Interpolation mit kubischen Splines

Fritz Philipp

Inhaltsverzeichnis

1 Definitionen

In diesem kleinen Artikel wollen wir uns mit kubischen Splines befassen. Splines sind reelle Funktionen auf einem Intervall $[a,b]$, das man derart in Teilintervalle aufteilen kann, daß die Funktion in jedem der Intervalle ein Polynom ist und an den Übergangsstellen mit einer gewissen Glätte versehen ist. Bei einem kubischen Spline ist der Grad dieser Polynome höchstens 3. Kubische Splines eignen sich besonders gut zur Interpolation, denn sie sind im interessanten Bereich zweimal stetig differenzierbar und oszillieren nicht an den Rändern, wie es Interpolationspolynome von hinreichend großem Grad zu tun pflegen.

Im Folgenden wollen wir uns stets mit dem Gitter

$$\Delta : \ -\infty < x_0 < x_1 < \cdots < x_n < \infty$$

beschäftigen. Die x_j stellen dabei die Randpunkte der oben genannten Teilintervalle dar.

Definition: *Eine Funktion* $s : \mathbb{R} \to \mathbb{R}$ *heißt ein* Spline *vom Grade* $m \in \mathbb{N}$ *über die Knoten* $x_0, \ldots, x_n$, *wenn*

(i) $s \in \Pi_m$ *auf jedem Intervall* $(-\infty, x_0), [x_0, x_1), \ldots, [x_{n-1}, x_n], (x_n, \infty)$ *;*
(ii) $s, s', \ldots, s^{(m-1)}$ *sind stetig auf ganz* $\mathbb{R}$.

Die Menge aller solcher Funktionen bezeichnen wir mit $\mathcal{S}_m(x_0, \ldots, x_m)$.

(ii) bedeutet, daß $s(x)$ auf $\mathbb{R}$ $(m-1)$–mal stetig differenzierbar ist. Wegen (i) ist dies auf den Teilintervallen (x_j, x_{j+1}) zwar bereits erfüllt, jedoch nicht an den Rändern. Man könnte (ii) auch wie folgt schreiben:

$$s^{(p)}(x_j-) = s^{(p)}(x_j+) \quad \text{für} \quad p = 0, \ldots, m-1 \quad \text{und} \quad j = 0, \ldots, n \ .$$

Da die Außenintervalle relativ unwichtig sind, berechnet man einen Spline nur im Intervall (x_0, x_n). Der Spline ist dann aber noch nicht eindeutig bestimmt, d.h., man kann noch weitere Bedingungen an den Spline stellen. Für kubische Splines sind dies 2 Bedingungen. Oft wird die Bedingung gestellt, daß s bei x_0 bzw. x_n zweimal stetig differenzierbar sein soll und auf $(-\infty, x_0)$ bzw. (x_n, ∞) linear fortsetzbar.

Definition: *Eine Funktion* $s : \mathbb{R} \to \mathbb{R}$ *heißt* natürlicher Spline *vom Grade* $m \in \mathbb{N}$ *über die Knoten* $x_0, \ldots, x_n$, *wenn*

(i) $s \in \mathcal{S}_m(x_0, \ldots, x_n)$ *;*
(ii) m *ist ungerade, d.h.* $m = 2k-1$ *mit* $k \in \mathbb{N}$ *;*
(iii) $s \in \Pi_{k-1}$ *auf* $(-\infty, x_0)$ *und* (x_n, ∞).

Die Menge aller solcher Funktionen bezeichnen wir mit $\mathcal{S}_m^{\mathcal{N}}(x_0, \ldots, x_n)$.

Lemma 1.1 *Es sei $s \in \mathcal{S}_m(\Delta)$ und $m = 2k-1$ mit $k \in \mathbb{N}$. s ist genau dann ein natürlicher Spline, wenn*

$$s^{(p)}(x_0) = s^{(p)}(x_n) = 0 \quad \textit{für} \quad p = k, k+1, \dots, 2k-2 \tag{1}$$

$$\textit{und} \quad s^{(m)}(x_0-) = s^{(m)}(x_n+) = 0 \ . \tag{2}$$

Beweis: a) Es sei s ein natürlicher Spline, d.h. $s \in \Pi_{k-1}$ auf den Intervallen $I_1 = (-\infty, x_0)$ und $I_2 = (x_n, \infty)$, also $s^{(k)} = \cdots = s^{(2k-2)} = 0$ auf diesen Intervallen. Da diese Funktionen aber wegen $2k-2 = m-1$ auf ganz $\mathbb{R}$ stetig sind, folgt mit der Grenzwertbetrachtung $x \to x_0$ bzw. $x \to x_n$ die Bedingung (1). Die zweite Bedingung folgt aus dem gleichen Grund.
b) Es seien die Bedingungen erfüllt. Wir wollen hier nur das Intervall I_1 betrachten. Die Vorgehensweise für I_2 ist analog. Wegen (1) ist x_0 eine $(k-1)$–fache Nullstelle für $s^{(k)}$. Da $s^{(k)}$ höchstens einen Grad von $(2k-1)-k = k-1$ hat, existiert somit ein $\lambda \in \mathbb{R}$, so daß $s^{(k)}(x) = \lambda(x-x_0)^{k-1}$ für $x \in I_1$. Leiten wir diesen Ausdruck $(k-1)$–mal ab, so erhalten wir $s^{(2k-1)}(x) = \lambda(k-1)!$. Mit (2) bedeutet das $\lambda = 0$ und somit $s^{(k)} \equiv 0$ auf I_1. Die Bedingung (iii) in der Definition eines natürlichen Splines ist daher erfüllt. □

