Interpolação

Leonardo F. Guidi

DMPA – IM
UFRGS

Cálculo Numérico

Índice

Interpolação

Neste capítulo estudaremos métodos que permitem encontrar um valor aproximado para uma função f calculada em um ponto x do intervalo I, através do conhecimento de uma coleção de pares ordenados (pontos) $\{(x_i, f(x_i))\}_{i=1}^{N}$tais que $x_i \in I$.
Seja g uma função que aproxima f no intervalo I. Se para um conjunto de pontos x_i, $i = 1, \ldots, N$

$$g(x_i) = f(x_i),$$

então dizemos que g *interpola* a função f nos valores x_1, $x_2, \ldots$, x_N.
Podemos utilizar a função g para encontrar uma aproximação para o valor de f no ponto $x \in [x_1, x_n]$, esse procedimento é denominado *interpolação*.
Se x estiver fora do intervalo $[x_1, x_n]$ e ainda assim utilizarmos a função g para encontrar o valor aproximado de f nesse ponto, o procedimento é denominado *extrapolação*.

Interpolação

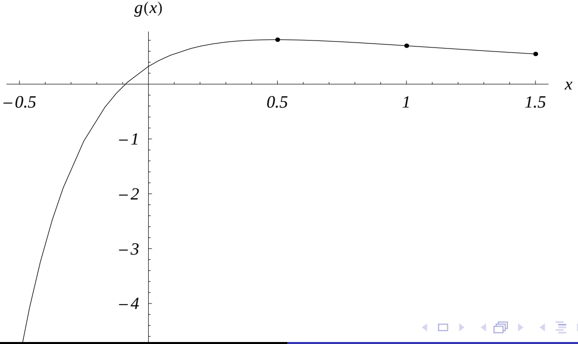
Vamos determinar uma função interpolante para o conjunto de pontos $\{(-0.5,-5.0);(0.5,0.81);$
$(1.0,0.7);(1.5,0.55)\}$ na forma $g(x)=a_1e^{a_2x}+a_3e^{a_4x}$. Determinar o valor dos coeficientes a_1, a_2, a_3 e a_4 significa determinar a interpolação.
Por definição, se g interpola o conjunto de pontos (entendido como $\{(x_i,f(x_i))\}_{i=1}^4$) então os coeficientes devem satisfazer as quatro equações $g(x_1)=f(x_1)$, ..., $g(x_4)=f(x_4)$, ou seja, devem ser solução do seguinte sistema de equações não lineares:

$$\left\{\begin{array}{ccc} a_1e^{-a_2 0.5}+a_3e^{-a_4 0.5} & = & -5.0 \\ a_1e^{a_2 0.5}+a_3e^{a_4 0.5} & = & 0.81 \\ a_1e^{a_2}+a_3e^{a_4} & = & 0.7 \\ a_1e^{a_2 1.5}+a_3e^{a_4 1.5} & = & 0.55 \end{array}\right. .$$

Interpolação

Esse sistema possui solução numérica dada por

$$\begin{aligned} a_1 &\approx 1.20334 \\ a_2 &\approx -0.519387 \\ a_3 &\approx -0.880292 \\ a_4 &\approx -4.01704 \end{aligned}$$



Interpolação

O sucesso em conseguir determinar a interpolação de um conjunto de pontos depende da escolha de função interpolante.

No exemplo anterior, a interpolação foi possível pois o "comportamento" dos pontos é compatível com a escolha realizada para a função interpolante. Essa "compatibilidade" se manifesta na existência de solução para o sistema de equações associado à interpolação. Se fosse escolhida uma função com comportamento muito distinto do manifestado pelos pontos, o sistema resultante poderia não possuir solução.

A escolha de polinômios como funções interpolantes é natural pelos seguintes motivos: é possível aproximar uma grande variedade de funções, os polinômios são de fácil manipulação matemática (principalmente derivação e integração) e o teorema de Weierstrass

Interpolação

Teorema Stone-Weierstrass

Seja f uma função contínua definida no intervalo fechado limitado $[a, b]$ e seja δ um número positivo. Então existe um polinômio p, tal que para todo $x \in [a, b]$,

$$|f(x) - p(x)| < \delta.$$

Interpolação

Teorema Stone-Weierstrass

Seja f uma função contínua definida no intervalo fechado limitado $[a, b]$ e seja δ um número positivo. Então existe um polinômio p, tal que para todo $x \in [a, b]$,

$$|f(x) - p(x)| < \delta.$$

No entanto, da mesma forma que o teorema de Weierstrass garante uma representação de f por um polinômio p tão "próximo" quanto queiramos, ele nada diz sobre o grau de p. Em algumas situações, o problema de encontrar p que desempenhe esse papel pode ser extraordinariamente difícil do ponto de vista numérico.

Interpolação

Antes de discutirmos o procedimento de interpolação por polinômios, vale a pena mencionar um algoritmo útil no cálculo do valor de p em um ponto x. Trata-se do algoritmo de Horner.

