$ (Fig. 7). Como $|1 + 0| = 1 \leq 1$, o método de Euler explícito é 0-estável.

Vamos verificar que método de Euler explícito também é consistente. Temos que o erro de truncamento satisfaz

$$\begin{aligned}\tau_n &= x(t_n) - [x(t_{n-1}) + hf(t_{n-1}, x(t_{n-1}))] \\ &= x(t_n) - [x(t_{n-1}) + hx'(t_{n-1})]\end{aligned}$$

Por outro lado, o termo entre colchetes corresponde ao polinômio de Taylor de grau 1 de $x(t)$ em torno do ponto t_{n-1} , logo

$$x(t_n) = x(t_{n-1}) + hx'(t_{n-1}) + \frac{h}{2}x''(\xi), \quad t_{n-1} \leq \xi \leq t_n.$$

Assumindo que $x(t)$ é de classe C^2 , obtemos $\tau_n = \mathcal{O}(h^2)$, logo $\tau_n/h = \mathcal{O}(h) \rightarrow 0$.

Segue do Teor. 14.1 que o método de Euler explícito é convergente.

Aula 15: Consistência dos métodos de passos múltiplos

Associamos ao método de passos múltiplos

$$\sum_{j=0}^k a_j X_{n+1-j} = h \sum_{j=0}^k b_j f_{n+1-j}, \quad a_0 \neq 0, \quad (15.1)$$

os seguintes polinômios:

$$p(x) = \sum_{j=0}^k a_j x^{k-j} \quad \text{e} \quad q(x) = \sum_{j=0}^k b_j x^{k-j}. \quad (15.2)$$

Vamos estudar a consistência do método (15.1). Segundo a notação utilizada na Aula 14,

$$X_n = \Phi([X_0, X_1, \dots, X_n]^T, t_0, T, h) = \frac{1}{a_0} \left[- \sum_{j=1}^k a_j X_{n-j} + h \sum_{j=0}^k b_j f_{n-j} \right],$$

e de (14.4),

$$\tilde{X}_n = \frac{1}{a_0} \left[- \sum_{j=1}^k a_j x(t_{n-j}) + h \sum_{j=0}^k b_j f(t_{n-j}, x_{n-j}) \right].$$

Assim, o erro de truncamento local τ_n satisfaz

$$\begin{aligned} a_0 \tau_n &= a_0 x(t_n) - a_0 \tilde{X}_n \\ &= a_0 x(t_n) + \sum_{j=1}^k a_j x(t_{n-j}) - h \sum_{j=0}^k b_j f(t_{n-j}, x(t_{n-j})) \\ &= \sum_{j=0}^k [a_j x(t_{n-j}) - h b_j x'(t_{n-j})], \end{aligned}$$

logo,

$$|\tau_n| = \frac{2}{|a_0|} \left| \sum_{j=0}^k a_j x(t_{n-j}) - h b_j x'(t_{n-j}) \right| \quad (15.3)$$

Pelo Teorema do Valor Médio, existe $t_{n-k} \leq \bar{t}_{1,j} \leq t_{n-j}$ tal que

$$x'(t_{n-j}) = x'(t_{n-k}) + h(k-j)x''(\bar{t}_{1,j}).$$

Além disso, a expansão de Taylor de segunda ordem de $x(t_{n-j})$ em torno de t_{n-k} satisfaz

$$\begin{aligned} x(t_{n-j}) &= x(t_{n-k}) + x'(t_{n-k})(t_{n-j} - t_{n-k}) + x''(\bar{t}_j^1)(t_{n-j} - t_{n-k})^2/2 \\ &= x(t_{n-k}) + x'(t_{n-k})[(n-j)h - t_0 - (n-k)h + t_0] \\ &+ x''(\bar{t}_j^1)[(n-j)h - t_0 - (n-k)h + t_0]^2/2 \\ &= x(t_{n-k}) + h(k-j)x'(t_{n-k}) + h^2(k-j)^2x''(\bar{t}_{2,j})/2 \end{aligned}$$

Para algum $t_{n-k} \leq \bar{t}_{2,j} \leq t_{n-j}$. Por outro lado, assumindo que $x \in C^2$, existe $M_2 > 0$ tal que

$$|x''(t)| \leq M_2 \quad \forall t \in [t_0, T].$$

Assim, segue de (15.3) que

$$\begin{aligned}
|\tau_n| &= \frac{2}{|a_0|} \left| \sum_{j=0}^k a_j [x(t_{n-k}) + h(k-j)x'(t_{n-k}) + \frac{h^2}{2}(k-j)^2 x''(\bar{t}_{2,j})] - hb_j [x'(t_{n-k}) + h(k-j)x''(\bar{t}_{1,j})] \right| \\
&= \frac{2}{|a_0|} \left| \left(\sum_{j=0}^k a_j \right) x(t_{n-k}) + h \left(\sum_{j=0}^k a_j (k-j) - b_j \right) x'(t_{n-k}) \right. \\
&\quad \left. + h^2 \sum_{j=0}^k \left[a_j \frac{(k-j)^2}{2} x''(\bar{t}_{2,j}) - b_j (k-j) x''(\bar{t}_{1,j}) \right] \right|
\end{aligned}$$

Escolhendo os coeficientes a_j e b_j tais que as duas primeiras parcelas acima se anulem, ou seja,

$$\sum_{j=0}^k a_j = 0, \quad \sum_{j=0}^k a_j (k-j) = \sum_{j=0}^{k-1} a_j (k-j) = \sum_{j=0}^k b_j,$$

ou em termos dos polinômios p e q definidos em (15.2),

$$p(1) = 0, \quad p'(1) = q(1), \tag{15.4}$$

encontramos a seguinte expressão para o erro de truncamento:

$$\begin{aligned}
|\tau_n| &= \frac{2}{|a_0|} \left| h^2 \sum_{j=0}^k a_j \frac{(k-j)^2}{2} x''(\bar{t}_{2,j}) - b_j (k-j) x''(\bar{t}_{1,j}) \right| \\
&\leq \frac{2h^2}{|a_0|} \sum_{j=0}^k \left[|a_j| \frac{(k-j)^2}{2} |x''(\bar{t}_{2,j})| + |b_j| (k-j) |x''(\bar{t}_{1,j})| \right] \\
&\leq \frac{2h^2}{|a_0|} M_2 \sum_{j=0}^k \left[|a_j| \frac{(k-j)^2}{2} + |b_j| (k-j) \right] \\
&\leq \frac{2h^2}{|a_0|} M_2 \sum_{j=0}^k \left[\max_{0 \leq j \leq k} |a_j| \frac{k^2}{2} + \max_{0 \leq j \leq k} |b_j| k \right] = Ch^2,
\end{aligned}$$

com $C = M_2(k+1) (\max_{0 \leq j \leq k} |a_j| k^2 + 2 \max_{0 \leq j \leq k} |b_j| k) / |a_0|$. Portanto, $\tau_n = \mathcal{O}(h^2)$ se (15.4) vale. Em particular, $\tau_n/h \rightarrow 0$, ou seja, o método (15.1) é consistente.

15.1 Condição alternativa de consistência

As fórmulas de Adams-Bashforth e Adams-Moulton foram obtidas nas Aulas 11 e 12 impondo-se $X_n = x(t_n)$ no caso em que $f(t, x(t)) \in \mathcal{P}_{k-1}$. Vamos verificar que esta propriedade também garante a consistência do método.

Assuma que $x \in C^{k+1}$ e seja $p_k(t)$ o polinômio de Taylor de grau k de $x(t)$ em torno de $t = t_{n-k}$. Como $x^{(j)}(t_{n-k}) = p_k^{(j)}(t_{n-k})$ para todo $0 \leq j \leq k$, temos que $r(t) = x(t) - p_k(t)$ satisfaz

$$r^{(j)}(t_{n-k}) = 0, \quad 0 \leq j \leq k.$$

Como $x, p_k \in C^{k+1}$ e $x(t) = p_k(t) + r(t)$, temos também que $r \in C^{k+1}$. Segue de (15.3) que

$$\begin{aligned}
|\tau_n| &= \frac{2}{|a_0|} \left| \sum_{j=0}^k a_j [p_k(t_{n-j}) + r(t_{n-j})] - hb_j [p_k'(t_{n-j}) + r'(t_{n-j})] \right| \\
&\leq \frac{2}{|a_0|} \left| \sum_{j=0}^k a_j p_k(t_{n-j}) - hb_j p_k'(t_{n-j}) \right| + \frac{2}{|a_0|} \left| \sum_{j=0}^k a_j r(t_{n-j}) - hb_j r'(t_{n-j}) \right|.
\end{aligned}$$

Tratemos primeiro da segunda parcela. A expansão de Taylor de $r(t_{n-j})$ em torno de $t = t_{n-k}$ resulta em

$$r(t_{n-j}) = \sum_{j=0}^k \frac{r^{(j)}(t_{n-k})}{j!} (t_{n-j} - t_{n-k})^j + \frac{r^{(k+1)}(\bar{t}_1)}{(k+1)!} (t_{n-j} - t_{n-k})^{k+1} = \frac{r^{(k+1)}(\bar{t}_1)}{(k+1)!} ((k-j)h)^{k+1},$$

para algum $t_{n-k} \leq \bar{t}_1 \leq t_{n-j}$. Por meio da expansão de Taylor de $r'(t_{n-j})$ em torno de $t = t_{n-k}$ obtemos, analogamente, que para algum $t_{n-k} \leq \bar{t}_2 \leq t_{n-j}$,

$$r'(t_{n-j}) = \sum_{j=0}^{k-1} \frac{r^{(j+1)}(t_{n-k})}{j!} (t_{n-j} - t_{n-k})^j + \frac{r^{(k+1)}(\bar{t}_2)}{k!} (t_{n-j} - t_{n-k})^k = \frac{r^{(k+1)}(\bar{t}_2)}{k!} ((k-j)h)^k.$$

Como $r \in C^{k+1}$, existe $M > 0$ tal que $|r^{(k+1)}(t)| \leq M$ para todo $t \in [t_0, T]$. Segue que

$$\left| \sum_{j=0}^k a_j r(t_{n-j}) - h b_j r'(t_{n-j}) \right| \leq \sum_{j=0}^k \frac{|a_j| M}{(k+1)!} j^{k+1} h^{k+1} + \frac{|b_j| M}{k!} j^k h^{k+1} \leq \left(\frac{k^{k+2} M}{(k+1)!} \max_{0 \leq j \leq k} \{|a_j|, |b_j|\} \right) h^{k+1},$$

de modo que

$$|\tau_n| = \frac{2}{|a_0|} \left| \sum_{j=0}^k a_j p_k(t_{n-j}) - h b_j p'_k(t_{n-j}) \right| + \mathcal{O}(h^{k+1}). \quad (15.5)$$

Por outro lado, como $p_k(t)$ é um polinômio de grau k , temos que $p'_k(t)$ é um polinômio de grau $k-1$, logo $p_k(t)$ é a solução do PVI

$$\begin{cases} x'(t) = \tilde{p}(t) \\ x(t_0) = p_k(t_0), \end{cases} \quad \tilde{p} \in \mathcal{P}_{k-1}.$$

Portanto, se o método de passos múltiplos

$$\sum_{j=0}^k a_j X_{n+1-j} = h \sum_{j=0}^k b_j f_{n+1-j}, \quad a_0 \neq 0$$

satisfaz $X_n = x(t_n)$ no caso em que $f(t, x(t)) \in \mathcal{P}_{k-1}$, então no caso particular em que $f(t, x(t)) = \tilde{p}$, temos $X_n = p_k(t_n)$, e assim,

$$\sum_{j=0}^k a_j p_k(t_{n+1-j}) = \sum_{j=0}^k a_j X_{n+1-j} = h \sum_{j=0}^k b_j \tilde{p}(t_{n+1-j}) = h \sum_{j=0}^k b_j p'_k(t_{n+1-j}),$$

ou seja, a primeira parcela em (15.5) se anula, comprovando que $\tau_n = \mathcal{O}(h^{k+1})$.

Aula 16: Convergência dos métodos de passos múltiplos

Na aula anterior, obtivemos as condições de consistência do método (15.1). Busquemos agora as condições de estabilidade, de modo a garantir a convergência.

Se $f(t, x) = \lambda x$ com $\lambda < 0$, então segue de (15.1) que

$$\sum_{j=0}^k a_j X_{n+1-j}^\lambda = h\lambda \sum_{j=0}^k b_j X_{n+1-j}^\lambda. \quad (16.1)$$

Busquemos soluções da equação acima na forma $X_k^\lambda = r^k$, com $r \neq 0$:

$$\sum_{j=0}^k a_j r^{n+1-j} = h\lambda \sum_{j=0}^k b_j r^{n+1-j}.$$

Multiplicando ambos os lados por r^{k-n-1} , obtemos

$$\sum_{j=0}^k a_j r^{k-j} = h\lambda \sum_{j=0}^k b_j r^{k-j} \implies p(r) = h\lambda q(r),$$

ou seja, r é uma raiz do polinômio $p_z(x) = p(x) - zq(x)$, $z = h\lambda$. Como p_z tem grau $\leq k$, existem $r_1, \dots, r_k \in \mathbb{C}$ tais que $p_z(r_k) = 0$. Se estas raízes forem distintas, a solução geral de (16.1) será

$$X_n^\lambda = \sum_{j=1}^k c_j X_n^{(j)}, \quad X_n^{(j)} = r_j^n. \quad (16.2)$$

Para que $|X_n^\lambda| \leq |X_0^\lambda| = |c_1 + \dots + c_k|$ para qualquer dado inicial $X_0^\lambda = x_0$, devemos ter $|r_j| \leq 1$ para todo j entre 1 e k . Vamos considerar o caso de raízes repetidas, por exemplo $r_k = r_{k-1}$. Neste caso, as soluções $X_n^{(k-1)} = r_{k-1}^n$ e $X_n^{(k)} = r_k^n$ são linearmente dependentes, logo uma delas deve ser substituída em (16.2) por outra solução. Veremos a seguir que esta solução é $\tilde{X}_n = nr_k^n$:

$$\begin{aligned} \sum_{j=0}^k a_j (n+1-j) r_k^{n+1-j} &= h\lambda \sum_{j=0}^k b_j (n+1-j) r_k^{n+1-j} \\ \sum_{j=0}^k a_j (n+1-j) r_k^{k-j} &= h\lambda \sum_{j=0}^k b_j (n+1-j) r_k^{k-j} \\ \sum_{j=0}^k (\textcolor{red}{n} + 1 - \textcolor{red}{k} + k - j) a_j r_k^{k-j} &= h\lambda \sum_{j=0}^k (\textcolor{red}{n} + 1 - \textcolor{red}{k} + k - j) b_j r_k^{k-j} \\ (n+1-k) \sum_{j=0}^k a_j r_k^{k-j} + \sum_{j=0}^k (k-j) a_j r_k^{k-j} &= h\lambda \left[(n+1-k) \sum_{j=0}^k b_j r_k^{k-j} + \sum_{j=0}^k (k-j) b_j r_k^{k-j} \right] \\ (n+1-k) \left[\sum_{j=0}^k a_j r_k^{k-j} - h\lambda \sum_{j=0}^k b_j r_k^{k-j} \right] &= - \sum_{j=0}^k (k-j) a_j r_k^{k-j} + h\lambda \sum_{j=0}^k (k-j) b_j r_k^{k-j} \\ (n+1-k) \left[\sum_{j=0}^k a_j r_k^{k-j} - h\lambda \sum_{j=0}^k b_j r_k^{k-j} \right] &= -\textcolor{red}{r}_k \left(\sum_{j=0}^k (k-j) a_j r_k^{k-j-\textcolor{red}{1}} - h\lambda \sum_{j=0}^k (k-j) b_j r_k^{k-j-\textcolor{red}{1}} \right) \\ (n+1-k)[p(r_k) - h\lambda q(r_k)] &= -r_k[p'(r_k) - h\lambda q'(r_k)] \\ (n+1-k)p_z(r_k) &= -r_k p'_z(r_k). \end{aligned}$$

Como r_k é uma raiz dupla, temos $p_z(r_k) = p'_z(r_k) = 0$, logo a equação acima é satisfeita. Assim, a solução geral passa a ser

$$X_n^\lambda = \sum_{j=1}^{k-1} c_k r_j^n + c_k n r_k^n. \quad (16.3)$$

Como $|n r_k^n| \rightarrow \infty$, devemos ter $|r_k| < 1$ para que $|X_n^\lambda| \leq |X_0^\lambda|$. Em geral, devemos ter $|r_j| < 1$ para qualquer raiz r_j de multiplicidade maior que 1.

Portanto, a região de estabilidade absoluta do método (15.1) é

$$R = \left\{ z \in \mathbb{C} \mid p_z(r) = p(r) - zq(r) = 0 \implies \begin{cases} |r| \leq 1, & \text{raiz simples;} \\ |r| < 1, & \text{raiz múltipla.} \end{cases} \right\}.$$

OBS: A solução $X_n^{(j)}$ correspondente uma raiz nula $r_j = 0$ é a solução trivial $X_n^{(j)} = 0$. Note que, neste caso, $|c_j X_n^{(j)}| \leq x_0$ para qualquer x_0 .

