in *Approximation and Function Spaces*
Z. Ciesielski (ed.)
North Holland (Amsterdam), 1981, pp. 163–175

On a max–norm bound for the least–squares spline approximant

Carl de Boor
University of Wisconsin-Madison, MRC, Madison, USA

0. Introduction

Let $\boldsymbol{\xi} = (\xi_i)_1^{\ell+1}$ be a partition of the interval $[a,b]$, i.e.,

$$a = \xi_1, < \cdots < \xi_{\ell+1} = b \quad ,$$

and let

$$S := \mathbb{P}^m_{k,\boldsymbol{\xi}} := \mathbb{P}_{k,\boldsymbol{\xi}} \cap C^{(m-1)}[a,b]$$

denote the collection of piecewise polynomial functions of order k (i.e., of degree $< k$) with (interior) breakpoints $\xi_2, \ldots, \xi_l$ and in $C^{(m-1)}[a,b]$, i.e., satisfying m continuity conditions at each of its interior breakpoints. We are interested in P_S, the orthogonal projector onto S with respect to the ordinary inner product

$$(f,g) := \int_a^b f(x)g(x)dx$$

on $[a,b]$. But, we are interested in P_S as a map on $C[a,b]$ or $\mathbb{L}_\infty[a,b]$. Specifically, we want to bound its norm

$$\|P_S\|_\infty := \sup_f \|P_S f\|_\infty \backslash \|f\|_\infty$$

with respect to the max–norm

$$\|f\|_\infty := \sup_{a \le x \le b} |f(x)| \quad .$$

Conjecture (de Boor [2]): $\sup_{\boldsymbol{\xi};m} \|P_S\|_\infty \le \mathrm{const}_k (< \infty)$.

This conjecture has been verified for $k = 1,2,3$. The case $k = 1$ is, of course, trivial and the case $k = 2$ was first done by Ciesielski [5]. It is the purpose of this talk to survey the current status of this conjecture, to correct a mistake in the verification of the case $k = 3$ in de Boor [1] and to verify the conjecture for $k = 4$.

For $k > 4$, the only results known prove boundedness of $\|P_S\|_\infty$ under some restriction on $\boldsymbol{\xi}$ and/or m. For example,

$$\sup_{\boldsymbol{\xi};m=0} \|P_S\|_\infty \le \mathrm{const}_k$$

is trivial since in this case the $\mathbb{L}_2$–approximation is found locally, on each interval $[\xi_i,\ \xi_{i+1}]$ separately, and so $\|P_S\|_\infty = \|P_{\mathbb{P}_k}\|_\infty$. It is also known (de Boor [4]) that

$$\sup_{\boldsymbol{\xi};m=1} \|P_S\|_\infty \le \mathrm{const}_k$$

but already the case $m = 2$ is open.

B. Mitiagin announced at a meeting at Kent State University in August 1979 that, for even k,

$$\sup_{\boldsymbol{\xi};m=k/2} \|P_S\|_\infty \le \mathrm{const}_k \quad ,$$

but he gave no proof.

Finally, there is the result of Douglas, Dupont and Wahlbin [8] to the effect that

(0.1)
$$\sup_{\Delta\xi_i/\Delta\xi_j \le c;m} \|P_S\|_\infty \le \mathrm{const}_{k,c}$$

in which the bound depends also on the global mesh ratio. This result subsumes Domsta's [7] earlier result for certain **dyadic** partitions $\boldsymbol{\xi}$.

1. A bound in terms of a global mesh ratio

In this section, I outline the proof of a slight strengthening of (0.1) in order to give an indication of some of the arguments that have been used for the general problem.

