11

Piecewise polynomial approximation

11.1 Introduction

Up to now, the focus of our discussion has been the question of approximation of a given function f, defined on an interval $[a, b]$, by a polynomial on that interval either through Lagrange interpolation or Hermite interpolation, or by seeking the polynomial of best approximation (in the ∞-norm or 2-norm). Each of these constructions was *global* in nature, in the sense that the approximation was defined by the same analytical expression on the whole interval $[a, b]$. An alternative and more flexible way of approximating a function f is to divide the interval $[a, b]$ into a number of subintervals and to look for a piecewise approximation by polynomials of low degree. Such piecewise-polynomial approximations are called **splines**, and the endpoints of the subintervals are known as the **knots**.

More specifically, a spline of degree n, $n \geq 1$, is a function which is a polynomial of degree n or less in each subinterval and has a prescribed degree of smoothness. We shall expect the spline to be at least continuous, and usually also to have continuous derivatives of order up to k for some k, $0 \leq k < n$. Clearly, if we require the derivative of order n to be continuous everywhere the spline is just a single polynomial, since if two polynomials have the same value and the same derivatives of every order up to n at a knot, then they must be the same polynomial. An important class of splines have degree n, with continuous derivatives of order up to and including $n - 1$, but as we shall see later, lower degrees of smoothness are sometimes considered.

To give a flavour of the theory of splines, we concentrate here on two simple cases: linear splines and cubic splines.

11.2 Linear interpolating splines

Definition 11.1 *Suppose that f is a real-valued function, defined and continuous on the closed interval $[a,b]$. Further, let $K = \{x_0, \dots, x_m\}$ be a subset of $[a,b]$, with $a = x_0 < x_1 < \dots < x_m = b$, $m \geq 2$. The* **linear spline** *s_L, interpolating f at the points x_i, is defined by*

$$s_L(x) = \frac{x_i - x}{x_i - x_{i-1}} f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} f(x_i)\,, \qquad x \in [x_{i-1}, x_i]\,, \quad i = 1, 2, \dots, m\,. \tag{11.1}$$

The points x_i, $i = 0, 1, \dots, m$, are the **knots** *of the spline, and K is referred to as the* **set of knots**.

As the function s_L interpolates the function f at the knots, i.e., $s_L(x_i) = f(x_i)$, $i = 0, 1, \dots, m$, and over each interval $[x_{i-1}, x_i]$, for $i = 0, 1, \dots, m$, the function s_L is a linear polynomial (and therefore continuous), we conclude that s_L is a continuous piecewise linear function on the interval $[a,b]$.

Given a set of knots $K = \{x_0, \dots, x_m\}$, we shall use the notation $h_i = x_i - x_{i-1}$, and let $h = \max_i h_i$. Also, for a positive integer k, we denote by $\mathrm{C}^k[a,b]$ the set of all real-valued functions, defined and continuous on the closed interval $[a,b]$, such that all derivatives, up to and including order k, are defined and continuous on $[a,b]$.

In order to highlight the accuracy of interpolation by linear splines we state the following error bound in the ∞-norm over the interval $[a,b]$.

Theorem 11.1 *Suppose that $f \in \mathrm{C}^2[a,b]$ and let s_L be the linear spline that interpolates f at the knots $a = x_0 < x_1 < \dots < x_m = b$; then, the following error bound holds:*

$$\|f - s_L\|_\infty \leq \frac{1}{8} h^2 \|f''\|_\infty\,,$$

where $h = \max_i h_i = \max_i (x_i - x_{i-1})$, and $\|\cdot\|_\infty$ denotes the ∞-norm over $[a,b]$, defined in (8.1).

Proof Consider a subinterval $[x_{i-1}, x_i]$, $1 \leq i \leq m$. According to Theorem 6.2, applied on the interval $[x_{i-1}, x_i]$,

$$f(x) - s_L(x) = \frac{1}{2} f''(\xi)(x - x_{i-1})(x - x_i)\,, \qquad x \in [x_{i-1}, x_i]\,,$$

where $\xi = \xi(x) \in (x_{i-1}, x_i)$. Thus,

$$|f(x) - s_{\mathrm{L}}(x)| \le \frac{1}{8} h_i^2 \max_{\zeta \in [x_{i-1}, x_i]} |f''(\zeta)|\,.$$

Hence,

$$|f(x) - s_{\mathrm{L}}(x)| \le \frac{1}{8} h^2 \|f''\|_\infty\,,$$

for each $x \in [x_{i-1}, x_i]$ and each $i = 1, 2, \ldots, m$. This gives the required error bound. $\square$

Figure 11.1 shows a typical example: a linear spline approximation to the function $f\colon x \mapsto \mathrm{e}^{-3x}$ over the interval $[0, 1]$, using two internal knots, $x_1 = \frac{1}{3}$, $x_2 = \frac{2}{3}$, together with the endpoints of the interval, $x_0 = 0$ and $x_3 = 1$.

