Cubic splines

Tino Kluge

2021–03–13

1 Interpolating cubic splines

Input: a set of ordered x, y coordinates[1]

$$\{(x_1, y_1), \dots, (x_n, y_n)\}, \quad \text{with } x_1 < x_2 < \cdots < x_n.$$

Output: a piecewise cubic function

$$\begin{aligned}
f(x) &= \begin{cases} f_0(x), & x < x_1, \\ f_1(x), & x \in [x_1, x_2), \\ \dots & \\ f_{n-1}(x), & x \in [x_{n-1}, x_n), \\ f_n(x), & x \geq x_n. \end{cases} \\
f_0(x) &= a_1 + b_1(x - x_1) + c_0(x - x_1)^2, & x < x_1, \\
f_i(x) &= a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, & x \in [x_i, x_{i+1}), \\
f_n(x) &= a_n + b_n(x - x_n) + c_n(x - x_n)^2, & x \geq x_n,
\end{aligned} \tag{1}$$

which satisfies

- $f(x_i) = y_i$, i.e. f interpolates all points (implies $a_i = y_i$),
- $f \in \mathrm{C}^1$, i.e. is continuously differentiable.

More conditions will be imposed to specify f uniquely but this depends on the type of spline as described below. Note, the derivatives of f_i are given by

$$\begin{aligned}
f_i'(x) &= 3d_i(x - x_i)^2 + 2c_i(x - x_i) + b_i, \\
f_i''(x) &= 6d_i(x - x_i) + 2c_i.
\end{aligned}$$

We also define $h_i := x_{i+1} - x_i$, $i \in \{1, \dots, n-1\}$.

[1] In this document we start with index 1 which differs from C-convention where indices start with 0. E.g. in this document x_1 equals to `x[0]` in code, $a_1 =$ `a[0]`, $c_n =$ `c[n-1]`, etc.

1.1 Cubic C^2 splines

They are defined by

- $f(x_i) = y_i$, i.e. f interpolates all points,
- $f \in \mathrm{C}^2$, i.e. f is twice continuously differentiable.

In detail this means:

$$
\begin{aligned}
\text{1) interpolates points: } f_i(x_i) = y_i : \quad & a_i = y_i, \\
\text{2) continuous } f \in \mathrm{C}^0\text{: } f_i(x_{i+1}) = y_{i+1} : \quad & d_i h_i^3 + c_i h_i^2 + b_i h_i = y_{i+1} - y_i, \\
\text{3) differentiable } f \in \mathrm{C}^1\text{: } f'_{i-1}(x_i) = f'_i(x_i) : \quad & 3d_{i-1}h_{i-1}^2 + 2c_{i-1}h_{i-1} + b_{i-1} = b_i, \\
\text{4) twice differentiable } f \in \mathrm{C}^2\text{: } f''_i(x_{i+1}) = f''_{i+1}(x_{i+1}) : \quad & 6d_i h_i + 2c_i = 2c_{i+1},
\end{aligned}
$$

Solving for d_i in the 4th condition (f'' continuous) gives

$$f \in \mathrm{C}^2 : \; d_i = \frac{c_{i+1} - c_i}{3h_i}.$$

Inserting this into the 2nd condition (f continuous) and solving for b_i gives

$$f \in \mathrm{C}^0 : \; b_i = \frac{y_{i+1} - y_i}{h_i} - \frac{1}{3}(2c_i + c_{i+1})h_i.$$

Finally, inserting these two equalities into the 3rd condition (f' continuous) gives a linear equation system for c:

$$
\boxed{
\begin{aligned}
\frac{1}{3}h_{i-1}c_{i-1} + \frac{2}{3}(h_{i-1} + h_i)c_i + \frac{1}{3}h_i c_{i+1} &= \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}, & i \in \{2, \dots, n-1\}, \\
d_i &= \frac{c_{i+1} - c_i}{3h_i}, & i \in \{1, \dots, n-1\}, \\
b_i &= \frac{y_{i+1} - y_i}{h_i} - \frac{1}{3}(2c_i + c_{i+1})h_i, & i \in \{1, \dots, n-1\}.
\end{aligned}
}
\tag{2}
$$