Experience has shown that it usually pays to express a spline problem, particularly a linear one, in terms of **B-splines** (see, e.g., de Boor [3]). These are spline functions whose support is as small as possible. Let $\boldsymbol{t}$ be a nondecreasing sequence constructed from $\boldsymbol{\xi}$ and m according to the recipe

$$\boldsymbol{t} = \big(\underbrace{a,\ldots,a}_{k},\ \underbrace{\xi_2,\ldots,\xi_2}_{k-m},\ \ldots,\ \underbrace{\xi_1,\ldots,\xi_1}_{k-m},\ \underbrace{b,\ldots,b}_{k}\big) =: (t_i)_1^{n+k} \quad .$$

Then there is a corresponding sequence $(N_{i,k})_1^n$ of elements of S, with $N_{i,k}$ depending on $t_i,\ldots,t_{i+k}$ only, having its support in $[t_i, t_{i+k}]$, and being positive on its support. In addition, these B-splines are normalized to sum to one. Hence

$$\|\sum \alpha_i N_{i,k}\|_\infty \le \|\boldsymbol{\alpha}\|_\infty \quad .$$

More generally, one can show (cf. de Boor [3]) that

$$\|\sum \alpha_i \kappa_i^{1/p} N_{i,k}\|_p \le \|\boldsymbol{\alpha}\|_p, \quad 1 \le p \le \infty \tag{1.1}$$

with

$$\kappa_i := k/(t_{i+k} - t_i)$$

and, in particular,

$$\|\kappa_i N_{i,k}\|_1 = \int \kappa_i N_{i,k} = 1 \quad . \tag{1.2}$$

Now consider $P_S f =: \sum \alpha_j(f) N_{j,k}$. Then

$$\sum_j \int N_{i,k} N_{j,k} \alpha_j(f) = \int N_{i,k} f, \quad \text{all} \quad i \quad .$$

But, since we wish to bound $\boldsymbol{\alpha}(f)$ in terms of $\|f\|_\infty$, we had better use the scale factors κ_i, since $\int \kappa_i N_{i,k} f \le \|f\|_\infty$, by (1.2).
This gives

$$\|\boldsymbol{\alpha}(f)\|_\infty \le \|A^{-1}\|_\infty \|f\|_\infty$$

with

$$A := \Big(\int \kappa_i N_{i,k} N_{j,k}\Big) \quad ;$$

and so

$$\|P_S\|_\infty \le \|A^{-1}\|_\infty \quad .$$

As it turns out, it is quite hard to bound A^{-1} in the max-row-sum norm $\|\cdot\|_\infty$, and one therefore wonders whether we have not replaced our original problem with a harder one. But that is not so. For, one can show (cf. de Boor [3]) that also

$$D_k^{-1} \|\boldsymbol{\alpha}\|_p \le \|\sum \alpha_i \kappa_i^{1/p} N_{i,k}\|_p \tag{1.3}$$

for some positive constant D_k which depends only on k, and this implies that

$$D_k^{-2} \|A^{-1}\|_\infty \le \|P_S\|_\infty \quad . \tag{1.4}$$

Hence, in bounding $\|P_S\|$ in the uniform norm, we are bounding $\|A^{-1}\|_\infty$ whether we want to or not.

Now, the same kind of argument shows that

$$D_k^{-2}\|A_2^{-1}\|_2 \le \|P_S\|_2 \le \|A_2^{-1}\|_2$$

with

$$A_2 := \Big(\int \kappa_i^{1/2} N_{i,k} N_{j,k} \kappa_j^{1/2}\Big) = E^{-1/2} A E^{1/2}$$

and

$$E := \operatorname{diag}\lceil \ldots,\ \kappa_i,\ \ldots \rfloor$$

from which we conclude that

$$\|A_2^{-1}\|_2 \le D_k^2 \quad .$$

If we now had to rely on the standard relationship between the 2–norm and the ∞–norm of a matrix, then the order n of the matrix A would come now in to spoil the bound. But, fortunately, A is $2k$–banded in the sense that $\int N_{i,k}N_{j,k} = 0$ for $|i-j| \ge 2k$. This allows us to make use of Demko's nice observation concerning the exponential decay of the inverse of a banded matrix.