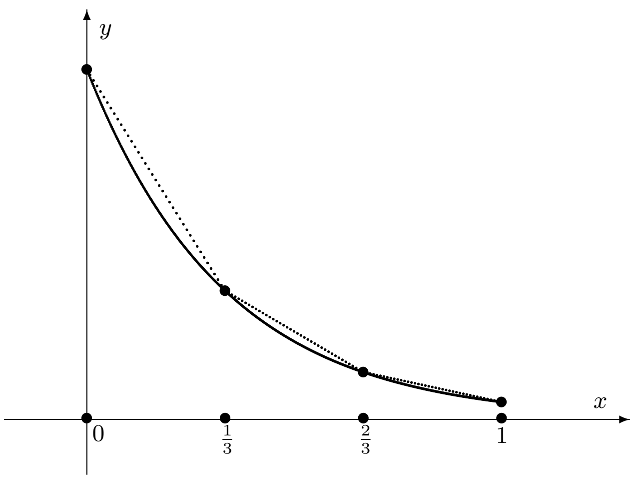

Fig. 11.1. The function $f\colon x \mapsto \mathrm{e}^{-3x}$ (full curve) and its linear spline approximation (dotted curve). The interval is $[0, 1]$, and the knots are at 0, $\frac{1}{3}$, $\frac{2}{3}$ and 1.

We conclude this section with a result that provides a characterisation of linear splines from the viewpoint of the calculus of variations.

A subset A of the real line is said to have **measure zero** if it can be contained in a countable union of open intervals of arbitrarily small total length; in other words, for every $\varepsilon > 0$ there exists a sequence of open intervals (a_i, b_i), $i = 1, 2, 3, \ldots$, such that

$$A \subset \bigcup_{i=1}^{\infty} (a_i, b_i) \quad \text{and} \quad \sum_{i=1}^{\infty} (b_i - a_i) < \varepsilon\,.$$

In particular, any finite or countable set $A \subset \mathbb{R}$ has measure zero. For example, the set of all rational numbers is countable, and therefore it has measure zero. Trivially, the empty set has measure zero.

Suppose that B is a subset of $\mathbb{R}$. We shall say that a certain property $\mathsf{P} = \mathsf{P}(x)$ holds for **almost every** x in B, if there exists a set $A \subset B$ of measure zero such that $\mathsf{P}(x)$ holds for *all* $x \in B \setminus A$.

A real-valued function v defined on the interval $[a, b]$ is said to be **absolutely continuous** on $[a, b]$ if it has finite derivative $v'(\xi)$ at almost every point ξ in $[a, b]$, v' is (Lebesgue-) integrable on $[a, b]$, and

$$\int_a^x v'(\xi)\mathrm{d}\xi = v(x) - v(a)\,, \qquad a \le x \le b\,.$$

Example 11.1 *Any $v \in \mathrm{C}^1[a, b]$ is absolutely continuous on the interval $[a, b]$. The function $x \mapsto \left|x - \frac{1}{2}(a + b)\right|$ is absolutely continuous on $[a, b]$, but it does not belong to $\mathrm{C}^1[a, b]$ as it is not differentiable at $x = \frac{1}{2}(a+b)$.*

Let us denote by $\mathrm{H}^1(a, b)$ the set of all absolutely continuous functions v defined on $[a, b]$ such that $v' \in \mathrm{L}^2(a, b)$, *i.e.*,

$$\|v'\|_2 = \left(\int_a^b |v'(\xi)|^2\mathrm{d}\xi\right)^{1/2} < \infty\,.$$

We observe in passing that any function $v \in \mathrm{H}^1(a, b)$ is uniformly continuous on the closed interval $[a, b]$. This follows by noting that, for any pair of points $x, y \in [a, b]$,

$$\begin{aligned} |v(x) - v(y)| &= \left|\int_x^y v'(\xi)\,\mathrm{d}\xi\right| \\ &\le |x - y|^{\frac{1}{2}} \left|\int_x^y |v'(\xi)|^2\mathrm{d}\xi\right|^{1/2} \\ &\le |x - y|^{\frac{1}{2}}\,\|v'\|_2\,. \end{aligned}$$

In the transition from the first line to the second we used the Cauchy–Schwarz inequality.

If $k \ge 1$, we shall denote by $\mathrm{H}^{k+1}(a, b)$ the set of all $v \in \mathrm{H}^k(a, b)$ such that $v^{(k)}$ is absolutely continuous on $[a, b]$ and $v^{(k+1)} \in \mathrm{L}^2(a, b)$. The set $\mathrm{H}^k(a, b)$ is called a **Sobolev space** of index k. We observe that

$$\mathrm{C}^k[a, b] \subset \mathrm{H}^k(a, b)$$

for any $k \ge 1$, with strict inclusion. For example, any linear spline on

$[a, b]$ belongs to $\mathrm{H}^1(a, b)$, but not to $\mathrm{C}^1[a, b]$ unless it is a linear function over the *whole of* the interval $[a, b]$.

Example 11.2 *Let $\alpha > 1/2$; the function $x \mapsto x^\alpha$ then belongs to $\mathrm{H}^1(0, 1)$, although it only belongs to $\mathrm{C}^1[0, 1]$ if $\alpha \geq 1$.*

As a second example, consider the function $x \mapsto x \ln x$ which belongs to $\mathrm{H}^1(0, 1)$, but not to $\mathrm{C}^1[0, 1]$.

The variational characterisation of linear splines stated in the next theorem expresses the fact that, among all functions $v \in \mathrm{H}^1(a, b)$ which interpolate a given continuous function f at a fixed set of knots in $[a, b]$, the linear spline s_{L} that interpolates f at these knots is the 'flattest', in the sense that its 'average slope' $\|s_{\mathrm{L}}'\|_2$ is smallest.

