Carl de Boor

Applied
Mathematical
sciences
27

A Practical Guide to Splines

Revised Edition

样条实用指南

修订版

Springer

世界图书出版公司
www.wpcbj.com.cn

Carl de Boor

A Practical Guide to Splines

Revised Edition

With 32 figures

Springer

Carl de Boor
Department of Computer Sciences
University of Wisconsin-Madison
Madison, WI 53706-1685
USA
deboor@cs.wisc.edu

Mathematics Subject Classification (2000): 65Dxx, 41A15

Library of Congress Cataloging-in-Publication Data
de Boor, Carl.
A practical guide to splines / Carl de Boor. — Rev. ed.
p. cm. — (Applied mathematical sciences ; 27)
Includes bibliographical references and index.
ISBN 0-387-95366-3 (alk. paper)
1. Spline theory. I. Title. II. Applied mathematical sciences (Springer-Verlag New York Inc.) ; v. 27
QA1 .A647 vol. 27 2001
[QA224]
510s—dc21
[511'.42] 2001049644

ISBN 0-387-95366-3 SPIN 10952603

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Preface

This book is a reflection of my limited experience with calculations involving polynomial splines. It stresses the representation of splines as linear combinations of B-splines, provides proofs for only some of the results stated but offers many Fortran programs, and presents only those parts of spline theory that I found useful in calculations. The particular literature selection offered in the bibliography shows the same bias; it contains only items to which I needed to refer in the text for a specific result or a proof or additional information and is clearly not meant to be representative of the available spline literature. Also, while I have attached names to some of the results used, I have not given a careful discussion of the historical aspects of the field. Readers are urged to consult the books listed in the bibliography (they are marked with an asterisk) if they wish to develop a more complete and balanced picture of spline theory.

The following outline should provide a fair idea of the intent and content of the book.

The first chapter recapitulates material needed later from the ancient theory of polynomial interpolation, in particular, divided differences. Those not familiar with divided differences may find the chapter a bit terse. For comfort and motivation, I can only assure them that every item mentioned will actually be used later. The rudiments of polynomial approximation theory are given in Chapter II for later use, and to motivate the introduction of piecewise polynomial (or, pp) functions.

Readers intent upon looking at the general theory may wish to skip the next four chapters, as these follow somewhat the historical development, with piecewise linear, piecewise cubic, and piecewise parabolic approximation discussed, in that order and mostly in the context of interpolation. Proofs are given for results that, later on in the more general context of splines of arbitrary order, are only stated. The intent is to summarize elementary spline theory in a practically useful yet simple setting.

The general theory is taken up again starting with Chapter VII, which, along with Chapter VIII, is devoted to the computational handling of pp functions of arbitrary order. B-splines are introduced in Chapter IX. It is only in that chapter that a formal definition of "spline" as a linear combination of B-splines is given. Chapters X and XI are intended to familiarize the reader with B-splines.

The remaining chapters contain various applications, all (with the notable exception of taut spline interpolation in Chapter XVI) involving B-splines. Chapter XII is the pp companion piece to Chapter II; it contains a discussion of how well a function can be approximated by pp functions. Chapter XIII is devoted to various aspects of spline interpolation as a particularly simple, computationally efficient yet powerful scheme for spline approximation in the presence of exact data. For noisy data, the smoothing spline and least-squares spline approximation are offered in Chapter XIV. Just one illustration of the use of splines in solving differential equations is given, in Chapter XV, where an ordinary differential equation is solved by collocation. Chapter XVI contains an assortment of items, all loosely connected to the approximation of a curve. It is only here (and in the problems for Chapter VI) that the beautiful theory of cardinal splines, i.e., splines on a uniform knot sequence, is discussed. The final chapter deals with the simplest generalization of splines to several variables and offers a somewhat more abstract view of the various spline approximation processes discussed in this book.

Each chapter has some problems attached to it, to test the reader's understanding of the material, to bring in additional material and to urge, at times, numerical experimentation with the programs provided. It should be understood, though, that Problem 0 in each chapter that contains programs consists of running those programs with various sample data in order to gain some first-hand practical experience with the methods espoused in the book.

The programs occur throughout the text and are meant to be read, as part of the text.