Algoritmo de Horner

Batizado com o nome do matemático inglês Willian George Horner mas já conhecido por Isaac Newton em 1669 e mesmo pelo matemático chinês Qin Jiunshao no séc. XIII, o algoritmo consiste em uma maneira otimizada de calcular $p(x) = a_m x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0$ através de m multiplicações e m adições.

Basta reescrever o polinômio na forma concatenada:

$$p(x) = ((\ldots((a_m x + a_{m-1})x + a_{m-a})x + \ldots + a_2)x + a_1)x + a_0.$$

Assim, $p(x)$ pode ser calculado iterativamente: se denominarmos $b_m = a_m$, $a_m x + a_{m-1} = b_m x + a_{m-1} = b_{m-1}$, então obtemos uma recursão para os b_i de modo que $p(x) = b_0$. Por exemplo, o polinômio $p(x) = 3x^3 + 8x^2 - x + 1 = ((3x+8)x - 1) + 1$. Nesse caso $b_3 = 3$, $b_2 = b_3 x + 8 = 3x + 8$, $b_1 = b_2 x - 1 = (3x+8)x - 1$ e finalmente $p(x) = b_0 = b_1 x + 1$.

Interpolação polinomial

Seja f_i, $i = 1, 2, \ldots, n$, o valor da função f calculada nos n pontos de interpolação x_i. Encontrar o polinômio de grau m que interpola f nesses pontos consiste em resolver o sistema de equações lineares $f_i \equiv f(x_i) = p(x_i)$, ou seja o sistema

$$\left\{\begin{array}{ccccccccccc} a_m x_1^m & + & a_{m-1} x_1^{m-1} & + & \ldots & + & a_1 x_1 & + & a_0 & = & f_1 \\ a_m x_2^m & + & a_{m-1} x_2^{m-1} & + & \ldots & + & a_1 x_2 & + & a_0 & = & f_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_m x_n^m & + & a_{m-1} x_n^{m-1} & + & & + & a_1 x_n & + & a_0 & = & f_n \end{array}\right.$$

As $m+1$ incógnitas são os coeficientes do polinômio, $a_0, a_1, \ldots, a_m$ e o sistema possui n equações. Portanto, tipicamente, o sistema não possui solução se $m+1 < n$, possui infinitas soluções se $m+1 > n$ e será unicamente determinado se $m+1 = n$.

Interpolação polinomial

Resolver esse sistema não é a maneira mais simples ou menos sujeita a erros de arredondamento quando desejamos determinar o polinômio interpolante.

O seguinte teorema garante a unicidade do polinômio interpolante, o que nos permite buscar maneiras alternativas de construí-lo. Por ser único, o resultado será independente da construção.

unicidade do polinômio interpolante

Sejam $x_1, \ldots, x_n$, pontos distintos. Para um conjunto arbitrário de valores $f_1, \ldots, f_n$ existe um e somente um polinômio p de grau menor ou igual a $n-1$ tal que

$$p(x_i) = f_i,$$

para $i = 1, 2, \ldots, n$.

Interpolação polinomial

Demonstração:

No caso em que temos n pontos distintos e procuramos um polinômio de grau menor ou igual a $n-1$, a matriz quadrada dos coeficientes do sistema de equações lineares assume a forma da seguinte matriz de Vandermonde,

$$\begin{pmatrix} x_1^{n-1} & x_1^{n-2} & \cdots & x_1^2 & x_1 & 1 \\ x_2^{n-1} & x_2^{n-2} & \cdots & x_2^2 & x_2 & 1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ x_n^{n-1} & x_n^{n-2} & \cdots & x_n^2 & x_n & 1 \end{pmatrix}.$$

Por hipótese os x_i são distintos, portanto o determinante da matriz, dado por

$$\prod_{1\leq i<j\leq n} (x_j - x_i)$$

é não nulo, consequentemente, a solução do sistema é única e o polinômio também

Método de Lagrange

De uma maneira geral, o polinômio interpolante pode ser determinado através a partir da escolha de uma base de polinômios de grau menor ou igual a $n-1$: $\{p_j\}_{j=1}^n$. Assim determinar a interpolação consiste em determinar as constantes c_j tais que

$$p(x) = \sum_{j=1}^{n} c_j p_j(x).$$

O método de Lagrande consistem em uma escolha particular para a base $\{p_j\}_{j=1}^n$.

.

Método de Lagrange

Vamos supor que para cada $1 \leq j \leq n$ exista um polinômio de grau $n-1$, $l_j(x)$ tal que para cada $1 \leq k \leq n$, o valor de l_j no ponto de interpolação x_k é tal que

$$l_j(x_k) = \delta_{j,k},$$

onde $\delta_{j,k}$ é o delta de Kronecker. .

Método de Lagrange

.
Essa escolha determina imediatamente as constantes c_j. Elas são dadas pelos valores f_j :

$$p(x) = f_1 l_1(x) + f_2 l_2(x) + \ldots + f_n l_n(x) = \sum_{j=1}^{n} f_j l_j(x),$$

podemos trivialmente verificar que $p(x_k) = \sum_{j=1}^{n} f_j l_j(x_k) = \sum_{j=1}^{n} f_j \delta_{j,k} = f_k$.