Theorem 11.2 *Suppose that s_{L} is the linear spline that interpolates $f \in \mathrm{C}[a, b]$ at the knots $a = x_0 < x_1 < \cdots < x_m = b$. Then, for any function v in $\mathrm{H}^1(a, b)$ that also interpolates f at these knots,*

$$\|s_{\mathrm{L}}'\|_2 \leq \|v'\|_2 \,.$$

Proof Let us observe that

$$\begin{aligned} \|v'\|_2^2 \quad &= \quad \int_a^b (v'(x) - s_{\mathrm{L}}'(x))^2 \mathrm{d}x + \int_a^b |s_{\mathrm{L}}'(x)|^2 \mathrm{d}x \\ &\quad\quad +2 \int_a^b (v'(x) - s_{\mathrm{L}}'(x)) s_{\mathrm{L}}'(x) \mathrm{d}x \,. \end{aligned} \tag{11.2}$$

We shall now use integration by parts to show that the last integral is equal to 0; the desired inequality will then follow by noting that the first term on the right-hand side is nonnegative and it is equal to 0 if, and only if, $v = s_{\mathrm{L}}$. Clearly,

$$\int_a^b (v'(x) - s_{\mathrm{L}}'(x)) s_{\mathrm{L}}'(x) \mathrm{d}x = \sum_{k=1}^m \int_{x_{k-1}}^{x_k} (v'(x) - s_{\mathrm{L}}'(x)) s_{\mathrm{L}}'(x) \mathrm{d}x$$

$$= \sum_{k=1}^m [(v(x_k) - s_{\mathrm{L}}(x_k)) s_{\mathrm{L}}'(x_k-) - (v(x_{k-1}) - s_{\mathrm{L}}(x_{k-1})) s_{\mathrm{L}}'(x_{k-1}+)$$

$$- \int_{x_{k-1}}^{x_k} (v(x) - s_{\mathrm{L}}(x)) s_{\mathrm{L}}''(x) \mathrm{d}x] \,. \tag{11.3}$$

Now $v(x_i) - s_{\mathrm{L}}(x_i) = f(x_i) - f(x_i) = 0$ for $i = 0, 1, \ldots, m$ and, since s_{L} is a linear polynomial over each of the open intervals (x_{k-1}, x_k), $k =$

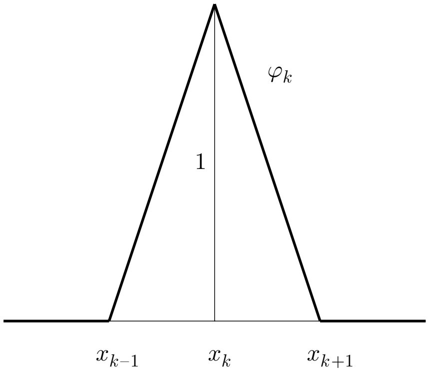

Fig. 11.2. The linear basis spline (or hat function) φ_k, $1 \le k \le m-1$.

$1, 2, \ldots, m$, it follows that s_{L}'' is identically 0 on each of these intervals. Thus, the expression in the square bracket in (11.3) is equal to 0 for each $k = 1, 2, \ldots, m$. $\square$

Sobolev spaces play an important role in approximation theory. We shall encounter them again in Chapter 14 which is devoted to the approximation of solutions to differential equations by piecewise polynomial functions.

11.3 Basis functions for the linear spline

Suppose that s_{L} is a linear spline with knots x_i, $i = 0, 1, \ldots, m$, interpolating the function $f \in \mathrm{C}[a, b]$. Instead of specifying the value of s_{L} on each subinterval $[x_{i-1}, x_i]$, $i = 1, 2, \ldots, m$, we can express s_{L} as a linear combination of suitable 'basis functions' φ_k as follows:

$$s_{\mathrm{L}}(x) = \sum_{k=0}^{m} \varphi_k(x) f(x_k)\,, \qquad x \in [a, b]\,.$$

Here, we require that each φ_k is itself a linear spline which vanishes at every knot except x_k, and $\varphi_k(x_k) = 1$. The function φ_k is often known as the **linear basis spline** or **hat function**, and is depicted in Figure 11.2.

The formal definition of φ_k is as follows:

$$\left.\begin{aligned}
&\varphi_k(x) = \begin{cases} 0 & \text{if } x \le x_{k-1}\,, \\ (x - x_{k-1})/h_k & \text{if } x_{k-1} \le x \le x_k\,, \\ (x_{k+1} - x)/h_{k+1} & \text{if } x_k \le x \le x_{k+1}\,, \\ 0 & \text{if } x_{k+1} \le x\,, \end{cases} \\
&\text{for } k = 1, \dots, m-1, \text{ and with} \\
&\varphi_0(x) = \begin{cases} (x_1 - x)/h_0 & \text{if } a = x_0 \le x \le x_1\,, \\ 0 & \text{if } x_1 \le x \end{cases} \\
&\text{and} \\
&\varphi_m(x) = \begin{cases} 0 & \text{if } x \le x_{m-1}\,, \\ (x - x_{m-1})/h_m & \text{if } x_{m-1} \le x \le x_m = b\,. \end{cases}
\end{aligned}\right\} \quad (11.4)$$

11.4 Cubic splines

Suppose that $f \in \mathrm{C}[a,b]$ and let $K = \{x_0, \dots, x_m\}$ be a set of $m+1$ knots in the interval $[a,b]$, $a = x_0 < x_1 < \cdots < x_m = b$. Consider the set $\mathcal{S}$ of all functions $s \in \mathrm{C}^2[a,b]$ such that

❶ $s(x_i) = f(x_i)$, $i = 0, 1, \dots, m$,
❷ s is a cubic polynomial on $[x_{i-1}, x_i]$, $i = 1, 2, \dots, m$.