Boundary conditions

Second order conditions give immediate equations for c_1 and c_n:

$$
\begin{aligned}
f''(x_1) = \gamma_1 : \quad c_0 = c_1 &= \frac{\gamma_1}{2}, \\
f''(x_n) = \gamma_n : \quad c_n &= \frac{\gamma_n}{2}.
\end{aligned}
$$

First order conditions give immediate relationships for b_1 and b_n which need to be translated to conditions for c:

$$
\begin{aligned}
f'(x_1) = \delta_1 : \quad b_1 = \delta_1 \Rightarrow \quad & \frac{2}{3}h_1 c_1 + \frac{1}{3}h_1 c_2 = \frac{y_2 - y_1}{h_1} - \delta_1, \\
f'(x_n) = \delta_n : \quad b_n = \delta_n \Rightarrow \quad & \frac{1}{3}h_{n-1}c_{n-1} + \frac{2}{3}h_{n-1}c_n = \delta_n - \frac{y_n - y_{n-1}}{h_{n-1}}.
\end{aligned}
$$

The equation for the left boundary condition follows directly by inserting $b_1 = \delta_1$ into the third equation of (2). The equation for the right boundary condition follows from the differentiability condition $f'_{n-1}(x_n) = f'_n(x_n)$, replacing b_n by δ_n and replacing b_{n-1} and d_{n-1} by the relations in Equation (2).

In all cases, d_n and b_n are then determined by:

$$\begin{aligned} \text{quadratic extrapolation:} &\quad d_n = 0, \\ f'_{n-1}(x_n) = f'_n(x_n) : &\quad b_n = 3d_{n-1}h_{n-1}^2 + 2c_{n-1}h_{n-1} + b_{n-1}. \end{aligned}$$

1.2 Hermite C^1 splines

They are defined by

- $f(x_i) = y_i$, i.e. f interpolates all points,
- $f \in \mathrm{C}^1$, i.e. f is continuously differentiable,
- $f'(x_i) = \delta_i$, i.e. derivatives are specified at each inner point, e.g. by three point finite differences

$$\delta_i = -\frac{h_i}{h_{i-1}(h_{i-1}+h_i)}y_{i-1} + \frac{h_i - h_{i-1}}{h_{i-1}h_i}y_i + \frac{h_{i-1}}{h_i(h_{i-1}+h_i)}y_{i+1}.$$

In detail this means:

$$\begin{aligned} \text{1) interpolates points: } f_i(x_i) = y_i : &\quad a_i = y_i, \\ \text{2) prescribed derivative: } f'_i(x_i) = \delta_i : &\quad b_i = \delta_i, \\ \text{3) continuous } f \in \mathrm{C}^0\text{: } f_i(x_{i+1}) = y_{i+1} : &\quad d_i h_i^3 + c_i h_i^2 + b_i h_i = y_{i+1} - y_i, \\ \text{4) differentiable } f \in \mathrm{C}^1\text{: } f'_i(x_{i+1}) = \delta_{i+1} : &\quad 3d_i h_i^2 + 2c_i h_i = \delta_{i+1} - \delta_i, \end{aligned}$$

which can be solved to obtain

$$\boxed{\begin{aligned} &\text{1) } a_i = y_i, \\ &\text{2) } b_i = \delta_i, \\ &\text{3,4) } c_i = -\frac{2b_i + b_{i+1}}{h_i} + 3\frac{a_{i+1} - a_i}{h_i^2} = -\frac{2\delta_i + \delta_{i+1}}{h_i} + 3\frac{y_{i+1} - y_i}{h_i^2}, \quad i \in \{1, \dots, n-1\}, \\ &\text{4) } d_i = -\frac{2c_i}{3h_i} + \frac{b_{i+1} - b_i}{3h_i^2} = \frac{\delta_i + \delta_{i+1}}{h_i^2} - 2\frac{y_{i+1} - y_i}{h_i^3}, \quad i \in \{1, \dots, n-1\}. \end{aligned}} \tag{3}$$