2 Darstellung von kubischen Splines

Nun wollen wir uns, wie der Titel dieses Artikels bereits besagt, mit kubischen Splines, also dem Fall $m = 3$ befassen. Man könnte den Spline in der Form

$$\begin{aligned} s(x) = d_{0j} + d_{1j}(x-x_{j-1}) + d_{2j}(x-x_{j-1})^2 + d_{3j}(x-x_{j-1})^3 \\ \text{für} \quad x \in [x_{j-1}, x_j) \quad \text{und} j = 1, \dots, n \end{aligned} \tag{3}$$

schreiben. Sind die Koeffizienten (d_{ij}) bekannt, so ist diese Form sehr gut zur Auswertung des Splines geeignet. Zur Bestimmung des Splines jedoch eignet sie sich recht schlecht, wie wir noch sehen werden, denn die Abhängigkeit der Koeffizienten voneinander wird hier nicht ausgenutzt. Wir wollen uns aber doch weiter mit der Form (3) beschäftigen, um eine andere Darstellung für s zu erhalten. Im Weiteren seien die folgenden Größen gegeben

$$\begin{aligned} s_j = s(x_j), \quad s'_j = s'(x_j), \quad s''_j = s''(x_j) \quad \text{für} \quad j = 0, \dots, n \\ h_j = x_j - x_{j-1} \quad \text{für} \quad j = 1, \dots, n \ . \end{aligned}$$

Damit gilt dann für $j = 1, \dots, n$:

$$\begin{array}{rclcl} s_{j-1} & = & s(x_{j-1}) & = & d_{0j} \\ s'_{j-1} & = & s'(x_{j-1}) & = & d_{1j} \\ s_j & = & s(x_j) & = & d_{0j} + d_{1j}h_j + d_{2j}h_j^2 + d_{3j}h_j^3 \\ s'_j & = & s'(x_j) & = & d_{1j} + 2d_{2j}h_j + 3d_{3j}h_j^2 \end{array} \quad . \tag{O}$$

In diesem System stecken bereits alle Bedingungen an die d_{ij} drin, die die Stetigkeit von s und s' gewährleisten. Die beiden letzten Gleichungen bilden ein

LGS mit den Unbekannten d_{2j} und d_{3j}. Dessen Lösung ist

$$\begin{aligned} d_{2j} &= \frac{3(s_j - s_{j-1}) - (s'_j + 2s'_{j-1})h_j}{h_j^2} \\ d_{3j} &= -\frac{2(s_j - s_{j-1}) - (s'_j + s'_{j-1})h_j}{h_j^3} \quad . \end{aligned} \tag{4}$$

Nach der obigen Darstellung (3) von s gilt weiter für $t \in [0,1)$:

$$s(x_{j-1} + th_j) = d_{0j} + d_{1j}h_j t + d_{2j}h_j^2 t^2 + d_{3j}h_j^3 t^3 \quad .$$

Da wir nun die Koeffizienten d_{ij} in Abhängigkeit von den Parametern s_j, s'_j, s_{j-1}, s'_{j-1} gesetzt haben, kann die letzte Gleichung nach eben diesen aufgelöst werden:

$$\begin{aligned} s(x_{j-1} + th_j) =& (1-t)^2((1+2t)s_{j-1} + th_j s'_{j-1}) \\ &+ t^2((3-2t)s_j + (t-1)h_j s'_j) \quad . \end{aligned} \tag{H}$$

Die Darstellung (H) wird die *Hermite–Darstellung* des kubischen Splines genannt. Der Vorteil gegenüber der ersten Darstellung (3) ist offensichtlich. Während diese nämlich $4n$ Unbekannte hat (4 in jedem Intervall), hat die Darstellung (H) nur noch $2n+2$ Unbekannte, nämlich die s_j und die s'_j. Für ein Interpolationsproblem hat (3) wegen $d_{0j} = s(x_{j-1})$, $j = 1, \ldots, n$, eine Anzahl von $3n$ Unbekannten. Die Hermite–Darstellung hat in diesem Fall nur noch $n+1$ Unbekannte.