Any element of $\mathcal{S}$ is referred to as an **interpolating cubic spline**. We note that, unlike linear splines which are uniquely determined by the interpolating conditions, there is more than one interpolating cubic spline $s \in \mathrm{C}^2[a,b]$ that satisfies the two conditions stated above; indeed, there are $4m$ coefficients of cubic polynomials (four on each subinterval $[x_{i-1}, x_i]$, $i = 1, 2, \dots, m$), and only $m+1$ interpolating conditions and $3(m-1)$ continuity conditions; since s belongs to $\mathrm{C}^2[a,b]$, this means that s, s' and s'' are continuous at the internal knots $x_1, \dots, x_{m-1}$. Hence, we have a total of $4m-2$ conditions for the $4m$ unknown coefficients. Depending on the choice of the remaining two conditions we can construct various interpolating cubic splines.

An important class of cubic splines is singled out by the following definition.

Definition 11.2 *The* **natural cubic spline***, denoted by s_2, is the element of the set $\mathcal{S}$ satisfying the end conditions*

$$s_2''(x_0) = s_2''(x_m) = 0\,.$$

We shall prove that this definition is correct in the sense that the two additional conditions in Definition 11.2 uniquely determine s_2: this will be done by describing an algorithm for constructing s_2.

Construction of the natural cubic spline. Let us begin by defining $\sigma_i = s_2''(x_i)$, $i = 0, 1, \ldots, m$, and noting that s_2'' is a linear function on each subinterval $[x_{i-1}, x_i]$. Therefore, s_2'' can be expressed as

$$s_2''(x) = \frac{x_i - x}{h_i}\,\sigma_{i-1} + \frac{x - x_{i-1}}{h_i}\,\sigma_i\,, \qquad x \in [x_{i-1}, x_i]\,.$$

Integrating this twice we obtain

$$\begin{aligned} s_2(x) \quad &= \quad \frac{(x_i - x)^3}{6h_i}\,\sigma_{i-1} + \frac{(x - x_{i-1})^3}{6h_i}\,\sigma_i \\ &\quad + \alpha_i(x - x_{i-1}) + \beta_i(x_i - x)\,, \quad x \in [x_{i-1}, x_i]\,, \end{aligned} \tag{11.5}$$

where α_i and β_i are constants of integration. Equating s_2 with f at the knots x_{i-1}, x_i yields

$$f(x_{i-1}) = \frac{1}{6}\sigma_{i-1}h_i^2 + h_i\beta_i\,, \qquad f(x_i) = \frac{1}{6}\sigma_i h_i^2 + h_i\alpha_i\,. \tag{11.6}$$

Expressing α_i and β_i from these, inserting them into (11.5) and exploiting the continuity of s_2' at the internal knots, (*i.e.*, using that $s_2'(x_i-) = s_2'(x_i+)$, $i = 1, \ldots, m-1$), gives

$$\begin{aligned} &h_i\sigma_{i-1} + 2(h_{i+1} + h_i)\sigma_i + h_{i+1}\sigma_{i+1} \\ &\qquad = 6\left(\frac{f(x_{i+1}) - f(x_i)}{h_{i+1}} - \frac{f(x_i) - f(x_{i-1})}{h_i}\right) \end{aligned} \tag{11.7}$$

for $i = 1, \ldots, m-1$, together with

$$\sigma_0 = \sigma_m = 0\,,$$

which is a system of linear equations for the σ_i. The matrix of the system is tridiagonal and nonsingular, since the conditions of Theorem 3.4 are clearly satisfied. By solving this linear system we obtain the σ_i, $i = 0, 1, \ldots, m$, and thereby all the α_i, β_i, $i = 1, 2, \ldots, m$, from (11.6).

We have seen in a previous section, in Theorem 11.2, that a linear spline can be characterised as a minimiser of the functional $v \mapsto \|v'\|_2$ over all $v \in \mathrm{H}^1(a, b)$ which interpolate a given continuous function at the knots of the spline. Natural cubic splines have an analogous property: among all functions $v \in \mathrm{H}^2(a, b)$ which interpolate a given continuous function f at a fixed set of knots in $[a, b]$, the natural cubic spline s_2 is smoothest, in the sense that it minimises $v \mapsto \|v''\|_2$, the 'average curvature' of v.

Theorem 11.3 *Let s_2 be the natural cubic spline that interpolates a function $f \in \mathrm{C}[a,b]$ at the knots $a = x_0 < x_1 < \cdots < x_m = b$. Then, for any function v in $\mathrm{H}^2(a,b)$ that also interpolates f at the knots,*

$$\|s_2''\|_2 \le \|v''\|_2 \,.$$

The proof is analogous to that of Theorem 11.2 and is left as an exercise.

The *smoothest interpolation property* expressed by Theorem 11.3 is the source of the name *spline*.[1] A spline is a flexible thin curve-drawing aid, made of wood, metal or acrylic. Assuming that its shape is given by the equation $y = v(x)$, $x \in [a,b]$, and is constrained by requiring that it passes through a finite set of prescribed points in the plane, v will take on a shape which minimises the strain energy

$$\mathrm{E}(v) = \int_a^b \frac{|v''(x)|^2}{(1 + |v'(x)|^2)^3} \mathrm{d}x$$

over all functions v which are constrained in the same way. If the function v is slowly varying, *i.e.*, $\max_{x\in[a,b]} |v'(x)| \ll 1$, this energy-minimisation property is very similar to the result in Theorem 11.3.