Se $z = 0$, então $p_z(x) = p(x)$. Portanto, para que $0 \in R$, devemos ter

$$p(r) = 0 \implies \begin{cases} |r| \leq 1, & \text{raiz simples;} \\ |r| < 1, & \text{raiz múltipla.} \end{cases} \quad (16.4)$$

Segue de (15.4) e (16.4) que, para que o método (15.1) seja convergente, devemos ter

$$\begin{cases} p(1) = 0 \\ p'(1) = q(1) \\ p(r) = 0 \implies \begin{cases} |r| \leq 1, & \text{raiz simples;} \\ |r| < 1, & \text{raiz múltipla.} \end{cases} \end{cases} \quad (16.5)$$

Exemplo 16.1 : Verifique que o método de Milne,

$$X_{n+1} - X_{n-1} = \frac{h}{3}(f_{n+1} + 4f_n + f_{n-1}),$$

é convergente.

O valor de k corresponde ao menor sub-índice presente na equação $(n-1)$, ou seja,

$$n+1-k = n-1 \implies k = 2.$$

Os coeficientes a_k e b_k são dados por

$$a_0 = 1, a_1 = 0, a_2 = -1, \quad b_0 = 1/3, b_1 = 4/3, b_2 = 1/3,$$

logo

$$\begin{aligned} p(x) &= a_0 x^2 + a_1 x + a_2 = x^2 - 1 = (x-1)(x+1), \\ q(x) &= b_0 x^2 + b_1 x + b_2 = \frac{1}{3}(x^2 + 4x + 1). \end{aligned}$$

Note que $p(1) = 0$, $p'(1) = q(1) = 2$ e que as duas raízes de $p(x)$, $r = \pm 1$, têm multiplicidade 1 (raízes simples) e satisfazem $|z| \leq 1$. Portanto, o método de Milne é convergente.

Aula 17: Sistemas de EDOs

Vamos considerar sistemas de equações diferenciais ordinárias (EDOs) na forma

$$\begin{cases} x'_1(t) = f_1(t, x_1(t), x_2(t), \dots, x_d(t)) \\ x'_2(t) = f_2(t, x_1(t), x_2(t), \dots, x_d(t)) \\ \vdots \\ x'_d(t) = f_d(t, x_1(t), x_2(t), \dots, x_d(t)) \end{cases} \quad \text{com condições iniciais} \quad \begin{cases} x_1(t_0) = x_{1,0} \\ x_2(t_0) = x_{2,0} \\ \vdots \\ x_d(t_0) = x_{d,0}. \end{cases}$$

Podemos escrever este problema na forma vetorial:

$$\begin{cases} \mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t)) \\ \mathbf{x}(t_0) = \mathbf{x}_0, \end{cases} \quad (17.1)$$

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_d(t) \end{bmatrix}, \quad \mathbf{f}(t, \mathbf{x}(t)) = \begin{bmatrix} f_1(t, x_1(t), x_2(t), \dots, x_d(t)) \\ f_2(t, x_1(t), x_2(t), \dots, x_d(t)) \\ \vdots \\ f_d(t, x_1(t), x_2(t), \dots, x_d(t)) \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} x_{1,0} \\ x_{2,0} \\ \vdots \\ x_{d,0} \end{bmatrix}.$$

A dimensão d destes sistemas depende do problema estudado. Por exemplo,

- No movimento de uma partícula, $1 \leq d \leq 3$;
- Em dinâmica populacional, $d \sim 10^0 - 10^1$;
- Em reações químicas, $d \sim 10^1 - 10^2$;

Os métodos numéricos vistos anteriormente são facilmente adaptados para sistemas de EDOs, quando escritos na forma (17.1). Por exemplo, o método de Euler explícito assume a forma

$$\mathbf{X}_{n+1} = \mathbf{X}_n + h\mathbf{f}(t_n, \mathbf{X}_n), \quad \mathbf{X}_n = \begin{bmatrix} X_{1,n} \\ \vdots \\ X_{d,n} \end{bmatrix},$$

enquanto o método de Euler modificado se torna

$$\begin{cases} \mathbf{F}_1 = h\mathbf{f}(t_n, \mathbf{X}_n) \\ \mathbf{F}_2 = h\mathbf{f}(t_n + (1/2)h, \mathbf{X}_n + (1/2)\mathbf{F}_1) \\ \mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{F}_2. \end{cases}$$

Exemplo 17.1 : Use uma iteração do método de Euler modificado para aproximar $\mathbf{x}(0.1)$, sendo $\mathbf{x}(t)$ a solução do PVI

$$\begin{cases} x'_1(t) = 2x_1(t) + x_2(t) \\ x'_2(t) = x_1(t) + 2x_2(t) \\ x_1(0) = 1 \\ x_2(0) = 1, \end{cases} \quad (17.2)$$

ou na forma matricial,

$$\begin{cases} \mathbf{x}(t) = A\mathbf{x}(t) \\ \mathbf{x}(0) = \mathbf{x}_0, \end{cases} \quad A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Temos que $h = 0.1$ e $\mathbf{f}(t, \mathbf{x}) = \mathbf{f}(t, x_1, x_2) = [2x_1 + x_2, x_1 + 2x_2]^T$, logo

$$\begin{aligned}\mathbf{F}_1 &= h\mathbf{f}(t_0, \mathbf{X}_0) = (0.1)\mathbf{f}(0, [1, 1]^T) = 0.1[2 + 1, 1 + 2]^T = [0.3, 0.3]^T \\ \mathbf{F}_2 &= h\mathbf{f}(t_0 + (1/2)h, \mathbf{X}_0 + (1/2)\mathbf{F}_1) = (0.1)\mathbf{f}(0.05, [1, 1]^T + (1/2)[0.3, 0.3]^T) \\ &= (0.1)\mathbf{f}(0.05, [1.15, 1.15]^T) = 0.1[2.3 + 1.15, 2.3 + 1.15]^T = [0.345, 0.345]^T \\ \mathbf{X}_1 &= \mathbf{X}_0 + \mathbf{F}_2 = [1, 1]^T + [0.345, 0.345]^T = [1.345, 1.345]^T.\end{aligned}$$

O PVI (17.2) é um exemplo de sistema linear de equações diferenciais ordinárias, utilizado por exemplo na análise de circuitos elétricos. Podemos encontrar a solução exata por técnicas semelhantes às aquelas utilizadas para uma única equação.

De fato, substituindo $x_1(t) = C_1 e^{rt}$ e $x_2(t) = C_2 e^{rt}$ nas equações diferenciais, encontramos

$$\begin{cases} rC_1 e^{rt} = 2C_1 e^{rt} + C_2 e^{rt} \\ rC_2 e^{rt} = C_1 e^{rt} + 2C_2 e^{rt} \end{cases} \implies \begin{cases} rC_1 = 2C_1 + C_2 \\ rC_2 = C_1 + 2C_2, \end{cases}$$

que corresponde ao problema de autovalores

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = r \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \quad (\text{ou } \mathbf{A}\mathbf{C} = r\mathbf{C}),$$

cujas soluções são

$$r_1 = 1, \quad \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{e} \quad r_2 = 3, \quad \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Assim, a solução geral é

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{1t} + \beta \begin{bmatrix} 1 \\ 1 \end{bmatrix} e^{3t} = \begin{bmatrix} \alpha e^t + \beta e^{3t} \\ -\alpha e^t + \beta e^{3t} \end{bmatrix} = \begin{bmatrix} e^t & e^{3t} \\ -e^t & e^{3t} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Escrevendo as condições iniciais na forma vetorial, encontramos

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \implies \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Portanto,

$$\begin{aligned}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} &= \begin{bmatrix} e^t & e^{3t} \\ -e^t & e^{3t} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} e^t & 0 \\ 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} e^t + e^{3t} & -e^t + e^{3t} \\ -e^t + e^{3t} & e^t + e^{3t} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned} \tag{17.3}$$

Segue de (17.3) que $\mathbf{x}(0.1) = [e^{0.3}, e^{0.3}]^T = [1.3498\dots, 1.3498\dots]^T$. Note que $\mathbf{X}_1$ foi uma boa aproximação para $\mathbf{x}(0.1)$.

Convém representar a solução (17.3) na forma matricial por $\mathbf{x}(t) = e^{At}\mathbf{x}_0$. Para justificar esta notação, observe que

$$\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1}, \quad \mathbf{Q} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} = \text{diag}(r_1, r_2),$$

com $r_1 = 1$ e $r_3 = 3$. Segue que $\mathbf{A}^K = \mathbf{Q}\mathbf{D}^K\mathbf{Q}^{-1}$, $\mathbf{D}^K = \text{diag}(r_1^K, r_2^K)$. Assim,

$$e^{At} = \sum_{k=0}^{\infty} \frac{1}{k!} (At)^k = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{Q}\mathbf{D}^K\mathbf{Q}^{-1}t^k = \mathbf{Q} \left(\sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{D}^K t^k \right) \mathbf{Q}^{-1} = \mathbf{Q} e^{Dt} \mathbf{Q}^{-1},$$

sendo $e^{Dt} = \text{diag}(e^{r_1 t}, e^{r_2 t}) = \text{diag}(e^t, e^{3t})$.

Aula 18: Sistemas rígidos de equações

Dado o sistema de EDOs $\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t))$, seja $J(t)$ a função matricial definida por

$$J(t) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(t, \mathbf{x}(t)) & \frac{\partial f_1}{\partial x_2}(t, \mathbf{x}(t)) & \dots & \frac{\partial f_1}{\partial x_d}(t, \mathbf{x}(t)) \\ \frac{\partial f_2}{\partial x_1}(t, \mathbf{x}(t)) & \frac{\partial f_2}{\partial x_2}(t, \mathbf{x}(t)) & \dots & \frac{\partial f_2}{\partial x_d}(t, \mathbf{x}(t)) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_d}{\partial x_1}(t, \mathbf{x}(t)) & \frac{\partial f_d}{\partial x_2}(t, \mathbf{x}(t)) & \dots & \frac{\partial f_d}{\partial x_d}(t, \mathbf{x}(t)) \end{bmatrix},$$

e sejam $\lambda_i(t) \in \mathbb{C}$ ($1 \leq i \leq d$) os autovalores de $J(t)$.

Definição 18.1 : Dizemos que o sistema de EDOs $\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t))$ é rígido (*stiff*) no intervalo $[t_0, T]$ se

1. $\text{Re}(\lambda_i(t)) < 0 \quad \forall t \in [t_0, T], \quad 1 \leq i \leq d;$
2. $\max_{1 \leq i \leq d} |\text{Re}(\lambda_i(t))| \gg 1 \quad \forall t \in [t_0, T].$

Se $d = 1$ e $f_1 = f$, ou seja, se $x(t) = f(t, x(t)) \in \mathbb{R}$, as condições acima se reduzem a

1. $\frac{\partial f}{\partial x}(t, x(t)) < 0 \quad \forall t \in [t_0, T];$
2. $|\frac{\partial f}{\partial x}(t, x(t))| \gg 1 \quad \forall t \in [t_0, T],$

o que acontece, por exemplo, se $f(t, x(t)) = \lambda x(t)$ com $\lambda < 0$ e $|\lambda| \gg 1$.

Para resolver sistemas rígidos de equações, em geral precisamos de métodos A-estáveis.

Definição 18.2 : Um método é A-estável se o conjunto

$$\mathbb{C}^- = \{z \in \mathbb{C} \mid \text{Re}(z) < 0\}$$

está contido na região de estabilidade absoluta do método.

Exemplo 18.1 : Mostre que o método de Euler implícito é A-estável.

O método de Euler implícito para $f(t, x) = \lambda x$ é dado por

$$X_{n+1} = X_n + hf(t_{n+1}, X_{n+1}) = X_n + h\lambda X_{n+1},$$

ou seja,

$$X_{n+1} = \frac{1}{1 - h\lambda} X_n = \dots = \left(\frac{1}{1 - h\lambda} \right)^{n+1} X_0.$$

Portanto, para que $|X_n| \leq |X_0|$, devemos ter $|1 - h\lambda| \geq 1$. Portanto, a região de estabilidade absoluta do método de Euler implícito é

$$R = \{z \in \mathbb{C} \mid |1 - z| \geq 1\}.$$

Seja $z = a + ib \in \mathbb{C}^-$. Temos que $a < 0$, logo

$$|1 - z|^2 = (1 - a)^2 + b^2 = (1 + |a|)^2 + b^2 \geq (1 + |a|)^2 > 1,$$

ou seja, $z \in R$. Portanto, o método de Euler implícito é A-estável.

OBS: Outro exemplo de método A-estável é o método trapezoidal.

Aula 19: Problemas de Valores de Contorno

Daqui em diante estudaremos métodos numéricos para problemas de valores de contorno (PVC)

$$\begin{cases} x''(t) = f(t, x(t), x'(t)), & t \in]a, b[\\ a_0x(a) - a_1x'(a) = \alpha \\ b_0x(b) + b_1x'(b) = \beta. \end{cases} \quad (19.1)$$

Temos as seguintes condições de existência e unicidade [9, Teor. 1.2.2]:

Teorema 19.1 : Seja D o conjunto dado por

$$D = \{(t, x, y) \in \mathbb{R}^3 \mid t \in [a, b]\}.$$

Seja $f : D \rightarrow \mathbb{R}$ tal que $f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \in C^0(D)$. Se existe $M > 0$ tal que

$$\frac{\partial f}{\partial x}(t, x, y) > 0, \quad |f(t, x_1, y) - f(t, x_2, y)| \leq M|x_1 - x_2|, \quad \left| \frac{\partial f}{\partial y}(t, x, y) \right| \leq M \quad (19.2)$$

para todo $t \in [a, b]$ e $x, y, x_1, x_2 \in \mathbb{R}$, e se os coeficientes das condições de contorno são tais que

1. $|a_0| + |a_1| \neq 0$ e $a_0a_1 \geq 0$;
2. $|b_0| + |b_1| \neq 0$ e $b_0b_1 \geq 0$;
3. $|a_0| + |b_0| \neq 0$,

então o PVC (19.1) tem solução única.

A segunda condição em (19.2) vale se $\left| \frac{\partial f}{\partial x}(t, x, y) \right| \leq M$. Além disso, note que as condições 1-3 acima são satisfeitas se $a_0 = b_0 = 1$ e $a_1 = b_1 = 0$, ou seja, se tivermos as *condições de Dirichlet*

$$x(a) = \alpha \quad \text{e} \quad x(b) = \beta.$$

Estas condições também são satisfeitas se substituirmos a condição de Dirichlet em uma das extremidades por uma *condição de Neumann*, por exemplo,

$$x(a) = \alpha \quad \text{e} \quad x'(b) = \beta.$$

Entretanto, o teorema é inconclusivo sobre condições de Neumann em ambas as extremidades,

$$x'(a) = \alpha \quad \text{e} \quad x'(b) = \beta,$$

pois teríamos $|a_0| + |b_0| = 0$. Os três exemplos a seguir mostram que a solução pode existir ou não, e ser ou não única.

Exemplo 19.1 : Estude a existência e unicidade de soluções do PVC

$$\begin{cases} x''(t) = 1, & t \in]0, 1[\\ x'(0) = 1 \\ x'(1) = 2. \end{cases}$$

Integrando a equação $x''(t) = 1$, encontramos a solução geral $x(t) = x^2/2 + C_1x + C_2$. Como $x'(t) = x + C_1$, as condições de contorno resultam em $C_1 = 1$, de modo que a solução do problema, $x(t) = x^2/2 + x + C_2$, sendo C_2 uma constante qualquer, não é única.

Exemplo 19.2 : Repita o exemplo anterior substituindo a segunda condição de contorno por $x'(0) = 0$.

Novamente, $x'(t) = x + C_1$ e as condições de contorno resultam no sistema impossível

$$\begin{cases} C_1 = 0 \\ 1 + C_1 = 2, \end{cases}$$

portanto neste caso o problema não tem solução.

Exemplo 19.3 : Estude a existência e unicidade de soluções do PVC

$$\begin{cases} x''(t) = x(t), & t \in]0, 1[\\ x'(0) = 1 \\ x'(1) = 2. \end{cases}$$

Neste problema, a solução geral é $x(t) = C_1 e^t + C_2 e^{-t}$. Como $x'(t) = C_1 e^t - C_2 e^{-t}$, as condições de contorno resultam no sistema linear

$$\begin{cases} C_1 + C_2 = 1 \\ C_1 e + C_2 e^{-1} = 2, \end{cases} \text{ ou seja, } \begin{bmatrix} 1 & 1 \\ e & e^{-1} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

que tem solução única, pois o determinante da matriz acima é igual a $1/e - e \neq 0$.

Exemplo 19.4 : Encontre condições sobre as funções u, v e w de modo a garantir a existência e unicidade de soluções do PVC linear

$$\begin{cases} x''(t) = u(t) + v(t)x(t) + w(t)x'(t), & t \in]a, b[\\ x(a) = \alpha \\ x(b) = \beta. \end{cases} \quad (19.3)$$

Temos que as condições 1-3 são satisfeitas. Além disso, $f(t, x, y) = u(t) + v(t)x + w(t)y$, logo

$$\frac{\partial f}{\partial x}(t, x, y) = v(t) \quad \text{e} \quad \frac{\partial f}{\partial y}(t, x, y) = w(t).$$

Assim, para que $f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \in C^0(D)$, basta que u, v, w sejam contínuas em $[a, b]$. Note que a continuidade garante em particular que v e w são limitadas em $[a, b]$, logo existe a constante $M > 0$ em (19.2). Finalmente, impondo $\frac{\partial f}{\partial x}(t, x, y) > 0$, obtemos a condição $v(t) > 0 \quad \forall t \in [a, b]$. Portanto, as condições são que u, v e w sejam contínuas em $[a, b]$ e que v seja positiva em $[a, b]$.

