Chapter 05.05
Spline Method of Interpolation

After reading this chapter, you should be able to:
1. interpolate data using spline interpolation, and
2. understand why spline interpolation is important.

What is interpolation?
Many times, data is given only at discrete points such as \((x_0, y_0), (x_1, y_1), \ldots, (x_{n-1}, y_{n-1}), (x_n, y_n)\). So, how then does one find the value of \(y\) at any other value of \(x\)? Well, a continuous function \(f(x)\) may be used to represent the \(n+1\) data values with \(f(x)\) passing through the \(n+1\) points (Figure 1). Then one can find the value of \(y\) at any other value of \(x\). This is called interpolation.

Of course, if \(x\) falls outside the range of \(x\) for which the data is given, it is no longer interpolation but instead is called extrapolation.

So what kind of function \(f(x)\) should one choose? A polynomial is a common choice for an interpolating function because polynomials are easy to
(A) evaluate,
(B) differentiate, and
(C) integrate
relative to other choices such as a trigonometric and exponential series.

Polynomial interpolation involves finding a polynomial of order \(n\) that passes through the \(n+1\) points. Several methods to obtain such a polynomial include the direct method, Newton’s divided difference polynomial method and the Lagrangian interpolation method.

So is the spline method yet another method of obtaining this \(n^{th}\) order polynomial. …… NO! Actually, when \(n\) becomes large, in many cases, one may get oscillatory behavior in the resulting polynomial. This was shown by Runge when he interpolated data based on a simple function of

\[ y = \frac{1}{1 + 25x^2} \]

on an interval of \([-1, 1]\). For example, take six equidistantly spaced points in \([-1, 1]\) and find \(y\) at these points as given in Table 1.
Figure 1  Interpolation of discrete data.

Table 1  Six equidistantly spaced points in \([-1, 1]\).

<table>
<thead>
<tr>
<th>(x)</th>
<th>(y = \frac{1}{1 + 25x^2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0</td>
<td>0.038461</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>1.0</td>
<td>0.038461</td>
</tr>
</tbody>
</table>

Now through these six points, one can pass a fifth order polynomial
\[ f_5(x) = 3.1378 \times 10^{-11} x^5 + 1.2019 x^4 - 3.3651 \times 10^{-11} x^3 - 1.7308 x^2 + 1.0004 \times 10^{-11} x + 5.6731 \times 10^{-1}, \]
\(-1 \leq x \leq 1\)
through the six data points. On plotting the fifth order polynomial (Figure 2) and the original function, one can see that the two do not match well. One may consider choosing more points in the interval \([-1, 1]\) to get a better match, but it diverges even more (see Figure 3), where 20 equidistant points were chosen in the interval \([-1, 1]\) to draw a 19th order polynomial. In fact, Runge found that as the order of the polynomial becomes infinite, the polynomial diverges in the interval of \(-1 < x < -0.726\) and \(0.726 < x < 1\).

So what is the answer to using information from more data points, but at the same time keeping the function true to the data behavior? The answer is in spline interpolation. The most common spline interpolations used are linear, quadratic, and cubic splines.
Figure 2  5th order polynomial interpolation with six equidistant points.

Figure 3  Higher order polynomial interpolation is a bad idea.
**Linear Spline Interpolation**

Given \((x_0,y_0),(x_1,y_1),\ldots,(x_{n-1},y_{n-1}),(x_n,y_n)\), fit linear splines (Figure 4) to the data. This simply involves forming the consecutive data through straight lines. So if the above data is given in an ascending order, the linear splines are given by \(y_i = f(x_i)\).

\[
\begin{align*}
  f(x) &= f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0), \quad x_0 \leq x \leq x_1 \\
  &= f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1), \quad x_1 \leq x \leq x_2 \\
  &\quad \vdots \\
  &= f(x_{n-1}) + \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}(x - x_{n-1}), \quad x_{n-1} \leq x \leq x_n
\end{align*}
\]

Note the terms of
\[
\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}
\]
in the above function are simply slopes between \(x_{i-1}\) and \(x_i\).
Example 1
To find how much heat is required to bring a kettle of water to its boiling point, you are asked to calculate the specific heat of water at $61^\circ C$. The specific heat of water is given as a function of time in Table 2.