11.5 Hermite cubic splines

In the previous section we took $f \in \mathrm{C}[a,b]$ and demanded that s belonged to $\mathrm{C}^2[a,b]$; here we shall strengthen our requirements on the smoothness of the function that we wish to interpolate and assume that $f \in \mathrm{C}^1[a,b]$; simultaneously, we shall relax the smoothness requirements on the associated spline approximation s by demanding that $s \in \mathrm{C}^1[a,b]$ only.

Let $K = \{x_0, \dots, x_m\}$ be a set of knots in the interval $[a,b]$ with $a = x_0 < x_1 < \cdots < x_m = b$ and $m \ge 2$. We define the **Hermite cubic spline** as a function $s \in \mathrm{C}^1[a,b]$ such that

❶ $s(x_i) = f(x_i)$, $s'(x_i) = f'(x_i)$ for $i = 0, 1, \dots, m$,
❷ s is a cubic polynomial on $[x_{i-1}, x_i]$ for $i = 1, 2, \dots, m$.

Writing the spline s on the interval $[x_{i-1}, x_i]$ as

$$s(x) = c_0 + c_1(x - x_{i-1}) + c_2(x - x_{i-1})^2 + c_3(x - x_{i-1})^3 \,, \qquad x \in [x_{i-1}, x_i] \,, \quad (11.8)$$

[1] See Carl de Boor: *A Practical Guide to Splines*, Revised Edition, Springer Applied Mathematical Sciences, 27, Springer, New York, 2001.

we find that $c_0 = f(x_{i-1})$, $c_1 = f'(x_{i-1})$, and

$$\begin{aligned} c_2 &= 3\frac{f(x_i) - f(x_{i-1})}{h_i^2} - \frac{f'(x_i) + 2f'(x_{i-1})}{h_i}, \\ c_3 &= \frac{f'(x_i) + f'(x_{i-1})}{h_i^2} - 2\frac{f(x_i) - f(x_{i-1})}{h_i^3}. \end{aligned} \tag{11.9}$$

Note that the Hermite cubic spline only has a continuous first derivative at the knots, and therefore it is *not* an interpolating cubic spline in the sense of Section 11.4.

Unlike natural cubic splines, the coefficients of a Hermite cubic spline on each subinterval can be written down explicitly without the need to solve a tridiagonal system.

Concerning the size of the interpolation error, we have the following result.

Theorem 11.4 *Let $f \in \mathrm{C}^4[a,b]$, and let s be the Hermite cubic spline that interpolates f at the knots $a = x_0 < x_1 < \cdots < x_m = b$; then, the following error bound holds:*

$$\|f - s\|_\infty \leq \frac{1}{384} h^4 \|f^{iv}\|_\infty,$$

where $f^{iv} = f^{(4)}$ is the fourth derivative of f with respect to its argument, x, $h = \max_i h_i = \max_i(x_i - x_{i-1})$, and $\|\cdot\|_\infty$ denotes the ∞-norm on the interval $[a,b]$.

The proof is analogous to that of Theorem 11.1, except that Theorem 6.4 is used instead of Theorem 6.2.

Both the linear spline and the Hermite cubic spline are local approximations; the value of the spline at a point x between two knots x_{i-1} and x_i depends only on the values of the function and its derivative at these two knots. On the other hand, the natural cubic interpolating spline is a global approximation and, in this respect, it is more typical of a generic spline: a change in just one of the values at a knot, $f(x_k)$, will alter the right-hand side of the system of equations (11.7), so the values of all the quantities σ_i will change. Thus, the spline will change throughout the whole interval $[x_0, x_m]$. We conclude this section with an example.

Example 11.3 *Figure 11.3 shows the Hermite cubic spline approximation to the function $f\colon x \mapsto 1/(1+x^2)$, using four equally spaced knots in the interval $[0,5]$.*

The accuracy of this approximation is in striking contrast to the Lagrange polynomial approximation of degree 10 in Figure 6.1. The approximation over $[-5, 5]$, using seven equally spaced knots, is obviously obtained by symmetry; here we show only half the range for clarity.

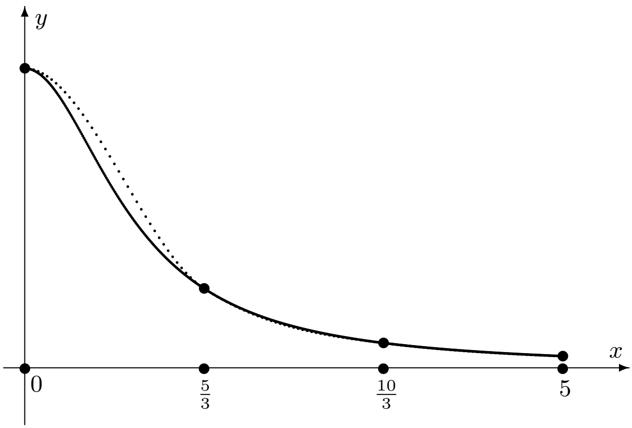

Fig. 11.3. The function $f\colon x \mapsto 1/(1+x^2)$ (full curve) and its Hermite cubic spline approximation (dotted curve). The interval is $[0, 5]$, and the knots are at 0, $\frac{5}{3}$, $\frac{10}{3}$ and 5.