Método de Lagrange

.
Portanto se formos capazes de construir os polinômios l_j a interpolação estará determinada.
Segundo a sua definição $l_j(x_k) = 0$ para todo x_k tal que $k \neq j$, então os valores x_k são raízes de l_j se $j \neq k$ e portanto, a menos de uma constante multiplicativa, C_j, o polinômio l_j é determinado pelo produtório

$$\begin{aligned} l_j(x) &= C_j(x-x_1)(x-x_2)\ldots(x-x_{j-1})(x-x_{j+1})\ldots(x-x_n) \\ &= C_j \prod_{\substack{i=1 \\ i\neq j}}^{n} (x-x_i). \end{aligned}$$

Por fim, a constante C_j pode ser determinada através da propriedade $l_j(x_j) = 1$:

$$l_j(x_j) = 1 \quad \Rightarrow \quad C_j \prod_{\substack{i=1 \\ i\neq j}}^{n} (x_j - x_i) = 1,$$

ou seja

Método de Lagrange

.

Dessa forma, os polinômios $l_j(x)$, denominados polinômios de Lagrange são determinados a partir do seguinte produtório

$$l_j(x) = \prod_{\substack{i=1 \\ i \neq j}}^{n} \frac{x - x_i}{x_j - x_i}$$

e a interpolação de Lagrange

$$p(x) = \sum_{j=1}^{n} f_j l_j(x).$$

Método de Lagrange

.

Exemplo:

Seja a função $f(x) = \text{sen}(x)$ a partir da qual construímos a interpolação nos três pontos $x_1 = 0$, $x_2 = 1$ e $x_3 = 2$. Será então um polinômio de segundo grau. Os pontos de interpolação são dados por

j	x_j	$f_j = \text{sen}(x_j)$
1	0	0
2	1	$\text{sen}(1)$
3	2	$\text{sen}(2)$

os polinômios de Lagrange são então dados por

$$l_1(x) = \frac{(x-1)(x-2)}{(0-1)(0-2)} = \frac{x^2 - 3x + 2}{2},$$

$$l_2(x) = \frac{(x-0)(x-2)}{(1-0)(1-2)} = -x^2 + 2x$$

Método de Newton

De acordo com o teorema da unicidade do polinômio interpolante, toda interpolação de n pontos por um polinômio de grau $n-1$ é única e pode ser obtida pelo método de Lagrange.

No entanto, existem outras maneiras de construir o polinômio $p(x)$ que podem ser mais convenientes.

Uma dessas maneiras é a interpolação de Newton, que permite a inserção de pontos adicionais de maneira simples e menos suscetível à deterioração por erros de arredondamento.

O método consiste em determinar o polinômio a partir da seguinte estrutura

$$p(x) = a_0 + a_1(x-x_1) + a_2(x-x_1)(x-x_2) + \ldots + a_{n-1}(x-x_1)\ldots(x-x_{n-1}).$$

Método de Newton

Por construção, o valor de p calculado em $x = x_1$ é

$$p(x_1) = a_0.$$

Além disso, como $p(x)$ é o polinômio interpolante, $p(x_1) = f_1$, portanto,

$$a_0 = f_1.$$

Método de Newton

Da mesma forma,

$$\begin{aligned} p(x_2) &= a_0 + a_1(x_2 - x_1) = f_2 \\ &= f_1 + a_1(x_2 - x_1) = f_2, \end{aligned}$$

ou seja,

$$a_1 = \frac{f_2 - a_0}{x_2 - x_1}$$

e assim por diante, os coeficientes são determinados recursivamente e o k-ésimo coeficiente é determinado em função dos pontos de interpolação e dos coeficientes anteriores pela expressão

$$a_k = \frac{f_{k+1} - a_0 - \sum_{j=1}^{k-1} a_j(x_{k+1} - x_1)\dots(x_{k+1} - x_j)}{\prod_{j=1}^{k}(x_{k+1} - x_j)}.$$

Método de Newton

Essa fórmula de recorrência pode ser convenientemente descrita através da notação de ***diferenças divididas***. Seja a função $f[x_k, x_{k+1}, \ldots, x_{l+1}]$ definida pela relação de recorrência

$$f[x_k, x_{k+1}, \ldots, x_l, x_{l+1}] = \frac{f[x_{k+1}, x_{k+1}, \ldots, x_{l+1}] - f[x_k, x_{k+1}, \ldots, x_l]}{x_{l+1} - x_k}$$

e

$$f[x_k] = f_k \doteq f(x_k).$$

Assim, podemos verificar que

$$f[x_k, x_{k+1}] = \frac{f[x_{k+1}] - f[x_k]}{x_{k+1} - x_k}$$

e

$$f[x_k, x_{k+1}, x_{k+2}] = \frac{f[x_{k+1}, x_{k+2}] - f[x_k, x_{k+1}]}{x_{k+2} - x_k}.$$