<table>
<thead>
<tr>
<th>Temperature, $T$ ($^\circ C$)</th>
<th>Specific heat, $C_p$ (J/(kg$\cdot^\circ C$))</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>4181</td>
</tr>
<tr>
<td>42</td>
<td>4179</td>
</tr>
<tr>
<td>52</td>
<td>4186</td>
</tr>
<tr>
<td>82</td>
<td>4199</td>
</tr>
<tr>
<td>100</td>
<td>4217</td>
</tr>
</tbody>
</table>

**Figure 5** Specific heat of water vs. temperature.

Determine the value of the specific heat at $T = 61^\circ C$ using linear splines.
Solution

Since we want to find the specific heat at $T = 61^\circ C$ and we are using linear splines, we need to choose the two data points closest to $T = 61^\circ C$ that also bracket $T = 61^\circ C$ to evaluate it.

The two points are $T_0 = 52$ and $T_1 = 82$.

Then

$$T_0 = 52, \quad C_p(T_0) = 4186$$
$$T_1 = 82, \quad C_p(T_1) = 4199$$

gives

$$C_p(T) = C_p(T_0) + \frac{C_p(T_1) - C_p(T_0)}{T_1 - T_0}(T - T_0)$$
$$= 4186 + \frac{4199 - 4186}{82 - 52}(T - 52)$$

Hence

$$C_p(T) = 4186 + 0.43333(T - 52), \quad 52 \leq T \leq 82$$

At $T = 61$,

$$C_p(61) = 4186 + 0.43333(61 - 52)$$
$$= 4189.9 \frac{J}{kg \cdot ^\circ C}$$

Linear spline interpolation is no different from linear polynomial interpolation. Linear splines still use data only from the two consecutive data points. Also at the interior points of the data, the slope changes abruptly. This means that the first derivative is not continuous at these points. So how do we improve on this? We can do so by using quadratic splines.

**Quadratic Splines**

In these splines, a quadratic polynomial approximates the data between two consecutive data points. Given $(x_0, y_0), (x_1, y_1), \ldots, (x_{n-1}, y_{n-1}), (x_n, y_n)$, fit quadratic splines through the data.

The splines are given by

$$f(x) = a_i x^2 + b_i x + c_i, \quad x_0 \leq x \leq x_i$$
$$= a_2 x^2 + b_2 x + c_2, \quad x_1 \leq x \leq x_2$$
$$\quad \vdots$$
$$= a_n x^2 + b_n x + c_n, \quad x_{n-1} \leq x \leq x_n$$

So how does one find the coefficients of these quadratic splines? There are $3n$ such coefficients

$$a_i, \quad i = 1, 2, \ldots, n$$
$$b_i, \quad i = 1, 2, \ldots, n$$
$$c_i, \quad i = 1, 2, \ldots, n$$
To find $3n$ unknowns, one needs to set up $3n$ equations and then simultaneously solve them. These $3n$ equations are found as follows.

1. Each quadratic spline goes through two consecutive data points

   \[ a_i x_i^2 + b_i x_i + c_i = f(x_i) \]

   \[ a_{i+1} x_{i+1}^2 + b_{i+1} x_{i+1} + c_{i+1} = f(x_{i+1}) \]

   \[ \vdots \]

   \[ a_n x_n^2 + b_n x_n + c_n = f(x_n) \]

   This condition gives $2n$ equations as there are $n$ quadratic splines going through two consecutive data points.

2. The first derivatives of two quadratic splines are continuous at the interior points. For example, the derivative of the first spline

   \[ a_i x^2 + b_i x + c_i \]

   is

   \[ 2a_i x + b_i \]

   The derivative of the second spline

   \[ a_{i+1} x^2 + b_{i+1} x + c_{i+1} \]

   is

   \[ 2a_{i+1} x + b_{i+1} \]

   and the two are equal at $x = x_i$ giving

   \[ 2a_i x_i + b_i = 2a_{i+1} x_i + b_{i+1} \]

   \[ 2a_i x_i + b_i - 2a_{i+1} x_i - b_{i+1} = 0 \]

   Similarly at the other interior points,

   \[ 2a_{i+1} x_{i+1} + b_{i+1} - 2a_i x_{i+1} - b_i = 0 \]

   \[ \vdots \]

   \[ 2a_{i+1} x_{i+1} + b_{i+1} - 2a_i x_{i+1} - b_i = 0 \]

   \[ \vdots \]

   \[ 2a_{n-1} x_{n-1} + b_{n-1} - 2a_n x_{n-1} - b_n = 0 \]
Since there are \((n - 1)\) interior points, we have \((n - 1)\) such equations. So far, the total number of equations is \((2n) + (n - 1) = (3n - 1)\) equations. We still then need one more equation.