As the error of this approximation is quite small, we show in Figure 11.4 graphs of the errors of three spline approximations, each using the same four knots. Note that in the first interval, $[0, \frac{5}{3}]$, the maximum error of the Hermite cubic spline is larger than that of the linear spline, but on the other two intervals it is much less. Both of these two splines are local approximations, as their values on any interval between two knots depend only on information about the function at those two knots. The natural cubic spline is a global approximation, as its value at any point depends on the values of the function at all the knots; on the first interval its error is much the same size as that of the Hermite cubic spline, but on the other two intervals its error is affected by this global coupling, and is a good deal bigger than that of the Hermite cubic spline. ◇

11.6 Basis functions for cubic splines

We have seen that the family of hat functions forms a basis for the linear space of linear splines corresponding to a certain fixed set of knots; we

https://doi.org/10.1017/CBO9780511801181.012 Published online by Cambridge University Press

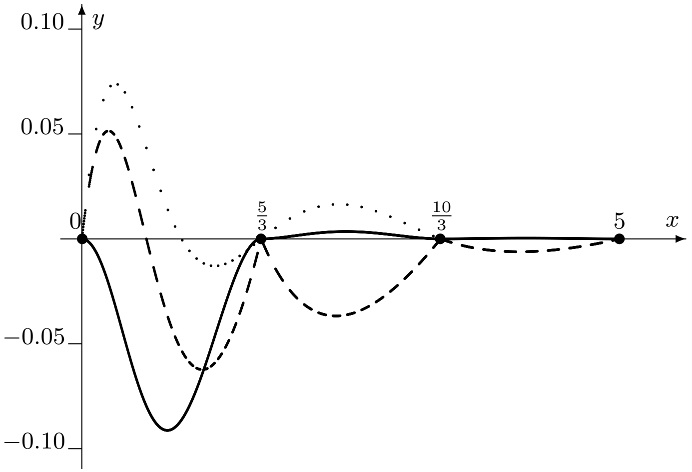

Fig. 11.4. Errors of three spline approximations to $f(x) = 1/(1+x^2)$: Hermite cubic (full curve), natural cubic (dotted curve) and linear spline (broken curve). The interval is $[0,5]$, and the knots are at 0, $\frac{5}{3}$, $\frac{10}{3}$ and 5.

shall now show how to construct a set of basis functions for cubic splines. The basis functions for splines are usually known as **B-splines**. Thus, the basis-splines constructed in Section 11.2 are referred to as linear B-splines. Here we shall be concerned with the construction of cubic B-splines. To simplify the notation we shall assume in this section that the knots are equally spaced, so that

$$x_k = kh\,, \qquad k = 0, 1, \ldots, n+1\,,$$

with $h > 0$.

We begin by introducing the idea of the positive part of a function.

Definition 11.3 *Suppose that $n \geq 1$. The* **positive part** *of the function $x \mapsto (x-a)^n$ is the function $x \mapsto (x-a)_+^n$ defined by*

$$(x-a)_+^n = \begin{cases} (x-a)^n\,, & x \geq a\,, \\ 0\,, & x < a\,. \end{cases}$$

Clearly the function $x \mapsto (x-x_k)_+^n$ is a spline of degree n; at the knot x_k the derivatives of order up to $n-1$ are zero, but the derivative of order n is not continuous at $x = x_k$.

Figure 11.5 shows the graphs of the functions $x \mapsto x_+$ and $x \mapsto x_+^3$ on the interval $[-1, 1]$.

We shall also need the following result.

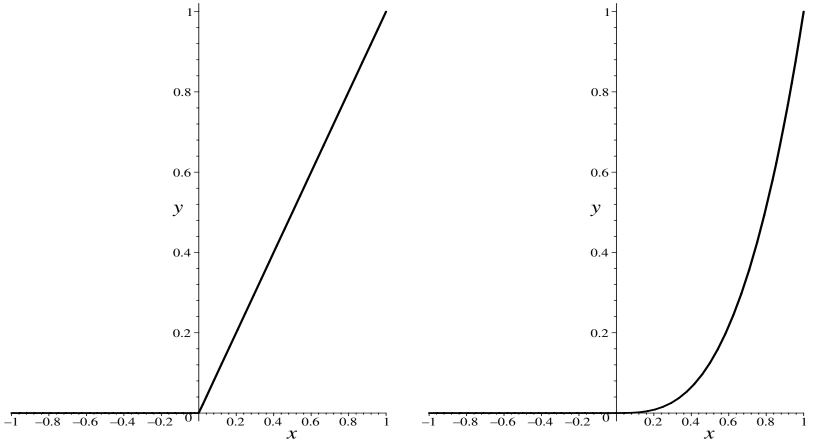

Fig. 11.5. The graph of the function $x \mapsto (x)_+^n$, for x in the interval $[-1, 1]$, with $n = 1$ (left) and $n = 3$ (right).

Lemma 11.1 *Suppose that P is a polynomial in x of degree $n \geq 1$. Then, for each $r = 1, \ldots, n$, the function $Q_{(r)}$ defined by*

$$Q_{(r)}(x) = \sum_{k=0}^{r} (-1)^k \begin{pmatrix} r \\ k \end{pmatrix} P(x - kh)$$

is a polynomial of degree $n - r$ and $Q_{(n+1)}(x) \equiv 0$, $x \in \mathbb{R}$.