Theorem (Demko [6]). If A is r–banded and $A^{-1} = (b_{ij})$, then there exist $\lambda \in [0,1)$, $K > 0$ depending only on r, $\|A\|$ and $\|A^{-1}\|$ so that

$$|b_{ij}| \le K\lambda^{|i-j|}, \quad \text{all} \quad i,j \quad .$$

Here, $\|A\|$, $\|A^{-1}\|$ are measured in any particular p–norm. But then, the result gives a bound on $\|A^{-1}\|_p$ for all p and dependent only on the numbers $\|A\|$, $\|A^{-1}\|$ and r. In particular, the order of A does not matter. In our case, $\|A_2\|_2 \le 1$ by (1.1), and so we conclude that

$$\|A_2^{-1}\|_\infty \le \mathrm{const}_k$$

for some const_k which depends on D_k. But then, since $A = E^{1/2}A_2E^{-1/2}$, we obtain

$$\|A^{-1}\|_\infty \le \max_{i,j}(\kappa_i/\kappa_j)^{1/2}\mathrm{const}_k$$

and so get de Boor's [4] strengthening

$$\sup_{\kappa_i/\kappa_j \le c; m} \|P_S\|_\infty \le \mathrm{const}_{k,c}$$

of (0.1).

This argument can also be used to give a bound on $\|P_S\|_\infty$ in terms of the **local** mesh ratio $\sup_{|i-j|=1}\kappa_i/\kappa_j =: \varrho$, as long as ϱ is sufficiently close to 1. In addition, as Güssmann [9] has recently pointed out, it gives a bound independent of l for the specific breakpoint sequence

$$\xi_i = \left(\frac{i-1}{l}\right)^\alpha, \quad i = 1, \ldots, l+1$$

for $[a,b] = [0,1]$ and for any $\alpha \ge 1$.

But, this kind of argument has as yet not yielded a bound in terms of an arbitrary local mesh ratio, let alone the conjectured mesh-independent bound. I therefore come now to the mesh-independent results for low order mentioned earlier.

2. Mesh-independent bounds for low order

For $k = 1$, $A = 1$. For $k = 2$, A is tridiagonal and strictly and uniformly row diagonally dominant. Specifically,

$$a_{ii} - |a_{i,i-1}| - |a_{i,i+1}| \geq 1/3, \quad \text{all} \quad i,$$

so that $\|A^{-1}\|_\infty \leq 3$ is immediate.

For $k = 3$, I published a proof mainly in response to a question from Schonefeld, then a student at Purdue University. He had read about Ciesielski's use of splines in the discussion of bases, and wanted to extend that work. Already for this case, A fails to be diagonally dominant, so a different argument has to be used.

The additional ingredient (in de Boor [1]) is the **total positivity** of A. This means that all minors of A are nonnegative. Actually, only very little of this is used, namely that $A^{-1} = (b_{ij})$ is **checkerboard**:

$$(-)^{i+j} b_{ij} \geq 0, \quad \text{all} \quad i, j \quad .$$

This is an immediate consequence of the total positivity of A since, by Cramer's rule

$$b_{ij} = (-)^{i+j} \det A \begin{pmatrix} 1, \ldots, j-1, j+1, \ldots, n \\ 1, \ldots, i-1, i+1, \ldots, n \end{pmatrix} / \det A \quad .$$

But, this checkerboard behavior of A^{-1} can be used to get a bound on $\|A^{-1}\|_\infty$ as follows. Let $\boldsymbol{x}$ be any vector for which $\boldsymbol{y} := A\boldsymbol{x}$ alternates, i.e., $(-)^{i+1} y_i > 0$, all i. Then

$$|x_i| = |\sum_j b_{ij} y_j| = \sum_j |b_{ij}| \, |y_j|$$

hence

$$\|\boldsymbol{x}\|_\infty = \max_i \sum_j |b_{ij}| \, |y_j| \geq \big(\max_i \sum_j |b_{ij}|\big) \min_j |y_j|$$

while $\|A^{-1}\|_\infty = \max_i \sum_j |b_{ij}|$. It follows that

$$\max_{i,j} |x_i / y_j| \geq \|A^{-1}\|_\infty \tag{2.1}$$

with equality iff $\min_j |y_j| = \|\boldsymbol{y}\|_\infty$.

