Splines Cúbicos

EP2 - CompIII - **Data de entrega: 21/10/2021**

1 Introdução

Em alguns problemas de interpolação com muitos pontos, o uso de interpolação polinomial leva a soluções que oscilam muito. Uma técnica muito usada nestes casos é interpolação polinomial por partes, sendo os splines cúbicos muito populares devido a boas propriedades de aproximação e permitir gerar curvas suaves a partir dos valores tabelados de uma função.

Os objetivos deste exercíco-programa são introduzir os splines cúbicos e apresentar um caso simples de sistemas lineares envolvendo matrizes com estruturas. Veremos que para os cálculos precisaremos resolver sistemas lineares cujas matrizes são tridiagonais, estritamente diagonais dominantes. Neste caso, o uso do método de eliminação de Gauss sem trocas de linhas e sem condensação pivotal é estável numericamente e pode ser implementado eficientemente.

2 Definições

Considere uma *partição* do intervalo $[a, b]$

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

em n subintervalos $[x_{i-1}, x_i]$, $1 \leq i \leq n$, e sejam $y_i = f(x_i)$, $0 \leq i \leq n$, os valores de uma função f nos pontos da partição. Um *spline cúbico* interpolador da tabela (x_i, y_i), $0 \leq i \leq n$, subordinado à partição, é uma função $S(x)$ tal que:

a) $S(x_i) = y_i$, $0 \leq i \leq n$.

b) A restrição de S a cada subintervalo $[x_{i-1}, x_i]$, $1 \leq i \leq n$, é um polinômio de grau menor ou igual a 3.

c) S, S' e S'' são contínuas em $[a, b]$.

O termo spline foi cunhado pelo matemático I. J. Schoenberg em conexão com alguns problemas de ajuste de dados estudados por ele. Spline é o nome em inglês de uma régua fina flexível usada para desenhar curvas suaves passando por pontos prescritos. Usando teoria da elasticidade linear, pode-se mostrar que estas curvas são aproximadamente polinômios cúbicos por partes com derivadas até ordem 2 contínuas.

3 Construindo splines cúbicos

Os splines cúbicos interpoladores podem ser carcterizados pelos valores y_i da função e pelos valores

$$m_i = S''(x_i)$$

da derivada segunda de S nos pontos da partição. Para isso, denote por

$$S_i(x) = S(x), \quad x \in [x_{i-1}, x_i]$$

as restrições de S aos intervalos $[x_{i-1}, x_i]$ e por $h_i = x_i - x_{i-1}$ os comprimentos destes intervalos, $1 \leq i \leq n$. Então,

$$S_i''(x) = \frac{x_i - x}{h_i} m_{i-1} + \frac{x - x_{i-1}}{h_i} m_i.$$

Integrando a expressão acima duas vezes e usando a propriedade de interpolação ($S_i(x_i) = S_{i+1}(x_i)$, $1 \leq i \leq n-1$) obtemos

$$\begin{aligned} S_i(x) &= \frac{x_i - x}{h_i} y_{i-1} + \frac{x - x_{i-1}}{h_i} y_i + \\ & \frac{h_i^2}{6} \left\{ \left[\left(\frac{x_i - x}{h_i} \right)^3 - \frac{x_i - x}{h_i} \right] m_{i-1} + \left[\left(\frac{x - x_{i-1}}{h_i} \right)^3 - \frac{x - x_{i-1}}{h_i} \right] m_i \right\} \end{aligned} \tag{1}$$

onde a fórmula é válida para $x \in [x_{i-1}, x_i]$. Note que ela já incorpora as condições S e S'' contínuas (por que?). Para determinarmos as incógnitas m_i, usamos a continuidade de S', ou seja, $S_i'(x_i) = S_{i+1}'(x_i)$, $1 \leq i \leq n-1$. Usando estas relações em (1) e após uma manipulação das expressões chegamos ao seguinte sistema de equações

$$\frac{h_i}{h_i + h_{i+1}} m_{i-1} + 2m_i + \frac{h_{i+1}}{h_i + h_{i+1}} m_{i+1} = d_i \tag{2}$$

para $1 \leq i \leq n-1$, onde

$$d_i = \frac{6}{h_i + h_{i+1}} \left(\frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i} \right). \tag{3}$$

Note que (2) é um sistema linear com $n-1$ equações e $n+1$ incógnitas. Pode-se mostrar que a matriz do sistema tem posto máximo, e portanto há uma infinidade de splines cúbicos interpoladores. Para caracterizar um único spline, é necessário impor condições adicionais. Veremos a seguir algumas possibilidades.