Proof It is easy to see that $Q_{(1)}(x) = P(x) - P(x - h)$, and therefore $Q_{(1)}$ is a polynomial of degree $n-1$. Suppose now that, for some $r > 0$, $Q_{(r)}$ is a polynomial in x of degree $n-r$; then, $x \mapsto Q_{(r)}(x) - Q_{(r)}(x-h)$ is a polynomial of degree $n - r - 1$. But

$$\begin{aligned} Q_{(r)}(x) - Q_{(r)}(x-h) \\ &= \sum_{k=0}^{r} (-1)^k \begin{pmatrix} r \\ k \end{pmatrix} [P(x-kh) - P(x-(k+1)h)] \\ &= P(x) + (-1)^{r+1} P(x-(r+1)h) \\ &\quad + \sum_{k=1}^{r} (-1)^k \left[\begin{pmatrix} r \\ k \end{pmatrix} + \begin{pmatrix} r \\ k-1 \end{pmatrix} \right] P(x-kh) \\ &= \sum_{k=0}^{r+1} (-1)^k \begin{pmatrix} r+1 \\ k \end{pmatrix} P(x-kh) \\ &= Q_{(r+1)}(x)\,, \end{aligned} \tag{11.10}$$

https://doi.org/10.1017/CBO9780511801181.012 Published online by Cambridge University Press

from the standard properties of binomial coefficients. Hence $Q_{(r+1)}$ is a polynomial in x of degree $n-r-1$, and the result follows by induction. Finally, this shows that $Q_{(n)}$ is a polynomial of degree 0, and is therefore constant on $\mathbb{R}$. Thus, by the same argument, $Q_{(n+1)}$ is identically 0 on $\mathbb{R}$. □

Theorem 11.5 *For each $n \geq 1$, the function $S_{(n)}$ defined by*

$$S_{(n)}(x) = \sum_{k=0}^{n+1} (-1)^k \left(\begin{array}{c} n+1 \\ k \end{array} \right) (x - kh)_+^n$$

is a spline of degree n with equally spaced knots kh, $k = 0, 1, \ldots, n+1$. It has a continuous derivative of order $n-1$ and is identically 0 outside the interval $(0, (n+1)h)$.

Proof The function $S_{(n)}$ is clearly a spline as stated, and $S_{(n)}(x)$ is identically 0 for $x \leq 0$. When $x \geq (n+1)h$ the arguments $x - kh$, $k = 0, 1, \ldots, n+1$, of the positive parts are all nonnegative, so that

$$S_{(n)}(x) = \sum_{k=0}^{n+1} (-1)^k \left(\begin{array}{c} n+1 \\ k \end{array} \right) (x - kh)^n \,,$$

and this is identically zero by Lemma 11.1. □

Taking $n = 1$ we find that

$$S_{(1)}(x) = x_+ - 2(x-h)_+ + (x-2h)_+ \,.$$

After normalisation by $1/h$ so as to have a maximum value of 1, and shifting $x = 0$ to $x = x_{k-1}$, this yields a representation of the linear hat function φ_k from (11.4) in the form

$$\varphi_k(x) = \frac{1}{h} S_{(1)}(x - x_{k-1}) \,,$$

which, for $1 \leq k \leq n$, is nonzero over two consecutive intervals: $(x_{k-1}, x_k]$ and $[x_k, x_{k+1})$.

In the same way we obtain a basis function for the cubic spline by taking $n = 3$:

$$S_{(3)}(x) = x_+^3 - 4(x-h)_+^3 + 6(x-2h)_+^3 - 4(x-3h)_+^3 + (x-4h)_+^3 \,.$$

Normalising so as to have a maximum value of 1 and shifting $x = 0$ to $x = x_{k-2}$, we get

$$\psi_k(x) = \frac{1}{4h^3} S_{(3)}(x - x_{k-2}) \,.$$

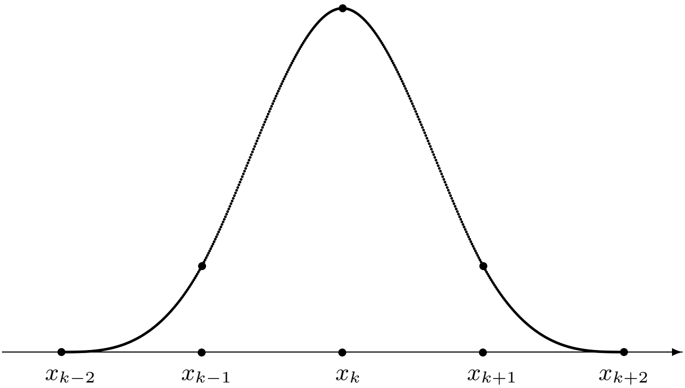

Fig. 11.6. Normalised cubic B-spline, $\psi_k(x)$, $2 \le k \le n-1$.

For $2 \le k \le n-1$, this function is nonzero over four consecutive intervals $(x_{k-2}, x_{k-1}]$, $[x_{k-1}, x_k]$, $[x_k, x_{k+1}]$ and $[x_{k+1}, x_{k+2})$, and is illustrated in Figure 11.6.

We see that both φ_k and ψ_k are nonnegative for all x; this is true for a spline basis function of any degree n, $n \ge 1$, constructed in this way, but we shall not prove it here (see Exercise 6).

For a finite set of knots $a = x_0 < x_1 < \cdots < x_{n+1} = b$ on the bounded and closed interval $[a, b]$ the normalised linear basis splines $x \mapsto \varphi_0(x)$ and $x \mapsto \varphi_{n+1}(x)$ are considered only for x in $[a, b]$, so as to avoid reference to nonexisting knots (such as x_{-1} or x_{n+2}) that lie outside $[a, b]$. A similar comment applies to the normalised cubic basis splines ψ_0, ψ_1, ψ_n and ψ_{n+1}.

11.7 Notes

There are many excellent texts covering the theory of piecewise polynomial approximation by splines. For a detailed survey of key results we refer to Chapters 18–24 of

- M.J.D. Powell, *Approximation Theory and Methods*, Cambridge University Press, Cambridge, 1996.

