Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Approximation Theory 121 (2003) 229–243
JOURNAL OF Approximation Theory
http://www.elsevier.com/locate/jat

Perfect spline approximation ☆

Y.-K. Hu[a,*,1] and X.M. Yu[b]

[a] *Department of Mathematics and Computer Science, Georgia Southern University, Statesboro, GA 30460-8093, USA*

[b] *Department of Mathematics, Southwest Missouri State University, Springfield, MO 65804, USA*

Received 26 November 2001; revised 6 October 2002; accepted 22 November 2002
Communicated by Amos Ron

Abstract

Our study of perfect spline approximation reveals: (i) it is closely related to $\Sigma\Delta$ modulation used in one-bit quantization of bandlimited signals. In fact, they share the same recursive formulae, although in different contexts; (ii) the best rate of approximation by perfect splines of order r with equidistant knots of mesh size h is h^{r-1}. This rate is optimal in the sense that a function can be approximated with a better rate if and only if it is a polynomial of degree $<r$.

The uniqueness of best approximation is studied, too. Along the way, we also give a result on an extremal problem, that is, among all perfect splines with integer knots on $\mathbb{R}$, (multiples of) Euler splines have the smallest possible norms.

MSC: 41A15; 41A29; 41A30; 41A50; 41A52; 49M25; 93C55

Keywords: Perfect splines; Spline approximation; Sigma–Delta modulation

1. Introduction

In the recent years, the research on one-bit quantization of bandlimited signals, in particular on the so-called $\Sigma\Delta$ modulation, has been active, see [2,6,7,10–12,15,16] and the references therein. The notion of perfect spline approximation is motivated

☆ This work began when the second author was doing her sabbatical with Professor R.A. DeVore at the University of South Carolina. A major part of the work was done when the first author was having his sabbatical with Professor DeVore.

*Corresponding author.

E-mail addresses: yhu@gasou.edu (Y.-K. Hu), xmy944f@smsu.edu (X.M. Yu).

[1] Partially supported by the Faculty Research Committee, Georgia Southern University.

0021-9045/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0021-9045(02)00062-X

by one-bit quantization. We will show later in this paper that they even have exactly the same recursive formulae, although in quite different contexts.

We begin with the definition of perfect splines. Given $h>0$, let $x_i := ih$, and $\mathbf{T}_h := \{x_i\}_i$. Denote by $\mathscr{S}_r = \mathscr{S}_r^h = \mathscr{S}_r(\mathbf{T}_h, A)$ the space of all splines of order $r>0$ on the knot sequence $\mathbf{T}_h$ on a finite or an infinite interval A. It is well known that the $(r-1)$st derivative of any spline of order r is piecewise constant. The splines whose $(r-1)$st derivative can only be either 1 or -1 on each subinterval $(x_i, \ x_{i+1})$ are called *perfect splines*. In general, all splines the absolute value of whose $(r-1)$st derivative is a constant $M>0$ are also called perfect splines. We denote the set of all such splines on $\mathbf{T}_h$ by

$$\mathscr{P}_r^M = \mathscr{P}_r^M(\mathbf{T}_h, A) := \{S \in \mathscr{S}_r(\mathbf{T}_h, A) \colon S^{(r-1)}(x) = \pm M, \ x \neq x_i\}, \quad M>0. \tag{1.1}$$

Note that $\mathscr{P}_r^M$ is not a linear space at all. It contains all polynomials of degree *exactly* $r-1$ whose leading coefficient is $M/(r-1)!$, but not any polynomials of degree $<r-1$. In particular, $0 \notin \mathscr{P}_r^M$. Euler splines are a special case of perfect splines with $h=1$, see Section 3 for more information. We will approximate functions $f \in \mathbf{C}(A)$ with $\omega_{r-1}(f,h) < \infty$ by perfect splines. Here and throughout the paper, ω_k is the usual kth modulus of smoothness of f, with $\omega_0(f,t)$ understood as $\|f\|$, and $\|\cdot\| = \|\cdot\|_{\mathbf{C}(A)}$ denotes the uniform norm on the interval A. If for a function $f \in \mathbf{C}(A)$, $\omega_k(f,t_0) < \infty$ for some $t_0>0$, then $\omega_k(f,t)<\infty$ for any $t>0$. From now on we shall simply say $\omega_k(f,1)<\infty$. The following is our main theorem of this paper.

Theorem 1. *Let $r \geqslant 2$ and $0<\alpha<1$ (called quantization parameter, see [2,7]) be given, and let $f \in \mathbf{C}(A)$ with $\omega_{r-1}(f,1)<\infty$. Then*

$$E(f, \mathscr{P}_r^M) := \inf_{S \in \mathscr{P}_r^M} \|f - S\| \leqslant \begin{cases} C\omega_{r-1}(f,h) & \text{if } \omega_{r-1}(f,h)>0, \\ CMh^{r-1} & \text{if } \omega_{r-1}(f,h)=0, \end{cases} \tag{1.2}$$

where C is a constant depending only on r and α. The size M of the $(r-1)$st derivative $S^{(r-1)}$ for any $S \in \mathscr{P}_r^M$ is chosen as a multiple of $h^{-r+1}\omega_{r-1}(f,h)$ in the c ase of $\omega_{r-1}(f,h)>0$ (see (2.7) below), and can be freely chosen in the case of $\omega_{r-1}(f,h)=0$.

Remark. In the second case of (1.2) the value of M can be *freely* chosen, as small as one wishes, thus the error can be as close to zero as one wishes. The only reason this error can not be zero is the requirement of $M>0$ in definition (1.1), which excludes polynomials of degree $<r-1$ from $\mathscr{P}_r^M$. Later we will show the inequality in the second case of (1.2) can be replaced by an *equality* with a specific value of C if $A = \mathbb{R}$, see the paragraph after Corollary 6.

Theorem 1 can be rewritten in a better-looking but less accurate form by adding the two terms in (1.2) together and replacing M by ε as follows.

