What is a multivariate spline?

C. de Boor

Abstract. The various concepts and ideas that have contributed to univariate spline theory are considered with a view to finding a suitable definition of a multivariate spline. In this way, an overview of the existing more or less complete univariate spline theory is given along with a survey of some of the high points of the current research in multivariate splines.

My very first paper dealt with multivariate (well, bivariate) splines and I was then quite certain of what a multivariate spline, i.e., a spline function of many variables, might be. Now, many years and several answers later, I am not so sure any more and therefore consider the question worth a forty-minute talk.

It is a worthwhile question since univariate splines have been phenomenally successful and one would wish to have available a similarly useful tool for the approximation of functions of *several* variables. This raises the question of just which features of the univariate spline to generalize. My talk will therefore be in part a survey of the more or less complete univariate spline theory with the aim of deciding which parts to take along into the multivariate context.

But before embarking on that discussion, I want to point out that there is available one way of generalization that is specifically designed to require no thought, no new idea (if this construction is satisfactory for you, I have nothing further to tell you). This is the **tensor product** construct. Here one takes one's favorite univariate spline class $ and fashion from it splines

$$\mathbb{R}^d \longrightarrow \mathbb{R} : (x, y, \dots, z) \longmapsto f(x)g(y)\cdots h(z)$$

This work was sponsored by the U.S. Army under Contract DAAG29-80-C-0041.
*CMS, University of Wisconsin, 610 Walnut St., Madison WI 53705

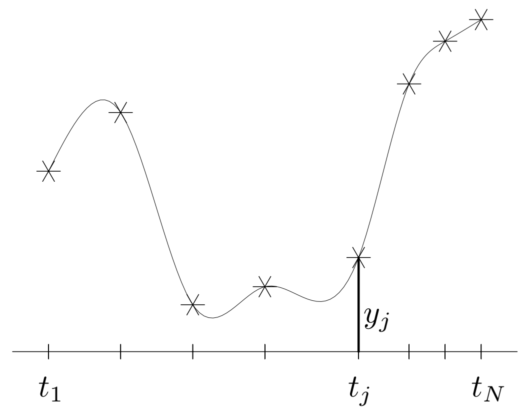

Figure 1. The ‘natural’ cubic spline interpolant. Are the two bottom bumps natural?

in the d variables $x, y, \ldots, z$ by taking (univariate) functions $f, g, \ldots, h$ in these variables from $ and multiplying them. One would take linear combinations of such functions, and the resulting approximation schemes are simply products of univariate schemes. This means that one can even use the univariate computer programs, and the resulting schemes are so efficient that it pays to force one's particular approximation problem into this form if one can do it. It does require that the data come in tensor product form, i.e., on a rectangular grid, and that raises questions. Is the proper multivariate version of an interval a (hyper)rectangle?Also, just how is one to deal with *scattered* data? This made me and others look for other ways of making up multivariate splines.

There are essentially two avenues to splines, the variational and the constructive. Although I have had the mathematical pleasure of writing papers using the variational approach, I am firmly in the constructive camp and so want to begin by doing a job on the variational approach.

The story is familiar since it is available wherever splines are sold, so I can be brief. If I am to fit data points $(t_j, y_j), j = 1, \ldots, N$, I ought to use the "natural" cubic spline interpolant, that is, the function which among all functions fitting the data has the smallest second derivative. This is a good thing, so the story goes, because in this way I am doing more or less what draftsmen have been doing even when they were still draughtsmen. More or less, because they would put a ‘spline’, i.e., a thin flexible rod, through the data, and this rod (if ideal) would take on the shape of that curve γ through the points which minimizes strain energy, i.e., the integral with respect to arclength of the squared curvature. Assuming now that curve γ to be a function, i.e., $\gamma = \{(t, f(t)) : a \le t \le b\}$, the integral being minimized can also be written

$$\int_\gamma \kappa^2 = \int_a^b \frac{\left(D^2 f\right)^2}{\left(1 + (Df)^2\right)^{3/2}},$$