4 Caracterizações de splines cúbicos

Veremos nesta seção algumas possibilidades para se obter um único spline cúbico interpolador [1].

a) **Spline cúbico natural:** é obtido impondo-se que as derivadas segundas de S em x_0 e x_n sejam nulas, isto é, $m_0 = m_n = 0$. Este é o spline gerado pela régua flexível.

[1]Veja a Seção 2.4.2 de J.Stoer, R.Bulirsch, *Introduction to Numerical Analysis*, Third Edition, Springer, 2002

b) **Spline cúbico completo:** se conhecermos os valores $y_0' = f'(x_0)$ e $y_n' = f'(x_n)$ da derivada de f nos extremos do intervalo, acrescentamos duas equações ao sistema (2) a partir de $S_1'(x_0) = y_0'$ e $S_n'(x_n) = y_n'$. Verifique como exercício quais são estas equações.

c) **Condição "not a knot":** Este spline, proposto por Carl de Boor, é obtido impondo-se que nos intervalos $[x_0, x_2]$ e $[x_{n-2}, x_n]$, S seja um polinômio de grau menor ou igual a 3. É como se "desligássemos os nós" x_1 e x_{n-1}. Estas condições são equivalentes a $S_1'''(x_1) = S_2'''(x_1)$ e $S_{n-1}'''(x_{n-1}) = S_n'''(x_{n-1})$, gerando o seguinte sistema $(n-1) \times (n-1)$: a primeira equação em (2) é modificada para

$$\left(2 + \frac{h_1}{h_2}\right) m_1 + \left(1 - \frac{h_1}{h_2}\right) m_2 = d_1,$$

a equação $n-1$ é modificada para

$$\left(1 - \frac{h_n}{h_{n-1}}\right) m_{n-2} + \left(2 + \frac{h_n}{h_{n-1}}\right) m_{n-1} = d_{n-1}$$

e para $2 \le i \le n-2$ usamos as mesmas equações de (2). Uma vez calculados $m_1, \dots, m_{n-1}$, obtemos m_0 e m_n por

$$m_0 = \frac{h_1 + h_2}{h_2} m_1 - \frac{h_1}{h_2} m_2; \;\; m_n = \frac{h_{n-1} + h_n}{h_{n-1}} m_{n-1} - \frac{h_n}{h_{n-1}} m_{n-2}.$$

d) **Spline cúbico periódico:** quando f é periódica de período $b - a$, temos $y_0 = y_n$. Podemos obter o spline cúbico periódico impondo-se $m_0 = m_n$ e $S'(x_0) = S'(x_n)$. As equações para $m_1, \dots, m_n$ ficam iguais a:

- $i = 1$: use $m_0 = m_n$ para obter

$$2m_1 + \frac{h_2}{h_1 + h_2} m_2 + \frac{h_1}{h_1 + h_2} m_n = d_1.$$

- $2 \le i \le n-1$: use as equações dadas por (2).
- $i = n$: Use $S_n'(x_n) = S_1'(x_0)$ para obter

$$\frac{h_1}{h_1 + h_n} m_1 + \frac{h_n}{h_1 + h_n} m_{n-1} + 2m_n = d_n$$

  onde a expressão (3) com $i = n$ deve ser avaliada usando-se $h_{n+1} = h_1$ e $y_{n+1} = y_1$.