Exemplo 19.5 : Dado $L > 0$, estude a existência e unicidade de soluções do PVC

$$\begin{cases} x''(t) = -x(t), & t \in]0, L[\\ x(0) = 3 \\ x(L) = 7. \end{cases} \quad (19.4)$$

Este é um caso do particular do Exemplo 19.4 com $u(t) = w(t) = 0$ e $v(t) = -1$. Temos que u, v, w são contínuas, mas como $v(t) < 0$, o Teorema 19.1 não permite concluir se a solução do problema (19.4) existe ou é única. De fato, observe que a solução geral da equação $x''(t) = -x(t)$ é

$$x(t) = A \cos(t) + B \sin(t),$$

e pelas condições de contorno, as constantes A e B devem satisfazer as equações

$$\begin{cases} A &= 3 \\ A \cos(L) + B \sin(L) &= 7, \end{cases} \quad \text{ou seja,} \quad \begin{bmatrix} 1 & 0 \\ \cos(L) & \sin(L) \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

A existência e/ou unicidade do sistema acima depende do seu determinante

$$\det \left(\begin{bmatrix} 1 & 0 \\ \cos(L) & \sin(L) \end{bmatrix} \right) = \sin(L)$$

Assim a solução existe e é única se $\sin(L) \neq 0$ (por exemplo, se $0 < L < \pi$), e não existe se $\sin(L) = 0$ (por exemplo se $L = \pi/2$).

Exemplo 19.6 : Repita o Exemplo 19.5 com a equação $x''(t) = x(t)$.

Agora $v(t) = 1 > 0$, e podemos garantir que a solução existe e é única. De fato, a solução geral é

$$x(t) = Ae^t + Be^{-t},$$

e pelas condições de contorno, as constantes A e B satisfazem as equações

$$\begin{cases} A + B &= 3 \\ Ae^L + Be^{-L} &= 7, \end{cases}$$

cuja solução é $A = (7e^L - 3)/(e^{2L} - 1)$, $B = (3e^{2L} - 7e^L)/(e^{2L} - 1)$. Portanto, para qualquer $L > 0$,

$$x(t) = \frac{(7e^L - 3)e^t - (3e^{2L} - 7e^L)e^{-t}}{e^{2L} - 1}.$$

Note que a condição $\frac{\partial f}{\partial x}(t, x, y) > 0$ do Teorema 19.1 é muito restritiva. O Teorema a seguir [10, Teor. 4, Sec 8.7] relaxa as condições sobre a função f :

Teorema 19.2 : Seja $D = [0, 1] \times \mathbb{R}$. Se $f \in C^0(D)$ e existe $k \leq 8$ tal que $|f(t, x_1) - f(t, x_2)| \leq k|x_2 - x_1|$ para todo $(t, x_1), (t, x_2) \in D$, então existe uma única solução para o PVC

$$\begin{cases} x''(t) &= f(t, x(t)), & t \in]0, 1[\\ x(0) &= 0 \\ x(1) &= 0 \end{cases} \quad (19.5)$$

Por outro lado, o Teorema 19.2 restringe o domínio para o intervalo $]0, 1[$ e as condições de contorno para condições homogêneas, conforme observamos em (19.5). Estas restrições podem ser amenizadas por meio de transformações de variáveis.

Exemplo 19.7 : Repita o Exemplo 19.5 utilizando o Teorema 19.2.

Por meio das transformações $t = Ls$ e $y(s) = x(t(s))$, encontramos

$$\frac{\partial y}{\partial s}(s) = \frac{\partial t}{\partial s} \frac{\partial x}{\partial t}(t(s)) = L \frac{\partial x}{\partial t}(t(s)) \quad \text{e} \quad \frac{\partial^2 y}{\partial s^2}(s) = L^2 \frac{\partial^2 x}{\partial t^2}(t(s)),$$

ou seja, $y''(s) = L^2 x''(t(s)) = L^2(-x(t(s))) = -L^2 y(s)$. Além disso, $y(0) = x(0) = 3$ e $y(1) = x(L) = 7$, portanto $y(s)$ é a solução do PVC auxiliar

$$\begin{cases} y''(s) &= -L^2 y(s), & s \in]0, 1[\\ y(0) &= 3 \\ y(1) &= 7. \end{cases}$$

Para tornar as condições de contorno homogêneas, podemos subtrair da solução $y(s)$ o polinômio $y_1(s) \in \mathcal{P}_1$ tal que $y_1(0) = 3$ e $y_1(1) = 7$, ou seja, $y_1(s) = 4s + 3$, resultando na nova variável $w(s) = y(s) - (4s + 3)$. Note que

$$w''(s) = y''(s) = -L^2 y(s) = -L^2(w(s) + 4s + 3),$$

logo w é solução do PVC (19.5) com $f(s, w) = -L^2(w + 4s + 3)$, ou seja, $f(t, x) = -L^2(x + 4t + 3)$. Agora sim podemos aplicar o Teorema 19.2, estudando a expressão $|f(t, x_1) - f(t, x_2)|$:

$$|f(t, x_1) - f(t, x_2)| = L^2|x_2 - x_1|.$$

Pelo Teorema 19.2, a solução é única se $L \leq \sqrt{8}$. Esta condição é mais restritiva que a condição $L < \pi$ obtida estudando a solução exata do problema (19.4). Apesar disso, o Teorema 19.2 é mais conclusivo que o Teorema 19.1 neste exemplo, embora seja mais trabalhoso.

Com respeito às transformações de variáveis, temos em geral o seguinte resultado: (Teoremas 2 e 3, Sec. 8.7, em [10]):

Teorema 19.3 : A solução do PVC

$$\begin{cases} x''(t) &= f_1(t, x(t)), & t \in]a, b[\\ x(a) &= \alpha \\ x(b) &= \beta \end{cases} \quad (19.6)$$

existe e é única se, se somente se, o PVC

$$\begin{cases} y''(t) &= f_2(s, y(s)), & s \in]0, 1[\\ y(0) &= \alpha \\ y(1) &= \beta, \end{cases} \quad (19.7)$$

com $f_2(s, y) = (b - a)^2 f_1(a + (b - a)s, y)$, também tem solução única. O PVC (19.7), por sua vez, tem solução única se, se somente se, o PVC (19.5), com $f(t, x) = f_2(t, x + \alpha + (\beta - \alpha)t)$, tem solução única.

OBS: O Teorema 19.3 não é suficiente para tornar o Teorema 19.2 aplicável a problemas que envolvam condições de Neumann ou problemas em que $f = f(t, x, y)$.

Aula 20: Fórmulas de diferenças finitas

Seja $a = t_0 < t_1 < \dots < t_N = b$ uma partição uniforme do intervalo $[a, b]$ com

$$t_n = a + nh, \quad h = \frac{b - a}{N}.$$

Vamos denominar uma aproximação de $f^{(m)}(t_n)$ escrita apenas em termos de valores de $x(t)$ em pontos da partição em torno de $t = t_n$, ou seja,

$$\frac{d^m f}{dx^m}(t_n) \approx \sum_{j=-k}^k \gamma_j x(t_{n+j}) = \gamma_{-k} x(t_{n-k}) + \gamma_{-k+1} x(t_{n-k+1}) + \dots + \gamma_0 x(t_n) + \dots + \gamma_k x(t_{n+k})$$

uma *aproximação por diferenças finitas*. Observando que a derivada de $x(t)$ em $t = t_n$ pode ser escrita como

$$x'(t_n) = \lim_{h \rightarrow 0} \frac{x(t_n + h) - x(t_n)}{h},$$

uma das aproximações de diferenças finitas mais simples é dada por

$$x'(t_n) \approx \frac{x(t_n + h) - x(t_n)}{h} = \frac{x(t_{n+1}) - x(t_n)}{h}.$$

Nesta fórmula, $k = 1$, $\gamma_{-1} = 0$, $\gamma_0 = -1/h$ e $\gamma_1 = 1/h$. Seguem outros exemplos:

$$x'(t_n) \approx \frac{x(t_n) - x(t_{n-1}))}{h}, \quad x'(t_n) \approx \frac{x(t_{n+1}) - x(t_{n-1}))}{2h}, \quad x''(t_n) \approx \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1}))}{h^2}.$$

A precisão das fórmulas de diferenças finitas é determinada por meio do Teorema de Taylor.

Exemplo 20.1 : Mostre que existe $k > 0$ tal que, se $x(t)$ é de classe C^k , então

$$x''(t_n) = \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1}))}{h^2} + \mathcal{O}(h^2).$$

Queremos mostrar que

$$h^2 x''(t_n) = x(t_{n+1}) - 2x(t_n) + x(t_{n-1})) + h^2 \mathcal{O}(h^2) = x(t_{n+1}) - 2x(t_n) + x(t_{n-1})) + \mathcal{O}(h^4).$$

O fator $\mathcal{O}(h^4)$ sugere que tomemos $k = 4$, ou seja, vamos assumir que x é de classe C^4 . Neste caso, temos as seguintes expansões de $x(t_{n\pm 1})$:

$$x(t_{n+1}) = x(t_n) + hx'(t_n) + (h^2/2)x''(t_n) + (h^3/6)x'''(t_n) + (h^4/24)x^{(4)}(\bar{t}_1), \quad (20.1)$$

$$x(t_{n-1}) = x(t_n) - hx'(t_n) + (h^2/2)x''(t_n) - (h^3/6)x'''(t_n) + (h^4/24)x^{(4)}(\bar{t}_2), \quad (20.2)$$

com $t_n \leq \bar{t}_1 \leq t_{n+1}$ e $t_{n-1} \leq \bar{t}_2 \leq t_n$. Somando (20.1) com (20.2), obtemos

$$x(t_{n+1}) + x(t_{n-1})) = 2x(t_n) + h^2 x''(t_n) + (h^4/24)[x^{(4)}(\bar{t}_1) + x^{(4)}(\bar{t}_2)],$$

e isolando $x''(t_n)$ da expressão acima,

$$x''(t_n) = \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1}))}{h^2} + \frac{h^2}{24}[x^{(4)}(\bar{t}_1) + x^{(4)}(\bar{t}_2)].$$

Como $x \in C^4$, $\exists M \geq 0$ tal que $|x^{(4)}(t)| \leq M$ para todo $t \in [a, b]$. Além disso, $\bar{t}_1 \in [t_n, t_{n+1}] \subset [a, b]$ e $\bar{t}_2 \in [t_{n-1}, t_n] \subset [a, b]$, logo

$$\left| x''(t_n) - \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1}))}{h^2} \right| = \frac{h^2}{24} |x^{(4)}(\bar{t}_1) + x^{(4)}(\bar{t}_2)| \leq \frac{h^2}{24} (M + M) = Ch^2,$$

com $C = M/12$, portanto,

$$x''(t_n) - \frac{x(t_{n+1}) - 2x(t_n) + x(t_{n-1}))}{h^2} = \mathcal{O}(h^2).$$

Aula 21: Métodos de diferenças finitas

O método de diferenças finitas para o PVC (19.1) associado às aproximações por diferenças finitas

$$x''(t_n) \approx \sum_{j=-k}^k \gamma_j x(t_{n+j}) \quad \text{e} \quad x'(t_n) \approx \sum_{j=-k_n}^{k_n} \gamma_j^{(n)} x(t_{n+j})$$

é o método deduzido a partir dos seguintes passos:

1. Extrair uma equação para cada ponto t_n da partição do intervalo $[a, b]$. As equações dos pontos extremos dadas pelas condições de contorno, enquanto as equações dos pontos interiores são obtidas avaliando-se a equação diferencial em $t = t_n$.

$$\begin{cases} x''(t_n) = f(t_n, x(t_n), x'(t_n)), & 1 \leq n \leq N-1, \\ a_0 x(t_0) - a_1 x'(t_0) = \alpha, \\ b_0 x(t_N) + b_1 x'(t_N) = \beta. \end{cases}$$

2. Aproximar $x''(t_n)$ e $x'(t_n)$ pelas fórmulas de diferenças finitas acima:

$$\begin{cases} \sum_{j=-k_1}^{k_1} \gamma_j x(t_{n+j}) \approx f\left(t_n, X_n, \sum_{j=-k_n}^{k_n} \gamma_j^{(n)} x(t_{n+j})\right), & 1 \leq n \leq N-1, \\ a_0 x(t_0) - a_1 \sum_{j=-k_0}^{k_0} \gamma_j^{(0)} x(t_j) \approx \alpha, \\ b_0 x(t_N) + b_1 \sum_{j=-k_N}^{k_N} \gamma_j^{(N)} x(t_{N+j}) \approx \beta. \end{cases}$$

3. Aproximar $x(t_n)$ por X_n ($0 \leq n \leq N$), sendo $[X_0, \dots, X_N]^T$ tais que

$$\begin{cases} \sum_{j=-k_1}^{k_1} \gamma_j X_{n+j} = f\left(t_n, X_n, \sum_{j=-k_n}^{k_n} \gamma_j^{(n)} X_{n+j}\right), & 1 \leq n \leq N-1, \\ a_0 X_0 - a_1 \sum_{j=-k_0}^{k_0} \gamma_j^{(0)} X_{0+j} = \alpha, \\ b_0 X_N + b_1 \sum_{j=-k_N}^{k_N} \gamma_j^{(N)} X_{N+j} = \beta. \end{cases} \quad (21.1)$$

Note que as somatórias acima podem envolver X_n com $n < 0$ ou $n > N$. Veremos posteriormente como contornar esta dificuldade.

Exemplo 21.1 : Seja $f \in C^0([a, b])$. Construa o sistema de equações referente ao método de diferenças finitas para o PVC

$$\begin{cases} x''(t) = f(t), & t \in]a, b[\\ x(a) = \alpha \\ x(b) = \beta. \end{cases}$$

utilizando a seguinte fórmula de diferenças finitas para $x''(t_n)$:

$$x''(t_n) \approx \frac{x(t_{n-1}) - 2x(t_n) + x(t_{n+1}))}{h^2}. \quad (21.2)$$

Note primeiro que pode-se mostrar, por meio dos Teoremas 19.2 e 19.3, que basta $f \in C^0([a, b])$ para que este problema tenha solução única.

Segue de (21.1) que as equações são as seguintes:

$$\begin{cases} \frac{X_{n-1} - 2X_n + X_{n+1}}{h^2} = f(t_n), & 1 \leq n \leq N-1, \\ X_0 = \alpha, \\ X_N = \beta, \end{cases}$$

de onde imediatamente encontramos $X_0 = \alpha$ e $X_N = \beta$, restando calcular $X_1, \dots, X_{N-1}$ por meio das $N-1$ primeiras equações:

$$X_{n-1} - 2X_n + X_{n+1} = h^2 f(t_n), \quad 1 \leq n \leq N-1.$$

Note que, na primeira equação,

$$\begin{aligned} X_0 - 2X_1 + X_2 &= h^2 f(t_1) \\ \alpha - 2X_1 + X_2 &= h^2 f(t_1) \\ -2X_1 + X_2 &= h^2 f(t_1) - \alpha, \end{aligned}$$

e de modo análogo, a última equação se reduz a $X_{N-2} - 2X_{N-1} = h^2 f(t_{N-1}) - \beta$. Temos portanto o seguinte sistema linear:

$$\begin{cases} -2X_1 + X_2 &= h^2 f(t_1) - \alpha \\ X_1 - 2X_2 + X_3 &= h^2 f(t_2) \\ \vdots & \\ X_{N-3} - 2X_{N-2} + X_{N-1} &= h^2 f(t_{N-2}) \\ X_{N-2} - 2X_{N-1} &= h^2 f(t_{N-1}) - \beta, \end{cases}$$

que na forma matriz-vetor corresponde a

$$\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{N-2} \\ X_{N-1} \end{bmatrix} = h^2 \begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_{N-2}) \\ f(t_{N-1}) \end{bmatrix} - \begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \\ \beta \end{bmatrix}.$$

Os autovalores desta matriz são $\lambda_n = -2 + 2\cos(n\pi/N)$, $1 \leq n \leq N-1$ [14], de modo que $\lambda_n < 0$, em particular nenhum autovalor é nulo. Portanto, a matriz é não-singular e o sistema linear tem solução única.