slide2at

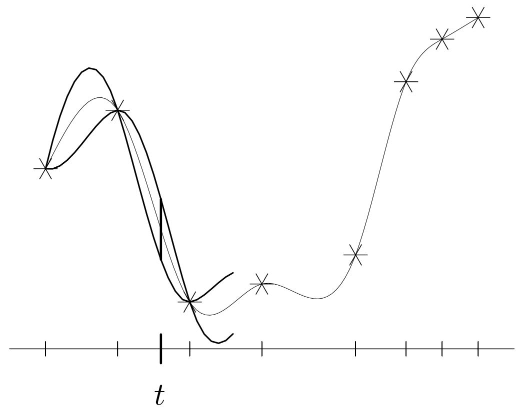

Figure 2. Part of the envelope in whose center the 'natural' cubic spline interpolant happens to lie.

and, for small Df, this is much like the integral

$$\int_a^b \left(D^2 f\right)^2$$

which is being minimized by the 'natural' cubic spline interpolant to the data.

You will discern several false notes in this story. For small Df, there is usually no call for any subtlety at all, a straight line or parabola will fit nicely. In any case, going from a curve to a function is a bit fishy. In fact, if we really believe in the draftsman's spline, then we should reject the cubic spline and compute the draftsman's spline instead. Of course, we will then run into some difficulties. For example, this minimization problem doesn't have a solution without further conditions, such as a bound on the length of the curve. Even with such a condition in place, the draftsman's spline (or *elastica*) is not easy to compute. This leads me to the conclusion that people use cubic splines, not because cubic splines provide them an automatic French curve, but because cubic splines are easy to compute.

There is a more serious variational approach to splines which these days goes under the name of *Optimal Recovery*. Here one starts with the worthwhile observation that, if we know nothing but the data points (t_j, y_j), then we can say nothing about the function between the data points. We need additional information. Suitable information could be a bound on some derivative. For example, to stay with our simple picture, we might also know that the L_2-norm of the second derivative $D^2 f$ is no bigger than some constant c. Then, for each t, the possible values of f at t form an interval, and we obtain in this way an envelope within which our function f must lie. Of course, this envelope depends on c. But, it so happens that, for each t, the midpoint of that interval lies, you guessed it, on our friend the 'natural' cubic spline interpolant, and this is so regardless of c. Thus, the cubic spline interpolant is rather central.

Yet I am not impressed, since all this depends on the decision to give a bound in terms of the L_2-norm and that decision seems arbitrary to me. Had we used, more reasonably to me, a bound on the maximum norm of $D^2 f$, we would again have found an envelope, but now the midpoint changes with c. It does converge, as $c \to \infty$, but not to the cubic spline interpolant, but to the broken line interpolant! Is that sufficient reason to reject the cubic spline in favor of the broken line?

In any case, if you look for the reason why splines occur as solutions to such extremal problems, you will find that it is so because they represent point evaluation with respect to bilinear forms involving some derivative, or, equivalently, they are sections of Green's functions (for D^4 or D^2 or whatever). In a multivariate variational approach, we would expect, correspondingly, to have sections of Green's functions of *partial* differential operators turn up. Such Green's functions are strongly domain dependent, i.e., the resulting multivariate 'splines' change in local detail as the domain of the minimization changes. This made me give up on this approach early on. It has recently been given a strong impetus by Duchon [D76] (see, e.g., [Me79]) who in effect declared that there is only one domain of interest, namely all of $\mathbb{R}^d$, and so created the **thin plate** splines which, for $d = 2$, are used in many places. They provide that interpolant f to given data points $(t_j, y_j), j = 1, \ldots, N$, which minimizes

$$\int_{\mathbb{R}^d} \sum_{i,j=1}^{d} (D_i D_j f)^2,$$

hence the name. But the resulting space of interpolants fails to have a *local* basis, hence the construction of the thin plate spline interpolant takes $O(N^3)$ effort, which is to be compared to the $O(N)$ effort required for the (univariate) spline interpolant.