The book grew out of orientation lectures on splines delivered at Redstone Arsenal in September, 1976, and at White Sands Missile Range in October, 1977. These lectures were based on a 1973 MRC report concerning a Fortran package for calculating with B-splines, a package put together in 1971 at Los Alamos Scientific Laboratories around a routine (now called BSPLVB) that took shape a year earlier during a workshop at Oberlin organized by Jim Daniel. I am grateful for advice received during those years, from Fred Dorr, Cleve Moler, Blair Swartz and others.

During the writing of the book, I had the benefit of detailed and copious advice from John Rice who read various versions of the entire manuscript. It owes its length to his repeated pleas for further elucidation. I owe him thanks also for repeated encouragement. I am also grateful to a group at Stanford, consisting of John Bolstad, Tony Chan, William Coughran, Jr., Alphons Demmler, Gene Golub, Michael Heath, Franklin Luk, and Marcello Pagano that, through the good offices of Eric Grosse, gave me much welcome advice after reading an early version of the manuscript. The

programs in the book would still be totally unreadable but for William Coughran's and Eric Grosse's repeated arguments in favor of comment cards. Dennis Jespersen read the final manuscript with astonishing care and brought a great number of mistakes to my attention. He also raised many questions, many of which found place among the problems at the end of chapters. Walter Gautschi, and Klaus Böhmer and his students, read a major part of the manuscript and uncovered further errors. I am grateful to them all.

Time for writing, and computer time, were provided by the Mathematics Research Center under Contract No. DAAG29–75–C–0024 with the U.S. Army Research Office. Through its visitor program, the Mathematics Research Center also made possible most of the helpful contacts acknowledged earlier. I am deeply appreciative of the mathematically stimulating and free atmosphere provided by the Mathematics Research Center.

Finally, I would like to thank Reinhold de Boor for the patient typing of the various drafts.

Carl de Boor
Madison, Wisconsin
February 1978

The present version differs from the original in the following respects. The book is now typeset (in *plain* TEX; thank you, Don Knuth!), the Fortran programs now make use of FORTRAN 77 features, the figures have been redrawn with the aid of MATLAB (thank you, Cleve Moler and Jack Little!), various errors have been corrected, and many more formal statements have been provided with proofs. Further, all formal statements and equations have been numbered by the *same* numbering system, to make it easier to find any particular item. A major change has occurred in Chapters IX–XI where the B-spline theory is now developed directly from the recurrence relations without recourse to divided differences (except for the derivation of the recurrence relations themselves). This has brought in knot insertion as a powerful tool for providing simple proofs concerning the shape-preserving properties of the B-spline series.

I gratefully acknowledge support from the Army Research Office and from the Division of Mathematical Sciences of the National Science Foundation.

Special thanks are due to Peter de Boor, Kirk Haller, and S. Nam for their substantial help, and to Reinhold de Boor for the protracted final editing of the TEX files and for all the figures.

Carl de Boor
Madison, Wisconsin
October 2000

Notation

Here is a detailed list of all the notation used in this book. Readers will have come across some of them, perhaps most of them. Still, better to bore them now than to mystify them later.

$:=$ is the sign indicating "equal by definition". It is asymmetric as such a sign should be (as none of the customary alternatives, such as $\equiv$, or $\stackrel{\text{def}}{=}$, or $\stackrel{\triangle}{=}$, etc., are). Its meaning: "$a := b$" indicates that a is the quantity to be defined or explained, and b provides the definition or explanation, and "$b =: a$" has the same meaning.

$\{x, y, z, \ldots\} :=$ the set comprising the elements x, y, z, ...

$\{x \in X : P(x)\} :=$ the set of elements of X having the property $P(x)$.

$(x, y, \ldots) :=$ the sequence whose first term is x, whose second term is y, ...

$\#S :=$ the number of elements (or terms) in the set (or sequence) S.

$\emptyset :=$ the empty set.

$\mathbb{N} := \{1, 2, 3, \ldots\}$.

$\mathbb{Z} := \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.

$\mathbb{R} :=$ the set of real numbers.

$\mathbb{C} :=$ the set of complex numbers.

$\overline{z} :=$ the complex conjugate of the complex number z.

$[a \,..\, b] := \{x \in \mathbb{R} : a \le x \le b\}$, a closed interval. This leaves $[a, b]$ free to denote a first divided difference (or, perhaps, the matrix with the two columns a and b).