We can assume that the first spline is linear, that is

\[ a_i = 0 \]

This gives us \(3n\) equations and \(3n\) unknowns. These can be solved by a number of techniques used to solve simultaneous linear equations.

**Example 2**

To find how much heat is required to bring a kettle of water to its boiling point, you are asked to calculate the specific heat of water at 61°C. The specific heat of water is given as a function of time in Table 3.

![Table 3 Specific heat of water as a function of temperature.](image)

<table>
<thead>
<tr>
<th>Temperature, (T) (°C)</th>
<th>Specific heat, (C_p) (J kg⁻¹ °C⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>4181</td>
</tr>
<tr>
<td>42</td>
<td>4179</td>
</tr>
<tr>
<td>52</td>
<td>4186</td>
</tr>
<tr>
<td>82</td>
<td>4199</td>
</tr>
<tr>
<td>100</td>
<td>4217</td>
</tr>
</tbody>
</table>

a) Determine the value of the specific heat at \(T = 61°C\) using quadratic splines. Find the absolute relative approximate error for the quadratic approximation.

b) The heat required to heat the water is given more accurately by

\[
Q = m \int_{T_r}^{T_b} C_p dT
\]

\(T_r\) = room temperature (°C)  
\(T_b\) = boiling temperature of water (°C)

Given

\(T_r = 22°C\)  
\(T_b = 100°C\)

find a better estimate of the heat required. What is the difference between the results from part (a) and part (b).

**Solution**

a) Since there are five data points, four quadratic splines pass through them.

\[
C_p(T) = a_i T^2 + b_i T + c_i, \quad 22 \leq T \leq 42
\]

\[
= a_i T^2 + b_i T + c_i, \quad 42 \leq T \leq 52
\]
The equations are found as follows

1. Each quadratic spline passes through two consecutive data points.

\[ a_i T^2 + b_i T + c_i, \quad 52 \leq T \leq 82 \]

\[ a_i T^2 + b_i T + c_i, \quad 82 \leq T \leq 100 \]

At \( T = 42 \)

\[ a_1 (42)^2 + b_1 (42) + c_1 = 4181 \]  
\[ a_1 (42)^2 + b_1 (42) + c_1 = 4179 \]

(1) \quad (2)

At \( T = 52 \)

\[ a_2 (42)^2 + b_2 (42) + c_2 = 4179 \]  
\[ a_2 (52)^2 + b_2 (52) + c_2 = 4186 \]

(3) \quad (4)

At \( T = 82 \)

\[ a_3 (52)^2 + b_3 (52) + c_3 = 4186 \]  
\[ a_3 (82)^2 + b_3 (82) + c_3 = 4199 \]

(5) \quad (6)

2. Quadratic splines have continuous derivatives at the interior data points.

At \( T = 42 \)

\[ 2a_1 (42) + b_1 - 2a_2 (42) - b_2 = 0 \]

(9)

At \( T = 52 \)

\[ 2a_2 (52) + b_2 - 2a_3 (52) - b_3 = 0 \]

(10)

At \( T = 82 \)

\[ 2a_3 (82) + b_3 - 2a_4 (82) - b_4 = 0 \]

(11)

Assuming the first spline \( a_1 T^2 + b_1 T + c_1 \) is linear,

\[ a_1 = 0 \]  

(12)
Solving the above 12 equations gives the 12 unknowns as

<table>
<thead>
<tr>
<th>i</th>
<th>(a_i)</th>
<th>(b_i)</th>
<th>(c_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-0.1</td>
<td>4183.2</td>
</tr>
<tr>
<td>2</td>
<td>0.08</td>
<td>-6.82</td>
<td>4324.3</td>
</tr>
<tr>
<td>3</td>
<td>-0.035556</td>
<td>5.1978</td>
<td>4011.9</td>
</tr>
<tr>
<td>4</td>
<td>0.090741</td>
<td>-15.515</td>
<td>4861.1</td>
</tr>
</tbody>
</table>