I hasten to add to this diatribe that I am all for the variational approach in case the smoothness measure being minimized has some *a priori* justification. For example, in planning the path of the arm of a painting robot, one wants the acceleration to be as small as possible, hences its minimization subject to the constraints imposed by the painting job makes very good sense. As another example, we might eventually understand in a mathematical sense just what we mean by a 'good' shape, and it would then be very desirable to look for interpolants of best possible shape. But, given the computational history of the *elastica* or the thin plate spline, we are not likely to compute such a 'best' or 'shapeliest' interpolant exactly. Rather, we are likely to follow the example set by D. Terzopoulos and others and compute such 'splines' only approximately, by minimizing over a suitably flexible, fine-meshed space of piecewise polynomial functions with a local basis.

This brings me to the *constructive* approach to splines. In this approach, a spline is, most simply, a pp (:= piecewise polynomial) function of degree $\le r$ with breakpoint sequence $t = (t_j)$; in symbols:

$$\$ = \pi_{r,t},$$

or, perhaps,

$$\$ = \pi_{r,t}^{\rho} := \pi_{r,t} \cap C^{\rho}.$$

Correspondingly, a d-variate spline would be any element of

$$\pi_{r,\Delta}^{\rho} := \pi_{r,\Delta} \cap C^{\rho},$$

figeuler

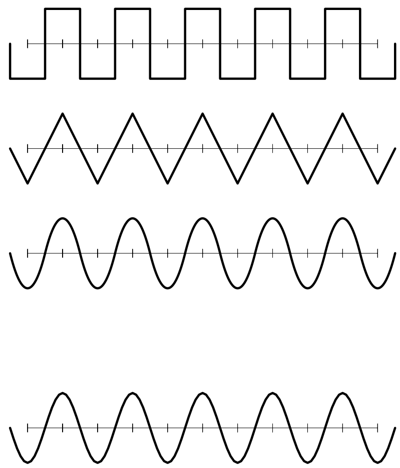

Figure 3. The first three Euler splines and their limit as their degree goes to infinity.

with $\pi_{r,\Delta}$ the collection of all functions which are pp of degree $\le r$ with respect to some partition Δ. If this satisfies you, let me try to convince you that there is more to splines than that.

Already in the very early papers on splines ([E28], [QC38], [S46]), there is much more structure than that. These early papers are concerned with what we now call **cardinal** splines

$$\$ = \pi^{k-2}_{k-1,\mathbb{Z}},$$

i.e., smooth piecewise polynomials with uniformly spaced breakpoints, for example at the integers. Although cardinal spline theory did not quite develop this way, you will find that you can understand cardinal splines most simply if you think of them as smoothed-out step functions, i.e., as obtained from step functions by repeated convolution with the characteristic function

$$M_1 := \chi_{[0,1]}$$

of the unit interval.

For example, that most beautiful of cardinal splines, the **Euler** spline, is obtained in this way. Starting with the (shifted) cardinal step function which is alternately ± 1, a first averaging brings the alternating broken line, while a second averaging (followed by a shift and multiplication by 2) gives the alternating parabolic cardinal spline which is already hard to distinguish (see Figure 3) from the function reached after infinitely many such steps, viz. the cosine. Schoenberg [S73] called this spline function 'Euler spline' since it is made up of Euler polynomials. But it had been put to good use long before that baptism. It had appeared as the solution of various variational problems. For example, it provides [F37] Favard's best constant in the bound on the distance of a function from trigonometric polynomials in terms of that function's k-th derivative. It also occurs [K62] as the slide4t

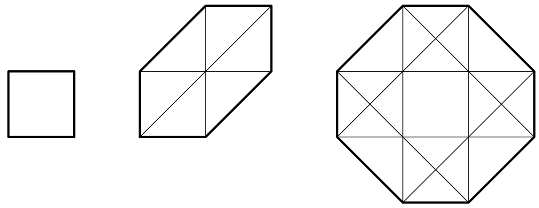

Figure 4. The support of the *bivariate* cardinal B-splines M_1, M_2, M_3.

simultaneous extremizer of the Landau-Kolmogorov inequalities in which the j-th derivative is bounded on $\mathbb{R}$ in terms of the zeroth and the kth.

Schoenberg's fundamental paper [S46] also introduced what became eventually the centerpiece of univariate spline theory, viz. the **B-spline**

$$M_k := \underbrace{M_1 * \cdots * M_1}_{k \text{ times}}.$$

Its integer translates provide a most suitable basis for the cardinal spline space of order k (i.e., of degree $k-1$).