Corollary 2. *Under the conditions of Theorem* 1, *we have*

$$\inf_{S\in\mathscr{P}_r^M} \|f - S\| \leqslant C(\omega_{r-1}(f,h) + \varepsilon h^{r-1}) \tag{1.3}$$

with properly chosen M *and any* $\varepsilon > 0$. *If* $f \in \mathbf{W}_\infty^{r-1}(A)$, *then*

$$\inf_{S\in\mathscr{P}_r^M} \|f - S\| \leqslant Ch^{r-1}(\varepsilon + \|f^{(r-1)}\|). \tag{1.4}$$

Theorem 1 will be proved in Section 2. In Section 3 we will show the rate in (1.2) is optimal. The uniqueness of best approximation will also be studied there.

2. Perfect spline approximation

We first introduce some properties of splines. The reader can find details in any book on splines, such as [3,4,13,14]. If $A = \mathbb{R}$, any $S \in \mathscr{S}_r$ can be written as a B-spline series

$$S(x) = \sum_{i=-\infty}^{\infty} c_i N_{ir}(x), \tag{2.1}$$

where

$$N_{ir}(x) := N(x; x_i, \ldots, x_{i+r}) := (x_{i+r} - x_i)[x_i, \ldots, x_{i+r}](\cdot - x)_+^{r-1}.$$

The derivative of S can be easily written in terms of lower order B-splines:

$$S' = (r-1) \sum_{i=-\infty}^{\infty} \frac{c_i - c_{i-1}}{x_{i+r-1} - x_i} N_{i,r-1} = \sum_{i=-\infty}^{\infty} \frac{\Delta c_i}{h} N_{i,r-1}$$

and in general,

$$S^{(j)} = \sum_{i=-\infty}^{\infty} \frac{\Delta^j c_i}{h^j} N_{i,r-j}, \quad j = 0, 1, \ldots, r-1, \tag{2.2}$$

where Δ is the difference operator defined by $\Delta c_i := c_i - c_{i-1}$, and Δ^j is the jth power of Δ. If A is a finite interval, we assume, without loss of generality, $A = [0, 1]$. Let n be the largest integer such that $nh < 1$, i.e., $n := \lceil h^{-1} \rceil - 1$, then (2.1) and (2.3) become

$$S = \sum_{i=1-r}^{n} c_i N_{ir}(x) \quad \text{and} \quad S^{(j)} = \sum_{i=1-r+j}^{n} \frac{\Delta^j c_i}{h^j} N_{i,r-j}. \tag{2.3}$$

Sometimes one considers the half-line $A = [0, \infty)$, on which the above become

$$S = \sum_{i=1-r}^{\infty} c_i N_{ir}(x) \quad \text{and} \quad S^{(j)} = \sum_{i=1-r+j}^{\infty} \frac{\Delta^j c_i}{h^j} N_{i,r-j}. \tag{2.4}$$

The coefficients of $S^{(j)}$ for different values of j are closely related. To reveal the relationship among them, we define auxiliary quantities

$$c_i^{(j)} := h^{-r+1} \Delta^j c_i, \quad j = 0, \ldots, r-1. \tag{2.5}$$

In particular, $c_i^{(r-1)} := h^{-r+1} \Delta^{r-1} c_i$ are the coefficients of $S^{(r-1)}$, which is a piecewise constant function. We can now rewrite

$$c_i^{(j)} = h^{-r+1} \Delta^j c_i = h^{-r+1} \Delta(\Delta^{j-1} c_i) = h^{-r+1} (\Delta^{j-1} c_i - \Delta^{j-1} c_{i-1}) = c_i^{(j-1)} - c_{i-1}^{(j-1)}$$

as

$$c_i^{(j-1)} = c_{i-1}^{(j-1)} + c_i^{(j)}, \quad c_{i-1}^{(j-1)} = c_i^{(j-1)} - c_i^{(j)}, \quad j = r-1, \ldots, 1. \tag{2.6}$$

This means if we know $c_{i_0}^{(j)}$ for some i_0 and $j = 0, \ldots, r-2$, and $c_i^{(r-1)}$ for all i, then we can calculate $c_i = h^{r-1} c_i^{(0)}$ for all i. In another word, if we know coefficients for $S, S', \ldots, S^{(r-2)}$ at some knot i_0, and know all coefficients $c_i^{(r-1)}$ of $S^{(r-1)}$, then we can recover S through recursive addition or subtraction by (2.6). This is a discrete analogue to the fact that any f in the Sobolev space $\mathbf{W}_1^{r-1}$ can be recovered from initial values $f^{(j)}(a)$ for some a and $j = 0, \ldots, r-2$, and $f^{(r-1)}(x)$ for all x through integration.

Let $f \in \mathbf{C}(A)$ with $\omega_{r-1}(f, 1) < \infty$. Then there exists a spline $G \in \mathscr{S}_r$ such that

$$\|f - G\| \leqslant C_1 \omega_r(f, h) \leqslant 2C_1 \omega_{r-1}(f, h),$$

therefore

$$\omega_{r-1}(G, h) \leqslant \omega_{r-1}(G - f, h) + \omega_{r-1}(f, h) \leqslant C_2 \omega_{r-1}(f, h) < \infty.$$

Hu and Yu [8] proved that $h^{r-1}\|G^{(r-1)}\|$ is equivalent to $\omega_{r-1}(G, h)$ with the equivalence constants depending only on r, that is,

$$\|G^{(r-1)}\| \leqslant C_3 h^{-r+1} \omega_{r-1}(G, h) \leqslant C_4 h^{-r+1} \omega_{r-1}(f, h) < \infty.$$

Let

$$M := \begin{cases} C_4 \alpha^{-1} h^{-r+1} \omega_{r-1}(f, h) & \text{if } \omega_{r-1}(f, h) > 0, \\ \text{any positive number} & \text{if } \omega_{r-1}(f, h) = 0. \end{cases} \tag{2.7}$$

We point out that if $\omega_{r-1}(f, h) > 0$ this M depends on h, and is bounded as $h \to 0$ only for $f \in \mathrm{Lip}^*(r-1, \mathbf{C})$; if $\omega_{r-1}(f, h) = 0$, M can be chosen arbitrarily small as long as it is positive, see the remark after Theorem 1. Define $g := G/M$, then

$$\|g^{(r-1)}\| = \frac{1}{M} \|G^{(r-1)}\| \leqslant \alpha. \tag{2.8}$$

Here we observe that (2.8) is still valid even if $\omega_{r-1}(f,h)=0$, since in this case $f=G$ and $f^{(r-1)}=G^{(r-1)}=0$. The key to prove Theorem 1 is the following theorem, which is significant by itself.