Boundary conditions

First order conditions are trivial:

$$\begin{aligned} f'(x_1) = \delta_1 : &\quad b_1 = \delta_1,\ c_0 = 0, \\ f'(x_n) = \delta_n : &\quad b_n = \delta_n,\ c_n = 0, \end{aligned}$$

Second order conditions imply a value for c but we can re-express this in terms of b:

$$f''(x_1) = \gamma_1 : \qquad c_0 = c_1 = \frac{1}{2}\gamma_1 \Leftrightarrow b_1 = \frac{1}{2}\left(-b_2 - \frac{\gamma}{2}h_1 + 3\frac{y_2 - y_1}{h_1}\right),$$

$$f''_n(x_n) = \gamma_n : \qquad c_n = \frac{1}{2}\gamma_n,$$

$$f''_{n-1}(x_n) = \gamma_n : \quad 6d_{n-1}h_{n-1} + 2c_{n-1} = \frac{1}{2}\gamma_n \Leftrightarrow b_n = \frac{1}{2}\left(-b_{n-1} + \frac{\gamma}{2}h_{n-1} + 3\frac{y_n - y_{n-1}}{h_{n-1}}\right).$$

1.3 Monotonic splines

The spline interpolating function f is monotonic increasing in a segment $[x_i, x_{i+1}]$, if and only if

$$f'_i(x) \geq 0, \quad \forall x \in [x_i, x_{i+1}].$$

Assuming the spline has already been build (without necessarily being monotonic) then f' satisfies

$$f'(x_i) = b_i,$$
$$f'(x_{i+1}) = b_{i+1},$$
$$\int_{x_i}^{x_{i+1}} f'(x)\,\mathrm{d}x = y_{i+1} - y_i.$$

Assuming further that if the input is monotonic, i.e. $y_{i-1} \leq y_i \leq y_{i+1} \leq y_{i+2}$, then b_i and b_{i+1} are both non-negative[2] then we can apply the results of Section 2.1: one sufficient condition for f' to be non-negative in the whole interval $[x_i, x_{i+1}]$ is

$$\sqrt{b_i^2 + b_{i+1}^2} \leq 3\frac{y_{i+1} - y_i}{h_i}. \tag{4}$$

If that conditions is not satisfied we can reduce the size of b_i and b_{i+1} until the condition is satisfied. Reducing the size of b_i will also not break monotonicity of the previous segment given the sufficient condition (4) was satisfied in that segment. One method[3] to create monotonic splines therefore is:

1. generate the spline without monotonicity requirement,
2. if input y-data is monotonic increasing but there is a $b_i < 0$ then set $b_i = 0$,
3. if input is increasing, i.e. $b_i \geq 0$, $b_{i+1} \geq 0$ and $y_i \leq y_{i+1}$, check condition (4) for each segment i: if not satisfied, rescale b_i and b_{i+1} so that it is satisfied,
4. re-calculate all coefficients c_i and d_i based on the new b_i using Equation (3).

Accordingly, monotonic decreasing splines can be created.

Warning: if any adjustment is made then the spline will not be C^2 anymore if it was before. If any of the boundary gradients (b_1, b_2, b_{n-1} or b_n) were adjusted then this breaks previously defined boundary conditions and extrapolation.

[2] This is automatically satisfied for the Hermite splines defined here as b_i are the finite differences. It is not necessarily true for the cubic C^2 spline.

[3] For Hermite splines this could be done more efficiently in a single pass.