Método de Newton

Nessa notação, os coeficientes do polinômio são dados por

$$\begin{aligned}
a_0 &= f[x_1],\\
a_1 &= f[x_1, x_2],\\
a_2 &= f[x_1, x_2, x_3],\\
\vdots \quad &\vdots \quad \vdots\\
a_{n-1} &= f[x_1, x_2, \ldots, x_n].
\end{aligned}$$

Método de Newton

Diagramaticamente, os coeficientes são calculados a partir da sequência de diferenças divididas calculadas recursivamente:

$$
\begin{array}{ccccccccccc}
x_1 & \rightarrow & f[x_1] & \rightarrow & f[x_1,x_2] & \rightarrow & f[x_1,x_2,x_3] & \cdots & f[x_1,\ldots,x_{n-1}] & \rightarrow & f[x_1,\ldots \\
 & & & \nearrow & & \nearrow & & & & \nearrow & \\
x_2 & \rightarrow & f[x_2] & \rightarrow & f[x_2,x_3] & \rightarrow & f[x_2,x_3,x_4] & \cdots & f[x_2,\ldots,x_n] & & \\
 & & & \nearrow & & \nearrow & & & & & \\
x_3 & \rightarrow & f[x_3] & \rightarrow & f[x_3,x_4] & \rightarrow & f[x_3,x_4,x_5] & \cdots & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & & & \\
x_{n-2} & \rightarrow & f[x_{n-2}] & \rightarrow & f[x_{n-2},x_{n-1}] & \rightarrow & f[x_{n-2},x_{n-1},x_n] & \cdots & & & \\
 & & & \nearrow & & \nearrow & & & & & \\
x_{n-1} & \rightarrow & f[x_{n-1}] & \rightarrow & f[x_{n-1},x_n] & & & & & & \\
 & & & \nearrow & & & & & & & \\
x_n & \rightarrow & f[x_n] & & & & & & & &
\end{array}
$$

Método de Newton

Exemplo 1:
Vamos realizar a interpolação da função $\text{sen}(x)$ no intervalo $x \in [0,2]$ através de um polinômio de segundo grau nos pontos $x_1 = 0$, $x_2 = 1$ e $x_3 = 2$. Neste caso,

j	x_j	$f_j = \text{sen}(x_j)$
1	0	0
2	1	$\text{sen}(1)$
3	2	$\text{sen}(2)$

e $f[x_1] = 0$, $f[x_2] = \text{sen}(1)$ e $f[x_3] = \text{sen}(2)$. As próximas diferenças divididas são dadas por

$$f[x_1,x_2] = \frac{f[x_2]-f[x_1]}{x_2-x_1} = \frac{\text{sen}(1)-0}{1-0} \text{ e } f[x_2,x_3] = \frac{f[x_3]-f[x_2]}{x_3-x_2} = \frac{\text{sen}(2)-\text{sen}(1)}{2-1}.$$

Finalmente,

$$f[x_1,x_2,x_3] = \frac{f[x_2,x_3]-f[x_1,x_2]}{x_3-x_1} = \frac{\text{sen}(2)-\text{sen}(1)-\text{sen}(1)}{2-0}.$$

Método de Newton

Portanto, o polinômio interpolante

$$p(x) = f[x_1] + f[x_1, x_2](x - x_1) + f[x_1, x_2, x_3](x - x_1)(x - x_2)$$

é

$$p(x) = \text{sen}(1)\, x + \frac{\text{sen}(2) - 2\text{sen}(1)}{2} x(x - 1)$$

Relação com a série de Taylor

O polinômio interpolante na forma de Newton

$$p(x) = f[x_1] + f[x_1, x_2](x - x_1) + \ldots + f[x_1, x_2, \ldots, x_n](x - x_1) \cdot \ldots \cdot (x - x_{n-1})$$

pode ser relacionado à série de Taylor para a função f através do seguinte limite:
Sejam as constantes positivas r_i, ordenadas de acordo com as desigualdades $0 < r_1 < r_2 < r_{n-1}$. A partir das constantes r_i, escolhemos os pontos de interpolção x_i dados por

$$x_i = x_0 + r_i h.$$

A intenção é tomar o limite $h \to 0$ de modo que todos os pontos $x_i \to x_0$ sem que cruzem em um valor finito de h. Desse modo a interpolação está sempre definida e além disso temos que

$$\lim_{h \to 0} f[x_1, x_2, \ldots, x_k] = \frac{f^{(k)}(x_0)}{k!}.$$

Relação com a série de Taylor

Portanto, quando $h \to 0$, o polinômio interpolante assume a forma de uma série de Taylor truncada:

$$f(x_0) + \frac{f'(x_0)}{2}(x - x_0) + \ldots + \frac{f^{(n-1)}(x_0)}{(n-1)!}(x - x_0)^{n-1}.$$

Qual é a importância desse fato? Veremos adiante que ele ajuda a compreender algumas dificuldades associadas à interpolação polinomial.