$(a \,..\, b) := \{x \in \mathbb{R} : a < x < b\}$, an open interval. This leaves (a, b) free to denote a particular sequence, e.g., a point in the plane (or, perhaps, the inner product of two vectors in some inner product space). Analogously, $[a \,..\, b)$ and $(a \,..\, b]$ denote half-open intervals.

$\text{const}_{\alpha,\ldots,\omega} :=$ a constant that may depend on $\alpha, \ldots, \omega$.

$f: A \to B: a \mapsto f(a)$ describes the function f as being defined on $A =: \operatorname{dom} f$ (called its **domain**) and taking values in the set $B =: \operatorname{tar} f$ (called its **target**), and carrying the typical element $a \in A$ to the element $f(a) \in B$. For example, $F: \mathbb{R} \to \mathbb{R}: x \mapsto \exp(x)$ describes the exponential function. I will use at times $f: a \mapsto f(a)$ if the domain and target of f are understood from the context. Thus, $\mu: f \mapsto \int_0^1 f(x)\,dx$ describes the linear functional μ that takes the number $\int_0^1 f(x)\,dx$ as its value at the function f, presumably defined on $[0\,..\,1]$ and integrable there.

$\operatorname{supp} f := \{\, x \in \operatorname{dom} f : f(x) \neq 0 \,\}$, the support of f. Note that, in Analysis, it is the *closure* of this set that is, by definition, the support of f.

$f_{|I} := \ g: I \to B: a \mapsto f(a)$, the restriction of f to I.

$h \to \begin{cases} 0^+ \\ 0^- \end{cases} \ := \ h$ approaches 0 through $\begin{matrix}\text{positive}\\ \text{negative}\end{matrix}$ values.

$f(a^+) := \lim_{h \to 0^+} f(a+h)$, $f(a^-) := \lim_{h \to 0^-} f(a+h)$.

$\operatorname{jump}_a f := \ f(a^+) - f(a^-)$, the jump in f across a.

$g(x) = \mathcal{O}(f(x))$ (in words, "$g(x)$ is of order $f(x)$") as x approaches a $:=$ $\limsup_{x \to a} \left| \frac{g(x)}{f(x)} \right| < \infty$. The lim sup itself is called the **order constant** of this order relation.

$g(x) = o(f(x))$ (in words, "$g(x)$ is of higher order than $f(x)$") as x approaches a $:= \lim_{x \to a} \frac{g(x)}{f(x)} = 0$.

$f(\cdot, y) :=$ the function of one variable obtained from the function $f: X \times Y \to Z$ by holding the second variable at a fixed value y. Also, $\int_a^b G(\cdot, y) g(y)\,dy$ describes the function that results when a certain integral operator is applied to the function g.

$(x)_+ := \ \max\{x, 0\}$, the **truncation function**.

$\ln x :=$ the natural logarithm of x.

$\lfloor x \rfloor := \ \max\{\, n \in \mathbb{Z} : n \leq x \,\}$, the **floor** function.

$\lceil x \rceil := \ \min\{\, n \in \mathbb{Z} : n \geq x \,\}$, the **ceiling** function.

$\ell_i :=$ a Lagrange polynomial (p. 2).

$(-)^r := \ (-1)^r$.

$\binom{k}{r} := \ \frac{k!}{r!(k-r)!}$, a binomial coefficient.

$\mathcal{E}_n :=$ The Euler spline of degree n (p. 65).

Boldface symbols denote sequences or vectors, the ith term or entry is denoted by the same letter in ordinary type and subscripted by i. Thus,

$\boldsymbol{\tau}$, (τ_i), $(\tau_i)_1^n$, $(\tau_i)_{i=1}^n$, $(\tau_i : i = 1, \ldots, n)$, and $(\tau_1, \ldots, \tau_n)$ are various ways of describing the same n-vector.

$\mathbf{m} := (m, \ldots, m)$, for $m \in \mathbb{Z}$.

$X^n := \{ (x_i)_1^n : x_i \in X, \text{ all } i \}$.

$\Delta\tau_i := \tau_{i+1} - \tau_i$, the **forward difference.**

$\nabla\tau_i := \tau_i - \tau_{i-1}$, the **backward difference.**

$\mathrm{S}^-\boldsymbol{\tau} :=$ number of strong sign changes in $\boldsymbol{\tau}$ (p. 138).