Exemplo 21.2 : Seja $u, v, w \in C^0([a, b])$, com $v(t) > 0$ para todo $t \in [a, b]$. Construa o sistema de equações referente ao método de diferenças finitas para o PVC linear (19.3) utilizando fórmula de diferenças finitas (21.2) para $x''(t_n)$, e a seguinte fórmula para $x'(t_n)$:

$$x'(t_n) \approx \frac{x(t_{n+1}) - x(t_{n-1}))}{2h}.$$

Conforme visto no Exemplo 21.2 as condições do enunciado são suficientes para que o PVC tenha solução única. Novamente de (21.1),

$$\begin{cases} \frac{X_{n-1} - 2X_n + X_{n+1}}{h^2} = u(t_n) + v(t_n)X_n + w(t_n)\frac{X_{n+1} - X_{n-1}}{2h}, & 1 \leq n \leq N-1, \\ X_0 = \alpha, \\ X_N = \beta. \end{cases} \quad (21.3)$$

Manipulando algebricamente a primeira equação, obtemos:

$$X_{n-1} - 2X_n + X_{n+1} = h^2 u(t_n) + h^2 v(t_n) X_n + (h/2) w(t_n) (X_{n+1} - X_{n-1})$$

$$\begin{aligned} (1 + (h/2)w(t_n))X_{n-1} - (2 + h^2 v(t_n))X_n + (1 - (h/2)w(t_n))X_{n+1} &= h^2 u(t_n) \\ a_n^+ X_{n-1} - d_n X_n + a_n^- X_{n+1} &= h^2 u(t_n), \end{aligned}$$

com $a_n^\pm = 1 \pm (h/2)w(t_n)$ e $d_n = 2 + h^2 v(t_n)$. Em particular se $n = 1, N-1$, e utilizando as condições de contorno, encontramos

$$-d_1 X_1 + a_1^- X_2 = h^2 u(t_1) - a_1^+ \alpha, \quad a_{N-1}^+ X_{N-2} - d_{N-1} X_{N-1} = h^2 u(t_{N-1}) - a_{N-1}^- \beta.$$

Na forma matriz-vetor,

$$\begin{bmatrix} -d_1 & a_1^- & & & \\ a_2^+ & -d_2 & a_2^- & & \\ & \ddots & \ddots & \ddots & \\ & & a_{N-2}^+ & -d_{N-2} & a_{N-2}^- \\ & & & a_{N-1}^+ & -d_{N-1} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{N-2} \\ X_{N-1} \end{bmatrix} = h^2 \begin{bmatrix} u(t_1) \\ u(t_2) \\ \vdots \\ u(t_{N-2}) \\ u(t_{N-1}) \end{bmatrix} - \begin{bmatrix} a_1^+ \alpha \\ 0 \\ \vdots \\ 0 \\ a_{N-1}^- \beta \end{bmatrix}.$$

Para mostrar que a matriz acima (vamos denominá-la A) é não-singular, basta mostrar que ela é estritamente diagonalmente dominante, ou seja,

$$|A_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{N-1} |A_{ij}|, \quad i = 1, 2, \dots, N-1. \quad (21.4)$$

Como $v(t_i) > 0$, temos que $|A_{ii}| = 2 + h^2 v(t_i)$. Além disso,

$$\sum_{\substack{j=1 \\ j \neq i}}^{N-1} |A_{ij}| \leq |a_i^+| + |a_i^-| = \left| 1 + \frac{h}{2} w(t_i) \right| + \left| 1 - \frac{h}{2} w(t_i) \right| = \left| 1 + \frac{h}{2} |w(t_i)| \right| + \left| 1 - \frac{h}{2} |w(t_i)| \right|$$

Há uma desigualdade na equação acima, que se deve ao fato que a_1^+ e a_{N-1}^- não constam na matriz A (veja a primeira e a última linha da matriz). Tomando h suficientemente pequeno (de modo que $1 - (h/2)|w(t_i)| > 0$ para todo $1 \leq i \leq N-1$), obtemos

$$\sum_{\substack{j=1 \\ j \neq i}}^{N-1} |A_{ij}| \leq 1 + \frac{h}{2} |w(t_i)| + 1 - \frac{h}{2} |w(t_i)| = 2 < 2 + h^2 v(t_i) = |A_{ii}|.$$

Aula 22: Métodos de diferenças finitas (continuação)

Voltemos ao Exemplo 21.2 agora com condições de Neumann em uma das extremidades:

$$\begin{cases} x''(t) &= u(t) + v(t)x(t) + w(t)x'(t), & t \in]a, b[\\ x'(a) &= \alpha \\ x(b) &= \beta. \end{cases}$$

Como no exemplo anterior, temos as equações

$$\begin{aligned} a_n^+ X_{n-1} - d_n X_n + a_n^- X_{n+1} &= h^2 u(t_n), & 2 \leq n \leq N-2 \\ a_{N-1}^+ X_{N-2} - d_{N-1} X_{N-1} &= h^2 u(t_{N-1}) - a_{N-1}^- \beta \\ X_N &= \beta, \end{aligned}$$

com $a_n^\pm = 1 \pm (h/2)w(t_n)$ e $d_n = 2 + h^2 v(t_n)$. Resta discretizar a condição de contorno $x'(a) = \alpha$. Se usássemos a fórmula de diferenças centradas

$$x'(t_n) \approx \frac{x(t_{n+1}) - x(t_{n-1}))}{2h}.$$

com $n = 0$ e $X_n \approx x(t_n)$, teríamos a equação

$$\frac{X_1 - X_{-1}}{2h} = \alpha \implies X_1 - X_{-1} = 2h\alpha. \quad (22.1)$$

Esta equação envolve o termo $X_{-1} \approx x(t_{-1}) = x(a - h)$, que não faz parte do conjunto de valores aproximados. Uma alternativa para evitar este termo é utilizar a fórmula de diferenças progressivas

$$x'(t_n) \approx \frac{x(t_{n+1}) - x(t_n)}{h}.$$

Inserindo $n = 0$ e $X_n \approx x(t_n)$, obtemos

$$\frac{X_1 - X_0}{h} = \alpha \implies X_1 - X_0 = h\alpha. \quad (22.2)$$

Ainda há uma dificuldade com a equação (22.2): o valor de X_0 não é obtido explicitamente. Neste caso, X_0 passa a ser uma incógnita do sistema, lembrando que devemos adicionar a aproximação da equação diferencial em $t = t_1$:

$$a_1^+ X_0 - d_1 X_1 + a_1^- X_2 = h^2 u(t_1). \quad (22.3)$$

Chegamos assim ao seguinte sistema linear:

$$\begin{bmatrix} -1 & 1 & & & & & \\ a_1^+ & -d_1 & a_1^- & & & & \\ & a_2^+ & -d_2 & a_2^- & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & a_{N-2}^+ & -d_{N-2} & a_{N-2}^- & \\ & & & & a_{N-1}^+ & -d_{N-1} & \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-2} \\ X_{N-1} \end{bmatrix} = h^2 \begin{bmatrix} 0 \\ u(t_1) \\ u(t_2) \\ \vdots \\ u(t_{N-2}) \\ u(t_{N-1}) \end{bmatrix} - \begin{bmatrix} -h\alpha \\ 0 \\ 0 \\ \vdots \\ 0 \\ a_{N-1}^- \beta \end{bmatrix}.$$

A matriz do sistema acima não é estritamente diagonalmente dominante, devido à sua primeira linha. Por isso, não podemos utilizar o argumento do exemplo anterior para justificar que o sistema tem solução única.

Porém, se eliminamos X_0 do sistema, substituindo $X_0 = X_1 - h\alpha$ em (22.3), ou seja,

$$(a_1^+ - d_1)X_1 + a_1^- X_2 = h^2 u(t_1) + a_1^+ h\alpha, \quad (22.4)$$

o sistema linear se torna

$$\begin{bmatrix} (a_1^+ - d_1) & a_1^- & & & \\ a_2^+ & -d_2 & a_2^- & & \\ & \ddots & \ddots & \ddots & \\ & & a_{N-2}^+ & -d_{N-2} & a_{N-2}^- \\ & & & a_{N-1}^+ & -d_{N-1} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{N-2} \\ X_{N-1} \end{bmatrix} = h^2 \begin{bmatrix} u(t_1) \\ u(t_2) \\ \vdots \\ u(t_{N-2}) \\ u(t_{N-1}) \end{bmatrix} - \begin{bmatrix} -a_1^+ h\alpha \\ 0 \\ \vdots \\ 0 \\ a_{N-1}^- \beta \end{bmatrix}. \quad (22.5)$$

Assumindo novamente que o espaçamento h é tal que $1 - (h/2)|w(t_i)| > 0$ para todo $1 \leq i \leq N-1$, temos que

$$|a_1^+ - d_1| = \left| 1 + \frac{h}{2}w(t_i) - 2 - h^2 v(t_i) \right| = 1 - \frac{h}{2}w(t_i) + h^2 v(t_i) > 1 - \frac{h}{2}w(t_i) = |a_1^-|,$$

o que garante que a matriz do sistema (22.5) é estritamente diagonalmente dominante.

Uma segunda alternativa é utilizar a equação (22.1) e adicionar uma equação da forma (22.3), mas com $t = t_0$:

$$a_1^+ X_{-1} - d_1 X_0 + a_1^- X_1 = h^2 u(t_0). \quad (22.6)$$

Temos de (22.1) que $X_{-1} = X_1 - 2h\alpha$. Substituindo esta expressão em (22.6), eliminamos o termo indesejado X_{-1} , encontrando a seguinte equação para X_0 e X_1 :

$$\begin{aligned} a_0^+ (X_1 - 2h\alpha) - d_0 X_0 + a_0^- X_1 &= h^2 u(t_0) \\ -d_0 X_0 + (a_0^- + a_0^+) X_1 &= h^2 u(t_0) + 2ha_0^+ \alpha \end{aligned}$$

Chegamos assim a um novo sistema linear:

$$\begin{bmatrix} -d_0 & a_0^- + a_0^+ & & & \\ a_1^+ & -d_1 & a_1^- & & \\ & a_2^+ & -d_2 & a_2^- & \\ & & \ddots & \ddots & \ddots \\ & & & a_{N-2}^+ & -d_{N-2} & a_{N-2}^- \\ & & & & a_{N-1}^+ & -d_{N-1} \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \\ X_{N-2} \\ X_{N-1} \end{bmatrix} = h^2 \begin{bmatrix} u(t_0) \\ u(t_1) \\ u(t_2) \\ \vdots \\ u(t_{N-2}) \\ u(t_{N-1}) \end{bmatrix} - \begin{bmatrix} -2ha_0^+ \alpha \\ 0 \\ 0 \\ \vdots \\ 0 \\ a_{N-1}^- \beta \end{bmatrix},$$

que desta vez é estritamente diagonalmente dominante.

OBS: Ao contrário dos demais métodos numéricos vistos anteriormente, não podemos garantir que $X_0 = x(t_0)$ nos dois métodos numéricos descritos neste exemplo.

OBS: A princípio, a equação (22.6) não é uma aproximação válida da equação

$$x''(t) = u(t) + v(t)x(t) + w(t)x'(t), \quad t \in]a, b[,$$

pois $t_0 = a \notin]a, b[$. A aproximação (22.6) pressupõe que a equação acima possa ser estendida para um intervalo aberto que contém o ponto $t = a$, ou seja,

$$x''(t) = u(t) + v(t)x(t) + w(t)x'(t), \quad t \in]c, d[, \quad \text{com } [a, b[\subset]c, d[.$$

Aula 23: Métodos de projeção

Seja $x(t)$ a solução do problema de Dirichlet homogêneo

$$\begin{cases} x''(t) = f(t, x(t), x'(t)), & t \in]a, b[\\ x(a) = 0 \\ x(b) = 0. \end{cases} \quad (23.1)$$

Seja V um espaço vetorial de funções tal que $x \in V$ (por exemplo, $V = C([a, b])$, pois x é contínua em $[a, b]$) e

$$v \in V \implies v(a) = v(b) = 0.$$

Os *métodos de projeção* [8] para o PVC (23.1) selecionam um subespaço vetorial $V_h \subset V$ de dimensão finita (por exemplo, $V_h = \mathcal{P}_m$) e buscam a função $x_h \in V_h$ mais próxima de $x \in V$:

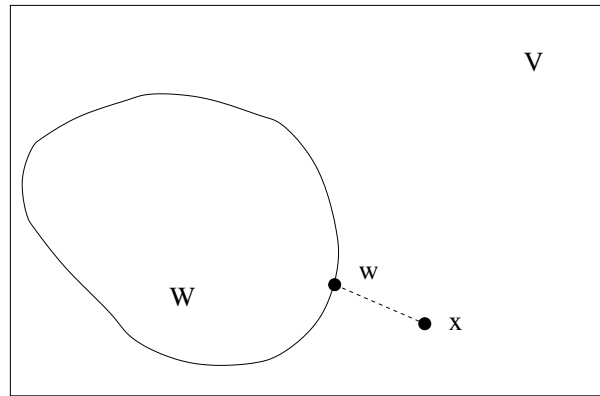


Figura 8: Descrição abstrata de um método de projeção.

A definição sobre o que significa ser “mais próxima” é a principal diferença entre cada método de projeção. Veremos a seguir dois tipos de métodos de projeção.

23.1 Método de Galerkin

Multiplicando os dois lados da equação $x'' = f$ por v e integrando de a até b , obtemos

$$\int_a^b x''(t)v(t) dt = \int_a^b f(t, x(t), x'(t))v(t) dt.$$

Assuma que V seja um espaço de funções diferenciáveis cujas derivadas são integráveis. Isto nos permite integrar por partes o lado esquerdo da equação acima:

$$\int_a^b x''(t)v(t) dt = x'(b)v(b) - x'(a)v(a) - \int_a^b x'(t)v'(t) dt.$$

Como $v \in V$, temos que $v(a) = v(b) = 0$, logo

$$\int_a^b x''(t)v(t) dt = - \int_a^b x'(t)v'(t) dt.$$

Portanto, x também é solução do seguinte problema (denominado *problema variacional*):

$$\langle x', v' \rangle = - \langle f, v \rangle \quad \forall v \in V, \quad (23.2)$$

em que $\langle \cdot, \cdot \rangle$ é o produto interno definido por

$$\langle u, v \rangle = \int_a^b u(t)v(t) dt.$$

O método de Galerkin corresponde a escolher $x_h \in V_h$ tal que, em analogia com (23.2),

$$\langle x'_h, v'_h \rangle = -\langle f, v_h \rangle \quad \forall v_h \in V_h, \quad (23.3)$$

observando que $f = f(t, x_h(t), x'_h(t))$. No caso particular em que $f = f(t)$, temos que, subtraindo (23.3) de (23.2) com $v = v_h$,

$$\langle x'_h - x', v'_h \rangle = 0, \quad \text{ou seja,} \quad \langle x'_h, v'_h \rangle = \langle x', v'_h \rangle \quad \forall v_h \in V_h. \quad (23.4)$$

Podemos escolher $v = v_h$ porque $V_h \subset V$ implica $v_h \in V$. A igualdade (23.4) permite interpretar a solução x_h de (23.3) como sendo a projeção de x sobre V_h com respeito ao operador $a(u, v) = \langle u', v' \rangle$.

Por outro lado, definindo a *semi-norma* $|u|_1 = \langle u', u' \rangle^{1/2}$, temos que, para todo $v_h \in V_h$,

$$\begin{aligned} |x - v_h|_1^2 &= \langle x' - v'_h, x' - v'_h \rangle \\ &= \langle x' - x'_h + x'_h - v'_h, x' - x'_h + x'_h - v'_h \rangle \\ &= \langle x' - x'_h, x' - x'_h \rangle + 2 \langle x' - x'_h, x'_h - v'_h \rangle + \langle x'_h - v'_h, x'_h - v'_h \rangle \\ &= |x - x_h|_1^2 + 2 \langle x' - x'_h, x'_h - v'_h \rangle + |x_h - v_h|_1^2 \\ &\geq |x - x_h|_1^2 + 2 \langle x' - x'_h, x'_h - v'_h \rangle. \end{aligned}$$

Como $x'_h - v'_h \in V_h$, segue de (23.4) que $\langle x' - x'_h, x'_h - v'_h \rangle = 0$, logo

$$|x - x_h|_1 \leq |x - v_h|_1 \quad \forall v_h \in V_h. \quad (23.5)$$

Assim, a solução x_h de (23.3) é o elemento de V_h “mais próximo” de $x \in V$ no sentido que $|x - x_h|_1$ é o menor valor possível para $|x - v_h|_1$ entre todos os elementos $v_h \in V_h$.

23.2 Método de colocação

Seja $N = \dim(V_h)$ e $t_1, \dots, t_N \in [a, b]$. De modo semelhante aos métodos de diferenças finitas, o método de colocação escolhe $x_h \in V_h$ tal que

$$x''_h(t_i) = f(t_i, x_h(t_i), x'_h(t_i)), \quad 1 \leq i \leq N. \quad (23.6)$$

Os pontos $t_1, \dots, t_N$ são denominados *pontos de colocação*, e não necessariamente são pontos de uma partição uniforme do intervalo $[a, b]$ (note que o primeiro ponto é t_1 , não t_0). No método de colocação, buscamos o elemento de V_h que satisfaz a equação diferencial nos N pontos escolhidos.

Novamente vamos assumir por simplicidade que $f = f(t)$. Neste caso, a medida de distância pode ser vista como o funcional

$$E[u] = \left(\sum_{i=1}^N |u''(t_i)|^2 \right)^{1/2}.$$

Dado $v_h \in V_h$, temos que

$$E[x - v_h]^2 = \sum_{i=1}^N |x''(t_i) - v''_h(t_i)|^2 = \sum_{i=1}^N |f(t_i) - v''_h(t_i)|^2.$$

Note que $E[x - v_h] \geq 0 \quad \forall v_h \in V_h$. Além disso, $E[x - x_h] = 0$ se x_h satisfaz (23.6). Portanto,

$$E[x - x_h] \leq E[x - v_h] \quad \forall v_h \in V_h. \quad (23.7)$$

Aula 24: Representação vetorial dos métodos de projeção

Esta seção enfatiza a importância de que V_h seja um subespaço vetorial de dimensão finita. Veremos que os métodos de Galerkin e de colocação se reduzem a um sistema de equações que depende da base escolhida para V_h .

