ÖAW AUSTRIAN ACADEMY OF SCIENCES

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

RICAM
JOHANN-RADON-INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

Filtered Legendre Expansion Method for Numerical Differentiation at the Boundary Point with Application to Blood Glucose Predictions

H.N. Mhaskar, V. Naumova, S. Pereverzyev

RICAM-Report 2013-07

Filtered Legendre Expansion Method for Numerical Differentiation at the Boundary Point with Application to Blood Glucose Predictions

Hrushikesh N. Mhaskar · Valeriya Naumova · Sergei V. Pereverzyev

Abstract Let $f : [-1, 1] \to \mathbb{R}$ be continuously differentiable. We consider the question of approximating $f'(1)$ from given data of the form $(t_j, f(t_j))_{j=1}^{M}$ where the points t_j are in the interval $[-1, 1]$. It is well known that the question is ill–posed, and there is very little literature on the subject known to us. We consider a summability operator using Legendre expansions, together with high order quadrature formulas based on the points t_j's to achieve the approximation. We also estimate the effect of noise on our approximation. The error estimates, both with or without noise, improve upon those in the existing literature, and appear to be unimprovable. The results are applied to the problem of short term prediction of blood glucose concentration, yielding better results than other comparable methods.

Keywords Numerical differentiation · Legendre Polynomials · Blood glucose prediction

Mathematics Subject Classification (2000) 65D25 · 42C10 · 65R30

1 Introduction

In the prediction of the blood glucose (BG) evolution in diabetes therapy management [13], [15], [25], several well-known and highly used predictors are based on linear extrapolation of current blood glucose trends. In turn, this requires an accurate approximation of the derivative of a function at the boundary point of an interval on which the BG-readings are available. Similar problems arise also in other areas of high practical interest in industrial applications. For example, the identification problem of the heat transfer function in the cooling process [8] relies on an accurate knowledge of the derivatives of functions describing temperature at the boundary points. In image completion, one seeks to extend the image data into a "hole" as a smooth function [3], [5]. Clearly, this problem also requires an estimation of derivatives of a function at the endpoint of the normal lines to the hole. In this paper, we are interested in proposing a method for numerical differentiation which is especially suitable for short–term prediction of blood glucose levels based on previous data.

H. N. Mhaskar · V. Naumova · S. V. Pereverzyev

The problem of numerical differentiation is the following. Let $f : [-1, 1] \to \mathbb{R}$ be a continuously differentiable function and $\{t_j\}_{j=1}^M \subset [-1, 1]$. Given information of the form $\{(t_j, f(t_j))\}_{j=1}^M$, find approximately the value of $f'(t)$. In practical problems, the data is often noisy, or at least given up to a fixed accuracy δ. This situation can be described by the so-called deterministic noise model. In this model, the noise intensity level is measured by a small positive number δ, and the available information has the form $\{(t_j, f_\delta(t_j))\}_{j=1}^M$, where f_δ is a continuous function on $[-1, 1]$ such that

$$|f(t_j) - f_\delta(t_j)| \leq \delta. \tag{1}$$

The problem of numerical differentiation is one of the classical ill-posed problems [7]. There are many papers spanning several years of research describing various numerical and analytical methods to address this problem in different contexts (for example, [4], [14], [28], [19], [34], just to mention a few). All these approaches differ greatly in implementation in dependence on a noise model and available data. Most of these deal with the question when the point $t \neq \pm 1$. The question of approximating $f'(1)$ (alternately, $f'(-1)$) is not investigated to the same extent.

Recently, a one-sided backward difference scheme equipped with an adaptive choice rule for the number of nodes $\{t_j\}$ [25] has been used to approximate the derivative at the boundary point with relevant application in diabetes technology.

In [30], Savitzky and Golay have proposed an approximation of $f'(t)$ by the derivative of a polynomial of least square fit. The degree of the polynomial acts as a regularization parameter. More specifically, for an integer parameter $n \geq 1$, one finds coefficients a_k^* such that

$$\sum_{j=0}^{M} (f_\delta(t_j) - \sum_{k=0}^{n} a_k^* t_j^k)^2 = \min_{a_0, \cdots, a_n \in \mathbb{R}} \left(\sum_{j=0}^{M} (f_\delta(t_j) - \sum_{k=0}^{n} a_k t_j^k)^2 \right), \tag{2}$$

and takes

$$\frac{d}{dt} \left(\sum_{k=0}^{n} a_k^* t^k \right)$$

as the approximation of $f'(t)$. This scheme could be easily applied for approximation of $f'(1)$, which is of our main interest in the current paper.

However, in addition to the intrinsic ill–conditioning of numerical differentiation, the solution of the least square problem as posed above involves a system of linear equations with the Hilbert matrix of order n, which is notoriously ill–conditioned. Therefore, it is proposed in [18] to use Legendre polynomials rather than the monomials as the basis for the space of polynomials of degree n. A procedure to choose n is given in [18], together with error bounds in terms of n and δ which are optimal up to a constant factor for the method in the sense of the oracle inequality.

In this paper, we propose two modifications of the approach [18]. Firstly, we propose the use of judiciously selected weights in the least square method as in (2) except for the use of Legendre basis. Secondly, we avoid the use of least square optimization altogether, using a summability method. We show that, by employing recent results [18] together with the modifications, we derive a method that yields lower noise propagation error than in other approach considered in [18].

We demonstrate numerically that these modifications lead to a performance superior to the Savitzky–Golay method as modified in [18] on a number of numerical examples.

In Section 2, we describe some background and notations. In particular, we elaborate on the choice of the weights as explained in Theorem 1. In Section 3, we develop our method, and prove the theoretical error bounds. An application to the problem of short–term prediction of blood glucose is described in Section 4. The proofs of the results in Section 3 are given in Section 5.

2 Background

The Legendre polynomials are defined by

$$P_k(x) = \frac{(-1)^k}{2^k k!}\left(\frac{d}{dx}\right)^k (1-x^2)^k, \qquad k = 0, 1, \cdots, \ x \in [-1, 1]. \tag{3}$$

For integers $k, m \geq 0$, they satisfy the orthogonality relations [32, Eqn. (4.3.3)]

$$\int_{-1}^{1} P_k(x)P_m(x)dx = \begin{cases} \dfrac{2}{2k+1}, & \text{if } k = m, \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

and the differential equation [32, Theorem 4.2.1]

$$2xP_k'(x) - (1-x^2)P_k''(x) = k(k+1)P_k(x). \tag{5}$$

In this paper, if $f : [-1, 1] \to \mathbb{R}$ is twice differentiable, we denote

$$\Delta(f)(x) := 2xf'(x) - (1-x^2)f''(x), \tag{6}$$

and observe that

$$f'(1) = \frac{1}{2}\Delta(f)(1). \tag{7}$$

The differential equation (5) can be rewritten in the form

$$\Delta(P_k)(x) = k(k+1)P_k(x), \qquad x \in [-1, 1], \ k = 0, 1, \cdots. \tag{8}$$

If $f : [-1, 1] \to \mathbb{R}$ is Lebesgue integrable, then it can be expanded formally in Legendre series

$$f \sim \sum \hat{f}(k)(k + 1/2)P_k,$$

where

$$\hat{f}(k) = \int_{-1}^{1} f(t)P_k(t)dt, \qquad k = 0, 1, \cdots, \tag{9}$$

are the Fourier-Legendre coefficients. We also introduce the Fourier partial sum operator that is given by

$$s_n(f)(x) = \sum_{k=0}^{n-1} (k + 1/2)\hat{f}(k)P_k(x), \qquad n = 1, 2, \cdots. \tag{10}$$