Theorem 3. *For any* $g\in\mathscr{S}_r(\mathbf{T}_h,A)$ *with* $||g^{(r-1)}||\leqslant\alpha$, $0<\alpha<1$, *there exists a perfect spline* $s\in\mathscr{P}_r^1$ *defined in* (1.1), *that is,* $s\in\mathscr{S}_r$ *with* $s^{(r-1)}(x)=\pm1$, *such that*

$$||g-s||\leqslant Ch^{r-1}=C(r,\alpha)h^{r-1}. \tag{2.9}$$

We write g, s, and $g-s$ as B-spline series:

$$g(x)=\sum_i a_iN_{ir}(x),\quad s(x)=\sum_i b_iN_{ir}(x), \tag{2.10}$$

$$g(x)-s(x)=\sum_i (a_i-b_i)N_{ir}(x)=:\sum_i u_iN_{ir}(x), \tag{2.11}$$

where the index i runs over the same range as in (2.1), (2.3), or (2.4), for $A=\mathbb{R}$, $[0,1]$, or $[0,\infty)$, respectively. Similar to (2.5), we define

$$a_i^{(j)}:=h^{-r+1}\Delta^j a_i,\quad b_i^{(j)}:=h^{-r+1}\Delta^j b_i,\quad u_i^{(j)}:=h^{-r+1}\Delta^j u_i,\qquad j=0,\dots,r-1. \tag{2.12}$$

Given $\{a_i\}_i$, we choose $b_0^{(j)}=a_0^{(j)}$, $j=0,\dots,r-2$, then

$$u_0^{(j)}=a_0^{(j)}-b_0^{(j)}=0,\quad j=0,\dots,r-2. \tag{2.13}$$

If we have schemes to determine $b_i^{(r-1)}$ for one i at a time, then $u_i^{(r-1)}=a_i^{(r-1)}-b_i^{(r-1)}$, and $u_i^{(0)}$ can be calculated by applying (2.16) to $u_i^{(j)}$:

$$\begin{array}{rclcrcl} u_i^{(r-2)} &=& u_{i-1}^{(r-2)}+u_i^{(r-1)}, && u_{i-1}^{(r-2)} &=& u_i^{(r-2)}-u_i^{(r-1)},\\ u_i^{(r-3)} &=& u_{i-1}^{(r-3)}+u_i^{(r-2)}, && u_{i-1}^{(r-3)} &=& u_i^{(r-3)}-u_i^{(r-2)},\\ &\vdots& && &\vdots& \\ u_i^{(0)} &=& u_{i-1}^{(0)}+u_i^{(1)}, && u_{i-1}^{(0)} &=& u_i^{(0)}-u_i^{(1)}, \end{array} \tag{2.14}$$

which in turn leads to u_i and then b_i. The schemes have to be such that the resulting $\{u_i\}$ is bounded by a constant depending only on r and α, which will guarantee (2.9). Such schemes are said *stable*. The reason we do not directly apply (2.6) to $b_i^{(j)}$ to find all b_i but calculate u_i first instead is that in all the schemes to be introduced, $b_i^{(r-1)}$ will depend on $u_{i-1}^{(j)}$, $j=0,\dots,r-2$, which give information on the errors in the function value and the derivatives in the previous step.

The first column of (2.14) turns out to be the same as the discrete dynamical system that arises from $\Sigma\Delta$ modulation, although in different notation and different context, see [2,7,15,14], etc. As mentioned at the beginning of this paper, our interest in perfect spline approximation was motivated by $\Sigma\Delta$ modulation using oversampled one-bit quantization, but we did not expect the two areas so closely related with each other that they share the same recursive relationship among their "state variables".

For this reason, a stable scheme in $\Sigma\Delta$ modulation can be directly used in perfect spline approximation without change, and vice versa.

Remark. One can readily implement a one-bit quantization scheme using perfect splines. This can be done by first using a quasi-interpolant operator to approximate the signal by a spline, which will be our G, and then obtaining a perfect spline approximation Ms to this G (recall that $g = G/M$ and s is an approximation to g) by an encoding/decoding scheme described here. The advantage is that the signal can be recovered by using highly efficient, widely available B-spline evaluation algorithms, see Chapter 5 of Schumaker's book [13], and Lyche and Schumaker [9].

In the following we will introduce some versions of schemes used in $\Sigma\Delta$ modulation.

Scheme for $r = 2$: This is a well known scheme. The forward part (on the left-hand side) of (2.14) becomes

$$u_i^{(0)} = u_{i-1}^{(0)} + u_i^{(1)} = u_{i-1}^{(0)} + a_i^{(1)} - b_i^{(1)}, \quad i = 1, 2, 3, \ldots .$$

We choose $b_i^{(1)}$ to minimize $u_i^{(0)}$:

$$b_i^{(1)} := \mathrm{Sign}(u_{i-1}^{(0)} + a_i^{(1)}), \tag{2.15}$$

where

$$\mathrm{Sign}(x) := \begin{cases} 1 & \text{if } x \geqslant 0, \\ -1 & \text{otherwise.} \end{cases}$$

It is trivial to verify that if $|a_i^{(1)}| \leqslant 1$ and $|u_{i-1}^{(0)}| \leqslant 1$, then $|u_i^{(0)}| \leqslant 1$, thus $|u_i^{(0)}| \leqslant 1$ for all i by induction.