Vamos iniciar com a representação vetorial do método de Galerkin. Seja $\{v_1, \dots, v_N\}$ uma base para V_h . Dado $v_h \in V_h$ arbitrário, existem $\alpha_1, \dots, \alpha_N$ tais que

$$v_h(t) = \alpha_1 v_1(t) + \dots, \alpha_N v_N(t) = \sum_{i=1}^N \alpha_i v_i(t).$$

Segue de (23.3) que

$$\left\langle x'_h, \left(\sum_{i=1}^N \alpha_i v_i \right)' \right\rangle = \left\langle x'_h, \sum_{i=1}^N \alpha_i v'_i \right\rangle = - \left\langle f, \sum_{i=1}^N \alpha_i v_i \right\rangle.$$

Pela linearidade do produto interno,

$$\sum_{i=1}^N \alpha_i \langle x'_h, v'_i \rangle = - \sum_{i=1}^N \alpha_i \langle f, v_i \rangle \implies \sum_{i=1}^N \alpha_i (\langle x'_h, v'_i \rangle + \langle f, v_i \rangle) = 0. \quad (24.1)$$

Temos que (24.1) (e portanto, (23.3)) é válido se, e somente se,

$$\langle x'_h, v'_i \rangle + \langle f, v_i \rangle = 0, \quad 1 \leq i \leq N. \quad (24.2)$$

Para verificar que (24.1) implica (24.2) quando $i = 1$, basta escolher $v_h \in V_h$ tal que $\alpha_1 = 1$ e $\alpha_i = 0$ para $2 \leq i \leq N$. A demonstração para $i = 2, \dots, N$ é análoga. Para demonstrar a recíproca, suponha que (24.2) seja falso, isto é, suponha que exista algum i^* entre 1 e N tal que $\langle x'_h, v'_{i^*} \rangle + \langle f, v_{i^*} \rangle \neq 0$. Escolhendo $v_h \in V_h$ tal que $\alpha_{i^*} = 1$ e $\alpha_i = 0$ para todo $i \neq i^*$, obtemos

$$\sum_{i=1}^N \alpha_i (\langle x'_h, v'_i \rangle + \langle f, v_i \rangle) = (\langle x'_h, v'_{i^*} \rangle + \langle f, v_{i^*} \rangle)^2 \neq 0,$$

o que contradiz (24.1).

Agora, levemos também em conta que $x_h \in V_h$. Logo, existem $X_1, \dots, X_N$ tais que

$$x_h(t) = \sum_{j=1}^N X_j v_j(t). \quad (24.3)$$

Substituindo (24.3) em (24.2) e usando novamente a linearidade do produto interno, obtemos

$$\sum_{j=1}^N X_j \langle v_j, v'_i \rangle + \langle f, v_i \rangle = 0, \quad 1 \leq i \leq N, \quad (24.4)$$

ou na forma vetorial,

$$K\mathbf{x} = \mathbf{F}(\mathbf{x}), \quad \mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}, \quad K = \begin{bmatrix} \langle v_1, v'_1 \rangle & \langle v_2, v'_1 \rangle & \dots & \langle v_N, v'_1 \rangle \\ \langle v_1, v'_2 \rangle & \langle v_2, v'_2 \rangle & \dots & \langle v_N, v'_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_1, v'_N \rangle & \langle v_2, v'_N \rangle & \dots & \langle v_N, v'_N \rangle \end{bmatrix}, \quad \mathbf{F}(\mathbf{x}) = \begin{bmatrix} -\langle f, v_1 \rangle \\ -\langle f, v_2 \rangle \\ \vdots \\ -\langle f, v_N \rangle \end{bmatrix},$$

observando que $\langle f, v_i \rangle$ depende de $\mathbf{x}$ para todo $1 \leq i \leq N$, ou seja:

$$\langle f, v_i \rangle = \int_a^b f(t, x_h(t), x'_h(t)) v_i(t) dt = \int_a^b f \left(t, \sum_{j=1}^N X_j v_j(t), \sum_{j=1}^N X_j v'_j(t) \right) v_i(t) dt.$$

No caso particular $f = f(t)$, as equações $K\mathbf{x} = \mathbf{F}(\mathbf{x})$ se tornam um sistema linear, como no Exemplo 21.1.

Para o método de colocação, obtemos de (23.6) uma expressão semelhante a (24.4):

$$\sum_{j=1}^N X_j v''_j(t_i) = f \left(t_i, \sum_{j=1}^N X_j v_j(t_i), \sum_{j=1}^N X_j v'_j(t_i) \right), \quad 1 \leq i \leq N. \quad (24.5)$$

Na forma vetorial, este sistema é dado por $K\mathbf{x} = \mathbf{F}(\mathbf{x})$, como no método de Galerkin, mas com

$$K = \begin{bmatrix} v''_1(t_1) & v''_2(t_1) & \dots & v''_N(t_1) \\ v''_1(t_2) & v''_2(t_2) & \dots & v''_N(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ v''_1(t_N) & v''_2(t_N) & \dots & v''_N(t_N) \end{bmatrix}, \quad \mathbf{F}(\mathbf{x}) = \begin{bmatrix} f \left(t_1, \sum_{j=1}^N X_j v_j(t_1), \sum_{j=1}^N X_j v'_j(t_1) \right) \\ f \left(t_2, \sum_{j=1}^N X_j v_j(t_2), \sum_{j=1}^N X_j v'_j(t_2) \right) \\ \vdots \\ f \left(t_N, \sum_{j=1}^N X_j v_j(t_N), \sum_{j=1}^N X_j v'_j(t_N) \right) \end{bmatrix}.$$

Exemplo 24.1 : Use o método de colocação para aproximar a solução do PVC

$$\begin{cases} x''(t) = t^3 - t^2 x(t), & t \in]0, 1[\\ x(0) = 0 \\ x(1) = 0, \end{cases}$$

sendo V_h o espaço gerado pela base $\{\sin(\pi t), \sin(2\pi t)\}$ e os pontos de colocação $t_1 = 1/4$ e $t_2 = 1/2$.

Como $f(t, x) = t^3 - t^2 x$ e $0 \leq t \leq 1$, temos que

$$|f(t, x_1) - f(t, x_2)| = |-t^2 x_1 + t^2 x_2| = |t^2| |x_1 - x_2| \leq k |x_1 - x_2|$$

com $k = 1 < 8$, logo segue do Teorema 19.2 que o PVC acima tem solução única.

Seja $x_h(t) = X_1 v_1(t) + X_2 v_2(t)$, com $v_1(t) = \sin(\pi t)$ e $v_2(t) = \sin(2\pi t)$. Temos que, para $i = 1, 2$,

$$\begin{aligned} x''_h(t_i) &= t_i^3 - t_i^2 x(t_i) \\ X_1 v''_1(t_i) + X_2 v''_2(t_i) &= t_i^3 - t_i^2 [X_1 v_1(t_i) + X_2 v_2(t_i)]. \end{aligned}$$

Note que $f(t, x, y)$ depende linearmente de x . Por isso, podemos isolar as incógnitas X_1 e X_2 :

$$\begin{aligned} X_1 v''_1(t_i) + X_2 v''_2(t_i) + t_i^2 [X_1 v_1(t_i) + X_2 v_2(t_i)] &= t_i^3 \\ [v''_1(t_i) + t_i^2 v_1(t_i)] X_1 + [v''_2(t_i) + t_i^2 v_2(t_i)] X_2 &= t_i^3 \end{aligned} \quad (24.6)$$

Vamos avaliar as derivadas de v_1 e v_2 :

$$\begin{aligned} v'_1(t) &= \pi \sin(\pi t) & v'_2(t) &= 2\pi \sin(2\pi t) \\ v''_1(t) &= -\pi^2 \cos(\pi t) & v''_2(t) &= -4\pi^2 \cos(2\pi t) \end{aligned}$$

Assim, segue que

$$\begin{aligned} v_1''(t_i) + t_i^2 v_1(t_i) &= -\pi^2 \text{sen}(\pi t_i) + t_i^2 \text{sen}(\pi t_i) = (t_i^2 - \pi^2) \text{sen}(\pi t_i) \\ v_2''(t_i) + t_i^2 v_2(t_i) &= -4\pi^2 \text{sen}(2\pi t_i) + t_i^2 \text{sen}(2\pi t_i) = (t_i^2 - 4\pi^2) \text{sen}(2\pi t_i), \end{aligned}$$

e (24.6) se reduz a

$$(t_i^2 - \pi^2) \text{sen}(\pi t_i) X_1 + (t_i^2 - 4\pi^2) \text{sen}(2\pi t_i) X_2 = t_i^3.$$

Finalmente, avaliamos esta expressão em $t_1 = 1/4$ e $t_2 = 1/2$:

$$\begin{cases} ((1/4)^2 - \pi^2) \text{sen}(\pi/4) X_1 + ((1/4)^2 - 4\pi^2) \text{sen}(2\pi/4) X_2 = (1/4)^3 \\ ((1/2)^2 - \pi^2) \text{sen}(\pi/2) X_1 + ((1/2)^2 - 4\pi^2) \text{sen}(2\pi/2) X_2 = (1/2)^3, \end{cases}$$

ou na forma vetorial,

$$\begin{bmatrix} (1/16 - \pi^2)\sqrt{2}/2 & 1/16 - 4\pi^2 \\ 1/4 - \pi^2 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 1/64 \\ 1/8 \end{bmatrix}.$$

Aula 25: Funções polinomiais de Lagrange por partes

Vimos que o método de diferenças finitas para o PVC

$$\begin{cases} x''(t) &= u(t) + v(t)x(t) + w(t)x'(t), & t \in]a, b[\\ x(a) &= \alpha \\ x(b) &= \beta \end{cases}$$

resulta em um sistema linear cuja matriz é tridiagonal. Isto não necessariamente ocorre com o métodos de Galerkin e de colocação, fazendo com que, em geral, a solução dos sistemas lineares vindos deste métodos seja mais lenta que no método de diferenças finitas.

Uma forma de garantir que a matriz resultante destes métodos seja tridiagonal é estabelecer uma partição uniforme do intervalo $[a, b]$ e escolher as funções de base $v_1, \dots, v_N$ tais que

$$v_j(t) = 0 \quad \forall t \notin]t_{j-1}, t_{j+1}[. \quad (25.1)$$

Note que (25.1) implica que $v'_j(t) = v''_j(t) = 0$ para todo $t \notin]t_{j-1}, t_{j+1}[$.

Temos que a matriz K do sistema $K\mathbf{x} = \mathbf{F}(\mathbf{x})$ da Aula 24 satisfaz, para $|i - j| > 1$,

$$\begin{aligned} K_{i,j} &= \int_a^b v'_j(t)v'_i(t) dt = 0 && \text{no método de Galerkin,} \\ K_{i,j} &= v''_j(t_i) = 0 && \text{no método de colocação.} \end{aligned}$$

Suponha que $i > j + 1$ ou $i < j - 1$. Neste caso, $]t_{i-1}, t_{i+1}[\cap]t_{j-1}, t_{j+1}[= \emptyset$ (Fig. 9). Em particular, $t_i \notin]t_{j-1}, t_{j+1}[$ e segue de (25.1) que $v''_j(t_i) = 0$.

Por outro lado, $v_i(t) = 0$ se $t \in]t_{j-1}, t_{j+1}[$ e reciprocamente, $v_j(t) = 0$ se $t \in]t_{i-1}, t_{i+1}[$. Portanto, $v_i(t)$ e $v_j(t)$ nunca são simultaneamente não-nulos (assim como $v'_i(t)$ e $v'_j(t)$), logo

$$v'_i(t)v'_j(t) = 0 \quad \forall t \in [a, b] \implies \int_a^b v'_j(t)v'_i(t) dt = 0.$$

Portanto, $K_{i,j} = 0$ para $|i - j| > 1$ em ambos os métodos. Assim, $K_{i,j} \neq 0$ somente se $j = i$ ou $j = i \pm 1$. Assim, apenas $K_{i,i}$ e $K_{i,i \pm 1}$ podem ser não-nulos, o que faz de K uma matriz tridiagonal.

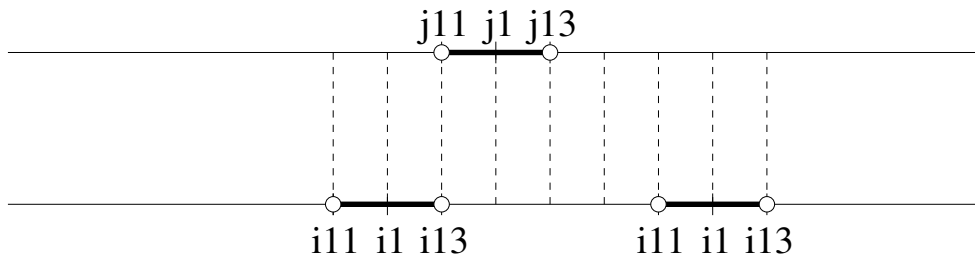


Figura 9: Comparação entre $]t_{i-1}, t_{i+1}[$ e $]t_{j-1}, t_{j+1}[$ com $j = i - 2$ e $j = i + 3$.

Seja $t_n = a + nh$, $0 \leq n \leq N$, como na Aula 20. A base mais popular que satisfaz (25.1) é formada pelas funções polinomiais por partes de Lagrange de grau 1. Para $1 \leq j \leq N - 1$, definimos $v_j(t)$ pelas seguintes condições:

1. v_j é um polinômio de grau 1 em cada intervalo $[t_n, t_{n+1}]$, para todo $0 \leq n \leq N - 1$, ou seja, existem $a_{j,n}$ e $b_{j,n}$ tais que $v_j(t) = a_{j,n}t + b_{j,n}$ para todo $t \in [t_n, t_{n+1}]$;
2. $v_j(t_n) = \delta_{n,j}$, ou seja, $v_j(t_j) = 1$ e $v_j(t_n) = 0$ se $n \neq j$, para $0 \leq n \leq N$.

O símbolo $\delta_{i,j}$ é conhecido como *delta de Kronecker*. Convém observar que nas condições acima associamos à partição $[a, b]$ apenas $N - 1$ funções de base, não N funções, como na Aula 23.

Vamos obter uma fórmula geral para a j -ésima função de base $v_j(t)$. Dado $t \in [t_n, t_{n+1}]$ com $j < n \leq N - 1$, temos que $t_n \neq t_j$ e $t_{n+1} \neq t_j$. Segue da condição 2 que

$$v_j(t_n) = v_j(t_{n+1}) = 0.$$

Por outro lado, temos da condição 1 que existem $a_{j,n}$ e $b_{j,n}$ tais que $v_j(t) = a_{j,n}t + b_{j,n}$ para todo $t \in [t_n, t_{n+1}]$. Assim,

$$0 = v_j(t_n) = a_{j,n}t_n + b_{j,n} \quad (25.2)$$

$$0 = v_j(t_{n+1}) = a_{j,n}t_{n+1} + b_{j,n} \quad (25.3)$$

Subtraindo (25.2) de (25.3), obtemos $a_{j,n}(t_{n+1} - t_n) = a_{j,n}h = 0$, ou seja, $a_{j,n} = 0$. Substituindo $a_{j,n} = 0$ em (25.2) ou (25.3), obtemos $b_{j,n} = 0$. Portanto,

$$v_j(t) = 0 \quad \forall t \in [t_n, t_{n+1}], \quad j < n \leq N - 1.$$

Analogamente, podemos mostrar que $v_j(t) = 0$ para todo $t \in [t_n, t_{n+1}]$ se $0 \leq n < j - 1$. Resta determinar $v_j(t)$ para $t \in [t_{j-1}, t_{j+1}]$.

Consideremos o caso $t \in [t_{j-1}, t_j]$. Sabemos da condição 1 que existem $a_{j,j-1}$ e $b_{j,j-1}$ tais que $v_j(t) = a_{j,j-1}t + b_{j,j-1}$ para todo $t \in [t_{j-1}, t_j]$. Além disso, $v_j(t_{j-1}) = 0$ e $v_j(t_j) = 1$, logo

$$0 = v_j(t_{j-1}) = a_{j,j-1}t_{j-1} + b_{j,j-1} \quad (25.4)$$

$$1 = v_j(t_j) = a_{j,j-1}t_j + b_{j,j-1} \quad (25.5)$$

Subtraindo (25.4) de (25.5), encontramos $a_{j,j-1}h = 1$, ou seja, $a_{j,j-1} = 1/h$. Substituindo $a_{j,j-1} = 1/h$ em (25.4), obtemos $b_{j,j-1} = -t_{j-1}/h$. Portanto,

$$v_j(t) = \frac{1}{h}t + \frac{-t_{j-1}}{h} = \frac{t - t_{j-1}}{h}, \quad t \in [t_{j-1}, t_j].$$

Analogamente, $v_j(t) = (t_{j+1} - t)/h$ se $t \in [t_j, t_{j+1}]$, de modo que a função v_j é definida pela seguinte expressão (Fig. 9):

$$v_j(t) = \begin{cases} 0, & t \leq t_{j-1}, \\ \frac{t - t_{j-1}}{h}, & t \in [t_{j-1}, t_j], \\ \frac{t_{j+1} - t}{h}, & t \in [t_j, t_{j+1}], \\ 0, & t > t_{j+1}. \end{cases} \quad (25.6)$$

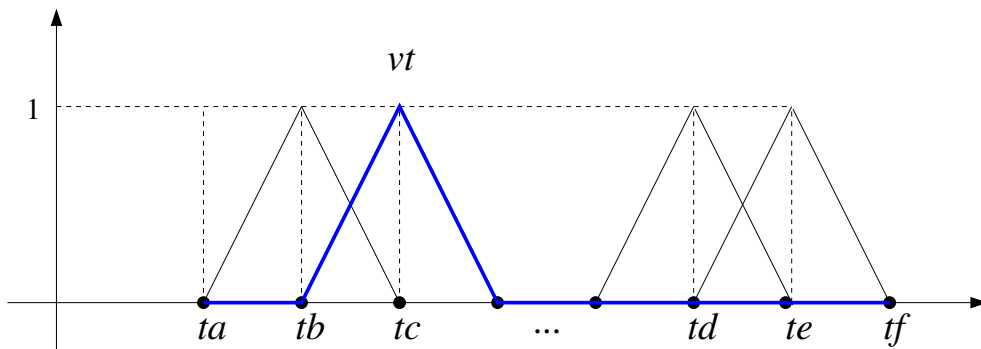


Figura 10: Funções polinomiais de Lagrange por partes de grau 1, com destaque para $v_2(t)$.