This structure is easily generalized to d variables, as I will now illustrate for $d = 2$. We get bivariate cardinal B-splines by starting with the characteristic function of the unit square,

$$M_1 := \chi_{[0,1]^2}.$$

From it, by averaging, e.g., in the direction $(1,1)$, we obtain

$$M_2(x) \ := \int_{-1}^{0} M_1(x+vt)dt, \qquad v := (1,1),$$

the familiar piecewise linear pyramid function already used by Courant [C43]. By following up with another averaging, this time in the direction $(1,-1)$, we obtain the C^1-quadratic finite element

$$M_3(x) \ := \int_{-1}^{0} M_2(x+wt)dt \qquad w := (1,-1)$$

of Zwart and Powell (e.g., [PS77]).

The space

$$\$:= \{ \sum_{j \in \mathbb{Z}^d} M(\cdot - j)a(j) : a : \mathbb{Z}^d \mapsto \mathbb{R}\}$$

spanned by integer translates of such functions M is rightly thought to be a multivariate cardinal spline space, and there is a complete theory of its approximation power and use available, as developed by mathematicians working in Finite Elements around 1970 and given final form by Strang & Fix [SF73].

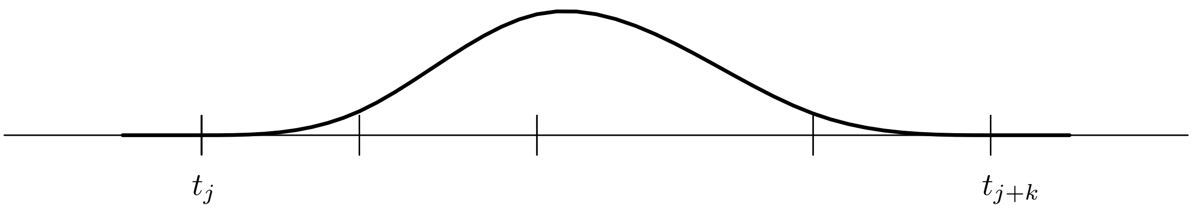

Figure 5. The B-spline $M(\cdot|t_j,\ldots,t_{j+k})$.

If these splines of *uniform* structure do it for you, here is yet another point to quit listening (except that there will be more of this later on). But if you have to deal with scattered data or other nonuniform problems, you know that you need more than cardinal splines.

It was Schoenberg's colleague, the logician H. B. Curry, who pointed out in a review of Schoenberg's '46 paper that, with the aid of divided differences, such B-splines could be constructed for an arbitrary spacing of breakpoints as follows

$$N_j(x) := ((t_{j+k}-t_j)/k)M(x|t_j,\ldots,t_{j+k}) := (t_{j+k}-t_j)[t_j,\ldots,t_{j+k}](\cdot - x)_+^{k-1},$$

and that these points t_i, now called **knots**, could even be repeated to control precisely the smoothness across the knot. In this way, one obtains [CS66] a convenient basis for any space of piecewise polynomials of degree $< k$ and of specified smoothness across breakpoints.

The list of useful properties of the univariate B-spline is quite impressive. Here are some of the items on that list (cf., e.g., [B76] or [Sch81] for details and references).

- N_j depends continuously on its knots $t_j,\ldots,t_{j+k}$.
- N_j has **minimal support**, is **nonnegative**, and $\sum N_j = 1$, i.e., (N_j) provides a good and local **partition of unity**.
- (N_j) provides a **stable basis**, i.e, $d_k^{-1}\|a\|_\infty \le \|\sum N_j a_j\|_\infty \le \|a\|_\infty$ for all coefficient sequences a and some knot-independent (positive) constant d_k.
- **Good quasi-interpolants** are available, of the form $f \sim Qf := \sum N_j \lambda_j f$, with λ_j locally supported, uniformly bounded linear functionals.
- These quasi-interpolants provide **optimal approximation order**, i.e., $\|f - Qf\| \le \mathrm{const}|t|^k\|D^k f\|$ (with $|t| := \sup \Delta t_j$).
- **Shape preserving** approximation schemes are available in the simple form $Vf := \sum N_j f(\tau_j)$.