Therefore, the splines are given by

\[ C_p(T) = \begin{cases} -0.1T + 4183.2, & 22 \leq T \leq 42 \\ 0.08T^2 - 6.82T + 4324.3, & 42 \leq T \leq 52 \\ -0.035556T^2 + 5.1978T + 4011.9, & 52 \leq T \leq 82 \\ 0.090741T^2 - 15.515T + 4861.1, & 82 \leq T \leq 100 \end{cases} \]

At \(T = 61\)

\[ C_p(61) = -0.035556(61)^2 + 5.1978(61) + 4011.9 \]

\[ = 4196.6 \frac{\text{J}}{\text{kg} \cdot ^\circ\text{C}} \]

The absolute relative approximate error \(|\varepsilon_a|\) obtained between the results from the linear and quadratic splines is

\[ |\varepsilon_a| = \left| \frac{4196.6 - 4189.9}{4196.6} \right| \times 100 \]

\[ = 0.16013\% \]

b) A more accurate calculation of the heat required to boil the water is given by

\[ Q = \int_{T_r}^{71.5} C_p(T) \, dT \]

\[ T_r = \text{room temperature} \ ({}^\circ\text{C}) \]
To find \( \int_C p \, d\mathbb{T} \), we can integrate the quadratic splines with respect to temperature.

\[
\int_C p \, dT = \int_{T_r}^{T_b} C_p(T) \, dT
\]

\[
= \int_{T_r}^{T_{22}} C_p(T) \, dT + \int_{T_{22}}^{T_{52}} C_p(T) \, dT + \int_{T_{52}}^{T_{100}} C_p(T) \, dT
\]

\[
= \int_{T_r}^{T_{22}} (-0.1T + 4183.2) \, dT + \int_{T_{22}}^{T_{52}} (0.08T^3 - 6.82T^2 + 4324.3) \, dT
\]

\[+ \int_{T_{52}}^{T_{100}} (-0.035556T^2 + 5.1978T + 4011.9) \, dT
\]

\[+ \int_{T_{100}}^{T_{82}} (0.090741T^2 - 15.515T + 4861.1) \, dT
\]

\[
= \left[ -0.1 \frac{T^2}{2} + 4183.2T \right]_{T_r}^{T_{22}} + \left[ 0.08 \frac{T^3}{3} - 6.82 \frac{T^2}{2} + 4324.3T \right]_{T_{22}}^{T_{52}}
\]

\[+ \left[ -0.035556 \frac{T^3}{3} + 5.1978 \frac{T^2}{2} + 4011.9T \right]_{T_{52}}^{T_{100}} + \left[ 0.090741 \frac{T^3}{3} - 15.515 \frac{T^2}{2} + 4861.1T \right]_{T_{100}}^{T_{82}}
\]

\[
= [83600] + [41812] + [125940] + [75656]
\]

\[
= 3.2700 \times 10^5 \frac{J}{\text{kg}}
\]

To compare this result with our results from part (a), we take the average specific heat over this interval, given by

\[
C_{p, \text{avg}} = \frac{\int_C p \, dT}{T_b - T_r}
\]

\[
= 3.2700 \times 10^5 \frac{J}{100 - 22} \text{kg} \cdot \text{°C}
\]

Knowing that part (b) is the more accurate way of calculating the heat transfer, we define the absolute relative approximate error \( \varepsilon_a \) by
\[ |\varepsilon_\alpha| = \left| \frac{4192.3 - 4196.6}{4192.3} \right| \times 100 = 0.10211\% \]

### INTERPOLATION

<table>
<thead>
<tr>
<th>Topic</th>
<th>Spline Method of Interpolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>Textbook notes on the spline method of interpolation</td>
</tr>
<tr>
<td>Major</td>
<td>Chemical Engineering</td>
</tr>
<tr>
<td>Authors</td>
<td>Autar Kaw, Michael Keteltas</td>
</tr>
<tr>
<td>Date</td>
<td>November 8, 2012</td>
</tr>
<tr>
<td>Web Site</td>
<td><a href="http://numericalmethods.eng.usf.edu">http://numericalmethods.eng.usf.edu</a></td>
</tr>
</tbody>
</table>