In the case $k = 2$, it is sufficient to take $x_i = (-)^i$, all i.
For then

$$(-)^i y_i = a_{ii} - a_{i,i-1} - a_{i,i+1} \geq 1/3$$

and we get once again $3 \geq \|A^{-1}\|_\infty$.

3. The case $k = 3$

In this case, it is sufficient to take the comparatively simple

$$(-)^j x_j = \big(1 + \frac{(\Delta t_{j+1})^2}{(t_{j+2} - t_j)\,(t_{j+3} - t_{j+1})}\big)/2 \quad .$$

Then $\|x\|_\infty \leq 1$ and

$$y_i = y_i(t_{i-2}, \ldots, t_{i+5})$$

since

$$a_{ij} = \int \kappa_i N_{i,3} N_{j,3} = 0 \quad \text{for} \quad |i - j| \geq 3$$

and x_j depends only on $t_j, \ldots, t_{j+3}$, i.e., on the same knots on which alone $N_{j,k}$ depends. We need to show that

$$\inf_{\boldsymbol{t}} \min_i (-)^i y_i > 0,$$

but progress has already been made since this does not require consideration of a knot sequence $\boldsymbol{t}$ of arbitrary length but only of length 8.

In de Boor [1] I made a mistake in the formula for $a_{i,i-1}$ (and in $a_{i,i+1}$, by symmetry), as was pointed out to me a year after publication by Lois Mansfield. I then corrected that mistake and went through the subsequent estimate to find that the end result, viz.

$$\inf_{\boldsymbol{t}} \min_i (-)^i y_i \geq 1/30 \tag{3.1}$$

remained unaffected. But, having once made such a mistake, how can I now be sure of having a correct argument?

In order to gain further assurance, I went through the following steps.

For general k, the (i,j) entry of A can be computed as

$$\begin{aligned} a_{ij} &= \int \kappa_i N_{i,k} N_{j,k} \\ &= \frac{(-)^k}{\binom{2k-1}{k}} (t_{j+k} - t_j)[t_i, \ldots, t_{i+k}]_x \quad \otimes [t_j, \ldots, t_{j+k}]_y (x-y)_+^{2k-1} \\ &= c_k (t_{j+k} - t_j) \sum_{r=i}^{i+k} \sum_{s=j}^{j+k} \frac{(t_r - t_s)_+^{2k-1}}{\prod_{\substack{\varrho=i \\ \varrho \neq r}}^{i+k} (t_r - t_\varrho) \prod_{\substack{\sigma=i \\ \sigma \neq s}}^{j+k} (t_s - t_\sigma)} \quad . \end{aligned} \tag{3.2}$$

Here

$$c_k := (-)^k / \binom{2k-1}{k}$$

and $[t_i, \ldots, t_{i+k}]_x f(x,y)$ indicates the operation of taking the k–th divided difference at the points $t_i, \ldots, t_{i+k}$ of the bivariate function f as a function of x for each fixed y, thus producing a function of y. Further, since

$$[t_i, \ldots, t_{i+k}]_x \otimes [t_j, \ldots, t_{j+k}]_y (x-y)^{2k-1} = 0$$

while

$$(x-y)_+^{2k-1} - (y-x)_+^{2k-1} = (x-y)^{2k-1},$$

the result in (3.2) will be the same whether the divided difference is taken of $(x-y)_+^{2k-1}$ or of $(y-x)_+^{2k-1}$. But, when $i > j$, then use of $(y-x)_+^{2k-1}$ will generate fewer nonzero summands in the double sum in (3.2).