Temos de (25.6) que v_j é uma função contínua. Além disso, segue da propriedade 2 que as funções $v_1(t), \dots, v_{N-1}(t)$ são linearmente independentes, e que as funções $v_j(t)$ satisfazem $v_j(a) = v_j(t_0) = 0$ e $v_j(b) = v_j(t_N) = 0$ para $1 \leq j \leq N - 1$, logo o espaço V_h gerado por $\{v_1(t), \dots, v_{N-1}(t)\}$ é apropriado para os métodos de projeção.

OBS: Como $v_j \notin C^2([a, b])$, o espaço V_h acima não pode ser utilizado no método de colocação.

Aula 26: Interpolação polinomial por partes

Definimos na aula passada a base $\{v_1(t), \dots, v_{N-1}(t)\}$, associada aos pontos $t_k = a + kh$ ($0 \leq k \leq N$), com $v_j(t)$ definida por (25.4). Vamos considerar também as funções

$$v_0(t) = \begin{cases} \frac{t_1-t}{h}, & t \in [t_0, t_1], \\ 0, & t > t_{j+1}, \end{cases} \quad v_N(t) = \begin{cases} 0, & t \leq t_{j-1}, \\ \frac{t-t_{N-1}}{h}, & t \in [t_{N-1}, t_N], \end{cases} \quad (26.1)$$

que também satisfazem as condições 1 e 2 da Aula 25. Estas funções são ilustradas na Fig. 11.

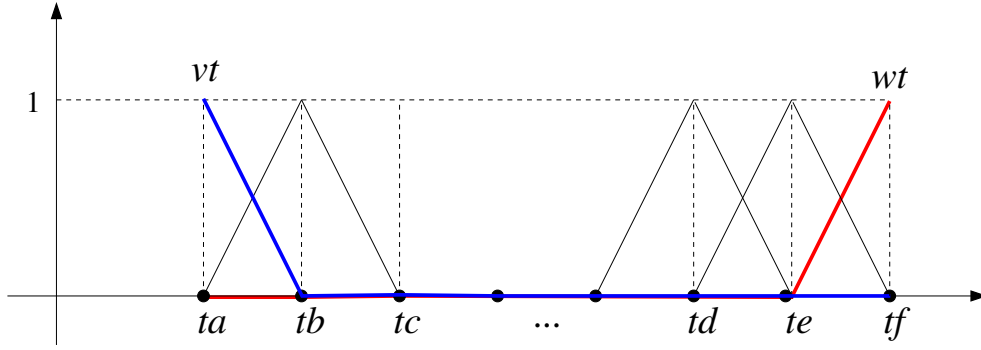


Figura 11: Funções $v_0(t)$ e $v_N(t)$ definidas em (26.1).

Vamos descrever o espaço W gerado por $\{v_0, v_1, \dots, v_{N-1}, v_N\}$. Dado $w \in W$, existem $W_0, \dots, W_N$ tais que

$$w(t) = \sum_{j=0}^N W_j v_j(t).$$

Como $v_0, \dots, v_N$ satisfazem a condição 1, temos que, para todo n entre 0 e $N-1$,

$$\exists a_{j,n}, b_{j,n} \quad \text{t.q.} \quad v_j(t) = a_{j,n}t + b_{j,n} \quad \forall t \in [t_n, t_{n+1}].$$

Segue que, para todo $t \in [t_n, t_{n+1}]$,

$$\begin{aligned} W_j v_j(t) &= W_j a_{j,n}t + W_j b_{j,n} \\ \sum_{j=0}^N W_j v_j(t) &= \sum_{j=0}^N W_j a_{j,n}t + W_j b_{j,n} \\ w(t) &= \left(\sum_{j=0}^N W_j a_{j,n} \right) t + \left(\sum_{j=0}^N W_j b_{j,n} \right) \\ w(t) &= a_n t + b_n, \quad a_n = \sum_{j=0}^N W_j a_{j,n}, \quad b_n = \sum_{j=0}^N W_j b_{j,n}, \end{aligned}$$

ou seja, w também satisfaz a condição 1. Além disso, como $v_0, \dots, v_N$ são contínuas, w também é contínua. Vamos verificar que a recíproca também é válida, isto é, se w é uma função contínua que satisfaz a condição 1, então $w \in W$.

De fato, se $\tilde{w}(t) = w(t_0)v_0(t) + \dots + w(t_N)v_N(t) \in W$, então

$$\tilde{w}(t_n) = \sum_{j=0}^N w(t_j)v_j(t_n) = \sum_{j=0}^N w(t_j)\delta_{j,n} = w(t_n)(1) = w(t_n), \quad 0 \leq n \leq N.$$

Assim, definindo $z(t) = \tilde{w}(t) - w(t)$, segue que $z(t_n) = \tilde{w}(t_n) - w(t_n) = 0$ para todo $0 \leq n \leq N$. Como $z(t)$ é contínua e satisfaz a condição 1 (pois w e $\tilde{w}$ também satisfazem estas propriedades), temos que $z(t) = 0$ para todo $t \in [a, b]$. Portanto, $w = \tilde{w}$, ou seja, $w \in W$.

Concluimos que o espaço W é exatamente o espaço das funções contínuas e polinomiais por partes de grau 1 associados à partição $\tau = \{t_0, t_1, \dots, t_N\}$ do intervalo $[a, b]$. Vamos denotar $W = \mathcal{PP}_1(\tau, [a, b])$, sendo que em geral,

$$\mathcal{PP}_m(\tau, [a, b]) = \{v \in C^0([a, b]) \mid v|_I \in \mathcal{P}_m(I), \ I = [t_n, t_{n+1}], \ 0 \leq n \leq N-1\},$$

sendo que $w = v|_I$, a *restrição de v em I* , é a função $w : I \rightarrow \mathbb{R}$ tal que $w(t) = v(t)$ para todo $t \in I$, e $\mathcal{P}_m(I)$ é o espaço dos polinômios de grau m restritos a I , ou seja,

$$\mathcal{P}_m(I) = \{p \in C^0(I) \mid p(t) = a_0 + a_1 t + \dots + a_m t^m \quad \forall t \in I\}.$$

Uma propriedade crucial da base $\{v_0, \dots, v_N\}$ definida por (25.6) e (26.1), usada nos argumentos acima, é que os coeficientes da combinação linear coincidem com o valor da função nos pontos da partição, ou seja,

$$w(t) = \sum_{j=0}^N W_j v_j(t) \implies w(t_n) = W_n \quad (0 \leq n \leq N).$$

Em particular, dada $f \in C^0([a, b])$ qualquer, a função $\tilde{f} \in \mathcal{PP}_1(\tau, [a, b])$ definida por

$$\tilde{f}(t) = \sum_{j=0}^N f(t_j) v_j(t) \tag{26.2}$$

satisfaz $\tilde{f}(t_n) = f(t_n)$, ou seja, coincide com $f(t)$ se $t = t_n$, para todo $0 \leq n \leq N$. A função $\tilde{f}(t)$ é denominada a *função de interpolação polinomial por partes* de grau 1 da função $f(t)$ na partição τ .

Exemplo 26.1 : Encontre $\tilde{f}$ para $f(t) = \sin(\pi t)$ em $[0, 1]$, sendo $\tau = \{0, 1/3, 2/3, 1\}$.

Temos de (26.2) que

$$\begin{aligned} \tilde{f}(t) &= f(t_0)v_0(t) + f(t_1)v_1(t) + f(t_2)v_2(t) + f(t_3)v_3(t) \\ &= \sin(0)v_0(t) + \sin(\pi/3)v_1(t) + \sin(2\pi/3)v_2(t) + \sin(\pi)v_3(t) \\ &= (\sqrt{3}/2)v_1(t) + (\sqrt{3}/2)v_2(t). \end{aligned}$$

Observando que $h = 1/3$, temos de (25.4) que

$$v_1(t) = \begin{cases} 0, & t \leq t_0, \\ \frac{t-t_0}{h} = \frac{t}{1/3}, & t \in]t_0, t_1], \\ \frac{t_2-t}{h} = \frac{2/3-t}{1/3}, & t \in]t_1, t_2], \\ 0, & t > t_2, \end{cases} \quad v_2(t) = \begin{cases} 0, & t \leq t_1, \\ \frac{t-t_1}{h} = \frac{t-1/3}{1/3}, & t \in]t_1, t_2], \\ \frac{t_3-t}{h} = \frac{1-t}{1/3}, & t \in]t_2, t_3], \\ 0, & t > t_3, \end{cases}$$

que podemos escrever na forma

$$v_1(t) = \begin{cases} 3t, & t \in [0, 1/3], \\ 2-3t, & t \in]1/3, 2/3], \\ 0, & t \in]2/3, 1], \end{cases} \quad v_2(t) = \begin{cases} 0, & t \in [0, 1/3], \\ 3t-1, & t \in]1/3, 2/3], \\ 3(1-t), & t \in]2/3, 1]. \end{cases}$$

Portanto,

$$\tilde{f}(t) = \begin{cases} (\sqrt{3}/2)3t + (\sqrt{3}/2)(0) = 3\frac{\sqrt{3}}{2}t, & t \in [0, 1/3], \\ (\sqrt{3}/2)(2-3t) + (\sqrt{3}/2)(3t-1) = \frac{\sqrt{3}}{2}, & t \in]1/3, 2/3], \\ (\sqrt{3}/2)(0) + (\sqrt{3}/2)(3(1-t)) = \frac{3\sqrt{3}}{2}(1-t), & t \in]2/3, 1]. \end{cases}$$

Aula 27: Funções splines cúbicas

O subespaço vetorial

$$S_m(\tau, [a, b]) = \mathcal{PP}_m(\tau, [a, b]) \cap C^{m-1}([a, b])$$

é o espaço das *funções splines de grau m*. Note em particular que

$$S_1(\tau, [a, b]) = \mathcal{PP}_1(\tau, [a, b]) \cap C^{1-1}([a, b]) = \mathcal{PP}_1(\tau, [a, b]).$$

Lembrando que, para avaliar $K_{i,j} = x_j''(t_i)$ no método de colocação, precisamos que V_h seja um espaço de funções de classe $C^2([a, b])$, consideremos o espaço $S_3(\tau, [a, b])$ das funções *splines cúbicas*:

$$S_3(\tau, [a, b]) = \mathcal{PP}_3(\tau, [a, b]) \cap C^2([a, b]) = \{v \in C^2([a, b]) \mid v|_I \in \mathcal{P}_3(I), I = [t_k, t_{k+1}], 0 \leq k \leq N-1\}.$$

Na Aula 29 exibiremos uma base para $S_3(\tau, [a, b])$. A fim de motivar estas funções de base, vamos estudar primeiro o problema de interpolação em $S_3(\tau, [a, b])$: dada uma partição $\tau = \{t_0, \dots, t_N\}$ e $f \in C^0([a, b])$, determinar a função

$$s(t) = \begin{cases} s_1(t) = a_1 + b_1 t + c_1 t^2 + d_1 t^3, & t \in [t_0, t_1], \\ s_2(t) = a_2 + b_2 t + c_2 t^2 + d_2 t^3, & t \in]t_1, t_2], \\ \vdots \\ s_N(t) = a_N + b_N t + c_N t^2 + d_N t^3, & t \in]t_{N-1}, t_N] \end{cases}$$

tal que $s \in C^2([a, b])$ e $s(t_i) = f(t_i)$, ou seja,

$$\begin{cases} s_1(t_0) = f(t_0) & (s_1(a) = f(a)) \\ s_i(t_i) = f(t_i), & 1 \leq i \leq N-1 \\ s_{i+1}(t_i) = f(t_i), & 1 \leq i \leq N-1 \\ s_N(t_N) = f(t_N) & (s_N(b) = f(b)), \end{cases} \quad \begin{cases} s'_i(t_i) = s'_{i+1}(t_i), & 1 \leq i \leq N-1, \\ s''_i(t_i) = s''_{i+1}(t_i), & 1 \leq i \leq N-1, \end{cases} \quad (27.1)$$

resultando em $1 + 2(N-1) + 1 + 2(N-1) = 4N - 2$ equações, contra $4N$ incógnitas $(a_1, b_1, \dots, d_N)$. Precisamos de mais duas equações, sendo que as mais comuns são

$$\begin{cases} s''_1(a) = 0 \\ s''_N(b) = 0 \end{cases} \quad (\text{spline natural}) \quad \begin{cases} s'''_1(t_1) = s'''_2(t_1) \\ s'''_{N-1}(t_{N-1}) = s'''_N(t_{N-1}) \end{cases} \quad (\text{not a knot})$$

As condições *not a knot* são a escolha padrão do Matlab/Octave. Para entender estas condições, vamos construir uma função auxiliar a partir dos polinômios que definem $s_1(t)$ e $s_2(t)$:

$$z(t) = (a_1 - a_2) + (b_1 - b_2)t + (c_1 - c_2)t^2 + (d_1 - d_2)t^3 \quad (27.2)$$

Temos de (27.1) que $s_1^{(k)}(t_1) = s_2^{(k)}(t_1)$ para $0 \leq k \leq 2$. Com a primeira condição *not a knot*, podemos estender k para $0 \leq k \leq 3$, ou seja,

$$s_1^{(k)}(t_1) = s_2^{(k)}(t_1), \quad 0 \leq k \leq 3. \quad (27.3)$$

Segue de (27.2) e (27.3) que $z(t_1) = z'(t_1) = z''(t_1) = z'''(t_1) = 0$, logo $z(t) = 0$, ou seja, $a_1 = a_2, \dots, d_1 = d_2$. Portanto, $s_1(t)$ e $s_2(t)$ são restrições do mesmo polinômio, como se o ponto t_1 não fosse um ponto (nó) da partição (*not a knot*). O mesmo raciocínio se aplica ao ponto t_{N-1} .

Exemplo 27.1 : Encontre a função spline cúbica natural que interpola $f(t) = \sin(\pi t)$ nos pontos $t_k = k/2$, $0 \leq k \leq 2$.