In view of (5), the derivative $f'(1)$ can be approximated by the derivative of the partial sums of the Fourier-Legendre series as

$$s_n'(f)(1) = \sum_{k=0}^{n-1}(k+1/2)\hat{f}(k)P_k'(1) = \frac{1}{2}\sum_{k=1}^{n-1}k(k+1/2)(k+1)\hat{f}(k). \tag{11}$$

Moreover, from (8) it also follows that the formal expansion of $\Delta(f)$ is given by

$$\Delta(f) \sim \sum_{k=0}^{\infty} k(k+1)(k+1/2)\hat{f}(k)P_k. \tag{12}$$

In the paper [18], the authors considered the following modification of (11) for approximating $f'(t)$

$$D_n f_\delta(t) = \sum_{k=0}^{n-1}(k+1/2)\bar{f}_\delta(k)P_k'(t), \tag{13}$$

where $\bar{f}_\delta(k)$ are the approximations of $\hat{f}(k)$ and found by the method of least squares from given noisy data. The authors have proved that the data noise propagates in the approximation D_n with an intensity $\mathcal{O}(n^3\delta)$. At the same time, it is important to mention that the noise model used in that paper is essentially different from the one considered in the current work. To be more precise, in [18] the authors considered additive square summable noise or that is the same as L_2-valued noise that is well-accepted within the framework of the regularization theory.

One of our innovations in this paper is to use the quadrature formulas proposed in [22] with special weights instead of least squares method in order to approximate the Fourier-Legendre coefficients. Such modification yields a lower noise propagation rate, namely $\mathcal{O}(n^2\delta)$, than the rate obtained in [18].

We review next the construction of these weights. The following discussion will involve many generic constants, whose specific value is of no interest to us. Therefore, before proceeding further, we make the following convention.

Constant convention:

In the sequel, the symbols $c, c_1, \cdots$ will denote generic constants independent of all the variables in the discussion, such as the functions involved, or the degree of the polynomial. They may depend upon fixed parameters in the discussion, such as the function h to be introduced later. Their values may be different at different occurrences, even within the same formula.

For each integer $n \ge 1$, let $\mathcal{C}_n = \{t_{M_n,n} < t_{M_n-1,n} < \cdots < t_{1,n}\} \subset (-1,1)$, $t_{j,n} =: \cos(\theta_{j,n})$, $j = 1, \cdots, M_n$, $\theta_{0,n} = 0$, $\theta_{M_n+1,n} = \pi$. Let

$$\delta_n := \max_{\theta\in[0,\pi]} \min_{1\le j\le M_n} |\theta_{j,n} - \theta|.$$

For integer $N \ge 1$, we denote by Π_N the class of all algebraic polynomials of degree $< N$, and define $\Pi_0 := \{0\}$. It is convenient to extend this notation to non–integer values of N by setting $\Pi_N = \Pi_{\lfloor N \rfloor}$. The following theorem is a consequence of [22, Theorem 4.1]:

Theorem 1 *Let $n \geq 1$. There exists a constant $\alpha > 0$ such that for $N_n = \lfloor \alpha \delta_n^{-1} \rfloor$, there exist real numbers $\{w_{j,n}\}_{j=1}^{M_n}$ with the following properties:*

$$\sum_{j=1}^{M_n} w_{j,n} P(t_{j,n}) = \int_{-1}^{1} P(t)dt, \qquad P \in \Pi_{2N_n}, \tag{14}$$

and

$$\sum_{j=1}^{M_n} |w_{j,n} P(t_{j,n})| \leq c \int_{-1}^{1} |P(t)|dt, \qquad P \in \Pi_{2N_n}. \tag{15}$$

In [22], we have described a constructive procedure to obtain the weights $\{w_{j,n}\}$. Note that in practice, taking $n_M < \lfloor 2^{-1} M \rfloor$, where M is the number of given data points $\{t_j\}$, one can always use least squares to solve an underdetermined system of the form

$$\sum_{j=1}^{M} w_j P_k(t_j) = \begin{cases} 2, & \text{if } k = 0, \\ 0, & \text{if } k = 1, \cdots, 2n_M, \end{cases} \tag{16}$$

to obtain the weights $\{w_j\}$ to satisfy

$$\sum_{j=1}^{M} w_j P(t_j) = \int_{-1}^{1} P(t)dt, \qquad P \in \Pi_{2n_M}. \tag{17}$$

Using the ideas in [9], it can be shown that the condition number of the Gram matrix involved in the least squares is of the same order of magnitude as the constant c appearing in (15). In the sequel, we mainly use (17) for our analysis and numerical experiments. We will also assume that

$$\sum_{j=1}^{M} |w_j P(t_j)| \leq A \int_{-1}^{1} |P(t)|dt, \qquad P \in \Pi_{2n_M}, \tag{18}$$

where the value of A depends only on the distribution of nodes $\{t_j\}$.

Moreover, as it should be clear from the noise model (1) we will deal exclusively with the space of continuous functions $C = C[-1, 1]$ that is equipped with the uniform norm

$$\|f\|_C := \max_{x \in [-1,1]} |f(x)|, \qquad f \in C[-1, 1].$$

It is also convenient to introduce the error of the best approximation of f by algebraic polynomials

$$E_n(f) := \min_{P \in \Pi_n} \|f - P\|_C.$$

3 Main results

First, we present the discrete analogue of (11), where the Fourier-Legendre coefficients are approximated from given noisy data $\{f_\delta(t_j)\}_{j=1}^M$ by means of a quadrature rule.

To be more precise, in general, if $\mathbf{y} = \{y_j\}_{j=1}^M \subset \mathbb{R}$ is the given data, we define the Fourier-Legendre coefficients

$$\tilde{\mathbf{y}}(k) = \sum_{j=1}^{M} w_j y_j P_k(t_j), \tag{19}$$

and the discrete analog of the summability operator s_n as

$$S_n(\mathbf{y})(x) = \sum_{k=1}^{n} \tilde{\mathbf{y}}(k)(k+1/2)P_k(x). \tag{20}$$

We will write $\mathbf{f} := (f(t_1), \cdots, f(t_M))$, and $\mathbf{f}_\delta = (f_\delta(t_1), \cdots, f_\delta(t_M))$ to denote the noise-free and noisy data respectively. On the basis of the above observations, we derive the following result, where A is the constant defined in (18).

Theorem 2 *Let $f : [-1,1] \to \mathbb{R}$, and $\Delta(f) \in C[-1,1]$. Then*

$$|f'(1) - S_n'(\mathbf{f}_\delta)(1)| \le cAn^{1/2}\left\{E_n(\Delta(f)) + n^2\delta\right\}. \tag{21}$$

The estimates in Theorem 2 can be improved further using summability methods. To describe this, we first make a definition.