Nos casos a), b) e c), o sistema linear resultante é tridiagonal e pode ser resolvido eficientemente pelo método de eliminação de Gauss, como descrito no Apêndice 1. No caso d), a matriz é *tridiagonal periódica*, e uma maneira de se resolver o sistema linear é descrita no Apêndice 2. Uma vez calculados os valores m_i, $0 \le i \le n$, podemos calcular $S(x)$ para qualquer $x \in [a, b]$. Determine o índice i tal que $x \in [x_{i-1}, x_i]$. Calcule então $A = (x_i - x)/h_i$ e $B = (x - x_{i-1})/h_i = 1 - A$. De (1) temos

$$S(x) = Ay_{i-1} + By_i + \frac{h_i^2}{6} \left[\left(A^3 - A\right) m_{i-1} + \left(B^3 - B\right) m_i \right].$$

5 Curvas bidimensionais suaves

Uma aplicação de interpolação com splines cúbicos é a construção de curvas suaves no plano passando por $n+1$ pontos P_k com coordenadas dadas (x_k, y_k), $0 \leq k \leq n$, onde uma representação geral na forma $y = f(x)$ não é possível. Portanto, devemos usar uma representação paramétrica

$$x = x(t), \quad y = y(t), \tag{4}$$

onde t denota o parâmetro. Podemos assumir que os valores t_0, t_1,...,t_n dos parâmetros correspondentes aos $n+1$ pontos estão em ordem crescente de magnitude. Construimos então dois splines cúbicos que interpolam as funções tabeladas (t_k, x_k) e (t_k, y_k), $0 \leq k \leq n$, por meio dos quais obtemos uma representação paramétrica (4).

O comprimento de arco da curva seria o parâmetro t mais apropriado. Como ele não é conhecido a priori, os valores t_k para o parâmetro são usualmente escolhidos como sendo as distâncias entre pontos consecutivos:

$$t_0 = 0, \quad t_k = t_{k-1} + \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}, \quad 1 \leq k \leq n.$$

Para gerar a curva entre P_{k-1} e P_k, os splines cúbicos $x(t)$ e $y(t)$ devem ser avaliados com o parâmetro t percorrendo o intervalo $[t_{k-1}, t_k]$.

6 Tarefa

Parte 1) Implemente um programa para construir os splines cúbicos interpoladores natural, completo e *not a knot*, dados os pontos x_i e os valores y_i de uma função nestes pontos, para $0 \leq i \leq n$. O programa deve calcular os valores m_i, $0 \leq i \leq n$, para cada um dos três casos, resolvendo-se os respectivos sistemas tridiagonais como descrito no Apêndice 1. O programa quando necessário deve permitir o cálculo dos valores dos splines em pontos diferentes dos pontos x_i.

Como teste para o seu programa, considere a função

$$f(x) = \frac{1}{2-x}, \quad x \in [0, 1].$$

Para cada valor de n = 10, 20, 40, 80 e 160, contrua os splines cúbicos interpoladores natural, completo e *not a knot* em relação aos pontos

$$x_i = \frac{i}{n} \quad \text{e} \quad y_i = f(x_i), \quad 0 \leq i \leq n$$

e estime o erro entre cada spline e f calculando

$$\max_{0 \leq k \leq 1000} |f(z_k) - S(z_k)|, \quad \text{com} \quad z_k = \frac{k}{1000},$$

onde S denota um dos três tipos de splines. Imprima n e os respectivos erros para cada spline. O que você observa? Estime para cada spline a potência de $h = \frac{1}{n}$ com a qual o erro tende a zero. Discuta os resultados.

Parte 2) Implemente um programa que usa splines cúbicos periódicos para a construção de curvas fechadas no plano, dadas as coordenadas de $n+1$ pontos P_0, P_1,…,P_n, onde $P_n = P_0$. O programa deve calcular e imprimir os valores t_k do parâmetro, $x''(t_k)$ e $y''(t_k)$, $0 \le k \le n$. Imprima também o valor de n e as coordenadas dos pontos P_k. Os sistemas lineares devem ser resolvidos pelo método descrito no Apêndice 2. Teste o seu programa com os seguintes dados:

k	0	1	2	3	4	5	6	7	8	9	10	11
x_k	25	19	13	9	5	2.2	1	3	8	13	18	25
y_k	5	7.5	9.1	9.4	9	7.5	5	2.1	2	3.5	4.5	5

Para ver a curva obtida, calcule $(x(t), y(t))$ para vários valores do parâmetro t além dos valores t_k, e plote a curva formada pelos pontos com estas coordenadas.