If $A = \mathbb{R}$, one needs the backward part (the second column) of (2.14), which takes the form:

$$u_{i-1}^{(0)} = u_i^{(0)} - u_i^{(1)} = u_i^{(0)} - a_i^{(1)} + b_i^{(1)}.$$

Then it is natural to modify the forward scheme (2.15) to

$$b_i^{(1)} := -\mathrm{Sign}(u_i^{(0)} - a_i^{(1)}). \tag{2.16}$$

The boundedness of $u_i^{(0)}$ can be verified the same way. We shall only discuss the forward version of each scheme for $r > 2$. The interested reader can work out a backward version for each of the schemes listed below in a way similar to (2.16). We should point out that we only discuss schemes in which $b_i^{(r-1)}$ depends only on $r-1$ state variables $u_{i-1}^{(j)}$, $j = 0, \ldots, r-2$, (and the input $a_i^{(r-1)}$). It is easy to construct a backward version for this class of schemes. But as the referee of this paper warns, it is not always straightforward (if possible at all) to do so out of a given general forward scheme. There are schemes that use more (maybe infinitely many) past state variables, and there is the question about whether a good initial condition for a forward scheme is also good for the corresponding backward scheme if exists. But these issues are far beyond the scope of this paper.

Scheme A *for* $r = 3$: The forward part of (2.14) becomes

$$u_i^{(1)} = u_{i-1}^{(1)} + u_i^{(2)} = u_{i-1}^{(1)} + a_i^{(2)} - b_i^{(2)},$$

$$u_i^{(0)} = u_{i-1}^{(0)} + u_i^{(1)} = u_{i-1}^{(0)} + u_{i-1}^{(1)} + a_i^{(2)} - b_i^{(2)}.$$

The idea of minimizing $|u_i^{(0)}|$ as in (2.15) by using

$$b_i^{(2)} := \mathrm{Sign}(u_{i-1}^{(0)} + u_{i-1}^{(1)} + a_i^{(2)})$$

does not work. In a numerical experiment, after we hand-picked about 10 values for $a_i^{(2)}$, $u_i^{(0)}$ began to oscillate with increasing amplitude. It turns out that minimizing

$$|u_i^{(0)} + u_i^{(1)}| = |u_{i-1}^{(0)} + 2u_{i-1}^{(1)} + 2a_i^{(2)} - 2b_i^{(2)}|$$

by choosing

$$b_i^{(2)} := \mathrm{Sign}(u_{i-1}^{(0)} + 2u_{i-1}^{(1)} + 2a_i^{(2)}) \tag{2.17}$$

is a much better idea. Daubechies and DeVore [2] showed that $\{u_i^{(0)}\}$ calculated by (2.17) is bounded if α is sufficiently small. In a numerical experiment, we used for $a_i^{(2)}$ in (2.17) random numbers from $[-0.5, 0.5]$, and found

$$|u_i^{(0)}| < 2.1, \quad |u_i^{(1)}| < 1.9, \quad |u_i^{(0)} + u_i^{(1)}| < 2.7 \quad \text{for } 0 \leqslant i \leqslant 10^8.$$

Scheme B *for* $r = 3$: (Daubechies and DeVore [2])

$$b_i^{(2)} := \mathrm{Sign}(u_{i-1}^{(1)} + M_1\, \mathrm{Sign}(u_{i-1}^{(0)})), \tag{2.18}$$

that is,

$$b_i^{(2)} := \begin{cases} \mathrm{Sign}(u_{i-1}^{(1)}) & \text{if } |u_{i-1}^{(1)}| > M_1, \\ \mathrm{Sign}(u_{i-1}^{(0)}) & \text{otherwise,} \end{cases}$$

where $M_1 := 2(1 + \alpha)$.

Remark. Özgür Yilmaz proved in a recent paper [16] the stability of a very general scheme for $r = 3$. While there are ad hoc schemes for $r \leqslant 6$ in electric engineering practice, the following scheme by Daubechies and DeVore [2] is the very first stability result for any r greater than 3 by our best knowledge.

Scheme for arbitrary order $r \geqslant 3$: One can generalize (2.18) to

$$\begin{aligned} b_i^{(r-1)} := \mathrm{Sign}(u_{i-1}^{(r-2)} &+ M_1\, \mathrm{Sign}(u_{i-1}^{(r-3)} + \cdots M_{r-3}\, \mathrm{Sign}(u_{i-1}^{(1)} \\ &+ M_{r-2}\, \mathrm{Sign}(u_{i-1}^{(0)}))\cdots)), \end{aligned} \tag{2.19}$$

where M_j, $j = 1, \ldots, r-2$, are constants depending on r and α.

Lemma 4 (Daubechies and DeVore [2]). *Let* $r \geqslant 3$,

$$K_1 := \left\lceil \frac{5(1+\alpha)}{1-\alpha} \right\rceil + 2 \quad \textit{and} \quad M_1 := 2(1+\alpha).$$

Then $\{u_i^{(0)}\}$ *generated by* (2.19) *is bounded*:

$$|u_i^{(0)}| \leqslant \frac{1}{2}(3K_1)^{r-2} 4^{(r-2)(r-3)} M_1. \tag{2.20}$$

We are now ready to prove Theorem 3.

Proof of Theorem 3. We only show the case $A = [0, \infty)$. It is similar for $A = \mathbb{R}$ or $[0, 1]$. Given $g(x) = \sum_{i=0}^{\infty} a_i N_{ir}(x)$ with $|a_i^{(r-1)}| \leqslant \alpha < 1$, we choose $b_0^{(j)} = a_0^{(j)}$, $j = 0, \dots, r-2$, and choose $b_i^{(r-1)}$ as in (2.15) if $r = 2$, or as in (2.19) if $r \geqslant 3$. Then $|u_i^{(0)}| \leqslant C$, where $C = 1$ if $r = 2$, or as in (2.20) if $r \geqslant 3$. This gives

$$|u_i| = |a_i - b_i| \leqslant Ch^{r-1} \quad \text{for } i = 1, 2, 3, \dots \ . \tag{2.21}$$

It is well known that the ℓ_∞ norm of $\{u_i\} = \{a_i - b_i\}$ is no less than the $\mathbf{L}_\infty$ norm of $g - s$:

$$||g - s|| \leqslant ||\{u_i\}||_{\ell_\infty}. \tag{2.22}$$

The Jackson inequality (2.9) follows immediately from this and (2.21). □

Proof of Theorem 1. We define $S := Ms \in \mathscr{S}_r$, where M is defined by (2.7). Then $||S^{(r-1)}|| = M||s^{(r-1)}|| = M$. From Theorem 3 we have

$$\begin{aligned} ||f - S|| &\leqslant ||f - G|| + ||G - S|| \leqslant C_1\omega_r(f, h) + M||g - s|| \\ &\leqslant \begin{cases} C\omega_{r-1}(f, h) & \text{if } \omega_{r-1}(f, h) > 0, \\ CMh^{r-1} & \text{if } \omega_{r-1}(f, h) = 0, \end{cases} \end{aligned}$$

with C depending only on r and α.