Definition 1 Let $h : [0,\infty) \to \mathbb{R}$ be a compactly supported function.
(a) The **summability kernel** with **filter** h is defined by

$$\Phi_n(h; x, t) := \sum_{k=0}^{\infty} h\left(\frac{k}{n}\right)(k+1/2)P_k(x)P_k(t), \qquad n > 0,\ x, t \in \mathbb{R}. \tag{22}$$

(b) We define the **summability operator** corresponding to the filter h by

$$\sigma_n(h; f)(x) := \int_{-1}^{1} f(t)\Phi_n(h; x, t)dt = \sum_{k=0}^{\infty} h\left(\frac{k}{n}\right)(k+1/2)\hat{f}(k)P_k(x), \tag{23}$$

for all $n > 0$, $f \in L^1[-1,1]$, and $x \in \mathbb{R}$.
(c) We denote the discretization of the operator σ_n by

$$\mathcal{S}_n(h; \mathbf{y})(x) := \sum_{j=1}^{M} w_j y_j \Phi_n(h; x, t_j) = \sum_{k=0}^{\infty} h\left(\frac{k}{n}\right)(k+1/2)\tilde{\mathbf{y}}(k)P_k(x), \qquad \mathbf{y} = \{y_j\}_{j=1}^M \subset \mathbb{R}. \tag{24}$$

If $y_j = f(t_j)$, $j = 1, \cdots, M$ for a function $f : [-1,1] \to \mathbb{R}$, we overload the notation by writing $\mathcal{S}_n(h; f)$.
(d) The function h will be called a **low pass filter** if $h(t) = 1$ for $0 \le t \le 1/2$, h is non–increasing on $[1/2, 1]$, and $h(t) = 0$ for all $t \ge 1$.

We remark that since h is compactly supported, the apparently infinite sums in (22), (23), and (24) are actually finite sums and the parameter n serves as a regularization parameter.

The difference between the exact value of the derivative $f'(1)$ and its estimate given by means of the discrete version of the summability operator (24) can be presented as follows

$$|f'(1) - \mathcal{S}'_n(h, \mathbf{f}_\delta)(1)| \leq |f'(1) - \mathcal{S}'_n(h, \mathbf{f})(1)| + |\mathcal{S}'_n(h, \mathbf{f})(1) - \mathcal{S}'_n(h, \mathbf{f}_\delta)(1)|, \quad (25)$$

where the first term in the right-hand side is the approximation error, whereas the second term is the noise propagation error.

The error bound on (25) is provided in the following theorem. It is shown in (28) that the use of the summability operator removes the factor $n^{1/2}$ in the estimate (21) of Theorem 2.

Theorem 3 *Let h be a twice continuously differentiable low pass filter. Let f admit two derivatives such that $\Delta(f) \in C[-1, 1]$. Then*

$$|f'(1) - \mathcal{S}'_n(h, \mathbf{f})(1)| \leq cAE_{n/2}(\Delta(f)). \quad (26)$$

Further, we have

$$|\mathcal{S}'_n(h, \mathbf{y})(1)| \leq Bn^2 \max_{1 \leq j \leq M} |y_j|, \quad (27)$$

with a positive constant B that depends only on the quantity A from (18).

Thus,

$$|f'(1) - \mathcal{S}'_n(h, \mathbf{f}_\delta)(1)| \leq cA \left\{ E_{n/2}(\Delta(f)) + n^2\delta \right\}. \quad (28)$$

Remark 1 If f is analytic, then it is well known (e.g., [24, Chapter 9, Section 3]) that $E_n(\Delta(f)) = \mathcal{O}(\rho^n)$ for some $\rho \in (0, 1)$. Thus, in the absence of noise, the upper bound $\rho^{n/2}$ in (28) is worse than the upper bound $n^{1/2}\rho^n$ indicated in (21). For functions of finite smoothness, both the bounds are of the same order of magnitude, but the summability method has other such advantages as localized approximation properties.

The smoothness of the function is rarely known in advance. Wavelet–like expansions based on Legendre expansions in particular are given in [22], [10], [23], where the terms of the expansion characterize the analyticity and various smoothness parameters at different points of the interval. In future work we intend to investigate an algorithm that would allow an adaptive choice of the method on the basis of the input data.

3.1 Adaptive parameter choice

In this section, we present an adaptive parameter choice rule for the method (24), as well as show its optimality up to a constant factor in the sense of the oracle inequality. As already mentioned, numerical differentiation of noisy data is one of the classical ill-posed problems [7] and, thus, a regularization mechanism is required. For instance, in Introduction we have seen that the parameter n in (13) as well as in (24) serves as a regularization parameter and should be correctly chosen depending on a noise level δ and smoothness of the function to be differentiated.

The importance of the proper parameter choice for the numerical differentiation problem is, for example, explicitly illustrated by numerical examples in [18].

Obviously, estimation (28) in Theorem 3 implies that increasing the values of n, the approximation error decreases. At the same time, from (28) we observe that with increase of the $n-$value the noise propagates in data with the rate $\mathcal{O}(n^2\delta)$. Thus, one needs to find a balance between the approximation and the noise propagation errors. This is achieved by presenting the a posteriori parameter choice rule, which is based on the so-called balancing principle that has been extensively studied (see, for example, [11], [20] and references therein).

Definition 2 Following [20], we say that a function $\varphi(n) = \varphi(n; f, \delta)$ is admissible for given f and δ if the following holds

1. $\varphi(n)$ is a non-increasing function on $[1, n_M]$, where n_M is the quantity involved in (16)-(18),
2. $\varphi(n_M) < Bn_M^2\delta$,
3. $\forall n \in \{1, \ldots, n_M\}$

$$|f'(1) - \mathcal{S}'_n(h, \mathbf{f})(1)| \leq \varphi(n). \tag{29}$$

For given $f,\ \delta$ the set of admissible functions is denoted by $\Phi(f, \delta)$.

From (26), (28) and Definition 2 the difference between $f'(1)$ and its approximation given by the Legendre filters can be bounded as follows

$$|f'(1) - \mathcal{S}'_n(h, \mathbf{f}_\delta)(1)| \leq \varphi(n) + Bn^2\delta. \tag{30}$$

We now present a principle for the adaptive choice of $n = n_+ \in [1, n_M]$ that allows us to reach the best possible error bound up to some multiplier.

Theorem 4 *Let $n = n_+$ be chosen as*

$$n_+ = \min\{n : |\mathcal{S}'_n(h, \mathbf{f}_\delta)(1) - \mathcal{S}'_m(h, \mathbf{f}_\delta)(1)| \leq 4Bm^2\delta,\ m = n, \ldots, n_M\}. \tag{31}$$

Then the following error bound holds true

$$|f'(1) - \mathcal{S}'_n(h, \mathbf{f}_\delta)(1)| \leq c \inf_{\varphi\in\Phi(f,\delta)} \min_{n=1,\ldots,n_M} \{\varphi(n) + Bn^2\delta\}, \tag{32}$$

where the right-hand side is, up to a constant factor, the best possible error bound that can be guaranteed for the approximation $f'(1)$ within the framework of the scheme (24) under Assumption (1) and (27).

Note that Theorem 4 can be proven similar to the one in [25]. Thus, we omit the proof here and refer to the papers [20], [25] for more details.

Remark 2 In general, the bound for the noise propagation error in numerical differentiation by algebraic polynomials can be, obtained in two steps. At first, we estimate the difference between polynomial approximants constructed for noisy and noise-free data. Then using the inequality of the form

$$\|P'_n\|_C \leq n^2\|P_n\|_C, \tag{33}$$

where the estimate for $\|P_n\|_C$ is obtained from the previous step, we estimate the difference between the derivatives of the approximants. Since the nature of a noise prevents us from any assumption on the properties of polynomials, we need to use the inequality of the form (33) that is valid for arbitrary polynomials of a degree n.