Erros de truncamento

Seja f uma função contínua e n vezes diferenciável no intervalo (a,b) que contém os pontos $x_1, x_2, \ldots, x_n$ e seja p o polinômio de grau $n-1$ que interpola f nesses pontos. Então é possível mostrar que para cada $x \in (a,b)$, existe um $\zeta(x) \in (a,b)$ tal que

$$f(x) - p(x) = \frac{1}{n!} f^{(n)}(\zeta) \prod_{i=1}^{n} (x - x_i).$$

Poderíamos supor que para uma f contínua e suficientemente suave, a sequência de polinômios interpolantes $\{p_n\}_{n\geq 1}$ convergiria para f conforme aumentássemos o número de pontos de interpolação no intervalo (a,b). No entanto, como o exemplo a seguir ilustra, isto nem sempre ocorre.

Erros de truncamento

Fenômeno de Runge

A seguinte função, proposta por Carle D. T. Runge ao estudar o comportamento dos erros na interpolação polinomial,

$$f(x) = \frac{1}{1+25x^2}, \qquad x \in [-1,1]$$

é tal que a sequência de polinômios interpolantes $\{p_n\}_n$ construídos a partir de pontos de interpolação igualmente espaçados não converge para $f(x)$ no intervalo de valores $x \in (-1,-0.727) \bigcup (0.727,1)$. Na realidade é possível demonstrar que

$$\lim_{n\to+\infty} \max_{-1\leq x\leq 1} |f(x) - p_n(x)| = +\infty.$$

Erros de truncamento

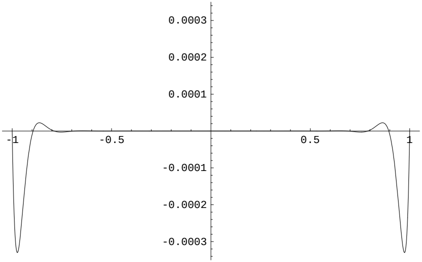
Podemos analisar esse comportamento não regular da interpolação a partir do termo

$$\prod_{i=1}^{n}(x - x_i)$$

contido na expressão . Esse produtório possui uma flutuação para os valores do argumento próximos à fronteira do intervalo $(-1, 1)$ que é progressivamente ampliada conforme aumentamos o número de pontos se os mesmos forem igualmente espaçados.

Os gráfico seguintes ajudam a ilustrar o comportamento do produtório.

Erros de truncamento



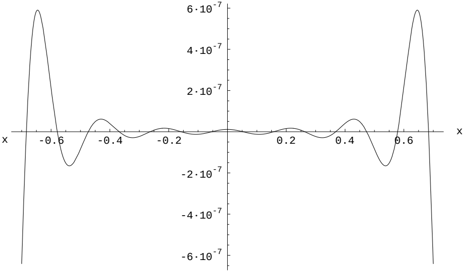


Figura: a) comportamento do produtório com 20 pontos igualmente espaçados no intervalo $[-1,1]$. b) recorte do mesmo produtório no intervalo $[-0.7,0.7]$.

Erros de truncamento

Esse comportamento é minimizado através da escolha de pontos não igualmente espaçados.
Na realidade, é possível demonstrar que a variação desse termo é mínima em valor absoluto quando os pontos x_i estão espaçados em um intervalo (a, b) segundo a seguinte expressão

$$x_i = \frac{a+b}{2} + \frac{a-b}{2} \cos\left(\frac{2i-1}{2n}\pi\right)$$

para $i = 1, 2, \ldots, n$. Esses pontos são denominados *pontos de Chebyshev*.
Utilizando os pontos de Chebyshev no intervalo $[-1, .1]$ podemos controlar o comportamento dos polinômios interpolantes para a função de Runge e garantir a convergência $p_{n-1}(x) \to f(x)$ quando $n \to +\infty$.

Erros de truncamento

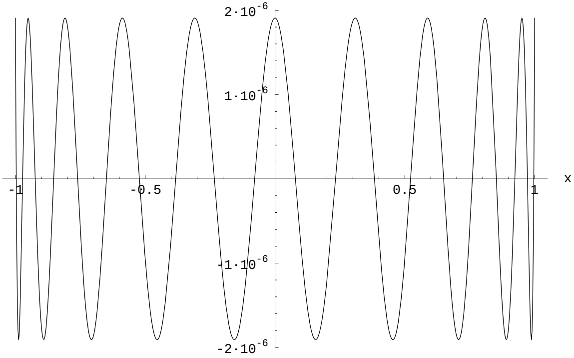


Figura: O produtório com 20 pontos de Chebyshev

Erros de truncamento

Ainda assim, existem funções contínuas que requerem um número impraticável de pontos para que a interpolação se aproxime da função original. Por exemplo, a função $\sqrt{|x|}$ no intervalo $[-1,1]$ requer um polinômio de grau maior que 10^6 para que a interpolação seja exata até 10^{-3}.
Em geral, quando utilizamos polinômios de grau maior ou igual a 100, a maior dificuldade é lidar com os erros de arredondamento.