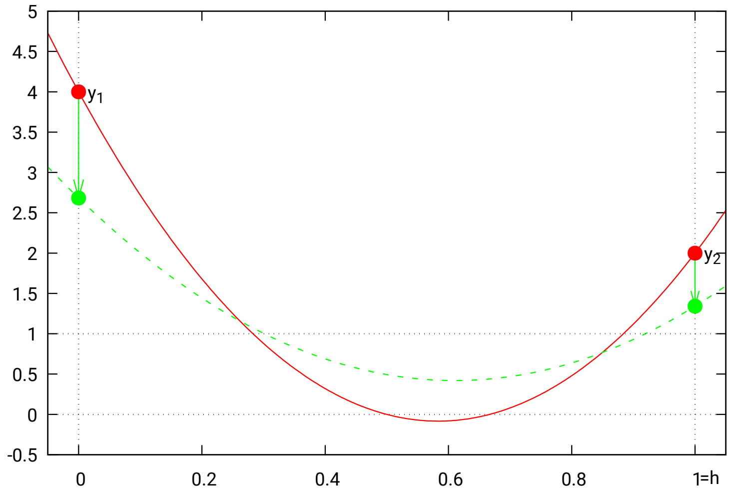

Figure 1: Interpolating two points $y_1 = 4$, $y_2 = 2$ whilst maintaining the average $avg = 1$ over the interval $[0, h]$ with $h = 1$. As can be seen the interpolation goes briefly negative. One sufficient (but not necessary) criteria for non-negativity is that $\sqrt{y_1^2 + y_2^2} \leq 3\,avg$ which the green points satisfy exactly.

2 Appendix

2.1 Non-negative average preserving interpolation

Here we are looking at interpolating two points whilst preserving a specified average. In particular we want to answer the question under which conditions the interpolation stays non-negative. See also [1]. The interpolation problem can be specified as follows and is shown in Figure 1:

$$\begin{aligned} f(x) &= a + bx + cx^2, \\ f(0) &= y_1, \\ f(h) &= y_2, \\ \frac{1}{h}\int_0^h f(x)\,\mathrm{d}x &= avg, \end{aligned} \tag{5}$$

which implies

$$\begin{aligned} f(0) = y_1: &\quad a = y_1, \\ f(h) = y_2: &\quad a + bh + ch^2 = y_2, \\ \frac{1}{h}\int_0^h f(x)\,\mathrm{d}x = avg: &\quad a + \frac{1}{2}bh + \frac{1}{3}ch^2 = avg, \end{aligned}$$

and can be solved to obtain the coefficients

$$\begin{aligned} a &= y_1, \\ b &= \frac{2}{h}\left(-2y_1 - y_2 + 3avg\right), \\ c &= \frac{3}{h^2}\left(y_1 + y_2 - 2avg\right). \end{aligned} \tag{6}$$

Proposition 2.1 (Non-negativity of the interpolation)
Assuming $y_1, y_2 \geq 0$ and $avg > 0$. The interpolating function f as defined in Equation (5) then satisfies

$$f(x) \geq 0, \qquad \forall x \in [0, h],$$

if and only if

$$z_1 + z_2 \leq 3 \quad \text{or} \quad z_1^2 + z_2^2 + z_1 z_2 - 6(z_1 + z_2) + 9 \leq 0, \tag{7}$$

with $z_1 := \frac{y_1}{avg}$, $z_2 := \frac{y_2}{avg}$. This is shown in red in Figure 2. A sufficient but not necessary condition is

$$\sqrt{z_1^2 + z_2^2} \leq 3, \tag{8}$$

which is shown in green in Figure 2.