With this, we now consider the specific expression

$$y = y(t_0, \ldots, t_7) = \sum_{j=0}^{4} x_j a_{2j} \quad .$$

It is our goal to bound this expression from below in terms of t_2, t_3, t_4, and t_5 only. Since we can assume, after a suitable translation and scaling, that, e.g., $t_3 = 0$, $t_4 = 1$, this would leave a problem with just two parameters.

For this, we first consider the term $x_0 a_{20}$. We have

$$a_{20} = c_3 (30) \frac{(32)^5}{(30)(31)(32) \cdot (23)(24)(25)} = -c_3 \frac{(32)^3}{(31)(42)(52)} \quad .$$

Here and below, we use the abbreviation

$$(ij) := t_i - t_j \quad .$$

In these terms, $x_0 = \frac{1}{2}(1 + (21)^2/[(20)(31)]) \geq 1/2$, hence

$$20x_0a_{20} \geq \frac{(32)^3}{(31)(42)(52)} \quad ,$$

using the fact that

$$c_3 = -1/10 \quad .$$

This lower bound for x_0a_{20} still involves t_1, but we will get rid of it in a moment, after combining this term with x_1a_{21}.

We have

$$x_1 = -(1+\beta)/2 \quad \text{with} \quad \beta := (32)^2/[(42)(31)] \quad .$$

Also,

$$\begin{aligned} 10a_{21} &= -(41)\Big\{\frac{(32)^5}{(12.\,4)\cdot(.\,345)} + \frac{(42)^5}{(123.\,)\cdot(.\,345)} + \frac{(43)^5}{(123.\,)\cdot(2.\,45)}\Big\} \\ &= -\frac{(32)}{(52)}\beta - \frac{(32)^3}{(42)(43)(52)} + \frac{(42)^3}{(43)(32)(52)} - \frac{(43)^3}{(42)(32)(53)} \quad . \end{aligned}$$

Here, I have used further abbreviations, such as

$$(12.\,4) := (31)(32)(34) \quad .$$

Thus

$$20(x_0a_{20} + x_1a_{21}) \geq \frac{(32)}{(52)}\beta - (1+\beta)\Big\{-\frac{(32)}{(52)}\beta + C\Big\} \tag{3.3}$$

with C independent of t_1, while β increases with t_1. The right side of (3.3) is convex in β, hence has a unique minimum which Calculus identifies as the point $\beta^{\min} := \frac{1}{2}C/\frac{(32)}{(52)} - 1$. But this number is bigger than the largest value which β can take, given that $t_1 \leq t_2$, viz., the value $\beta|_{t_1=t_2} = (32)/(42)$. Hence

$$x_0a_{20} + x_1a_{21} \geq (x_0a_{20} + x_1a_{21})\big|_{\substack{t_0=-\infty \\ t_1=t_2}} \quad .$$

Using symmetry, we conclude that

$$y(t_0, \ldots, t_7) \geq y(-\infty, t_2, t_2, t_3, t_5, t_5, \infty) =: \widetilde{y}(t_2, t_3, t_4, t_5) \quad . \tag{3.4}$$

The various sums of products of terms of the form $(ij)/(pq)$ which make up $\widetilde{y}$ have the common denominator

$$D := (42)(53) \cdot \prod_{2 \leq i < i \leq 5} (ji) \quad . \tag{3.5}$$

With this,

$$\begin{aligned} 20D\widetilde{y} = {} & (42)(53)\cdot(32)(52)(43)(54)\cdot(32)^2 - \\ & - (53)\cdot(54)\cdot[(42)+(32)]\{-(32)^3(42)(53) + \\ & + (42)^4(53) - (43)^4(52)\} + \\ & + [(42)(53) + (43)^2](52)\{(32)^4(54) - (42)^4(53) + (52)^4(43) + \\ & + (43)^4(52) - (53)^4(42) + (54)^4(32)\} + \\ & + \text{two more terms obtainable by symmetry} \quad . \end{aligned} \tag{3.6}$$