3. The rate of perfect spline approximation

In this section we prove the approximation rate (1.2) is optimal. First we prove that if $A = \mathbb{R}$ then h^{r-1} is optimal in the case of $\omega_{r-1}(f, h) = 0$ by showing that even for $f(x) \equiv 0$ in (1.2), or $g(x) \equiv 0$ in Theorem 3, the approximation error by perfect splines is still CMh^{r-1} with an explicit value for C. In fact, we will identify best approximations to any $f \in \mathbf{C}(\mathbb{R})$ with $\omega_{r-1}(f, h) = 0$, (or in another word, for any $f \in \mathbf{P}_{r-2}$), resulting in *exact errors* rather than error upper bounds. Euler splines $\mathscr{E}_m$, which are well documented in the literature, or their multiples, turn out to be the best approximation. Some of their properties are listed below.

(i) $\mathscr{E}_m$ is a spline of order $m+1$ and has simple knots at the integers if m is odd, and at the half integers $i+1/2$, $i\in\mathbb{Z}$, if m is even;
(ii) $\|\mathscr{E}_m\| = 1$, with $\mathscr{E}_m(i) = (-1)^i$, $i\in\mathbb{Z}$;
(iii) $\mathscr{E}_m(x) > 0$ for $x\in(-1/2, 1/2)$;
(iv) $\mathscr{E}_m$ is a periodic function of period 2, with $\mathscr{E}_m(x+1) = -\mathscr{E}_m(x)$;
(v) $\mathscr{E}_0(x) = (-1)^i$, $i-1/2 < x < i+1/2$, $i\in\mathbb{Z}$.

The derivatives of Euler splines are easy to calculate. For $0\leqslant k\leqslant m$,

$$\mathscr{E}_m^{(k)}(x) = \frac{K_{m-k}}{K_m}\pi^k \begin{cases} (-1)^{k/2}\mathscr{E}_{m-k}(x) & \text{if } k \text{ is even,} \\ (-1)^{(k-1)/2}\mathscr{E}_{m-k}(x+1/2) & \text{if } k \text{ is odd,} \end{cases} \tag{3.1}$$

where K_ℓ, $\ell = 0, 1, 2, \ldots$, are the so-called Favard numbers. We have

$$K_0 = 1,\ \ K_1 = \frac{\pi}{2},\ \ K_2 = \frac{\pi^2}{8},\ \ K_3 = \frac{\pi^3}{24},\quad \ldots \quad \text{and} \quad \lim_{\ell\to\infty} K_\ell = \frac{4}{\pi},$$
$$K_0 < K_2 < K_4 < \cdots < \frac{4}{\pi} < \cdots < K_5 < K_3 < K_1.$$

In particular,

$$\mathscr{E}_m^{(m)}(x) = A_m \begin{cases} (-1)^{m/2}\mathscr{E}_0(x) & \text{if } m \text{ is even} \\ (-1)^{(m-1)/2}\mathscr{E}_0(x+1/2) & \text{if } m \text{ is odd,} \end{cases}$$
$$\mathscr{E}_m^{(m-1)}(x) = \frac{1}{2}A_m \begin{cases} (-1)^{(m-1)/2}\mathscr{E}_1(x) & \text{if } m \text{ is odd,} \\ (-1)^{(m-2)/2}\mathscr{E}_1(x+1/2) & \text{if } m \text{ is even,} \end{cases}$$

where $A_m := \pi^m/K_m$. The reader can see that they are perfect splines with knots at the integers or half-integers depending on m, and $\|\mathscr{E}_m^{(m)}\| = 2\|\mathscr{E}_m^{(m-1)}\| = A_m$.

Euler splines appear as the solution for many extremal problems such as in the following version of the Kolmogorov–Landau inequalities [4, Chapter 5, Theorem 7.2].

Theorem A. *If $F\in\mathbf{W}_\infty^m(\mathbb{R})$ satisfies $\|F\|\leqslant\|\mathscr{E}_m\| = 1$ and $\|F^{(m)}\|\leqslant\|\mathscr{E}_m^{(m)}\| = A_m$, then*

$$\|F^{(k)}\|\leqslant\|\mathscr{E}_m^{(k)}\|,\quad k = 1, \ldots, m-1. \tag{3.2}$$

Denote by

$$\tilde{\mathscr{P}} := \mathscr{P}_{m+1}^{A_m}(\mathbf{T}_1, \mathbb{R}) = \{s\in\mathscr{S}_{m+1}(\mathbf{T}_1, \mathbb{R}):\ |s^{(m)}(x)| = A_m,\ \ x\notin\mathbb{Z}\} \tag{3.3}$$

the class of all cardinal perfect splines with the mth derivative being $\pm A_m$. We now prove the Euler spline $\mathscr{E}_m$ is the best approximation to 0 from $\tilde{\mathscr{P}}$.

Theorem 5. *Let $m \geqslant 1$ and $A = \mathbb{R}$. The best approximation to $f(x) \equiv 0$ from $\tilde{\mathscr{P}}$ is, up to a sign, the Euler spline $\mathscr{E}_m(x)$ if m is odd, or $\mathscr{E}_m(x+1/2)$ if m is even.*

Remark. If $m = 0$, the best approximation is not unique. In fact, any spline in $\tilde{\mathscr{P}}$ has the same approximation error $A_0 = 1/K_0 = 1$.