$\mathrm{S}^+\boldsymbol{\tau} :=$ number of weak sign changes in $\boldsymbol{\tau}$ (p. 232).

$$\sum_{i=r}^{s} \tau_i := \begin{cases} \tau_r + \tau_{r+1} + \cdots + \tau_s, & \text{if } r \le s; \\ 0, & \text{if } r > s. \end{cases}$$

$\sum_i \tau_i := \sum_{i=r}^{s} \tau_i$, with r and s understood from the context.

$$\prod_{i=r}^{s} \tau_i := \begin{cases} \tau_r \cdot \tau_{r+1} \cdots \tau_s, & \text{if } r \le s; \\ 1, & \text{if } r > s. \end{cases}$$

For nondecreasing sequences or meshes $\boldsymbol{\tau}$, we use

$|\boldsymbol{\tau}| := \max_i \Delta\tau_i$, the **mesh size.**

$M_{\boldsymbol{\tau}} := \max_{i,j} \Delta\tau_i/\Delta\tau_j$, the **global mesh ratio.**

$m_{\boldsymbol{\tau}} := \max_{|i-j|=1} \Delta\tau_i/\Delta\tau_j$, the **local mesh ratio.**

$\tau_{i+1/2} := (\tau_i + \tau_{i+1})/2$.

Matrices are usually denoted by capital letters, their entries by corresponding lower case letters doubly subscripted. A, (a_{ij}), $(a_{i,j})_1^{m,n}$,
$(a_{ij})_{i=1}^{m}{}_{;j=1}^{\,n}$, $\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$ are various ways of describing the same matrix.

$A^T := (a_{ji})_{j=1}^{n}{}_{;i=1}^{\,m}$, the transpose of A.

$A^H := (\overline{a}_{ji})_{j=1}^{n}{}_{;i=1}^{\,m}$, the conjugate transpose or Hermitian of A.

$\det A :=$ the determinant of A.

$\delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$, the Kronecker Delta.

$\mathrm{span}(\varphi_i) := \{ \sum_i \alpha_i\varphi_i : \alpha_i \in \mathbb{R} \}$, the linear combinations of the sequence $(\varphi_i)_1^n$ of elements of a linear space X. Such a sequence is a basis for its span in case it is linearly independent, that is, in case $\sum_i \alpha_i\varphi_i = 0$ implies that $\boldsymbol{\alpha} = \mathbf{0}$. We note that the linear independence of such a sequence $(\varphi_i)_1^n$ is almost invariably proved by exhibiting a corresponding sequence $(\lambda_i)_1^n$ of linear functionals on X for which the matrix $(\lambda_i\varphi_i)$ is invertible, e.g., $\lambda_i\varphi_j = \delta_{ij}$, for all i, j. In such a case, $\dim \mathrm{span}(\varphi_i)_1^n = n$.

$\Pi_{<k} := \Pi_{k-1} :=$ linear space of polynomials of order k (p. 1).

$\Pi_{<k,\boldsymbol{\xi}} := \Pi_{k-1,\boldsymbol{\xi}} :=$ linear space of pp functions of order k with break sequence $\boldsymbol{\xi}$ (p. 70).

$D^j f :=$ jth derivative of f; for $f \in \Pi_{<k,\boldsymbol{\xi}}$, see p. 70.

$\Pi_{<k,\boldsymbol{\xi},\boldsymbol{\nu}} := \Pi_{k-1,\boldsymbol{\xi},\boldsymbol{\nu}} :=$ linear subspace of $\Pi_{<k,\boldsymbol{\xi}}$ consisting of those elements that satisfy continuity conditions specified by $\boldsymbol{\nu}$ (p. 82).

$\$_{k,\mathbf{t}} := \mathrm{span}(\mathrm{B}_{i,k,\mathbf{t}})$, linear space of splines of order k with knot sequence $\mathbf{t}$ (p. 93).

$\mathrm{B}_i := \mathrm{B}_{i,k,\mathbf{t}} :=$ ith B-spline of order k with knot sequence $\mathbf{t}$ (p. 87).

$\$_{k,n} := \cup\{\, f \in \$_{k,\mathbf{t}} : t_1 = \cdots = t_k = a, t_{n+1} = \cdots = t_{n+k} = b \,\}$ (pp. 163, 239).