Apêndice 1: Sistemas lineares tridiagonais

Considere o sistema linear tridiagonal

$$\begin{array}{lcll} b_1x_1 + c_1x_2 & = & d_1 & \\ a_ix_{i-1} + b_ix_i + c_ix_{i+1} & = & d_i, & 2 \le i \le n-1 \\ a_nx_{n-1} + b_nx_n & = & d_n & \end{array}$$

Ao resolvermos este sistema pelo método de eliminação de Gauss, a matriz triangular superior U é bidiagonal, sendo que $U_{i,i+1} = c_i$, e os únicos multiplicadores necessários são os $l_{i+1,i}$ (VERIFIQUE!). Usando este fato, podemos resolver o sistema com $O(n)$ operações aritméticas. O algoritmo pode ser descrito da seguinte forma:

Triangularização

$u_1 = b_1$
$y_1 = d_1$
para $i = 2, \cdots, n$ **faça**
  $l = a_i/u_{i-1}$ (multiplicador)
  $u_i = b_i - l\, c_{i-1}$
  $y_i = d_i - l\, y_{i-1}$
fim

Após a execução deste laço, a diagonal de U e o lado direito modificado ficam armazenados em u e y, respectivamente. A solução do sistema linear é então calculada por:

Substituição regressiva

$x_n = y_n/u_n$
para $i = n-1, \cdots, 1$ **faça**
  $x_i = (y_i - c_i\, x_{i+1})/u_i$
fim

Para matrizes diagonais dominantes, este algoritmo é numericamente estável, não sendo necessário usar condensação pivotal.

Apêndice 2: Sistemas tridiagonais periódicos

Considere o sistema linear

$$\begin{array}{lcll} b_1x_1 + c_1x_2 + a_1x_n & = & d_1 & \\ a_ix_{i-1} + b_ix_i + c_ix_{i+1} & = & d_i, & 2 \le i \le n-1 \\ c_nx_1 + a_nx_{n-1} + b_nx_n & = & d_n & \end{array}$$

Sistemas lineares deste tipo aparecem na construção de splines cúbicos periódicos e em várias outras situações envolvendo periodicidade. Eles podem ser resolvidos a partir do método de eliminação de Gauss para sistemas lineares tridiagonais da seguinte forma.

Denote por $\tilde{A}$ seguinte submatriz de ordem $n-1$ da matriz A do sistema:

$$\tilde{A} = \begin{bmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-2} & b_{n-2} & c_{n-2} \\ & & & a_{n-1} & b_{n-1} \end{bmatrix}$$

Então, o sistema $Ax = d$ pode ser escrito na forma

$$\tilde{A}\tilde{x} + x_n\tilde{u} = \tilde{d}$$

$$\tilde{v}^T\tilde{x} + b_nx_n = d_n$$

onde $\tilde{x} = [x_1, x_2, \ldots, x_{n-2}, x_{n-1}]^T$, $\tilde{u} = [a_1, 0, \ldots, 0, c_{n-1}]^T$ ($n-1$ dimensional), $\tilde{v} = [c_n, 0, \ldots, 0, a_n]^T$ ($n-1$ dimensional) e $\tilde{d} = [d_1, d_2, \ldots, d_{n-2}, d_{n-1}]^T$. Logo,

$$x_n = \frac{d_n - \tilde{v}^T\tilde{z}}{b_n - \tilde{v}^T\tilde{y}} \quad \text{e} \quad \tilde{x} = \tilde{z} - x_n\tilde{y},$$

onde $\tilde{y}$ é a solução do sistema linear $\tilde{A}\tilde{y} = \tilde{u}$ e $\tilde{z}$ é a solução do sistema linear $\tilde{A}\tilde{z} = \tilde{d}$, ambos com matriz tridiagonal de ordem $n-1$. Estes sistemas podem ser resolvidos com o método do Apêndice 1.

Bônus: Como temos que resolver dois sistemas lineares com a mesma matriz tridiagonal $\tilde{A}$, é melhor obtermos a decomposição LU de $\tilde{A}$ e depois resolver os dois sistemas. Implemente esta decomposição LU e resolva os sistemas, *explorando a estrutura* de $\tilde{A}$. Não use a matriz cheia. Armazene as informações em vetores convenientes e trabalhe apenas com eles.