Interpolação segmentada

Splines são funções formadas por diferentes polinômios de grau menor ou igual a um m, definidos para cada intervalo entre os pontos de interpolação de modo que em cada ponto de interpolação o spline é contínuo,assim como todas as derivadas até ordem $m-1$.

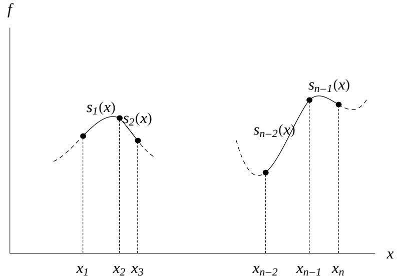


Figura: Interpolação spline

Interpolação segmentada

Nas situações em que o número de pontos de interpolação é grande (por exemplo, em aplicações CAD – *computer-aided design*), a inexatidão na aproximação obtida com um polinômio de grau elevado é dominada pelos erros de arredondamento. Ou então, quando a função que se quer interpolar possui derivadas de valor numérico elevado em alguma região do intervalo de interpolação, a aproximação é prejudicada em todo o intervalo. Nessas situações, a interpolação por spline pode auxiliar a tarefa de interpolação.

Interpolação segmentada

Nas situações em que o número de pontos de interpolação é grande (por exemplo, em aplicações CAD – *computer-aided design*), a inexatidão na aproximação obtida com um polinômio de grau elevado é dominada pelos erros de arredondamento. Ou então, quando a função que se quer interpolar possui derivadas de valor numérico elevado em alguma região do intervalo de interpolação, a aproximação é prejudicada em todo o intervalo. Nessas situações, a interpolação por spline pode auxiliar a tarefa de interpolação.

O procedimento de construir splines é análogo qualquer que seja o grau dos polinômios utilizados, como o spline de maior interesse (veremos porque) é aquele formado por polinômios de grau 3, nos concentraremos nesse caso apenas.

Spline cúbico

Sejam $x_1 < x_2 < \ldots < x_n$ os pontos e interpolação. um spline cúbico é uma função $s(x)$, definida no intervalo $[x_1, x_n]$ com as seguintes propriedades:

1. $s(x)$, $s'(x)$ e $s''(x)$ são funções contínuas no intervalo (x_1, x_n).
2. Em cada subintervalo $[x_i, x_{i+1}]$, $s(x)$ é um polinômio cúbico tal que $s(x_i) = f_i \doteq f(x_i)$ para $i = 1, 2, \ldots, n$.

Portanto, s é composto por $n-1$ polinômios cúbicos, cada polinômio é determinado por 4 coeficientes (a_i, b_i, c_i e d_i) o que dá um total de $4n-4$ coeficientes a determinar, ou seja $4n-4$ incógnitas.

Spline cúbico

Sejam $x_1 < x_2 < \ldots < x_n$ os pontos e interpolação. um spline cúbico é uma função $s(x)$, definida no intervalo $[x_1, x_n]$ com as seguintes propriedades:

1. $s(x)$, $s'(x)$ e $s''(x)$ são funções contínuas no intervalo (x_1, x_n).
2. Em cada subintervalo $[x_i, x_{i+1}]$, $s(x)$ é um polinômio cúbico tal que $s(x_i) = f_i \doteq f(x_i)$ para $i = 1, 2, \ldots, n$.

Portanto, s é composto por $n-1$ polinômios cúbicos, cada polinômio é determinado por 4 coeficientes (a_i,b_i, c_i e d_i) o que dá um total de $4n-4$ coeficientes a determinar, ou seja $4n-4$ incógnitas.
Cada polinômio deve satisfazer a condição de continuidade nos pontos de interpolação além, é claro, de interpolar o ponto x_i, ou seja,

$$s_i(x_i) = f_i \quad \text{(interpolação)},$$

para $i = 1, 2, \ldots, n-1$ e

$$s_{n-1}(x_n) = f_n.$$

Spline cúbico

Cada polinômio deve satisfazer a condição de continuidade nos pontos de interpolação além, é claro, de interpolar o ponto x_i, ou seja,

$$s_i(x_i) = f_i \quad \text{(interpolação)},$$

para $i = 1, 2, \ldots, n-1$ e

$$s_{n-1}(x_n) = f_n.$$

A continuidade é satisfeita se

$$s_i(x_{i+1}) = f_{i+1} \ \ \text{(continuidade de } s \text{)},$$

para $i = 1, 2, \ldots, n-2$.. As condições acima implicam $2(n-1)$ equações. Faltam ainda as continuidades de $s'(x)$ e $s''(x)$:

$$s_i'(x_{i+1}) = s_{i+1}'(x_{i+1}) \ \ \text{(continuidade de } s' \text{)},$$

$$s_i''(x_{i+1}) = s_{i+1}''(x_{i+1}) \ \ \text{(continuidade de } s'' \text{)},$$

para $i = 1, 2, \ldots, n-2$. Cada condição equivale a $n-2$ equações.

spline natural e spline (quase) completo

Portanto temos até agora um total de $4n - 6$ equações. Restam duas equações para que seu número seja igual ao número de incógnitas. Essas duas últimas equações relacionam-se com as condições de fronteira do spline.

spline natural e spline (quase) completo

Portanto temos até agora um total de $4n-6$ equações. Restam duas equações para que seu número seja igual ao número de incógnitas. Essas duas últimas equações relacionam-se com as condições de fronteira do spline.