Therefore, within the framework of (1) one may not expect that the noise will propagate with the rate lower than n^2. This reasoning can be seen as support for the adequacy of the bound (28).

4 Numerical experiments

The main aim of this section is to discuss the performance of the method (24) equipped with the adaptive parameter choice rule (31) in predicting the blood glucose (BG) evolution.

Mathematically the problem of the BG-prediction can be formulated as follows. Assume that at the time moment $t = t_0$ we are given m preceding estimates $g_\delta(t_i),\ i = -m+1, \dots, 0$, of a patient's BG-concentration sampled correspondingly at the time moments $t_0 > t_{-1} > t_{-2} > \dots > t_{-m+1}$ within the sampling horizon $SH = t_0 - t_{-m+1}$. The goal is to construct a predictor that uses these past measurements to predict the BG-concentration as a function of time $g = g(t)$ for k subsequent future time moments $\{t_j\}_{j=1}^k$ within the prediction horizon $PH = t_k - t_0$ such that $t_0 < t_1 < t_2 < \dots < t_k$.

There are several prediction techniques, and a variety of the glucose predictors has been recently proposed, see, for example, [26] and references therein. In this section we discuss the predictors based on the numerical differentiation [13]. Such predictors estimate the rate of change of the BG-concentration at the prediction moment $t = t_0$ from current and past measurements and the future BG-concentration at any time moment $t \in [t_0, t_k]$ is given as follows

$$g(t) = g'(t_0) \cdot (t - t_0) + g_\delta(t_0), \tag{34}$$

where $g'(t_0)$ is approximated from the given noisy data $\{(t_i, g_\delta(t_i))\},\ i = -m + 1, \dots, 0,\ SH = 30$ (min), $n_M = m = 7$. We have chosen $m = 7$, because for $\Delta t = 5$ (min) the sampling horizon $SH = 30 = 6\Delta t$ (min) has been suggested in [13] as the optimal one for BG prediction.

At this point it is important to stress the fact that to approximate the derivative $g'(t_0)$ by means of (24) the given data points $\{t_i\}_{i=-m+1}^0$ should at first be transformed from the interval $[t_{-m+1}, t_0]$ into the interval $[-1, 1]$. For this reason, a simple linear transformation of the form

$$t_{-m+i} \mapsto \hat{t}_i = (2\frac{t_{-m+i} - t_{-m+1}}{t_0 - t_{-m+1}} - 1)$$

maps each point from the original interval into $\hat{t}_i \in [-1, 1],\ i = 1, 2, \dots, m$.

To employ now the method (24) we at first approximate the Fourier-Legendre coefficients of the function

$$f_\delta(\hat{t}) = g_\delta(t_{-m+1} + 2^{-1}SH(1 + \hat{t})).$$

by means of the quadrature rule (16). Once the vector of the quadrature weights (w_i) is determined we obtain the reconstruction of the derivative of a function at the boundary point $\hat{t} = \hat{t}_m = 1$ by means of

$$\mathcal{S}_n(h;\mathbf{f}_\delta)(1) = \frac{d}{d\hat{t}} \left(\sum_{i=1}^{m} w_i f_\delta(\hat{t}_i) \Phi_n(h;\hat{t},\hat{t}_i) \right) \Bigg|_{\hat{t}=1} \tag{35}$$

where $n \in \{1,2,\ldots,n_M\}$ is the adaptively chosen by means of the balancing principle and $h(x) \in \mathbb{R}^+$ is the filter function of the form

$$h(x) = \begin{cases} 1, & x \in [0,1/2], \\ \exp\left(\dfrac{-\exp(2/(1-2x))}{1-x} \right), & x \in (1/2,1), \\ 0, & x \in [1,\infty). \end{cases}$$

In order to apply the balancing principle (31), one at first needs to specify the value of the constant B that appears in the estimate on the noise propagation error. This constant could be found as follows: we form a training set that consists of BG-measurements of one patient and find a value of B that leads to a good performance of the principle (31) on simulated data. Then this value of B is used for all other test cases. As a result of such an adjustment procedure, we have found $B = 0.004$.

Once, the estimate (35) is calculated, we can construct the predictor (34) with

$$g'(t_0) \approx \frac{2}{t_0 - t_{-m+1}} \mathcal{S}_n(h;\mathbf{f}_\delta)(1). \tag{36}$$

Recall that at the beginning we transformed the data points from the interval $[t_{-m+1}, t_0]$ into the interval $[-1,1]$, with (36) we perform the inverse transformation.

To illustrate how these predictors work we use data set of 100 virtual subjects which are obtained from Padova/University of Virginia simulator [16]. For each in silico patient BG-measurements have been simulated and sampled with a frequency of 5 (min) during 3 days. These simulated measurements have been corrupted by random white noise with the standard deviation δ of 6 (mg/dL). We perform our illustrative tests with data of the same 10 virtual subjects that have been considered in [25], [31].

To quantify the clinical accuracy of the considered predictors, we use the Prediction Error-Grid Analysis (PRED-EGA) [31], which has been designed especially for the blood glucose predictors. This assessment methodology records reference glucose estimates paired with the estimates predicted for the same moments. As a result, the PRED-EGA reports the numbers (in percent) of Accurate (Acc.), Benign (Benign) and Erroneous (Error) predictions in hypoglycemic (0–70 mg/dL), euglycemic (70–180 mg/dL) and hyperglycemic (180–450 mg/dL) ranges. This stratification is of great importance because consequences caused by a prediction error in the hypoglycemic range are very different from ones in the euglycemic range. We would like to stress that the assessment has been done with respect to the references given as simulated noise-free BG-readings.

Table 1 demonstrates the performance assessment matrix given by the PRED-EGA for 15 (min) ahead glucose predictions by the linear extrapolation predictors, where the derivative is estimated by means of (35), (36) with the parameter chosen in accordance with (31), operating on simulated noisy data with $SH = 30$ (min).

Table 1 The performance assessment matrix given by the PRED-EGA for the linear extrapolation predictors, where the derivative is found by (35), (36) with a truncation level chosen by the balancing principle (31), operating on simulated noisy data with $PH = 15$ (min) and $SH = 30$ (min)

Patient	BG $\leq$ 70 (mg/dL) (%)			BG 70-180 (mg/dL) (%)			BG $\geq$180 (mg/dL) (%)		
Vir. ID	Acc.	Benign	Error	Acc.	Benign	Error	Acc.	Benign	Error
1	-	-	-	99.88	0.12	-	100	-	-
2	-	-	-	99.88	0.12	-	-	-	-
3	-	-	-	99.88	0.12	-	-	-	-
17	99.69	0.31	-	100	-	-	-	-	-
18	99.71	0. 29	-	100	-	-	-	-	-
24	100	-	-	99.81	0.19	-	-	-	-
33	99.71	0.29	-	99.21	0.79	-	100	-	-
34	99.60	0.40	-	97.32	2.34	0.33	100	-	-
42	100	-	-	99.84	0.16	-	100	-	-
47	99.47	0.53	-	98.13	1.67	0.20	100	-	-
Avg.	**99.74**	**0.26**	**-**	**99.40**	**0.55**	**0.05**	**100**	**-**	**-**

We perform the comparison of the constructed predictors with the predictors considered in [13, **?**], where the derivative in (34) is estimated by means of the modified version of the Savitzky-Golay filtering technique [30], which is also based on the differentiation of algebraic polynomials approximating the function that has to be differentiated. In our experiments, to choose the degree of these polynomials we employ the balancing principle (see [18] for further details) in the same modification as above. The performance of such predictors is displayed in Table 2. The comparison of both tables allows us to conclude that the predictors (34), (35), (36) outperform the predictors based on the modified version of the Savitzky-Golay technique.