slide5

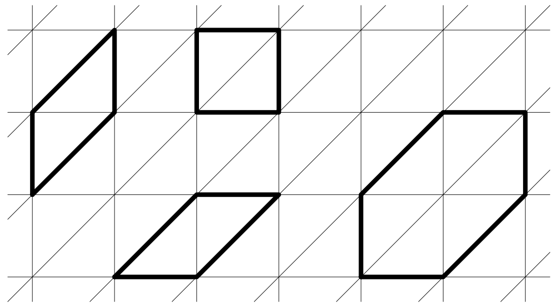

Figure 6. The minimally supported elements in $\pi^0_{2,\Delta}$ (on the left) have to be augmented by elements of far-from-minimal support (such as on the right) in order to obtain a basis for $\pi^0_{2,\Delta}$.

- Determination of the B-spline coefficients of a spline approximation leads to **banded systems**.
- Evaluation of the B-splines can be accomplished by a **stable recurrence**.
- ...

In fact, this list is so impressive that I have come to the conclusion that, in the univariate context, splines are, *by definition*, linear combinations of B-splines.

Once this is accepted, it is obvious what a multivariate spline is; it is a linear combination of multivariate B-splines. All that is now required is the construction of multivariate B-splines. This turned out to be a nontrivial task.

A generalization via divided differences turned out to be difficult since the divided difference $[t_j, \ldots, t_{j+k}]f$ is customarily defined as the leading coefficient of the polynomial of degree $\le k$ which agrees with f at $t_j, \ldots, t_{j+k}$ and this definition becomes doubtful in the multivariate context because interpolating polynomials are only defined for certain pointsets and, even if defined, have several 'leading' coefficients.

While it is possible to develop the univariate B-splines entirely from their recurrence relation, there was no obvious way to extend these to a multivariate context. In fact, when multivariate B-splines were ultimately defined, it took two years of intense effort to find stable recurrence relations for them.

A very tempting approach was via the minimal support property. In this approach, one defines multivariate B-splines to be those functions in a given class of smooth piecewise polynomials whose support is as small as possible. Unfortunately, already very simple examples, such as C^0-parabolics on the 'three-direction mesh' (see Figure 6), show that the resulting functions may not be plentiful enough to staff a basis. (There is an alternative definition of minimal support in terms of the Bernstein-Bézier net for these pp's, but that idea has never been fully explored.)

The approach finally used in [B76] relied on yet another B-spline property, already found in [CS66] where it is shown that

$$M(y|t_j, \ldots, t_{j+k}) = \mathrm{vol}_{k-1}\sigma \cap P^{-1}y,$$

slide6

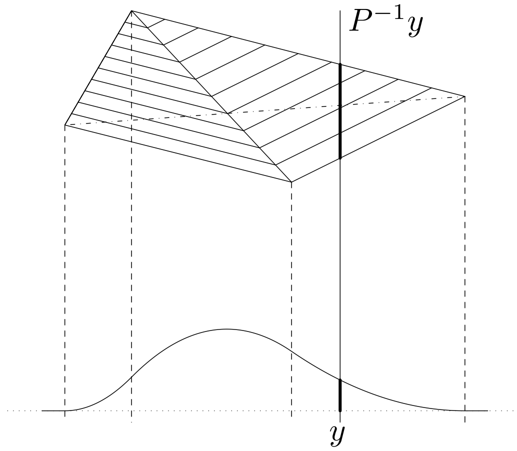

Figure 7. Curry-Schoenberg construction of the (univariate) B-spline as a simplex spline.

with P the canonical projector $P : x \mapsto x(1)$ on $\mathbb{R}^k$ to $\mathbb{R}$ and σ any appropriately scaled simplex $[v_0, \ldots, v_k]$ in $\mathbb{R}^k$ oriented in such a way that $Pv_i = t_{j+i}$, all i. This construction changes only in minor detail when P is taken to be the canonical projector onto $\mathbb{R}^d$ and then provides what is now called the d-variate **simplex** spline

$$M(y|\sigma) \;:=\; \mathrm{vol}_{k-d}\, \sigma \cap P^{-1}y$$

(in order to distinguish it from other multivariate B-splines; see below).