Now, finally, observe that $\widetilde{y}$ is invariant under linear changes in its variables. In particular, under the linear substitution

$$t_2 = -a, \quad t_3 = 0, \quad t_4 = 1, \quad t_5 = c \tag{3.7}$$

$\widetilde{y}$ goes over into a rational function of just two variables

$$\widetilde{\widetilde{y}}(a, c) := \widetilde{y}(-a, 0, 1, c)$$

whose minimum we are to determine as $\Delta t_2 = a$ and $\Delta t_4 = c$ vary over the nonnegative quadrant.

For this, I wrote a computer program which would generate symbolically D and $20D\widetilde{y}$ as polynomials in a and c from the information (3.5)–(3.7). This produced the coefficient tables

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	3	3	1
2	0	3	8	7	2
3	0	3	7	5	1
4	0	1	2	1	0

	0	1	2	3	4
0	0	0	0	0	0
1	0	2	5	4	1
2	0	5	8	4	2
3	0	4	4	2	1
4	0	1	2	1	0

for D and $20D\widetilde{y}$, respectively, from which it is evident that

$$20D\widetilde{y}/D \geq 2/3$$

for a, $c \geq 0$. In fact, this lower bound could be improved just slightly. In any event, this proves (3.1) once again.

4. The case $k = 4$

In this, the cubic case, I found by numerical experiment that the comparatively simple choice

$$(-)^j x_j = (3 + 4\frac{(t_{j+3} - t_{j+1})^2}{(t_{j+3} - t_j)(t_{j+4} - t_{j+1})})/7, \quad \text{all} \quad j,$$

works, giving

$$\inf_{\boldsymbol{t}} \min_i (-)^i \sum_j x_j a_{ij} \geq 3/245$$

for this case. The methodical verification of this lower bound along the lines just given for the parabolic case reduces the problem to one of minimizing a rational function of just three variables over the nonnegative orthant. The details of this extended calculation will be given elsewhere.

References

[1] C. de Boor, *On the convergence of odd-degree spline interpolation*, J. Approx. Theory, 1 (1968), pp. 452–463.

[2] C. de Boor, *The quasi-interpolant as a tool in elementary polynomial spline theory*, Approximation Theory (G. G. Lorentz *et al.*, eds), Academic Press, New York, 1973, pp. 269–276.

[3] C. de Boor, *Splines as linear combinations of B–splines, a survey*, Approximation Theory, II (G. G. Lorentz, C. K. Chui, and L. L. Schumaker, eds), Academic Press, New York, 1976, pp. 1–47.

[4] C. de Boor, *A bound on the L_∞-norm of L_2-approximation by splines in terms of a global mesh ratio*, Math. Comp., 30(136) (1976), pp. 765–771.

[5] A. Ciesielski, *Properties of the orthonormal Franklin system*, Studia Math., 23 (1963), pp. 141–157.

[6] Stephen Demko, *Inverses of band matrices and local convergence of spline projections*, SIAM J. Numer. Anal., 14 (1977), pp. 616–619.

[7] J. Domsta, *A theorem on B-splines*, Studia Math., 41 (1972), pp. 291–314.

[8] Jim Douglas Jr., Todd Dupont, and Lars Wahlbin, *Optimal L_∞ error estimates for Galerkin approximations to solutions of two-point boundary value problems*, Math. Comp., 29(130) (1975), pp. 475–483.

[9] B. Güssmann, *L_∞-bounds of L_2-projections on splines*, Quantitative Approximation (R. DeVore and K. Scherer, eds), Academic Press, New York, 1980, pp. 20–24.