As mentioned in Introduction, one could also consider one-sided finite difference formulas for approximating the derivative $g'(t_0)$. We do not do so here, since it is clearly demonstrated in [18] that the Savitzky–Goldy filtering technique already yields superior perfromance than that obtained by the use of these formulas.

5 Proofs

We will organize the proofs of the results in Section 3 as follows. First, we prove a number of preparatory results, which are independent of the data set and the choice of the weight functions. This is done in Section 5.1. The proofs of the results in Section 3 are then completed in Section 5.2

5.1 Preparatory results

It is convenient to prove first the results preparatory for Theorem 3. The proof of Theorem 3 consists of three major steps. The first step is to prove the analogues of the classical Favard and Bernstein inequalities. These inequalities are not new, but we believe that the proofs presented in the current paper are new and interesting. The second step in the proof of Theorem 3 is to obtain a simultaneous

Table 2 The performance assessment matrix given by the PRED-EGA for the linear extrapolation predictors, where the derivative is found by the modified version of the Savitzky-Golay filtering technique with a truncation level chosen by the balancing principle (31), operating on simulated noisy data with $PH = 15$ (min) and $SH = 30$ (min) [18]

Patient	BG $\leq$ 70 (mg/dL) (%)			BG 70-180 (mg/dL) (%)			BG $\geq$180 (mg/dL) (%)		
Vir. ID	Acc.	Benign	Error	Acc.	Benign	Error	Acc.	Benign	Error
1	-	-	-	99.88	0.12	-	100	-	-
2	-	-	-	99.88	0.12	-	-	-	-
3	-	-	-	99.88	0.12	-	-	-	-
17	99.69	0.31	-	100	-	-	-	-	-
18	99.71	0. 29	-	100	-	-	-	-	-
24	100	-	-	99.81	0.19	-	-	-	-
33	99.71	0.29	-	99.80	-	0.20	100	-	-
34	99.60	0.40	-	95.32	4.18	0.50	57.14	42.86	-
42	100	-	-	98.35	1.65	-	100	-	-
47	99.73	0.27	-	96.88	2.92	0.21	100	-	-
Avg.	**99.78**	**0.22**	**-**	**98.98**	**0.93**	**0.091**	**91.43**	**8.57**	**-**

approximation theorem. Finally, in Section 5.2, we will obtain an estimate on the norm and approximation capabilities of the operators $\mathcal{S}_n$. These three results will be combined to yield a proof of Theorem 3.

In order to prove the Favard and Bernstein type inequalities, we prove first the bounds and approximation properties of the operators σ_n (23). To this end, for any sequence $(a_k)_{k=0}^{\infty}$ of real numbers we define Fejér summation

$$F_n\left((a_k)_{k=0}^{\infty}\right) := \frac{1}{n}\sum_{m=1}^{n}\sum_{k=0}^{m-1} a_k = \sum_{k=0}^{n}\left(1-\frac{k}{n}\right) a_k.$$

We note the following simple proposition, obtained using a summation by parts arguments (cf. [12, Theorem 71, p. 128]).

Proposition 1 *Let $(a_k)_{k=0}^{\infty}$ and $(h_k)_{k=0}^{\infty}$ be real sequences with $h_k = 0$ for all sufficiently large k. Then*

$$\sum_{k=0}^{\infty} h_k a_k = \sum_{\ell=1}^{\infty} \ell(h_{\ell+1} - 2h_\ell + h_{\ell-1}) F_\ell\left((a_k)_{k=0}^{\infty}\right). \tag{37}$$

In the sequel we abbreviate $F_n(((k+1/2)\hat{f}(k)P_k)_{k=0}^{\infty})$ by $F_n(f)$.

The next well-known result [1],[32] shows that the norms of the operators $f \mapsto F_n(f)$ are bounded in n.

Proposition 2 *Let $f \in C$. Then*

$$\|F_n(f)\|_C \leq c\|f\|_C, \qquad n = 1, 2, \cdots.$$

With the propositions above, we can now prove the following theorem that guarantees boundedness of the summability operator (23).

Theorem 5 *Let $h : [0, \infty) \to \mathbb{R}$ be twice continuously differentiable, and $h(t) = 0$ if $t \geq 1$. Then for any $f \in C$, the following holds*

$$\|\sigma_n(h, f)\|_C \leq c \max_{t\in[0,\infty)} |h''(t)| \|f\|_C, \tag{38}$$

Proof We use Proposition 1 with $h_k = h(k/n)$ and $a_k = (k+1/2)\hat{f}(k)P_k$ to obtain

$$\sigma_n(h, f) = \sum_{k=0}^{\infty} h\left(\frac{k}{n}\right) \hat{f}(k)(k + 1/2)P_k = \sum_{\ell=1}^{\infty} \ell(h_{\ell+1} - 2h_\ell + h_{\ell-1})F_\ell(f).$$

Therefore, in view of Proposition 2, we deduce that

$$\|\sigma_n(h, f)\|_C \leq c \sum_{\ell=1}^{\infty} \ell \left| h\left(\frac{\ell+1}{n}\right) - 2h\left(\frac{\ell}{n}\right) + h\left(\frac{\ell-1}{n}\right)\right| \|f\|_C. \tag{39}$$

We use Taylor's theorem to estimate the sum above. Since h is supported on $[0, 1]$,

$$\sum_{\ell=1}^{\infty} \ell \left| h\left(\frac{\ell+1}{n}\right) - 2h\left(\frac{\ell}{n}\right) + h\left(\frac{\ell-1}{n}\right)\right| \leq \max_{t\in[0,\infty)} |h''(t)| \sum_{\ell=1}^{n+1} \frac{\ell}{n^2} \leq c \max_{t\in[0,\infty)} |h''(t)|.$$

Together with (39), this leads to (38).

Remark 3 Theorem 5 was proved in [21] with the additional condition that h is a constant in a neighborhood of 0.

As a corollary of Theorem 5, we note the following [22, Proposition 3.1].

Corollary 1 *Let $h : [0, \infty) \to \mathbb{R}$ be twice continuously differentiable low pass filter and let $f \in C$.*

(a) For any $P \in \Pi_{n/2}$, $\sigma_n(h, P) = P$.

(b) There exists $c = c(h)$ such that

$$E_n(f) \leq \|f - \sigma_n(h, f)\|_C \leq cE_{n/2}(f). \tag{40}$$

With this preparation, we are ready to prove the following Favard and Bernstein estimates.