$\$^{\mathrm{nat}}_{k,\mathbf{x}} :=$ "natural" splines of order k for the sites $\mathbf{x}$ (p. 207).

$C[a\,..\,b] := \{\, f\colon [a\,..\,b] \to \mathbb{R} : f \text{ continuous} \,\}$.

$\|f\| := \max\{\, |f(x)| : a \le x \le b \,\}$, the uniform norm of $f \in C[a\,..\,b]$. (We note that $\|f+g\| \le \|f\| + \|g\|$ and $\|\alpha f\| = |\alpha|\|f\|$ for $f, g \in C[a\,..\,b]$ and $\alpha \in \mathbb{R}$).

$\omega(f;h) := \max\{\, |f(x) - f(y)| : x, y \in [a\,..\,b], |x-y| \le h \,\}$, the modulus of continuity for $f \in C[a\,..\,b]$ (p. 25).

$\mathrm{dist}\,(g,S) := \inf\{\, \|g-f\| : f \in S \,\}$, the distance of $g \in C[a\,..\,b]$ from the subset S of $C[a\,..\,b]$.

$C^{(n)}[a\,..\,b] := \{\, f\colon [a\,..\,b] \to \mathbb{R} : f$ is n times continuously differentiable $\}$.

$[\tau_i, \ldots, \tau_j]f :=$ divided difference of order $j-i$ of f, at the sites $\tau_i, \ldots, \tau_j$ (p. 3). In particular, $[\tau_i]\colon f \mapsto f(\tau_i)$.

Special spline approximation maps:

$I_k :=$ interpolation by splines of order k (p. 182),

$L_k :=$ Least-squares approximation by splines of order k (p. 220),

$V :=$ Schoenberg's variation diminishing spline approximation (p. 141).

Contents

III · Piecewise Linear Approximation

IV · Piecewise Cubic Interpolation

V · Best Approximation Properties of Complete Cubic Spline Interpolation and Its Error 51

VI · Parabolic Spline Interpolation 59

VII · A Representation for Piecewise Polynomial Functions

VIII · The Spaces $\Pi_{<k,\xi,\nu}$ and the Truncated Power Basis

IX · The Representation of PP Functions by B-Splines

X · The Stable Evaluation of B-Splines and Splines

XI · The B-Spline Series, Control Points, and Knot Insertion

XII · Local Spline Approximation and the Distance from Splines

XIII · Spline Interpolation

XIV · Smoothing and Least-Squares Approximation

XV · The Numerical Solution of an Ordinary Differential Equation by Collocation

XVI · Taut Splines, Periodic Splines, Cardinal Splines and the Approximation of Curves

XVII · Surface Approximation by Tensor Products

I

Polynomial Interpolation

In this introductory chapter, we state, mostly without proof, those basic facts about polynomial interpolation and divided differences needed in subsequent chapters. The reader who is unfamiliar with some of this material is encouraged to consult textbooks such as Isaacson & Keller [1966] or Conte & de Boor [1980] for a more detailed presentation.

One uses polynomials for approximation because they can be evaluated, differentiated, and integrated easily and in finitely many steps using the basic arithmetic operations of addition, subtraction, and multiplication. A polynomial of **order** n is a function of the form

$$(1) \qquad p(x) = a_1 + a_2 x + \cdots + a_n x^{n-1} = \sum_{j=1}^{n} a_j x^{j-1},$$

i.e., a polynomial of degree $< n$. It turns out to be more convenient to work with the *order* of a polynomial than with its degree since the set of all polynomials of degree n fails to be a linear space, while the set of all polynomials of order n forms a linear space, denoted here by

$$\Pi_{<n} = \Pi_{\le n-1} = \Pi_{n-1}.$$

Note that a polynomial of order n has exactly n degrees of freedom.

Note also that, in MATLAB, hence in the SPLINE TOOLBOX (de Boor [1990]$_2$), the coefficient sequence `a = [a(1), ... ,a(n)]` of a polynomial of order n starts with the *highest* coefficient. In particular, if `x` is a scalar, then the MATLAB command `polyval(a,x)` returns the number

```
a(1)*x^(n-1) + a(2)*x^(n-2) + ... + a(n-1)*x + a(n)
```