Proof. We assume m is odd. The proof for even m is almost identical. We first show if the mth derivative $s^{(m)}$ of any $s \in \tilde{\mathscr{P}}$ fails to change sign at every integer i, then $||s^{(m-1)}|| > \frac{1}{2} A_m = ||\mathscr{E}_m^{(m-1)}||$, thus by Theorem A, $||0 - s|| = ||s|| > ||\mathscr{E}_m|| = ||0 - \mathscr{E}_m||$, which means this s is a worse approximation than $\mathscr{E}_m \in \tilde{\mathscr{P}}$, thus is not a best approximation to 0. Indeed, let σ_i be the sign of $s^{(m)}$ on $(i, i+1)$, then from

$$s^{(m-1)}(x) = s^{(m-1)}(i) + \int_i^x s^{(m)}(t)\,\mathrm{d}t \tag{3.4}$$

we have

$$s^{(m-1)}(i-1) = s^{(m-1)}(i) - \sigma_{i-1} A_m, \tag{3.5}$$

$$s^{(m-1)}(i+1) = s^{(m-1)}(i) + \sigma_i A_m. \tag{3.6}$$

If $|s^{(m-1)}(i)| > \frac{1}{2} A_m$, then $||s^{(m-1)}|| > \frac{1}{2} A_m$ and s is not a best approximation to 0. Now we suppose $|s^{(m-1)}(i)| \leqslant \frac{1}{2} A_m$. By (3.6) the only situation in which $|s^{(m-1)}(i+1)| \leqslant \frac{1}{2} A_m$ can happen is $|s^{(m-1)}(i)| = \frac{1}{2} A_m$ and $\sigma_i = -\mathrm{Sign}(s^{(m-1)}(i))$, in which case $s^{(m-1)}(i+1) = -s^{(m-1)}(i)$. Similarly, it has to be the case that $s^{(m-1)}(i-1) = -s^{(m-1)}(i)$ and $\sigma_{i-1} = \mathrm{Sign}(s^{(m-1)}(i)) = -\sigma_i$. That is, a best approximation to 0 must change the sign of its mth derivative $s^{(m)}$ at each knot i. In another word, only splines s in the form of

$$s(x) = p_{m-1}(x) \pm \mathscr{E}_m(x), \tag{3.7}$$

where $p_{m-1} \in \mathbf{P}_{m-1}$ is a polynomial of degree $< m$, can possibly approximate 0 better than $\mathscr{E}_m$.

In the second part of the proof, we find a polynomial p_{m-1} in (3.7) such that $||s|| = ||\mathscr{E}_m - (-p_{m-1})||$ is minimal, which will give a best approximation from $\tilde{\mathscr{P}}$. Since $\mathscr{E}_m$ is bounded, p_{m-1} has to be bounded too, which means it is a constant polynomial. From the fact $-\min_x \mathscr{E}_m(x) = \max_x \mathscr{E}_m(x) = 1$, one immediately sees that $p_{m-1}(x) \equiv 0$ makes $||s||$ minimal. □

Corollary 6. *Let $m \geqslant 1$ and $A = \mathbb{R}$. The best approximation to any $p_{m-1} \in \mathbf{P}_{m-1}$ from $\tilde{\mathscr{P}}$ is, up to a sign, $p_{m-1}(x) \pm \mathscr{E}_m(x)$ if m is odd, or $p_{m-1}(x) \pm \mathscr{E}_m(x+1/2)$ if m is even.*

Proof. The proof is almost the same as that of Theorem 5. □

By Corollary 6 the best approximation to any polynomial $p_{r-2}\in\mathbf{P}_{r-2}$, $r\geqslant 2$, by perfect splines from $\mathscr{P}_r^M=\mathscr{P}_r^M(\mathbf{T}_h,\mathbb{R})$ is

$$s(x)=\begin{cases} p_{r-2}(x)\pm\dfrac{Mh^{r-1}}{A_{r-1}}\mathscr{E}_{r-1}\left(\dfrac{x}{h}\right) & \text{if } m \text{ is odd,}\\ p_{r-2}(x)\pm\dfrac{Mh^{r-1}}{A_{r-1}}\mathscr{E}_{r-1}\left(\dfrac{x}{h}+\dfrac{1}{2}\right) & \text{if } m \text{ is even}\end{cases}$$

with an error of Mh^{r-1}/A_{r-1}. This is also true for $r=1$ with $p_{-1}:=0$ by direct verification although the best approximation is not unique. In fact, any $s\in\mathscr{P}_1^M$ is a best approximation to 0 with the same error M.

Theorem 5 also solves an extremal problem, quite classical in nature, namely among all perfect splines with the $(r-1)$st derivative ±1, which ones have the smallest norm. This goes back long time. As a matter of fact, the second part of the proof of Theorem 5 is a special case of Cavaretta [1], (also see Schoenberg [14, Lecture 9]). Cavaretta's results are valid for some other knot sequences, but with the *assumption* that $s^{(r-1)}$ changes sign at every knot (thus are only related to the second part of our proof). Our set $\tilde{\mathscr{P}}$ allows any sign pattern in $s^{(r-1)}$. We rewrite Theorem 5 in classical language below as a corollary.

Corollary 7. *Among all perfect splines in* $\mathscr{P}^1_{m+1}(\mathbf{T}_1,\mathbb{R})=\{s\in\mathscr{S}_{m+1}(\mathbf{T}_1,\mathbb{R})\colon |s^{(m)}(x)|=1,\ x\notin\mathbb{Z}\}$, *only* $\pm A_m^{-1}\mathscr{E}_m(x)$, *if m is odd, or* $\pm A_m^{-1}\mathscr{E}_m(x+1/2)$, *if m is even, have the smallest possible uniform norm* A_m^{-1}.

In the following theorem we show the order $r-1$ is optimal in all cases.

Theorem 8. *The approximation order given by inequalities* (1.2) *is optimal for any finite or infinite interval* A.