Da die von den Größen s_j und s'_j abhängigen Koeffizienten d_{ij}, wie oben erwähnt, die Stetigkeit von s und s' gewährleisten, ist der Spline in der Darstellung (H) bereits stetig differenzierbar, was man auch leicht direkt zeigt. Nun muß s per Definition aber zweimal stetig differenzierbar sein. Dazu betrachten wir die zweiten Ableitungen der Darstellung (3) von s in den beiden Intervallen $[x_{j-1}, x_j)$ und $[x_j, x_{j+1})$. Diese sind

$$s''(x) = 2d_{2j} + 6d_{3j}(x - x_{j-1}) \quad \text{bzw.} \quad s''(x) = 2d_{2,j+1} + 6d_{3,j+1}(x - x_j) \quad .$$

Daraus ergeben sich dann die Bedingungen

$$s''_j = 2d_{2,j+1} = 2d_{2j} + 6d_{3j}h_j \quad \text{für} \quad j = 1, \ldots, n-1 \quad .$$

Wir setzen nun wieder gemäß (4) ein und erhalten nach einer leichten Rechnung die Gleichungen

$$\begin{aligned} & h_{j+1}s'_{j-1} + 2(h_j + h_{j+1})s'_j + h_j s'_{j+1} \\ & = 3\left(\frac{h_j}{h_{j+1}}(s_{j+1} - s_j) + \frac{h_{j+1}}{h_j}(s_j - s_{j-1})\right) \quad \text{für} \quad j = 1, \ldots, n-1 \quad . \end{aligned} \tag{5}$$

3 Interpolation

Beim Interpolationsproblem ist die rechte Seite von (5) bekannt, und die Matrix zu diesem linearen Gleichungssystem sieht folgendermaßen aus:

$$\mathbf{A} = \begin{pmatrix} h_2 & 2h_{12} & h_1 & & \dots & 0 \\ & h_3 & 2h_{23} & h_2 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & \dots & & h_n & 2h_{n-1,n} & h_{n-1} \end{pmatrix} \in \mathbb{R}^{(n-1)\times(n+1)} \quad ,$$

wobei $h_{ij} := h_i + h_j$ für $i, j = 1, \dots, n$. Stellt $\mathbf{x} = (s'_0, \dots, s'_n)^T \in \mathbb{R}^{n+1}$ den Unbekannten-Vektor dar, so kann das LGS (5) auch als

$$\mathbf{Ax} = \mathbf{b}$$

geschrieben werden. Die $(n-1)$ Komponenten von $\mathbf{b}$ sind dann die rechte Seite von (5). Aufgrund ihrer Dreiecksform hat die Matrix $\mathbf{A}$ vollen Rang, also $Rang\mathbf{A} = n - 1$. Der Lösungsraum von $\mathbf{Ax} = \mathbf{b}$ ist also 2–dimensional. Wollen wir einen natürlichen Spline haben, so müssen nach dem Lemma 1.1 zusätzlich die beiden Bedingungen

$$s''(x_0) = s''(x_n) = 0$$

erfüllt sein. Umgerechnet in das System mit dem Vektor $\mathbf{x}$ lauten diese Gleichungen wie folgt:

$$2s'_0 + \quad s'_1 = \frac{3}{h_1}(s_1 - s_0) \qquad =: l_0 \qquad\qquad (s''(x_0) = 0)$$

$$2s'_n + s'_{n-1} = \frac{3}{h_n}(s_n - s_{n-1}) \quad =: l_n \qquad\qquad (s''(x_n) = 0)$$

Mit $\mathbf{c} = (l_0, \mathbf{b}, l_n)^T \in \mathbb{R}^{n+1}$ und

$$\mathbf{B} = \begin{pmatrix} 2 & 1 & \mathbf{0} \\ — & \mathbf{A} & — \\ \mathbf{0} & 2 & 1 \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

ergibt sich dann das wegen $Rang\mathbf{B} = n+1$ eindeutig lösbare lineare Gleichungssystem

$$\mathbf{Bx} = \mathbf{c} \quad . \tag{NS}$$

Wäre $\mathbf{B}$ nicht invertierbar, dann gäbe es ein $\lambda \in \mathbb{R}^*$, so daß eine der folgenden Gleichungen erfüllt wäre

$$(h_2, 2h_{12}, h_1) = \lambda(2, 1, 0) \quad \text{oder} \quad (h_n, 2h_{n-1,n}, h_{n-1}) = \lambda(0, 1, 2) \quad .$$

Aus diesen Gleichungen folgt $h_1 = 0$ oder $h_n = 0$, und das kann nach den Definitionen von h_j und dem Gitter Δ nicht sein. Wir wollen nun schließlich die Ergebnisse in einem Satz formulieren.

Satz 3.1 *Es sei das Gitter Δ von oben gegeben und weiterhin die $n + 1$ Daten $(s_0, \dots, s_n) \in \mathbb{R}^{n+1}$. Dann gibt es genau einen natürlichen Spline $s \in \mathcal{S}_3^{\mathcal{N}}(\Delta)$, der das Interpolationsproblem $s(x_j) = s_j$ für $j = 0, \dots, n$ löst. Seine Ableitungen s'_j in den Knoten x_j für $j = 0, \dots, n$ ergeben sich aus dem System (NS). Mit diesen kann die Auswertung durch die Darstellung (H) vorgenommen werden.*$\square$