Theorem 6 *(a) Let f and $\Delta(f)$ be continuous on $[-1, 1]$. Then*

$$E_n(f) \leq \frac{c}{n^2} E_n(\Delta(f)), \qquad n \geq 1. \tag{41}$$

(b) Let $n \geq 1$ and $P \in \Pi_n$. Then

$$\|\Delta(P)\|_C \leq cn^2 \|P\|_C. \tag{42}$$

Proof In this proof, let $h : [0,\infty) \to \mathbb{R}$ be a fixed, twice continuously differentiable low pass filter, and $n \geq 1$ be an integer.

We first prove the part (b) of the theorem. Let $h_{1,n}(t) = t(t+1/n)h(t)$. Since h is supported on $[0,1]$, so is $h_{1,n}$ and

$$\max_{t\in[0,\infty)} |h''_{1,n}(t)| = \max_{t\in[0,1]} |h''_{1,n}(t)| \leq c, \qquad n \geq 1. \tag{43}$$

For any $P \in \Pi_n$ we can use the representation

$$P(t) = \sum_{k=1}^{n-1} \hat{P}(k)(k+1/2)P_k(t).$$

Then using (8), (12) and the definition of $h(t)$ we can express $\Delta(P)$ as follows

$$\begin{aligned}\Delta(P) &= \sum_{k=0}^{n-1} k(k+1)(k+1/2)\hat{P}(k)P_k = \sum_{k=0}^{n-1} h\left(\frac{k}{2n}\right) k(k+1)(k+1/2)\hat{P}(k)P_k \\ &= 4n^2 \sum_{k=0}^{n} h_{1,2n}\left(\frac{k}{2n}\right)(k+1/2)\hat{P}(k)P_k = 4n^2\sigma_{2n}(h_{1,2n}, P).\end{aligned}$$

Using Theorem 5 with $h_{1,2n}$ in place of h and (43), we obtain

$$\|\Delta(P)\|_C = 4n^2\|\sigma_{2n}(h_{1,2n}, P)\|_C \leq cn^2 \max_{t\in[0,\infty)} |h''_{1,2n}(t)| \|P\|_C \leq cn^2\|P\|_C.$$

This proves part (b).

To prove part (a), let $g(t) = h(t) - h(2t)$. Then for any $k \geq 0$, and integers $m \geq \nu + 1 \geq 1$, the following equality holds

$$\sum_{\ell=\nu+1}^{m} g(k2^{-\ell}) = \sum_{\ell=\nu+1}^{m} h(k2^{-\ell}) - \sum_{\ell=\nu+1}^{m} h(k2^{-\ell+1}) = h(k2^{-m}) - h(k2^{-\nu}).$$

Consequently,

$$\sigma_{2^m}(h,f) - \sigma_{2^\nu}(h,f) = \sum_{\ell=\nu+1}^{m} \sigma_{2^\ell}(g,f).$$

In view of (40), it is clear that $\sigma_{2^m}(h,f) \to f$ as $m \to \infty$. Thus,

$$f - \sigma_{2^\nu}(h,f) = \sum_{\ell=\nu+1}^{\infty} \sigma_{2^\ell}(g,f), \tag{44}$$

Now, let

$$g_{1,n}(t) = \frac{g(t)}{t(t+1/n)}, \qquad t > 0, n \geq 1.$$

Since $g(t) = h(t) - h(2t)$, and $h(t)$ is supported on $t \in [0,1]$ such that it is a constant for $t \in [0,1/2]$, it is clear that g is supported on $[1/4,1]$. Moreover, $g_{1,n}$ is twice continuously differentiable on $[0,\infty)$, and

$$\max_{t\in[0,\infty)} |g''_{1,n}(t)| < c, \qquad n \geq 1. \tag{45}$$

We note for any $\nu \geq 1$ and $\ell \geq \nu+1$,

$$\begin{aligned}\sigma_{2^\ell}(g,f) &= \sum_{k=0}^{2^\ell} g\left(\frac{k}{2^\ell}\right)(k+1/2)\hat{f}(k)P_k \\ &= \sum_{k=0}^{2^\ell} \frac{g(k2^{-\ell})}{k(k+1)} k(k+1)(k+1/2)\hat{f}(k)P_k \\ &= 2^{-2\ell}\sum_{k=0}^{2^\ell} g_{1,2^\ell}\left(\frac{k}{2^\ell}\right)(k+1/2)\widehat{\Delta(f)}(k)P_k = 2^{-2\ell}\sigma_{2^\ell}(g_{1,2^\ell},\Delta(f)).\end{aligned} \tag{46}$$

Therefore, from the estimates (40), (44), (46), (38), and (45), we conclude that

$$\begin{aligned}E_{2^\nu}(f) &\leq \|f-\sigma_{2^\nu}(h,f)\|_C \leq \sum_{\ell=\nu+1}^{\infty}\|\sigma_{2^\ell}(g,f)\|_C = \sum_{\ell=\nu+1}^{\infty} 2^{-2\ell}\|\sigma_{2^\ell}(g_{1,2^\ell},\Delta(f))\|_C \\ &\leq c\|\Delta(f)\|_C \sum_{\ell=\nu+1}^{\infty} 2^{-2\ell} \leq c2^{-2\nu}\|\Delta(f)\|_C.\end{aligned} \tag{47}$$

Since the sequence $\{E_j(f)\}_{j=0}^{\infty}$ is non–increasing, this leads to the estimate

$$E_n(f) \leq cn^{-2}\|\Delta(f)\|_C, \qquad n \geq 1. \tag{48}$$

Now, without loss of generality, we can choose $R_1 \in \Pi_n$ so that $\|\Delta(f)-R_1\|_C \leq 2E_n(\Delta(f))$. Since $\widehat{\Delta(f)}(0)=0$, we may estimate the first Fourier-Legendre coefficient $\hat{R}_1(0)$ of R_1 such that

$$|\hat{R}_1(0)| = |\hat{R}_1(0)-\widehat{\Delta(f)}(0)| \leq \int_{-1}^{1}|R_1(u)-\Delta(f)(u)|du \leq c\|R_1-\Delta(f)\|_C \leq cE_n(\Delta(f)).$$

We also define the polynomial $R = R_1 - \hat{R}_1(0)$ that satisfies $\hat{R}(0)=0$. Using the above estimations, it is easy to see that $\|R-\Delta(f)\|_C \leq cE_n(\Delta(f))$. Let

$$P = \sum_{k=1}^{n-1}\frac{k+1/2}{k(k+1)}\hat{R}(k)P_k$$

be a polynomial in Π_n. Then it follows that $\Delta(P)=R$, and using (48), we obtain the following estimate

$$E_n(f) = E_n(f-P) \leq cn^{-2}\|\Delta(f-P)\|_C = cn^{-2}\|R-\Delta(f)\|_C \leq c_1n^{-2}E_n(\Delta(f)).$$

The second step in our proof of Theorem 3 is the following simultaneous approximation theorem.