Proof. Without loss of generality, we assume $0\in A$. We first consider the case $\omega_{r-1}(f,h)>0$. By the well-known Bernstein-type inequality

$$\|s^{(k)}\|\leqslant C_5h^{-k}\|s\|,\quad s\in\mathscr{S}_r,\ \ 1\leqslant k<r,$$

(see Theorem 1.2 of Chapter 5 in [4] for the case $A=[0,1]$, which can be easily extended to $\mathbb{R}$ for $p=\infty$), we have

$$\|s\|\geqslant\frac{1}{C_5}h^{r-1}\|s^{(r-1)}\|=:C_6h^{r-1}\|s^{(r-1)}\|\geqslant C_7\omega_{r-1}(s,h),\quad s\in\mathscr{S}_r,\tag{3.8}$$

where C_5–C_7 are constants depending on r. Here again we have used the fact that $h^{r-1}\|s^{(r-1)}\|$ is equivalent to $\omega_{r-1}(s,h)$ [8]. Let f be an $(r-1)$-fold integral of

$$f^{(r-1)}(x):=\begin{cases}(-1)^i & x\in(ih,(i+1)h),\quad i\neq 0,\\ 0 & x\in(0,h).\end{cases}$$

Then f belongs to $\mathscr{S}_r \subset \mathbf{C}(A)$, and so does $f - s$ for any $s \in \mathscr{P}_r^M \subset \mathscr{S}_r$ with any $M > 0$. We apply (3.8) to $f - s$ and f and obtain

$$\|f - s\| \geqslant C_6 h^{r-1} \|f^{(r-1)} - s^{(r-1)}\| \geqslant \tfrac{1}{2} C_6 h^{r-1} \|f^{(r-1)}\| \geqslant \tfrac{1}{2} C_7 \omega_{r-1}(f, h), \tag{3.9}$$

where in the second inequality we have used the facts $\|f^{(r-1)} - s^{(r-1)}\| = \max(M, |1 - M|) \geqslant 1/2$ and $\|f^{(r-1)}\| = 1$. This shows $\omega_{r-1}(f, h)$ is the optimal order for functions f in $\mathbf{C}(A)$ with $\omega_{r-1}(f, h) > 0$.

Now let $f \in \mathbf{C}(A)$ with $\omega_{r-1}(f, h) = 0$, that is, $f \in \mathbf{P}_{r-2}$. Then the first part of (3.9), which is valid for all $f \in \mathscr{S}_r$, becomes

$$\|f - s\| \geqslant C_6 h^{r-1} \|f^{(r-1)} - s^{(r-1)}\| = C_6 h^{r-1} \|s^{(r-1)}\| = C_6 M h^{r-1},$$

which shows the second part of (1.2) is optimal. □

In Theorem 8 we showed (1.2) is optimal, in the so-called worst scenario sense for the case $\omega_{r-1}(f, h) > 0$. We now show $r - 1$ is the best order in the sense that f has an approximation order $o(h^{r-1})$ if and only if $f \in \mathbf{P}_{r-1}$.

Theorem 9. *Let A be any interval and $f \in \mathbf{C}(A)$. Then the following statements are equivalent*:

(i) *for each $h > 0$, there exists a spline $S_h \in \mathscr{P}_r^{M_h}(\mathbf{T}_h, A)$ for some $M_h > 0$ such that*

$$\|f - S_h\| = o(h^{r-1}); \tag{3.10}$$

(ii) *f is a polynomial of degree $< r$ on A.*

Proof. We first prove (i) implies (ii). Let A be finite with a length comparable to 1 and let $\tilde{S}_h$ be a best spline approximation to f from $\mathscr{S}_r(\mathbf{T}_h, A)$. We have

$$\|f - \tilde{S}_h\| \leqslant \|f - S_h\| = o(h^{r-1}). \tag{3.11}$$

The inverse theorem for spline approximation from $\mathscr{S}_r(\mathbf{T}_h, A)$ ([5], also see Section 12.2 of [4]) gives

$$\omega_r(f, h) \leqslant C \max_{n \leqslant m \leqslant 2n} \|f - \tilde{S}_{1/m}\| = o(h^{r-1}), \tag{3.12}$$

where $n := [1/h]$. Therefore

$$\omega_r(S_h, h) \leqslant \omega_r(f - S_h, h) + \omega_r(f, h) \leqslant C\|f - S_h\| + \omega_r(f, h) = o(h^{r-1}), \tag{3.13}$$

and, by Hu and Yu [8],

$$\omega(S_h^{(r-1)}, h) \sim h^{-r+1} \omega_r(S_h, h) = o(1). \tag{3.14}$$

We observe that

$$\omega(S_h^{(r-1)}, h) = \begin{cases} 2M_h & \text{if } S_h^{(r-1)} \text{ changes sign at least once} \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

Note here "otherwise" means $S_h \in \mathbf{P}_{r-1}$. We now prove f is a polynomial in two cases.

Case 1: There exists an $h_0 > 0$ such that $S_h^{(r-1)}$ changes sign at least once for every $h < h_0$. Then by (3.14) and (3.15)

$$o(1) = \omega(S_h^{(r-1)}, h) = 2M_h = 2||S_h^{(r-1)}||, \quad h < h_0.$$

Hence

$$\omega_{r-1}(S_h, h) \leqslant h^{r-1}||S_h^{(r-1)}|| = o(h^{r-1}).$$

From (3.10)

$$\omega_{r-1}(f, h) \leqslant \omega_{r-1}(f - S_h, h) + \omega_{r-1}(S_h, h) = o(h^{r-1}), \tag{3.16}$$

which implies that $f \in \mathbf{P}_{r-2}$, (see for example, Theorem 2.59 of Schumaker [13]).

Case 2: There exists a sequence $\{h_n\}$ with $\lim_{n\to\infty} h_n = 0$ such that $S_{h_n}^{(r-1)}$ have no sign changes, that is, $\{S_{h_n}\}$ is a sequence of polynomials of degree $r-1$. Since S_{h_n} converge to f in norm, f is in the closure of $\mathbf{P}_{r-1}$, which is a closed subspace. Therefore $f \in \mathbf{P}_{r-1}$.