It is obvious that $M(\cdot|\sigma)$ is a compactly supported nonnegative function. It is not hard to see that $M(\cdot|\sigma) \in \pi_{k-d,\Delta}$, with the partition Δ generated by the $(d-1)$-dimensional images under P of faces of σ. With somewhat more effort, one can establish that $M(\cdot|\sigma)$ is as smooth as possible , i.e., that $M(\cdot|\sigma) \in C^{k-d-1}$ in case σ is in general position. It is also easy to construct enough simplex splines to provide a (local) partition of unity: If the (essentially disjoint) simplices σ are so chosen that $\bigcup \sigma \;=\; \mathbb{R}^d \times G$, then $\sum_\sigma M(y|\sigma) = \mathrm{vol}_{k-d} G$ is constant.

But it took some time before Micchelli [M78] came up with stable recurrence relations for these simplex splines. For their proof, Micchelli described the simplex spline equivalently as the distribution which carries the smooth test function φ to the number $\int_\sigma \varphi(Px)dx$. This formulation made it easy to prove [BH82] similar results for the more general multivariate B-spline $M(\cdot|B, P)$ which is defined as the distribution on $\mathbb{R}^d$ which carries the test function φ to the number $\int_B \varphi(Px)dx$, with B, more generally, a (convex) polytope and P, more generally, some linear map on $\mathbb{R}^k$ to $\mathbb{R}^d$.

The most recent summary of material about multivariate B-splines is [H86]; see also [Ch88]. These results show that several of these multivariate B-splines have almost all the properties we listed earlier for the univariate B-spline, i.e., all the properties that we can expect them to have. (For example, we cannot hope for a 'shape-preserving' map to parallel Schoenberg's map V since there is as yet no satisfactory multivariate definition of 'shape preservation'.) Going down our list

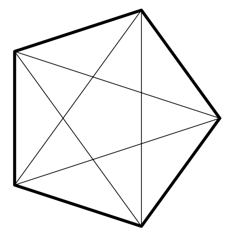

Figure 8. The partition for a simplex spline cannot be made to fit an arbitrary partition.

of good properties, we find stable recurrence relations and good quasi-interpolants. We also find much beautiful mathematics (see, e.g, Dahmen and Micchelli [DM84]), particularly when we use bodies other than simplices, and use projectors other than orthogonal projectors. For example, we obtain the so-called **box splines** [BH83] when we use the unit cube as the body. Such a box spline is, offhand, the distribution defined by

$$\int_{\mathbb{R}^d} M(x|V)\varphi(x)dx := \int_{[0,1]^V} \varphi(\sum_{v\in V} vt_v)dt$$

for some sequence V in $\mathbb{R}^d$, hence can be obtained recursively by

$$M(x|V) := \int_{-1}^{0} M(x+vt|V\backslash v)dt,$$

with $M(\cdot|V)$ the characteristic function of the convex hull of $0 \cup V$ in case $\#V = d$. This shows the multivariate cardinal B-splines introduced earlier to be box splines.

But the initial enthusiasm for these multivariate B-splines has somewhat abated for the simple reason that they are not entitled to the prefix 'B': they fail to be basic. Since they are obtained as shadows of polyhedra (or polytopes), their partition or mesh depends on the structure of those polyhedra. E.g., the 2-dimensional shadow of a simplex has any line connecting any two of the projected vertices as meshlines. This makes it in general impossible to suit multivariate B-splines to a given partition. Even if we restrict attention to partitions generated by such shadows, the collection of all pp functions of the appropriate degree and smoothness is usually larger than the span of all these B-splines. This puts into question the ultimate usefulness of these multivariate B-splines for **practical** work, except, perhaps, for the box splines (if a regular partition is satisfactory).