Theorem 7 *Let f and $\Delta(f)$ be in C, and $P \in \Pi_n$, $n \geq 1$. Then*

$$\|\Delta(f)-\Delta(P)\|_C \leq c\left\{E_{n/2}(\Delta(f)) + n^2\|f-P\|_C\right\}. \tag{49}$$

Proof Here h is a fixed function as in the proof of Theorem 6. The key observation that can be verified from (12) is that $\Delta(\sigma_n(h,f)) = \sigma_n(h,\Delta(f))$. Since $\sigma_n(h,f) \in \Pi_n$, we obtain from the Bernstein inequality (42), (40), and Favard inequality (41) that

$$\begin{aligned}
\|\Delta(f)-\Delta(P)\|_C &\le \|\Delta(f)-\Delta(\sigma_n(h,f))\|_C + \|\Delta(\sigma_n(h,f))-\Delta(P)\|_C\\
&= \|\Delta(f)-\sigma_n(h,\Delta(f))\|_C + cn^2\|\sigma_n(h,f)-P\|_C\\
&\le c_1E_{n/2}(\Delta(f)) + cn^2\|\sigma_n(h,f)-f\|_C + cn^2\|f-P\|_C\\
&\le c_1E_{n/2}(\Delta(f)) + c_2n^2E_{n/2}(f) + cn^2\|f-P\|_C\\
&\le c_1E_{n/2}(\Delta(f)) + c_3E_{n/2}(\Delta(f)) + cn^2\|f-P\|_C,
\end{aligned}$$

that implies (49).

Next, we prove some results preparatory for the proof of Theorem 2. The main difference here is that when we use the operators $s_n(f)$, the analogue of Corollary 1 is weaker. The analogue of Corollary 1 is the following statement.

Proposition 3 *Let $f \in C$, $n \ge 1$.*
(a) *For any $P \in \Pi_n$, $s_n(P) = P$.*
(b) *We have*

$$\|s_n(f)\|_C \le cn^{1/2}\|f\|_C. \tag{50}$$

Consequently,

$$E_n(f) \le \|f - s_n(f)\|_C \le cn^{1/2}E_n(f). \tag{51}$$

Proof Part (a) is clear from the definitions. To prove part (b), we take as the starting point the integral representation

$$s_n(f)(x) = \int_{-1}^{1} f(t)K_n(x,t)dt, \qquad x \in \mathbb{R}, \tag{52}$$

where the *Christoffel–Darboux kernel* K_n is defined by

$$K_n(x,t) := \sum_{k=0}^{n}(k+1/2)P_k(x)P_k(t), \qquad x,t \in \mathbb{R}. \tag{53}$$

It is well known [29], [17] that

$$\max_{x\in[-1,1]}\int_{-1}^{1}|K_n(x,t)|dt = \max_{t\in[-1,1]}\int_{-1}^{1}|K_n(x,t)|dx = \sqrt{\frac{2n}{\pi}} + o(n^{-1/2}). \tag{54}$$

Together with (52), this leads to

$$\|s_n(f)\|_C \le cn^{1/2}\|f\|_C. \tag{55}$$

The first inequality in (51) is clear. If $P \in \Pi_n$ is arbitrary, we use part (a) and (50) to conclude that

$$\|f - s_n(f)\|_C = \|f - P - s_n(f-P)\|_C \le cn^{1/2}\|f-P\|_C.$$

This leads to (51).

The analogue of Theorem 7 with $E_n(f)$ in place of $E_{n/2}(f)$, and an extra mulitiplicative factor of $n^{1/2}$ is the following.

Proposition 4 *Let f and $\Delta(f)$ be in C, $n > 1$, and $P \in \Pi_n$. Then*

$$\|\Delta(f) - \Delta(P)\| \le cn^{1/2}\left\{E_n(\Delta(f)) + n^2\|f - P\|\right\}. \tag{56}$$

The proof is verbatim the same as that of Theorem 7, except that $s_n(f)$ is used in place of $\sigma_n(h; f)$, and the bounds on the norms are used from Proposition 3 rather than Corollary 1. We omit this proof.

5.2 Proofs of the results in Section 3

In this section, we assume the set up as in Section 3. Thus, we assume that $\{t_j\}_{j=1}^M \subset [-1, 1]$, and an integer $n \ge 1$ and real numbers w_j are found so as to satisfy (16). We assume further that (18) is satisfied. As the final ingredient in the proof of Theorem 3, we state the analogues of Theorem 5 and Corollary 1, proved in [22, Proposition 3.1].

Proposition 5 *Let $h : [0, \infty) \to \mathbb{R}$ be twice continuously differentiable low pass filter, $\mathbf{y} = (y_1, \cdots, y_M) \in \mathbb{R}^M$, and $f \in C[-1, 1]$.*
(a) *We have*

$$\|\mathcal{S}_n(h; \mathbf{y})\|_C \le cA \max_{1\le j\le M} |y_j|. \tag{57}$$

(b) *For any $P \in \Pi_{n/2}$, $\mathcal{S}_n(h; P) = P$.*
(c) *There exists $c = c(h)$ such that*

$$E_n(f) \le \|f - \mathcal{S}_n(h; f)\|_C \le cAE_{n/2}(f). \tag{58}$$

We are now in a position to prove Theorem 3.

Proof (Proof of Theorem 3) From (58) and (41), we obtain

$$\|f - \mathcal{S}_n(h; f)\|_C \le cAE_{n/2}(f) \le \frac{cA}{n^2} E_{n/2}(\Delta(f)).$$

Consequently, Theorem 7 implies that

$$\|\Delta(f) - \Delta(\mathcal{S}_n(h; f))\|_C \le cAE_{n/2}(\Delta(f)). \tag{59}$$

Since $\Delta(f)(1) = 2f'(1)$ for any f (7), (59) implies (26).
In order to prove (27), we use (42) and (57) to deduce that

$$|\mathcal{S}_n'(h; \mathbf{y})(1)| = \frac{1}{2}|\Delta(\mathcal{S}_n(h; \mathbf{y}))(1)| \le \frac{1}{2}\|\Delta(\mathcal{S}_n(h; \mathbf{y}))\|_C \tag{60}$$

$$\le cn^2\|\mathcal{S}_n(h; \mathbf{y})\|_C \le cAn^2 \max_{1\le j\le M} |y_j|. \tag{61}$$

The estimate (28) follows easily by applying (27) with $\mathbf{y} = \mathbf{f} - \mathbf{f}_\delta$ and using the resulting estimate together with (26) and triangle inequality.

The proof of Theorem 2 is very similar. In place of Proposition 5, we need the following weaker analogue, which is proved in exactly the same way. We will sketch the proof for the sake of completeness.

Proposition 6 *Let* $\mathbf{y} = (y_1, \cdots, y_M) \in \mathbb{R}^M$*, and* $f \in C[-1,1]$.
(a) *We have*

$$\|S_n(h;\mathbf{y})\|_C \le cA\sqrt{n} \max_{1\le j\le M} |y_j|. \tag{62}$$

(b) *For any* $P \in \Pi_n$*,* $S_n(h;P) = P$.
(c) *We have*

$$E_n(f) \le \|f - S_n(h;f)\|_C \le cA\sqrt{n}E_n(f). \tag{63}$$

Proof In light of (18) and (54), we deduce that

$$\begin{aligned}|S_n(\mathbf{y})(x)| = \left|\sum_{j=1}^{M} w_j y_j K_n(x,t_j)\right| &\le \left(\max_{1\le j\le M}|y_j|\right)\sum_{j=1}^{M}|w_j||K_n(x,t_j)| \\ &\le A\left(\max_{1\le j\le M}|y_j|\right)\int_{-1}^{1}|K_n(x,t)|dt \le cAn^{1/2}\max_{1\le j\le M}|y_j|. \end{aligned}\tag{64}$$

Next, let $P \in \Pi_n$. Using (14), valid for $PK_n(x,\cdot) \in \Pi_{2n}$, we deduce that

$$S_n(P)(x) = \sum_{j=1}^{M} w_j P(t_j) K_n(x,t_j) = \int_{-1}^{1} P(t)K_n(x,t)dt = s_n(P)(x) = P(x).$$

This proves part (b).