Temos que $[a, b] = [0, 1]$, $N = 2$, e $\tau = \{0, 1/2, 1\}$. Assim,

$$s(t) = \begin{cases} s_1(t) &= a_1 + b_1 t + c_1 t^2 + d_1 t^3, & t \in [0, 1/2], \\ s_2(t) &= a_2 + b_2 t + c_2 t^2 + d_2 t^3, & t \in]1/2, 1], \end{cases}$$

sendo que $a_1, \dots, d_8$ satisfazem as seguintes equações:

$$\begin{cases} s_1(0) = f(0) = 0, \\ s_1(1/2) = f(1/2) = 1, \\ s_2(1/2) = f(1/2) = 1, \\ s_2(1) = f(1) = 0, \end{cases} \quad \begin{cases} s'_1(1/2) = s'_2(1/2), \\ s''_1(1/2) = s''_2(1/2), \end{cases} \quad \begin{cases} s''_1(0) = 0, \\ s''_2(1) = 0. \end{cases}$$

Note que, para $i = 1, 2$,

$$\begin{aligned} s'_i(t) &= b_i + 2c_i t + 3d_i t^2, \\ s''_i(t) &= 2c_i + 6d_i t, \end{aligned}$$

de modo que as equações acima podem ser escritas na forma

$$\begin{cases} a_1 &= 0 \\ a_1 + b_1/2 + c_1/4 + d_1/8 &= 1 \\ a_2 + b_2/2 + c_2/4 + d_2/8 &= 1 \\ a_2 + b_2 + c_2 + d_2 &= 0 \\ b_1 + c_1 + 3d_1/4 &= b_2 + c_2 + 3d_2/4 \\ 2c_1 + 3d_1 &= 2c_2 + 3d_2 \\ 2c_1 &= 0 \\ 2c_2 + 6d_2 &= 0. \end{cases}$$

A solução deste sistema é dada por $a_1 = c_1 = 0$, $b_1 = 3$, $d_1 = -4$, $a_2 = -1$, $b_2 = 9$, $c_2 = -12$ e $d_2 = 4$, o que resulta na função

$$s(t) = \begin{cases} 3t - 4t^3, & t \in [0, 1/2], \\ -1 + 9t - 12t^2 + 4t^3, & t \in]1/2, 1]. \end{cases}$$

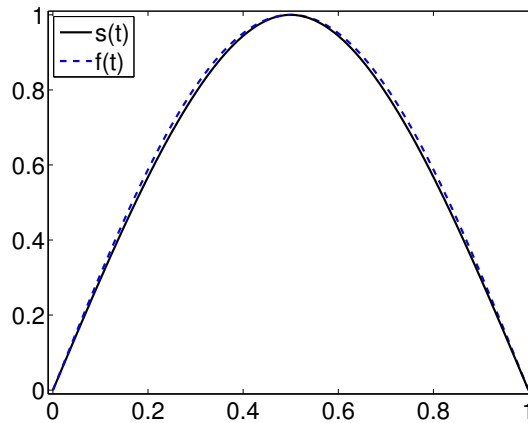


Figura 12: Interpolação de $f(t) = \sin(\pi t)$ pela spline cúbica $s(t)$ com dois intervalos, comparada a solução do Exemplo 26.1

Aula 28: Splines cúbicas naturais

Vamos obter o sistema linear correspondente à spline natural que interpola $f \in C^0([a, b])$ na partição $\tau = \{t_0, \dots, t_N\}$, ou seja,

$$s(t) = \begin{cases} s_1(t) = a_1 + b_1 t + c_1 t^2 + d_1 t^3, & t \in [t_0, t_1], \\ s_2(t) = a_2 + b_2 t + c_2 t^2 + d_2 t^3, & t \in]t_1, t_2], \\ \vdots \\ s_N(t) = a_N + b_N t + c_N t^2 + d_N t^3, & t \in]t_{N-1}, t_N] \end{cases}$$

tal que $s \in C^2([a, b])$, $s(t_i) = f(t_i)$ para todo $0 \leq i \leq N$ e $s''(a) = s''(b) = 0$. Seja $z_i = s''(t_i)$, $0 \leq i \leq N$. Temos que $z_0 = z_N = 0$, $s'_i(t_{i-1}) = z_{i-1}$ e $s''_i(t_i) = z_i$. Como $s''_i(t) = 2c_i + 6d_i t$,

$$\begin{cases} 2c_i + 6d_i t_{i-1} = z_{i-1} \\ 2c_i + 6d_i t_i = z_i \end{cases} \implies \begin{cases} d_i = (z_i - z_{i-1})/6h \\ c_i = z_i/2 - (z_i - z_{i-1})t_i/2h, \end{cases}$$

ou seja,

$$\begin{aligned} s''_i(t) &= z_i - \frac{(z_i - z_{i-1})t_i}{h} + \frac{z_i - z_{i-1}}{h}t = z_{i-1} \frac{t_i - t}{h} + z_i \frac{h - t_i + t}{h} = z_{i-1} \frac{t_i - t}{h} + z_i \frac{t - (t_i - h)}{h} \\ s''_i(t) &= z_{i-1} \frac{t_i - t}{h} + z_i \frac{t - t_{i-1}}{h}. \end{aligned} \quad (28.1)$$

Vamos integrar a expressão (28.1) duas vezes:

$$s'_i(t) = z_{i-1} \frac{-(t_i - t)^2}{2h} + z_i \frac{(t - t_{i-1})^2}{2h} + C_1 \quad (28.2)$$

$$s_i(t) = z_{i-1} \frac{(t_i - t)^3}{6h} + z_i \frac{(t - t_{i-1})^3}{6h} + C_1 t + C_2. \quad (28.3)$$

Como $s_i(t_{i-1}) = f(t_{i-1})$ e $s_i(t_i) = f(t_i)$, segue de (28.3) que

$$\begin{cases} s_i(t_{i-1}) = \frac{z_{i-1}h^3}{6h} + C_1 t_{i-1} + C_2 = f(t_{i-1}) \\ s_i(t_i) = \frac{z_i h^3}{6h} + C_1 t_i + C_2 = f(t_i). \end{cases}$$

A solução deste sistema é

$$C_1 = \frac{f(t_i) - f(t_{i-1})}{h} - h \frac{z_i - z_{i-1}}{6}, \quad C_2 = f(t_i) - t_i \frac{f(t_i) - f(t_{i-1})}{h} + h t_i \frac{z_i - z_{i-1}}{6} - \frac{z_i h^2}{6}. \quad (28.4)$$

Vamos colocar em evidência os termos em $C_1 t + C_2$ com coeficientes z_{i-1} , z_i , $f(t_{i-1})$ e $f(t_i)$:

$$C_1 t + C_2 = z_{i-1} \left(-h \frac{t_i - t}{6} \right) + z_i \left(-h \frac{t - t_{i-1}}{6} \right) + f(t_{i-1}) \left(\frac{t_i - t}{h} \right) + f(t_i) \left(\frac{t - t_{i-1}}{h} \right).$$

Substituindo esta expressão em (28.3), obtemos

$$s_i(t) = z_{i-1} \left[\frac{(t_i - t)^3}{6h} - h \frac{t_i - t}{6} \right] + z_i \left[\frac{(t - t_{i-1})^3}{6h} - h \frac{t - t_{i-1}}{6} \right] + f(t_{i-1}) \frac{t_i - t}{h} + f(t_i) \frac{t - t_{i-1}}{h}. \quad (28.5)$$

A expressão (28.5) nos permite determinar, uma vez calculados $z_1, \dots, z_{N-1}$, as funções $s_i(t)$ para todos os valores de i entre 1 e N . Para achar $z_1, \dots, z_{N-1}$, usamos as N equações provenientes da condição de continuidade da derivada primeira:

$$s'_i(t_i) = s'_{i+1}(t_i), \quad 1 \leq i \leq N. \quad (28.6)$$

Segue de (28.2) e (28.4) que

$$s'_i(t_i) = \frac{z_i h}{2} + \frac{f(t_i) - f(t_{i-1})}{h} - h \frac{z_i - z_{i-1}}{6} = z_{i-1} \frac{h}{6} + z_i \left(\frac{h}{2} - \frac{h}{6} \right) + \frac{f(t_i) - f(t_{i-1})}{h}$$

$$s'_i(t_i) = z_{i-1} \frac{h}{6} + z_i \frac{2h}{6} + \frac{f(t_i) - f(t_{i-1})}{h}. \quad (28.7)$$

Analogamente,

$$s'_i(t_{i-1}) = -z_{i-1} \frac{2h}{6} - z_i \frac{h}{6} + \frac{f(t_i) - f(t_{i-1})}{h},$$

e substituindo i por $i + 1$,

$$s'_{i+1}(t_i) = -z_i \frac{2h}{6} - z_{i+1} \frac{h}{6} + \frac{f(t_{i+1}) - f(t_i)}{h}. \quad (28.8)$$

Segue de (28.6), (28.7) e (28.8) que

$$z_{i-1} \frac{h}{6} + z_i \frac{4h}{6} + z_{i+1} \frac{h}{6} = \frac{f(t_{i+1}) - f(t_i)}{h} - \frac{f(t_i) - f(t_{i-1})}{h} \quad (1 \leq i \leq N),$$

ou seja,

$$z_{i-1} \frac{1}{6} + z_i \frac{4}{6} + z_{i+1} \frac{1}{6} = \frac{f(t_{i+1}) - 2f(t_i) + f(t_{i-1}))}{h^2} \quad (1 \leq i \leq N).$$

Lembrando que $z_0 = z_N = 0$, chegamos ao seguinte sistema tridiagonal:

$$\frac{1}{6} \begin{bmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & 4 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_{N-2} \\ z_{N-1} \end{bmatrix} = \frac{1}{h^2} \begin{bmatrix} f(t_2) - 2f(t_1) + f(t_0) \\ f(t_3) - 2f(t_2) + f(t_1) \\ \vdots \\ f(t_{N-1}) - 2f(t_{N-2}) + f(t_{N-3}) \\ f(t_N) - 2f(t_{N-1}) + f(t_{N-2}) \end{bmatrix}. \quad (28.9)$$

Convém notar que, no caso particular em que $f \in C^2([a, b])$,

$$\frac{z_{i-1} + 4z_i + z_{i+1}}{6} = \frac{s''(t_{i-1}) + 4s''(t_i) + s''(t_{i+1}))}{6} \approx s''(t_i), \quad \frac{f(t_{i+1}) - 2f(t_i) + f(t_{i-1}))}{h^2} \approx f''(t_i),$$

e analogamente, a média entre as equações (28.7) e (28.8) resulta na aproximação $s'(t_i) \approx f'(t_i)$. Assim, podemos interpretar que a interpolação de $f(t)$ por uma spline natural não somente coincide com $f(t)$ nos pontos da partição, como também aproxima as derivadas de primeira e segunda ordem de $f(t)$ nestes pontos.

Em geral, podemos ter $z_0 \neq 0$ ou $z_N \neq 0$. Supondo que z_0 e z_1 são conhecidos, devemos alterar o sistema (28.9) adicionando o vetor $\mathbf{z}_0 = [-z_0/6, 0, \dots, 0, -z_N/6]^T$ ao lado direito do sistema. Neste sentido, podemos determinar uma spline cúbica $s(t)$ a partir dos valores z_0 , z_N e $f(t_i)$ ($0 \leq i \leq N$). Além disso, $s(t)$ pode ser escrita como uma combinação linear das funções splines

- $s^{(1)}(t)$ definida por $z_0 = 1$, $z_N = 0$ e $f(t_i) = 0$ ($0 \leq i \leq N$),
- $s^{(j+2)}(t)$ definida por $z_0 = z_N = 0$, $f(t_j) = 1$ e $f(t_i) = 0$ ($0 \leq i \leq N$, $i \neq j$), com $0 \leq j \leq N$,
- $s^{(N+3)}(t)$ definida por $z_0 = 0$, $z_N = 1$ e $f(t_i) = 0$ ($0 \leq i \leq N$).

Portanto, a dimensão do espaço $S_3(\tau, [a, b])$ é no máximo $N + 3$. Na Aula 29 vamos construir outro conjunto de $N + 3$ splines (as B-splines) e veremos que este conjunto é linearmente independente, confirmando que $\dim(S_3(\tau, [a, b])) = N + 3$.

Aula 29: B-Splines cúbicas

Seja $B(t)$ a spline natural associada à partição $\tau = \{-2, -1, 0, 1, 2\}$ que interpola uma função $f(t)$ tal que

$$f(t_i) = \begin{cases} 0, & i = 0, \\ 1/6, & i = 1, \\ 4/6, & i = 2, \\ 1/6, & i = 3, \\ 0, & i = 4. \end{cases}$$

Sejam $z_i = B''(t_i)$ ($0 \leq i \leq 4$). Como $B(t)$ é uma spline natural, $z_0 = z_4 = 0$. Conforme vimos na Aula 28, os parâmetros z_1, z_2 e z_3 são soluções do sistema linear

$$\frac{1}{6} \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \frac{1}{h^2} \begin{bmatrix} f(t_2) - 2f(t_1) + f(t_0) \\ f(t_3) - 2f(t_2) + f(t_1) \\ f(t_4) - 2f(t_3) + f(t_2) \end{bmatrix} = \frac{1}{6h^2} \begin{bmatrix} 4 - 2(1) + 0 \\ 1 - 2(4) + 1 \\ 0 - 2(1) + 4 \end{bmatrix}.$$

Observando que $h = 1$, temos que

$$\begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -6 \\ 2 \end{bmatrix} \implies \begin{cases} z_1 = 1 \\ z_2 = -2 \\ z_3 = 1 \end{cases} \quad (29.1)$$

Substituindo estes valores em (28.5), determinamos $B(t)$ em cada intervalo da partição:

- Se $-2 \leq t \leq -1$,

$$\begin{aligned} B(t) = B_1(t) &= (0) \left[\frac{(t_1 - t)^3}{6h} - h \frac{t_1 - t}{6} \right] + (1) \left[\frac{(t - t_0)^3}{6h} - h \frac{t - t_0}{6} \right] + 0 \frac{t_1 - t}{h} + \frac{1}{6} \frac{t - t_0}{h} \\ &= \frac{(t+2)^3}{6} - \frac{t+2}{6} + \frac{t+2}{6} = \frac{(t+2)^3}{6} \end{aligned}$$

- Se $-1 < t \leq 0$,

$$\begin{aligned} B(t) = B_2(t) &= (1) \left[\frac{(t_2 - t)^3}{6h} - h \frac{t_2 - t}{6} \right] + (-2) \left[\frac{(t - t_1)^3}{6h} - h \frac{t - t_1}{6} \right] + \frac{1}{6} \frac{t_2 - t}{h} + \frac{4}{6} \frac{t - t_1}{h} \\ &= \frac{(-t)^3}{6} - \frac{-t}{6} - 2 \frac{(t+1)^3}{6} + 2 \frac{t+1}{6} + \frac{-t}{6} + 4 \frac{t+1}{6} = t + 1 - \frac{t^3 + 2(t+1)^3}{6} \end{aligned}$$

- Se $0 < t \leq 1$,

$$\begin{aligned} B(t) = B_3(t) &= (-2) \left[\frac{(t_3 - t)^3}{6h} - h \frac{t_3 - t}{6} \right] + (1) \left[\frac{(t - t_2)^3}{6h} - h \frac{t - t_2}{6} \right] + \frac{4}{6} \frac{t_3 - t}{h} + \frac{1}{6} \frac{t - t_2}{h} \\ &= -2 \frac{(1-t)^3}{6} + 2 \frac{1-t}{6} + \frac{t^3}{6} - \frac{t}{6} + 4 \frac{1-t}{6} + \frac{t}{6} = 1 - t + \frac{t^3 - 2(1-t)^3}{6} \end{aligned}$$

- Se $1 < t \leq 2$,

$$\begin{aligned} B(t) = B_4(t) &= (1) \left[\frac{(t_4 - t)^3}{6h} - h \frac{t_4 - t}{6} \right] + (0) \left[\frac{(t - t_3)^3}{6h} - h \frac{t - t_3}{6} \right] + \frac{1}{6} \frac{t_4 - t}{h} + 0 \frac{t - t_3}{h} \\ &= \frac{(2-t)^3}{6} - \frac{2-t}{6} + \frac{2-t}{6} = \frac{(2-t)^3}{6}. \end{aligned}$$

Note que $B(t_i) = B'(t_i) = B''(t_i) = 0$ se $i = 0$ e se $i = 4$. Portanto, se estendermos o domínio da spline $B(t)$ para $|t| > 2$ definindo $B(t) = 0$ para $|t| > 2$, ou seja,

$$B(t) = \begin{cases} 0, & t \leq -2, \\ \frac{(2+t)^3}{6}, & t \in]-2, -1], \\ 1 + t - \frac{t^3 + 2(1+t)^3}{6}, & t \in]-1, 0], \\ 1 - t + \frac{t^3 - 2(1-t)^3}{6}, & t \in]0, 1], \\ \frac{(2-t)^3}{6}, & t \in]1, 2], \\ 0, & t > 2, \end{cases} \quad (29.2)$$

a função $B(t)$ ainda é de classe C^2 e cúbica por partes, ou seja, ainda é uma spline cúbica. Note também que $B(t)$ é uma spline cúbica em qualquer partição do tipo

$$\tau = \{-M, \dots, -1, 0, 1, \dots, M\}, \quad M \geq 2.$$

Vamos considerar agora o intervalo $[0, N]$ dividido em subintervalos de comprimento $h = 1$, ou seja, $\tau = \{0, 1, \dots, N\}$. A partir da função $B(t)$ em (29.2), definimos

$$B_j(t) = B(t - j + 2)|_{[0, N]}. \quad (29.3)$$

Por exemplo, $B_1(t) = B(t + 1)|_{[0, N]}$ corresponde a uma translação de $B(t)$ de uma unidade para a esquerda seguida de um corte em $[0, N]$ (Fig. 13).

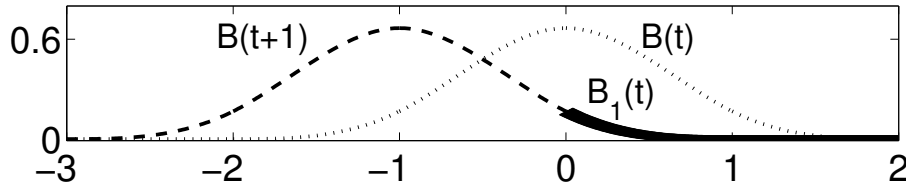


Figura 13: Função $B_1(t) = B(t + 1)|_{[0, N]}$ (linha sólida), comparada com $B(t + 1)$ e $B(t)$.

Lema 29.1 : O conjunto $\{B_1, \dots, B_{N+3}\}$ é linearmente independente.