But it also puts into question that naive definition of a spline as a pp function of some degree and some smoothness on some partition. For this class can often be shown not to have a locally supported basis. I.e., even if the class contains locally supported elements, it also contains functions which cannot be represented by them. On the other hand, these non-local elements are usually not useful for approximation, i.e., it can often be shown that

$$\text{dist}\,(f, \$) \sim \text{dist}\,(f, \$_{\text{loc}}),$$

with

$$\$_{\text{loc}} := \text{span}\{s \in \$: \text{supp}\, s \text{ compact}\}.$$

If this leaves you a bit wondering what multivariate splines might be, I am pleased. For I don't know myself. I am coming to the realization, though, that it will be necessary to separate the various roles the univariate spline plays simultaneously. My guess is that, if there is ultimately a satisfactory definition of a multivariate spline as a tool for approximation, it will capture the best features of the univariate spline, i.e., it will refer to classes of functions probably pp of controllable smoothness which are spanned by a stable, locally supported basis which is not too hard to handle in computations. Such classes will also be used to provide suitable and entirely satisfactory approximations to 'splines' in the variational sense.

References

[B76] C. de Boor, *Splines as linear combinations of B-splines*, in *Approximation Theory II*, G. G. Lorentz, C. K. Chui and L. L. Schumaker, eds., Academic Press, (1976), 1–47.

[BH82] C. de Boor and K. Höllig, *Recurrence relations for multivariate B-splines*, Proc. Amer. Math. Soc. **85** (1982), 397–400.

[BH83] C. de Boor and K. Höllig, *B-splines from parallelepipeds*, J. d'Anal. Math. **42** (1982/83), 99–115.

[Ch88] C. Chui, *Multivariate Splines. Theory and Applications*, CMBS-NSF Lectures, SIAM, Philadelphia PA 1988.

[C43] R. Courant, *Variational methods for the solution of problems of equilibrium and vibrations*, Bull. Amer. Math. Soc. **49** (1943), 1-23.

[CS66] H. B. Curry and I. J. Schoenberg, *Pólya frequency functions.IV, The fundamental spline functions and their limits*, J. d'Anal. Math. **17** (1966), 71-107.

[DM84] W. Dahmen and C.A. Micchelli, *Some results on box splines*, Bull. Amer. Math. Soc. **11** (1984), 147–150.

[D76] J. Duchon, *Interpolation des functions de deux variables suivant le principe de la flexion des plaques minces*, R.A.I.R.O. Analyse Numérique **10** (1976), 5-12.

[E28] A. Eagle, *On the relation between Fourier constants of a periodic function and the coefficients determined by harmonic analysis*, Phil.Mag. **5** (1928), 113–132.

[F37] J. Favard, *Sur les meilleurs procédés d'approximation de certaines classes de fonctions par des polynômes trigonométriques*, Bull. Sci. Math. Sér. 2, **61** (1937), 209–224, 243–256.

[H86] K. Höllig, *Multivariate splines*, in *Approximation Theory*, C. de Boor ed., Amer. Mathem. Soc., Providence RI 1986.

[K62] A. Kolmogorov, *On inequalities between upper bounds of the successive derivatives of an arbitrary function on an infinite interval*, Amer. Math. Soc. Translations, Ser.1,2 (1962), 233-243.

[Me79] J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, J. Appl. Math. Phys. (ZAMP) **30** (1979), 292-304.

[M78] C. A. Micchelli, *A constructive approach to Kergin interpolation in* $\mathbf{R}^k$*: Multivariate B-splines and Lagrange interpolation*, Rocky Mountain J. Math. **10** (1980), 485–497.

[PS77] M. J. Powell and M. A. Sabin, *Piecewise quadratic approximations on triangles*, ACM Trans. on Mathematical Software **3** (1977), 316–325.

[QC38] W. Quade & L. Collatz, *Zur Interpolationstheorie der reellen periodischen Funktionen*, Akad.Wiss., Math.-Phys. Klasse **30** (1938), 383-429.

[S46] I. J. Schoenberg, *Contributions to the problem of approximation of equidistant data by analytic functions, Parts A & B*, Quarterly Appl.Math. **IV** (1946), 45-99, 112-141.

[S73] I. J. Schoenberg, *Cardinal Spline Interpolation*, SIAM, Philadelphia 1973.

[Sch81] L. L. Schumaker, *Spline Functions: Basic Theory*, Wiley, New York 1981.

[SF73] G. Strang and G. Fix, *A Fourier analysis of the finite element variational method*, C.I.M.E., II Ciclo 1971, in *Constructive Aspects of Functional Analysis*, G. Geymonat, ed., (1973), 793–840.