The proof of part (c) is verbatim the same as that of Proposition 3 (c).

We are now in a position to prove Theorem 2.

Proof (Proof of Theorem 2) The first part of this theorem is proved in exactly the same way as Theorem 3. From (63) and (41), we obtain

$$\|f - S_n(f)\|_C \le cAn^{1/2}E_n(f) \le \frac{cAn^{1/2}}{n^2}E_n(\Delta(f)).$$

Consequently, Proposition 4 implies that

$$\|\Delta(f) - \Delta(S_n(f))\|_C \le cAn^{1/2}E_n(\Delta(f)). \tag{65}$$

Since $\Delta(f)(1) = 2f'(1)$ for any f (7), (65) implies

$$|f'(1) - S_n'(f)(1)| \le cAn^{1/2}E_n(\Delta(f)). \tag{66}$$

We estimate $|S_n'(f)(1) - S_n'(\mathbf{f}_\delta)(1)|$ by (62). Together with (66), this estimate leads to (21).

Acknowledgements The research of the first author was supported, in part, by grant DMS-0908037 from the National Science Foundation and grant W911NF-09-1-0465 from the U.S. Army Research Office. The second and third authors are supported by the Austrian Fonds Zur Förderung der Wissenschaftlichen Forschung (FWF), Grant P25424.

References

1. Askey, R., Wainger, S.: A convolution structure for Jacobi series, American Journal of Mathematics 91 (2), 463–485 (1969)
2. Bergh, J., Löfström, J.: Interpolation Spaces, an Introduction. Springer Verlag, Berlin (1976)
3. Chan, T.F., Shen, J.: Mathematical models for local nontexture inpaintings, SIAM J. Appl. Math. 62, 1019–1043 (2002)
4. Cheng, J., Jia, X. Z., Wang, Y. B.: Numerical differentiation and its applications, Inverse Problems Sci. Eng. 15, 339–357(2007).
5. Chui, C. K., Mhaskar, H. N.: MRA Contextual-Recovery Extension of Smooth Functions on Manifolds, Appl. Comput. Harmon. Anal. 28, 104–113 (2010)
6. DeVore, R. A., Lorentz, G. G.: Constructive Approximation. Springer Verlag, Berlin (1993)
7. Engl, H., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer, Dordrecht (1996)
8. Engl, H., Fusek, P., Pereverzev, S. V.: Natural linearization for the identification of nonlinear heat transfer laws, Journal of Inverse and Ill-posed Problems 13, 567–582 (2005)
9. Filbir, F., Mhaskar, H.N.: Marcinkiewicz–Zygmund measures on manifolds, Journal of Complexity 27, 568–596 (2011)
10. Filbir, F., Mhaskar, H.N., Prestin, J.: On a filter for exponentially localized kernels based on Jacobi polynomials, Journal of Approximation Theory 160, 256–280 (2009)
11. Hämarik, U., Raus, T.: About the balancing principle for choice of the regularization parameter, Numer. Functional Anal. and Optimiz. 30, 951–970 (2009)
12. Hardy, G.H.: Divergent series. AMS Chelsea Publ., Amer. Math. Soc., Providence (1991)
13. Hayes, A., Mastrototaro, J. J., Moberg, S. B., Mueller Jr., J. C., Bud Clark, H., Charles Vallet Tolle, M., Williams, G. L., Wu, B., Steil, G. M.: Algorithm sensor augmented bolus estimator for semi-closed loop infusion system, US Patent Application Publication US2009/0234213 (2009)
14. Hu, B., Lu, S.: Numerical differentiation by a Tikhonov regularization method based on the discrete cosine transform, Applicable Analysis 91, 719-736 (2012)
15. Kovatchev, B., Clarke, W.: Peculiarities of the Continuous Glucose Monitoring Data Stream and Their Impact on Developing Closed-Loop Control Technology, J. Diabetes Sci Technol. 2, 158–163 (2008)
16. Kovatchev, B., Breton, M., Dalla Man, C., Cobelli, C.: In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes, J. Diabetes Sci Technol. 3, 44–55 (2009)
17. Lorch, L.: The Lebesgue constants for Jacobi series I, Proc. Amer. Math. Soc. 10, 756–761 (1959)
18. Lu, S., Naumova, V., Pereverzev, S. V.: Legendre polynomials as a recommended basis for numerical differentiation in the presence of stochastic white noise, Journal of Inverse and Ill-posed Problems 20, 1–22 (2012)
19. Lu, S., Pereverzev, S. V.: Numerical differentiation from a viewpoint of regularization theory, Math. Comp. 75, 1853–1870 (2006)
20. Mathé, P.; Pereverzev, S. V.: Regularization of some linear ill-posed problems with discretized random noisy data, Mathematics of Computation 75, 1913–1929 (2006)
21. Mhaskar, H. N.: Polynomial operators and local smoothness classes on the unit interval, J. Approx. Theory 131, 243–267 (2004)
22. Mhaskar, H. N.: Polynomial operators and local smoothness classes on the unit interval II, Jaén J. of Approx. 1, 1–25 (2005)
23. Mhaskar, H. N., Prestin, J.: Polynomial operators for spectral approximation of piecewise analytic functions, Appl. Comput. Harmon. Anal. 26, 121–142 (2009)
24. Natanson, I. P.: Constructive function theory, vol. I. Frederick Ungar, New York (1964)
25. Naumova, V., Pereverzyev, S.V., Sivananthan, S.: Adaptive Parameter Choice for One-Sided Finite Difference Schemes and its Application in Diabetes Technology, Journal of Complexity 28, 524–538 (2012)
26. Naumova, V., Pereverzyev, S.V., Sivananthan, S.: A meta-learning approach to the regularized learning – Case study: Blood glucose prediction, Neural Networks 33, 181–193 (2012)
27. Nevai, P.: Orthogonal Polynomials. Mem. Amer. Math. Soc. 203. Providence, Amer. Math. Soc. (1979)
28. Ramm, A. G., Smirnova, A. B.: On Stable Numerical Differentiation, Math. Comp. 70, 1131–1153 (2001).

29. Rau, H.: Über die Lebesgueschen Konstanten der Reihenentwicklungen nach Jacobischen Polynomen, J. Reine Angew. Math. 161, 237–254 (1929)
30. Savitzky, A., Golay M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry 36, 1627–1639 (1964)
31. Sivananthan, S., Naumova, V., Dalla Man, C., Facchinetti, A., Renard, E., Cobelli, C., Pereverzyev, S.V.: Assessment of Blood Glucose Predictors: The Prediction-Error Grid Analysis, Diabetes Technol Ther 13, 787–796 (2011)
32. Szegö, G: Orthogonal polynomials. Amer. Math. Soc. Colloq. Publ. 23, Amer. Math. Soc., Providence(1975)
33. Winch, D. E., Roberts P. H.: Derivatives of additions theorems for Legendre functions, J. Austral. Math. Soc. Ser. B 37, 212–234 (1995)
34. Zhao, Z.: A truncated Legendre spectral method for solving numerical differentiation, International Journal of Computer Mathematics 87, 3209–3217 (2010)