Sejam $a_1, \dots, a_{N+3}$ tais que $v(t) = a_1 B_1(t) + \dots + a_{N+3} B_{N+3}(t) = 0 \quad \forall t \in [0, N]$. Temos que

$$v(t) = \sum_{j=0}^{N+3} a_j B_j(t) = \sum_{j=0}^{N+3} a_j B(t - j + 2)$$

Lembrando que $B(t) = 0$ se $|t| \geq 2$, temos que $B(0 - j + 2) = 0$ se $j \geq 4$. Portanto,

$$v(0) = a_1 B(1) + a_2 B(0) + a_3 B(-1) = a_1/6 + 4a_2/6 + a_3/6.$$

Além disso, $v''(0) = a_1 B''(1) + a_2 B''(0) + a_3 B''(-1) = a_1(1) + a_2(-2) + a_3(1)$, conforme (29.1). Como $v(0) = v''(0) = 0$, obtemos as equações

$$\begin{cases} a_1 + 4a_2 + a_3 = 0 \\ a_1 - 2a_2 + a_3 = 0, \end{cases} \quad (29.4)$$

que nos fornece $a_2 = 0$. Analogamente, segue das equações $v(N) = v''(N) = 0$ que

$$\begin{cases} a_{N+1} + 4a_{N+2} + a_{N+3} = 0 \\ a_{N+1} - 2a_{N+2} + a_{N+3} = 0, \end{cases} \quad (29.5)$$

que resulta em $a_{N+2} = 0$. Vamos determinar $a_3, \dots, a_{N+1}$ por meio das equações $v(i) = 0$ para $1 \leq i \leq N-1$, ou seja,

$$\begin{aligned} \sum_{j=0}^{N+3} a_j B(i-j+2) &= a_{i+1} B(i-i-1+2) + a_{i+2} B(i-i-2+2) + a_{i+3} B(i-i-3+2) \\ &= a_{i+1} B(1) + a_{i+2} B(0) + a_{i+3} B(-1) \\ &= a_{i+1}/6 + 4a_{i+2}/6 + a_{i+3}/6 = 0, \quad 1 \leq i \leq N-1. \end{aligned}$$

Lembrando que $a_2 = a_{N+2} = 0$, as equações acima correspondem ao sistema linear

$$\frac{1}{6} \begin{bmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & 4 \end{bmatrix} \begin{bmatrix} a_3 \\ a_4 \\ \vdots \\ a_N \\ z_{N+1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Como este é um sistema linear homogêneo cuja matriz é estritamente diagonalmente dominante (portanto, inversível), sua única solução é $a_3 = a_4 = \dots = a_N = a_{N+1} = 0$. Substituindo $a_2 = a_3 = 0$ em (29.4) e $a_{N+1} = a_{N+2} = 0$ em (29.4), encontramos $a_1 = a_{N+3} = 0$. Portanto $v(t) = 0$ implica $a_j = 0$, $1 \leq j \leq N+3$, ou seja, $B_1(t), \dots, B_{N+3}(t)$ são linearmente independentes.

Analogamente, dada uma partição uniforme $\tau = \{t_0, \dots, t_N\}$ do intervalo $[a, b]$ com espaçamento h , as funções

$$B_j(t) = B\left(\frac{t-a}{h} - j + 2\right) \Big|_{[a,b]}, \quad 1 \leq j \leq N+3, \quad (29.6)$$

são linearmente independentes. Como $B_j(t) \in S_3(\tau, [a, b])$ e $\dim(S_3(\tau, [a, b])) = N+3$, temos que as funções $B_j(t)$ em (29.6) formam uma base do espaço $S_3(\tau, [a, b])$. Estas funções são denominadas *B-splines* cúbicas. Em geral, B-splines de grau m podem ser definidas de modo recursivo [10, Sec. 6.5].

Aula 30: Método de Colocação com B-Splines cúbicas

Voltemos ao exemplo de um PVC apresentado na Aula 24:

$$\begin{cases} x''(t) = t^3 - t^2x(t), & t \in]0, 1[\\ x(0) = 0 \\ x(1) = 0, \end{cases}$$

Dada a partição uniforme $\tau = \{t_j = jh, 0 \leq j \leq N\}$ do intervalo $[0, 1]$ com espaçamento $h = 1/N$, buscamos uma aproximação $x_h(t)$ da solução $x(t)$ pelo método de colocação com as funções de base

$$\phi_j(t) = B_j(t) = B\left(\frac{t}{h} - j + 2\right) \Big|_{[0,1]}, \quad 1 \leq j \leq N + 3,$$

sendo $B(t)$ definida em (29.2) O inconveniente desta escolha é que as funções do espaço gerado pelas funções acima não necessariamente se anulam em $t = 0$ e $t = 1$. Por isto, devemos impor as condições de contorno a $x_h(t)$. O método de colocação neste caso é descrito como a seguir:

$$\text{Encontrar } x_h(t) = \sum_{j=1}^{N+3} X_j \phi_j(t) \text{ tal que } \begin{cases} x_h''(t_i) = t_i^3 - t_i^2 x_h(t_i), & 0 \leq i \leq N, \\ x_h(0) = 0, \\ x_h(1) = 0. \end{cases} \quad (30.1)$$

Lembremos que, dado um número inteiro k ,

$$B(k) = \begin{cases} 0, & k \leq -2, \\ 1/6, & k = -1, \\ 4/6, & k = 0, \\ 1/6, & k = 1, \\ 0, & k \geq 2, \end{cases} \quad B''(k) = \begin{cases} 0, & k \leq -2, \\ 1, & k = -1, \\ -2, & k = 0, \\ 1, & k = 1, \\ 0, & k \geq 2. \end{cases} \quad (30.2)$$

Como $\phi_j(t_i) = \phi_j(ih) = B(i - j + 2)$ e $\phi_j''(t_i) = (1/h^2)B''(t_i/h - j + 2) = B''(i - j + 2)/h^2$,

$$\phi_j(t_i) = \begin{cases} 0, & j \geq i + 4, \\ 1/6, & j = i + 3, \\ 4/6, & j = i + 2, \\ 1/6, & j = i + 1, \\ 0, & j \leq i, \end{cases} \quad \phi_j''(t_i) = \begin{cases} 0, & j \geq i + 4, \\ 1/h^2, & j = i + 3, \\ -2/h^2, & j = i + 2, \\ 1/h^2, & j = i + 1, \\ 0, & j \leq i. \end{cases} \quad (30.3)$$

As expressões (30.3) permitem especificar as equações que envolvem os coeficientes $X_1, \dots, X_{N+3}$ em (30.1). Da condição de contorno em $t = 0$, segue de (30.3) que

$$\sum_{j=1}^{N+3} X_j \phi_j(t_0) = X_1(1/6) + X_2(4/6) + X_3(1/6) = 0.$$

Por outro lado, segue da primeira equação de (30.1), com $0 \leq i \leq N$, que

$$\begin{aligned} x_h''(t_i) + t_i^2 x_h(t_i) &= t_i^3 \\ \sum_{j=1}^{N+3} X_j [\phi_j''(t_i) + t_i^2 \phi_j(t_i)] &= t_i^3 \\ \sum_{j=i+1}^{i+3} X_j [\phi_j''(t_i) + t_i^2 \phi_j(t_i)] &= t_i^3 \\ X_{i+1} \left[\frac{1}{h^2} + \frac{t_i^2}{6} \right] + X_{i+2} \left[\frac{-2}{h^2} + \frac{4t_i^2}{6} \right] + X_{i+3} \left[\frac{1}{h^2} + \frac{t_i^2}{6} \right] &= t_i^3, \end{aligned}$$

e da condição de contorno em $t = 1$,

$$\sum_{j=1}^{N+3} X_j \phi_j(t_N) = X_{N+1}(1/6) + X_{N+2}(4/6) + X_{N+3}(1/6) = 0.$$

Temos portanto o sistema linear

$$\begin{bmatrix} \frac{1}{6} & \frac{4}{6} & \frac{1}{6} & & & \\ \frac{t_1^2}{6} + \frac{1}{h^2} & \frac{4t_1^2}{6} - \frac{2}{h^2} & \frac{t_1^2}{6} + \frac{1}{h^2} & & & \\ & \frac{t_2^2}{6} + \frac{1}{h^2} & \frac{4t_2^2}{6} - \frac{2}{h^2} & \frac{t_2^2}{6} + \frac{1}{h^2} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{t_{N-1}^2}{6} + \frac{1}{h^2} & \frac{4t_{N-1}^2}{6} - \frac{2}{h^2} & \frac{t_{N-1}^2}{6} + \frac{1}{h^2} \\ & & & & \frac{t_N^2}{6} + \frac{1}{h^2} & \frac{4t_N^2}{6} - \frac{2}{h^2} & \frac{t_N^2}{6} + \frac{1}{h^2} \\ & & & & & \frac{1}{6} & \frac{4}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{N+1} \\ X_{N+2} \\ X_{N+3} \end{bmatrix} = \begin{bmatrix} 0 \\ t_0^3 \\ t_1^3 \\ \vdots \\ t_{N-1}^3 \\ t_N^3 \\ 0 \end{bmatrix}.$$

Embora este sistema não seja tridiagonal devido aos coeficientes $A_{1,3}$ e $A_{N-2,N}$ da primeira e da última linha da matriz acima, ele apresenta praticamente a mesma esparsidade que o sistema linear resultante do método de diferenças finitas.

OBS: A primeira equação de (30.1) também é imposta aos pontos extremos t_0 e t_N , por isso é preciso estender a equação diferencial do intervalo aberto $]0, 1[$ para o intervalo fechado $[0, 1]$.

Apendice A: Erro de truncamento global

Vamos retomar o estudo do erro de aproximação $e_n = x(t_n) - X_n$ abordado na Aula 5. Vimos que $e_n = \tilde{e}_n + e_n^{loc}$, sendo $\tilde{e}_n = x(t_n) - \tilde{x}(t)$ e $e_n^{loc} = \tilde{x}(t) - X_n$ com $\tilde{x}(t) = \tilde{x}(t, t_{n-1}, X_{n-1})$ definido como a solução do PVI auxiliar (5.1).

Temos que $e_0 = 0$ e, como $\tilde{x}(t, t_0, X_0) = x(t)$, temos $\tilde{e}_1 = 0$ e assim $e_1 = e_1^{loc}$. Entretanto, em geral $\tilde{e}_n \neq 0$ para $n \geq 2$. Neste caso,

$$\begin{aligned} \tilde{e}_n = x(t_n) - \tilde{x}(t_n) &= x(t_{n-1}) + \int_{t_{n-1}}^{t_n} f(t, x(t)) dt - \tilde{x}(t_{n-1}) - \int_{t_{n-1}}^{t_n} f(t, \tilde{x}(t)) dt \\ &= x(t_{n-1}) - \tilde{x}(t_{n-1}) + \int_{t_{n-1}}^{t_n} f(t, x(t)) - f(t, \tilde{x}(t)) dt \\ &= x(t_{n-1}) - \textcolor{red}{X}_{n-1} + \int_{t_{n-1}}^{t_n} f(t, x(t)) - f(t, \tilde{x}(t)) dt \\ &= e_{n-1} + \int_{t_{n-1}}^{t_n} \frac{\partial f}{\partial x}(t, \bar{x}(t)) [x(t) - \tilde{x}(t)] dt, \end{aligned}$$

sendo $\bar{x} = \bar{x}(t)$ entre $x(t)$ e $\tilde{x}(t)$. Seja $M > 0$ tal que

$$\frac{\partial f}{\partial x}(t, x) \leq M \quad \forall (t, x) \in [t_0, T] \times \mathbb{R}.$$

Neste caso, existe uma função $\alpha(t) \geq 0$ tal que

$$\frac{\partial f}{\partial x}(t, \bar{x}(t)) = M - \alpha(t).$$

Assim, definindo a função $\tilde{e}(t) = x(t) - \tilde{x}(t)$, encontramos

$$\tilde{e}(t_n) = e_{n-1} + \int_{t_{n-1}}^{t_n} [M - \alpha(t)] \tilde{e}(t) dt,$$

ou seja, $\tilde{e}(t)$ é a solução do PVI

$$\begin{cases} v'(t) = [M - \alpha(t)]v(t) \\ v(t_{n-1}) = e_{n-1} \end{cases} \quad (\text{A.1})$$

Vamos aplicar uma separação de variáveis no PVI (A.1):

$$\begin{aligned} \int_{t_{n-1}}^{t_n} \frac{dv}{v} &= \int_{t_{n-1}}^{t_n} M - \alpha(t) dt \\ \log |v(t_n)| - \log |v(t_{n-1})| &= Mh - \int_{t_{n-1}}^{t_n} \alpha(t) dt \leq Mh \\ \log(|v(t_n)|/|v(t_{n-1})|) &\leq Mh. \end{aligned}$$

Como $e^a \geq e^b$ para todo $a \geq b$, temos que $|v(t_n)|/|v(t_{n-1})| \leq e^{Mh}$, ou em termos de $\tilde{e}_n$,

$$|\tilde{e}_n| \leq |e_{n-1}| e^{Mh}. \quad (\text{A.2})$$

Dado

$$\delta = \max_{0 \leq j \leq n} |e_n^{loc}|,$$

vamos iterar a equação recursiva (A.2) até atingirmos $\tilde{e}_1 = 0$:

$$\begin{aligned}
|\tilde{e}_n| &\leq |e_{n-1}|e^{Mh} = |\tilde{e}_{n-1} + e_{n-1}^{loc}|e^{Mh} \leq (|\tilde{e}_{n-1}| + |e_{n-1}^{loc}|)e^{Mh} \leq |\tilde{e}_{n-1}|e^{Mh} + \delta e^{Mh} \\
&\leq (|\tilde{e}_{n-2}|e^{Mh} + \delta e^{Mh})e^{Mh} + \delta e^{Mh} = |\tilde{e}_{n-2}|e^{2Mh} + \delta(e^{2Mh} + e^{Mh}) \\
&\leq (|\tilde{e}_{n-3}|e^{Mh} + \delta e^{Mh})e^{2Mh} + \delta(e^{2Mh} + e^{Mh}) = |\tilde{e}_{n-3}|e^{3Mh} + \delta(e^{3Mh} + e^{2Mh} + e^{Mh}) \\
&\leq \dots \leq |\tilde{e}_{n-(n-1)}|e^{(n-1)Mh} + \delta(e^{(n-1)Mh} + \dots + e^{2Mh} + e^{Mh}).
\end{aligned}$$

Lembrando que $e_n = \tilde{e}_n + e_n^{loc}$ e que $\tilde{e}_1 = 0$,

$$\begin{aligned}
|e_n| &\leq |\tilde{e}_n| + |e_n^{loc}| \leq |\tilde{e}_n| + \delta \\
&\leq |\tilde{e}_1|e^{(n-1)Mh} + \delta \sum_{j=1}^{n-1} e^{jMh} + \delta = 0 + \delta \sum_{j=0}^{n-1} e^{jMh} = \delta \frac{e^{nMh} - 1}{e^{Mh} - 1} \\
&= \delta \frac{e^{M(t-t_0)} - 1}{[1 + Mh + (Mh)^2/2 + \dots] - 1} \\
&\leq \delta \frac{e^{M(t-t_0)} - 1}{(1 + Mh) - 1} = \delta \frac{e^{M(t-t_0)} - 1}{Mh} \leq \frac{\delta}{h} \frac{e^{M(T-t_0)} - 1}{M}.
\end{aligned}$$

Vimos que, para os métodos da série de Taylor e de Runge-Kutta, $e_n^{loc} = \mathcal{O}(h^{k+1})$ com $k \geq 1$ para todo n , ou seja,

$$\exists C > 0, \epsilon > 0 \quad | \quad |e_n^{loc}| \leq C|h^{k+1}| = Ch^{k+1} \quad \forall |h| \leq \epsilon,$$

com C independente de n , logo podemos escolher $\delta = Ch^{k+1}$ e assim, para $|h| \leq \epsilon$,

$$|e_n| \leq \frac{Ch^{k+1}}{h} \frac{e^{M(T-t_0)} - 1}{M} = \tilde{C}h^k, \quad \tilde{C} = C \frac{e^{M(T-t_0)} - 1}{M}$$

ou seja, $e_n = \mathcal{O}(h^k)$.

OBS: 1) Os passos realizados para obter a desigualdade (A.2) correspondem, essencialmente, à demonstração do **lema de Gronwall**;

2) ao contrário do erro de truncamento local τ_n , o erro local e_n^{loc} não requer a condição de estabilidade para garantir a convergência;

3) Lambert [11] mostra que τ_n e e_n^{loc} têm a mesma ordem de convergência para uma família de métodos de passos múltiplos (que inclui os métodos de Adams-Bashforth e Adams-Moulton).

Referências

- [1] P. Bogacki and L. F. Shampine. A 3 (2) pair of Runge-Kutta formulas. *Applied Mathematics Letters*, 2(4):321–325, 1989.
- [2] J. C. Butcher. Coefficients for the study of Runge-Kutta integration processes. *Journal of the Australian Mathematical Society*, 3(2):185–201, 1963.
- [3] J. C. Butcher. On the attainable order of Runge-Kutta methods. *Mathematics of Computation*, 19(91):408–417, 1965.
- [4] J. C. Butcher. A history of Runge-Kutta methods. *Applied Numerical Mathematics*, 20(3):247–260, 1996.
- [5] G. Dahlquist and A. Bjorck. *Numerical Methods*. Prentice-Hall, Englewood Cliffs, 1974.
- [6] J. R. Dormand and P. J. Prince. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980.
- [7] E. Fehlberg. Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems. Technical Report R315, NASA, 1969.
- [8] G. H. Golub and J. M. Ortega. *Scientific computing and differential equations: an introduction to numerical methods*. Academic Press, New York, 1992.
- [9] H. Keller. *Numerical Methods for Two-Point Boundary-Value Problems*. Blaisdell, Waltham, 1968.
- [10] D. Kincaid and W. Cheney. *Numerical Analysis: Mathematics of Scientific Computing*. American Mathematical Society, Providence, 2002.
- [11] J. D. Lambert. On the local error and the local truncation error of linear multistep methods. *BIT*, 30(4):673–681, 1990.
- [12] E. L. Lima. *Curso de análise, Volume 1*. IMPA, Rio de Janeiro, 1995.
- [13] F. Mendonça, M. Domingues, and E. Macau. Controle temporal e adaptabilidade espacial na resolução numérica de uma equação tipo KdV. *TEMA*, 9(2):265–274, 2008.
- [14] J. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*. Springer, New York, 1995.