You may have noticed that we have given bounds on the error in linear spline approximation in Theorem 11.1, and in Hermite cubic spline approximation in Theorem 11.4, but not for the natural cubic spline. The analysis of the error in the natural cubic spline approximation is quite complicated; Powell gives full details in his book.

The following are classical texts on the theory of splines.

- J.H. Ahlberg, E.N. Nilson, and J.L. Walsh, *The Theory of Splines and Their Applications*, Mathematics in Science and Engineering, 38, Academic Press, New York, 1967.
- C. de Boor, *A Practical Guide to Splines*, Revised Edition, Springer Applied Mathematical Sciences, 27, Springer, New York, 2001.
- Larry L. Schumaker, *Spline Functions: Basic Theory,* John Wiley & Sons, New York, 1981.

The variational characterisations of splines stated in Sections 11.1 and 11.3 stem from the work of J.C. Holladay, Smoothest curve approximation, *Math. Comput.* **11**, 233–243, 1957.

Our definition of the Sobolev space $\mathrm{H}^k(a,b)$ in Section 11.1, based on the concept of absolute continuity, is specific to functions of a single variable. More generally, for functions of several real variables one needs to invoke the theory of weak differentiability or the theory of distributions to give a rigorous definition of the Sobolev space $H^k(\Omega)$ with $\Omega \subset \mathbb{R}^n$; alternatively, one can define $H^k(\Omega)$ by completion of the set of smooth functions in a suitable norm. For the sake of simplicity of exposition we have chosen to avoid such general approaches.

Exercises

11.1 An interpolating spline of degree n is required to have continuous derivatives of order up to and including $n-1$ at the knots. How many additional conditions are required to specify the spline uniquely?

11.2 (i) Suppose that f is a polynomial of degree 1. Show that the linear spline s_{L} which interpolates f at the knots x_i for $i = 0, 1, \dots, m$ is identical to f, so that $s_{\mathrm{L}} \equiv f$.

(ii) Suppose that f is a polynomial of degree 3. Show that the Hermite cubic spline s_{H} which interpolates f at the knots x_i, $i = 0, 1, \dots, m$, is identical to f, so that $s_{\mathrm{H}} \equiv f$.

(iii) Suppose that f is a polynomial of degree 3. Show that the natural cubic spline s_2 which interpolates f at the knots x_i, $i = 0, 1, \dots, m$, is not in general identical to f.

11.3 Suppose that the natural cubic spline s_2 interpolates the function $f\colon x \mapsto x^3$ on the interval $[0,1]$, the knots being equally spaced, so that $x_i = ih$, $i = 0, 1, \ldots, m$, with $h = 1/m$, $m \geq 2$. Write down the equations which determine the quantities σ_i. If the two additional conditions are $\sigma_0 = \sigma_m = 0$, show that these equations are not satisfied by $\sigma_i = f''(x_i)$, $i = 1, \ldots, m-1$, so that s_2 and f are not identical. If, however, these two additional conditions are replaced by $\sigma_0 = f''(0)$, $\sigma_m = f''(1)$, show that $\sigma_i = f''(x_i)$, $i = 0, 1, \ldots, m$, and deduce that s_2 and f are identical.

11.4 A linear spline on the interval $[0,1]$ is expressed in terms of the basis functions as

$$s(x) = \sum_{k=0}^{m} \alpha_k \varphi_k(x)\,.$$

Instead of being required to interpolate the function f at the knots, the spline s is required to minimise $\|f - s\|_2$. Show that the coefficients α_k satisfy the system of equations

$$A\boldsymbol{\alpha} = \boldsymbol{b}\,,$$

where the elements of the matrix A are

$$A_{ij} = \int_0^1 \varphi_j(x)\varphi_i(x)\mathrm{d}x$$

and the elements of $\boldsymbol{b}$ are

$$b_i = \int_0^1 f(x)\varphi_i(x)\mathrm{d}x\,.$$

Now suppose that the knots are equally spaced, so that $x_k = kh$, $k = 0, 1, \ldots, m$, where $h = 1/m$, $m \geq 2$. Show that the matrix A is tridiagonal, with $A_{ii} = \frac{2}{3}h$ for $i = 1, \ldots, m-1$, and determine the other nonzero elements of A. Show also that A has the properties required for the use of the Thomas algorithm described in Section 3.3.

11.5 In the notation of Exercise 4, suppose that $f(x) = x$. Verify that the system of equations is satisfied by $\alpha_k = kh$, so that $s = f$.

Now suppose that $f(x) = x^2$. Verify that the equations are satisfied by $\alpha_k = (kh)^2 + Ch^2$, where C is a constant to be determined. Deduce that $s(x_k) = f(x_k) + Ch^2$.

https://doi.org/10.1017/CBO9780511801181.012 Published online by Cambridge University Press

11.6 In the notation of Theorem 11.5, the spline basis function $S_{(n)}$ of degree n is defined by

$$S_{(n)}(x) = \sum_{k=0}^{n+1} (-1)^k \left(\begin{array}{c} n+1 \\ k \end{array} \right) (x - kh)_+^n .$$

Explain why, for any value of a,

$$(x-a)_+^n (x-a) = (x-a)_+^{n+1} .$$

Show that

$$xS_{(n)}(x) + [(n+2)h - x] S_{(n)}(x-h) = S_{(n+1)}(x) .$$

Hence show by induction that $S_{(n)}(x) \geq 0$ for all x.

11.7 Use the result of Exercise 6 to show by induction that each basis function $S_{(n)}$ is symmetric; that is,

$$S_{(n)}(p+x) = S_{(n)}(p-x)$$

for all x, where $p = \frac{1}{2}(n+1)h$.