Com relação ao comportamento de $s(x)$ no extremo do intervalo, existem várias possibilidades a ser consideradas. Vamos nos concentrar em duas:

spline natural e spline (quase) completo

Com relação ao comportamento de $s(x)$ no extremo do intervalo, existem várias possibilidades a ser consideradas. Vamos nos concentrar em duas:

spline natural

$$\begin{aligned} s_1''(x_1) &= 0 \\ s_{n-1}''(x_n) &= 0 \end{aligned}$$

possui esse nome por ser a condição equivalente à aproximação por réguas elásticas (uso mais tradicional do spline).

spline natural e spline (quase) completo

splines completo e quase completo

$$
\begin{aligned}
s_1'(x_1) &= f'(x_1) \\
s_{n-1}'(x_n) &= f'(x_n)
\end{aligned}
$$

essa escolha pressupõe que a informação sobre o valor da derivada de f nos extremos do intervalo seja conhecida (exatamente no caso completo e como aproximação numérica a partir dos demais pontos no caso quase completo). A aproximação obtida com essa escolha possui uma maior acurácia do que a obtida com o spline natural.

spline natural e spline (quase) completo

Nos próximos parágrafos montaremos o sistema de equações lineares para determinarmos $4n-4$ os coeficientes a_i,b_i, c_i e d_i dos $n-1$ polinômios que compõe o spline:

$$s_i(x) = a_i + b_i(x-x_i) + c_i(x-x_i)^2 + d_i(x-x_i)^3.$$

Por ser uma interpolação, a cada x_i, temos que $s(x_i) = f_i$, ou seja, $s_i(x_i) = f_i$ o que implica

$$f_i = s_i(x_i) = a_i$$

para $i = 1, 2, \ldots, n-1$. Isto determina o valor dos coeficientes a_i.

spline natural e spline (quase) completo

Nos próximos parágrafos montaremos o sistema de equações lineares para determinarmos $4n-4$ os coeficientes a_i,b_i, c_i e d_i dos $n-1$ polinômios que compõe o spline:

$$s_i(x) = a_i + b_i(x-x_i) + c_i(x-x_i)^2 + d_i(x-x_i)^3.$$

Por ser uma interpolação, a cada x_i, temos que $s(x_i) = f_i$, ou seja, $s_i(x_i) = f_i$ o que implica

$$f_i = s_i(x_i) = a_i$$

para $i = 1, 2, \ldots, n-1$. Isto determina o valor dos coeficientes a_i.
A continuidade do spline $s(x)$ nos pontos de interpolação implica a equação $s_i(x_{i+1}) = s_{i+1}(x_{i+1})$ para $i = 1, 2, \ldots, n-2$, ou seja,

$$\begin{aligned} a_i + b_i(x_{i+1}-x_i) + c_i(x_{i+1}-x_i)^2 + d_i(x_{i+1}-x_i)^3 &= a_{i+1}, \\ f_i + b_i(x_{i+1}-x_i) + c_i(x_{i+1}-x_i)^2 + d_i(x_{i+1}-x_i)^3 &= f_{i+1}. \end{aligned}$$

Para aliviar a notação, vamos introduzir a notação $h_i = (x_{i+1} - x_i)$.

spline natural e spline (quase) completo

Dessa forma, a equação anterior pode ser reescrita como

$$f_i + b_i h_i + c_i h_i^2 + d_i h_i^3 = f_{i+1}$$

spline natural e spline (quase) completo

Dessa forma, a equação anterior pode ser reescrita como

$$f_i + b_i h_i + c_i h_i^2 + d_i h_i^3 = f_{i+1}$$

A continuidade na primeira e na segunda derivadas implicam

$$b_i + 2c_i h_i + 3d_i h_i^2 = b_{i+1}$$

e

$$c_i + 3d_i h_i = c_{i+1}$$

para $i = 1, 2, \ldots, n-2$.

spline natural e spline (quase) completo

Isolando d_i na última equação e substituindo o resultado nas demais encontramos respectivamente

$$f_{i+1} = f_i + b_i h_i + \frac{h_i^2}{3}(2c_i + c_{i+1})$$

e

$$b_{i+1} = b_i + h_i(c_i + c_{i+1})$$

para $i = 1, 2, \ldots, n-2$.
Isolando b_i na equação anterior podemos determiná-lo em termos dos valores conhecidos f_i, h_i e da incógnita c_i:

$$b_i = \frac{f_{i+1} - f_i}{h_i} - \frac{h_i}{3}(2c_i + c_{i+1}),$$

para $i = 1, 2, \ldots, n-2$.

spline natural e spline (quase) completo

A partir da equação para a continuidade de s e das expressões para b_i e d_i dadas em termos de c_i, h_i e f_i, encontramos uma equação para os coeficientes c_i em termos dos valores conhecidos f_i e h_i:

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} = 3\left(\frac{f_{i+1} - f_i}{h_i}\right) - 3\left(\frac{f_i - f_{i-1}}{h_{i-1}}\right),$$

para $i = 2, 3, \ldots, n-1$.

spline natural e spline (quase) completo

A partir da equação para a continuidade de s e das expressões para b_i e d_i dadas em termos de c_i, h_i e f_i, encontramos uma equação para os coeficientes c_i em termos dos valores conhecidos f_i e h_i:

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} = 3\left(\frac{f_{i+1} - f_i}{h_i}\right) - 3\left(\frac{f_i - f_{i-1}}{h_{i-1}}\right),$$

para $i = 2, 3, \ldots, n-1$.
A equação anterior define um sistema de equações lineares para as incógnitas c_i. Note que além dos coeficientes $c_1, c_2, \ldots, c_{n-1}$, o sistema envolve um coeficiente c_n que não está diretamente relacionado a algum dos $n-1$ polinômios s_i . Na realidade, c_n está relacionado às condições no extremo do intervalo de interpolação e sua determinação depende do tipo de spline que estamos construindo, se é um spline natural ou um spline (quase) completo

spline natural e spline (quase) completo

Spline natural

O spline natural deve satisfazer as condições $s''(x_1) = 0$ e $s''(x_n) = 0$, estas duas equações implicam respectivamente

$$c_1 = 0$$

e

$$2c_{n-1} + 6d_{n-1}h_{n-1} = 0.$$

A equação para d_i em termos de c_i implica $c_n = 0$.

spline natural e spline (quase) completo

Colecionando esses resultados temos então a seguinte situação: resolvendo o sistema de equações

$$\begin{cases} c_1 & = & 0 \\ h_{i-1}c_{i-1} + 2(h_{i-1}+h_i)c_i + h_i c_{i+1} & = & 3\left(\frac{f_{i+1}-f_i}{h_i}\right) - 3\left(\frac{f_i - f_{i-1}}{h_{i-1}}\right), \quad i = 2, \ldots, \\ c_n & = & 0 \end{cases}$$

encontramos o valor dos coeficientes c_i. A partir desses coeficientes determinamos o valor dos coeficientes b_i através das equações

$$b_i = \frac{f_{i+1}-f_i}{h_i} - \frac{h_i}{3}(2c_i + c_{i+1}),$$

para $i = 1, 2, \ldots, n-1$; e o valor dos coeficientes d_i através da equações

$$d_i = \frac{c_{i+1} - c_i}{3h_i},$$

spline natural e spline (quase) completo

Splines completo e quase completo

Nesse caso o spline deve satisfazer as condições $s'(x_1) = f'(x_1) \equiv f_1'$ e $s'(x_n) = f'(x_n) \equiv f_n'$.Para determinar o spline, f_1' e f_n' devem ser valores conhecidos. As condições implicam respectivamente

$$\begin{cases} 2h_1 c_1 + h_1 c_2 & = & 3\left(\frac{f_2 - f_1}{h_1}\right) - 3f_1' \\ h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_i c_{i+1} & = & 3\left(\frac{f_{i+1} - f_i}{h_i}\right) - 3\left(\frac{f_i - f_{i-1}}{h_{i-1}}\right), \quad i = 2, \ldots, \\ h_{n-1}c_{n-1} + 2h_{n-1}c_n & = & -3\left(\frac{f_n - f_{n-1}}{h_{n-1}}\right) + 3f_n' \end{cases}$$

e então determinar os coeficientes b_i e d_i através das equações

$$b_i = \frac{f_{i+1} - f_i}{h_i} - \frac{h_i}{3}(2c_i + c_{i+1}),$$

$$d_i = \frac{c_{i+1} - c_i}{3h_i},$$

spline natural e spline (quase) completo

Vamos determinar a interpolação spline cúbica (natural e com as mesmas condições de f) para a função seno, no intervalo $[0,2\pi]$ a partir de 5 pontos igualmente espaçados. Ou seja,

$$\begin{aligned}\{(x_i,f_i)\}_{i=1}^{5} &= \left\{\left(\frac{(i-1)}{2}\pi, \sin\left(\frac{(i-1)}{2}\pi\right)\right)\right\}_{i=1}^{5} \\ &= \left\{(0,0),\left(\frac{\pi}{2},1\right),(\pi,0),\left(\frac{3\pi}{2},-1\right),(2\pi,0)\right\}\end{aligned}$$

Como o conjunto de pontos possui o mesmo espaçamento na coordenada x, então todos os h_i são iguais a $\dfrac{\pi}{2}$.

O spline cúbico é uma função $s(x)$ da forma

$$s(x) = \begin{cases} s_1(x), & 0 \le x < \dfrac{\pi}{2} \\ s_2(x), & \dfrac{\pi}{2} \le x < \pi \\ s_3(x), & \pi \le x < \dfrac{3\pi}{2} \end{cases},$$