Proof Note, if $c \neq 0$, f has a local extrema x^* at

$$x^* = -\frac{b}{2c}, \quad f(x^*) = a - \frac{b^2}{4c}.$$

Note also that the minimum of a function is either assumed on its boundaries or at a local minimum and we have assumed that on the boundary f is non-negative ($y_1, y_2 \geq 0$). Therefore, the function f is non-negative in $[0, h]$ if and only if at least one of the following three criteria is satisfied:

1. f has no local minima, i.e. $c \leq 0$, and given c in Equation (6) this is equivalent to

$$\frac{y_1}{avg} + \frac{y_2}{avg} \leq 2.$$

2. f has a local minima x^* but $x^* \notin (0, h)$ which is equivalent to $b \geq 0$ or $b \leq -2ch$ (since $c > 0$). With b and c as given in Equation (6) this is equivalent to

$$\frac{y_1}{avg} + 2\frac{y_2}{avg} \leq 3 \quad \text{or} \quad 2\frac{y_1}{avg} + \frac{y_2}{avg} \leq 3.$$

3. f has a local minima but the minima is greater or equal zero, i.e. $f(x^*) \geq 0$ which is equivalent to $a - \frac{b^2}{4c} \geq 0 \Leftrightarrow 4ac - b^2 \geq 0$, since $c > 0$. Given a, b and c as in Equation (6) this is equivalent (after a little re-arranging) to

$$\left(\frac{y_1}{avg}\right)^2 + \left(\frac{y_2}{avg}\right)^2 + \frac{y_1 y_2}{avg^2} - 6\frac{y_1 + y_2}{avg} + 9 \leq 0.$$

... □

Note, the inequality under item 3 is the inside of an ellipse for $z_1 := \frac{y_1}{avg}$ and $z_2 := \frac{y_2}{avg}$, rotated by 90° and origin shifted to the coordinate $(2, 2)$:

- Ellipse

$$\frac{z_1^2}{6} + \frac{z_2^2}{2} = 1.$$

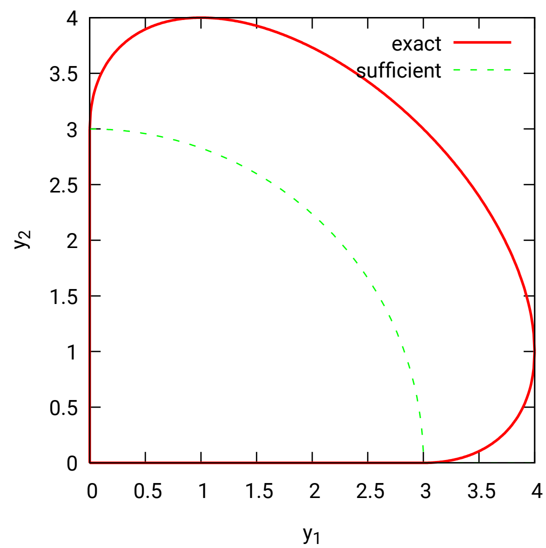

Figure 2: Non-negativity of the interpolating function is guaranteed if and only if (y_1, y_2) are inside the red line (in the case of $avg = 1$). A sufficient but not necessary condition is drawn by the dashed green line.

- Rotated[4] by 90°, i.e.

$$\frac{\frac{1}{2}(z_1 - z_2)^2}{6} + \frac{\frac{1}{2}(z_1 + z_2)^2}{2} = 1,$$
$$\Updownarrow$$
$$z_1^2 + z_2^2 + z_1 z_2 = 3.$$

- Origin shifted to $(2, 2)$, i.e.

$$(z_1 - 2)^2 + (z_2 - 2)^2 + (z_1 - 2)(z_2 - 2) = 3,$$
$$\Updownarrow$$
$$z_1^2 + z_2^2 + z_1 z_2 - 6(z_1 + z_2) + 9 = 0.$$

References

[1] F.N. Fritsch and R.E. Carlson. Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.*, 17(2):238–246, 1980.

[4]Rotation matrix $A = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$ and inverse $A^{-1} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ for a 90° rotation. A set defined by an implicit equation $\{x \in \mathbb{R}^2 : g(x) = 0\}$ rotated is $\{Ax \in \mathbb{R}^2 : g(x) = 0\} = \{x \in \mathbb{R}^2 : g(A^{-1}x) = 0\}$.