We now consider the case $A = \mathbb{R}$ or $A = [0, \infty)$. Since (3.10)) is true on every interval $[i, i+1]$ that is contained in A, therefore f is a piecewise polynomial of degree $< r$ on A with possible knots on every integer i. But if we consider every interval $[i-1, i+1]$ that is contained in A, we know no such i is actually a knot, thus f is a polynomial on the whole interval A.

We prove that (ii) implies (i) also in two cases.

Case 1: f is a polynomial of degree exactly $r-1$. Since $f \in \mathscr{P}_r^M(\mathbf{T}_h, A)$ with $M = f^{(r-1)}$ for every h, one can choose $S_h = f$ for every h, which results in $||f - S_h|| = 0$, of course.

Case 2: f is a polynomial of degree $< r-1$. Then $f \notin \mathscr{P}_r^M(\mathbf{T}_h, A)$ no matter what M or h is. By Theorem 1, the error is $\leqslant CMh^{r-1}$, where C is independent of h, while M can be freely chosen. One can choose M_h such that $M_h \to 0$ as $h \to 0$ (for example $M_h = h$), and use for S_h a (near) best approximation to f from $\mathscr{P}_r^{M_h}(\mathbf{T}_h, A)$. □

Remark. It is unusual that polynomials of lower degrees have nonzero errors while those of degree $r-1$ are approximated exactly. This again reflects the fact that the only polynomials belonging to $\bigcup_{M,h>0} \mathscr{P}_r^M(\mathbf{T}_h, A)$ are those of degree exactly $r-1$.

We conclude this paper by some word on the existence of best approximation from $\mathscr{P}_r^M = \mathscr{P}_r^M(\mathbf{T}_h, A)$ defined in (1.1). Let $\sigma = \{\sigma_i M\}_{i\in\Lambda}$ be a finite or (bi)infinite sequence with $\sigma_i = 1$ or -1, and $\Lambda := \{i : (i, i+1) \cap A \neq \emptyset\}$. Let $S_{\boldsymbol{\sigma}} \in \mathscr{P}_r^M$ be the spline such that

$$S_\sigma^{(r-1)}(x) = \sigma_i M, \quad x \in (i, i+1) \quad \text{and} \quad S_\sigma(0) = S_{\boldsymbol{\sigma}}'(0) = \cdots = S_\sigma^{(r-2)}(0) = 0.$$

Then

$$\mathscr{P}_r^M = \bigcup (S_{\boldsymbol{\sigma}} + \mathbf{P}_{r-2}),$$

where the union runs over all possible sequences σ defined above. Each affine space $S_\sigma + \mathbf{P}_{r-2}$ is a closed set, so is $\mathscr{P}_r^M$ if A is finite. We have just proved

Theorem 10. *If A is a finite interval, then best approximation from $\mathscr{P}_r^M$ to any $f \in \mathbf{C}(A)$ exists.*

Acknowledgments

The authors feel deeply indebted to Professor Ronald A. DeVore, who encouraged us to pursue, and was actively involved in, this work. In fact, his joint work with Professor Ingrid Daubechies on Sigma–Delta modulation inspired this research while the second author was doing her sabbatical with him at the University of South Carolina. He contributed good ideas to the work through various discussions with us. A major part of the work was done when he hosted the first author's sabbatical.

We would like to thank Sinan Güntürk for helpful discussions, and thank him and Özgür Yilmaz for kindly providing us with their dissertation and unpublished manuscripts. We also would like to thank the referee of this paper for remarks and suggestions that helped improve the manuscript very much, especially the one that enabled us to shorten the proof of Theorem 5. The referee also pointed out that it is not always straightforward (if possible at all) to obtain a backward version for $\Sigma\Delta$ modulation schemes that use more than $r-1$ state variables, which resulted in the caution after (2.16).

References

[1] A. Cavaretta, On cardinal perfect splines of least sup-norm on the real axis, Doctoral Thesis, University of Wisconsin, Madison, 1970.

[2] I. Daubechies, R.A. DeVore, Reconstructing a bandlimited function from very coarsely quantized data: a family of stable sigma–delta modulators of arbitrary order, Ann. Math., to appear.

[3] C. de Boor, Practical Guide to Splines, Springer, New York, 4th printing, 1987.

[4] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Springer, Berlin, 1993.

[5] R.A. DeVore, F. Richards, Saturation and inverse theorems for spline approximation, in: A. Meir, A. Sharma (Eds.), Spline Functions and Approximation Theory, ISNM 21, Birkhäuser, Basel, 1972, pp. 73–82.

[6] R.M. Gray, Spectral analysis of quantization noise in a single-loop Sigma–Delta modulator with dc input, IEEE Trans. Comm. 37 (1989) 588–599.

[7] Sinan Güntürk, Harmonic analysis of two problems in signal quantization and compression, Ph.D Dissertation, Princeton University, 2000.

[8] Y.-K. Hu, X.M. Yu, Discrete modulus of smoothness of splines with equally spaced knots, SIAM J. Numer. Anal. 32 (1995) 1428–1435.

[9] T. Lyche, L. Schumaker, Local spline approximation methods, J. Approx. Theory 15 (1975) 294–325.

[10] S.R. Norsworthy, R. Schreier, G.C. Themes (Eds.), Delta–Sigma data converters, IEEE Press, New York, 1997.

[11] S.J. Park, R.M. Gray, Sigma–Delta modulation with leaky integration and constant input, IEEE Trans. Inform. Theory 38 (1992) 1512–1533.

[12] S.C. Pinault, P.V. Lopresti, On the behavior of the double-loop Sigma–Delta modulator, IEEE Trans. Circuits Systems-II 40 (1993) 467–479.
[13] L.L. Schumaker, Spline Functions: Basic Theory, Wiley, New York, 1981.
[14] I.J. Schoenberg, Cardinal Spline Interpolation, SIAM, Philadelphia, 1973.
[15] N. Thao, Quadratic one-bit second-order Sigma–Delta modulators, preprint.
[16] Ö. Yilmaz, Stability analysis for several second-order Sigma–Delta methods of coarse quantization of bandlimited functions, Constr. Approx. 18 (2002) 599–623.