J. Inverse Ill-Posed Probl., Ahead of Print
DOI 10.1515/jip-2012-0050

Legendre polynomials as a recommended basis for numerical differentiation in the presence of stochastic white noise

Shuai Lu, Valeriya Naumova and Sergei V. Pereverzev

Abstract. In this paper, we consider the problem of estimating the derivative y' of a function $y \in C^1[-1,1]$ from its noisy version y_δ contaminated by a stochastic white noise and argue that in certain relevant cases the reconstruction of y' by the derivatives of the partial sums of Fourier–Legendre series of y_δ has advantage over some standard approaches. One of the interesting observations made in the paper is that in a Hilbert scale generated by the system of Legendre polynomials the stochastic white noise does not increase, as it might be expected, the loss of accuracy compared to the deterministic noise of the same intensity. We discuss the accuracy of the considered method in the spaces L_2 and C and provide a guideline for an adaptive choice of the number of terms in differentiated partial sums (note that this number is playing the role of a regularization parameter). Moreover, we discuss the relation of the considered numerical differentiation scheme with the well-known Savitzky–Golay derivative filters, as well as possible applications in diabetes technology.

Keywords. Numerical differentiation, Legendre polynomials, stochastic white noise, adaptive parameter choice, Savitzky–Golay method, diabetes technology.

2010 Mathematics Subject Classification. 65D25, 42C10, 65R30, 60H40.

1 Introduction and brief overview

Numerical differentiation is known as a classical ill-posed problem [4, 11, 32]. It consists in approximating the derivative y' of a function y, which is defined and differentiable on, say, the interval $[-1,1]$, from information about its noisy counterpart

$$y_\delta = y + \delta\xi, \tag{1.1}$$

where δ is a small positive number used for measuring noise intensity level, and ξ is a noise element, which is assumed to be normalized somehow.

If ξ is supposed to be a continuous function on $[-1,1]$, then the problem of numerical differentiation is well studied (see, e.g., [15] and references therein).

At the same time, in classical regularization theory the problem of numerical differentiation has been considered mostly within the Hilbert space setting, where

ξ is assumed to belong to the Hilbert space $L_2 = L_2(-1,1)$ of square integrable functions, such that $\|\xi\| \le 1$ and

$$\|y - y_\delta\|_{L_2} \le \delta. \tag{1.2}$$

Then one is sometimes advised to find $x(t) = y'(t)$ from the ill-posed integral equation

$$Ax(t) := \int_{-1}^{t} x(\tau)d\tau = y_\delta(t) - y_\delta(-1). \tag{1.3}$$

Note that the above mentioned noise model do not cover the case of the so-called random "white" noise, which is widely considered in statistical analysis of physical measurements. Recall that this noise is assumed to be generated by an L_2-valued (weak) random variable ξ defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\xi = \xi(\omega, t)$, $\omega \in \Omega$, $t \in [-1, 1]$, is centered, has second weak moments and for any $g \in L_2$ it holds

$$\begin{aligned}\mathbb{E}\langle \xi, g\rangle &:= \int_\Omega \langle \xi(\omega,\cdot), g\rangle d\mathbb{P}(\omega)\\ &= \int_\Omega \int_{-1}^{1} \xi(\omega,t)g(t)dt\,d\mathbb{P}(\omega) = 0, \end{aligned}\tag{1.4}$$

$$\mathbb{E}|\langle \xi, g\rangle|^2 = \|g\|^2_{L_2}. \tag{1.5}$$

Note that random (stochastic) white noise ξ is not expected to belong to L_2. Indeed, for any orthonormal system $\{e_k\}_{k=1}^\infty$ in L_2 we have $\mathbb{E}|\langle \xi, e_k\rangle|^2 = 1$, which means that the expected value of the sum $\sum_{k=1}^n |\langle \xi, e_k\rangle|^2$ tends to infinity as $n \to \infty$, whereas for any element of L_2 such a sum should be bounded.

This last remark means that in the case of stochastic white noise ξ, instead of complete data, generically expressed as y_δ, we should limit ourselves to a finite set of noisy Fourier coefficients

$$y_k^\delta = \langle y_\delta, e_k\rangle := \int_{-1}^{1} y_\delta(\tau)e_k(\tau)d\tau, \quad k = 1, 2, \dots, N, \tag{1.6}$$

with respect to some L_2-orthonormal system $\{e_k\}_{k=1}^\infty$, which is sometimes called a design. Note that from (1.4), (1.5) it follows that the integrals (1.6) exist, at least with probability one.

Remark that each design system $\{e_k\}$ potentially generates a variable Hilbert scale $\{W^\psi\{e_k\}\}$ which is a family of spaces

$$W^\psi\{e_k\} := \left\{ y \in L_2 : \|y\|_\psi^2 := \sum_{k=1}^{\infty} \psi^2(k)|\langle e_k, y\rangle|^2 < \infty \right\}$$

labeled by the so-called index functions $\psi : [1,\infty) \to (0,\infty)$, as introduced in [9] (see also, [18]). In particular, for the case of $\psi(k) = k^r$, $r > 0$, we obtain a commonly introduced Hilbert scale $\{W^r\{e_k\}\}$, where r plays the role of a regularity index [3].

Starting from the work [34] the concept of Hilbert scales $\{W^r\{e_k\}\}$ has been used to quantify the ill-posedness of problems presented in the form of operator equations $Ax = y_\delta$. Such quantification is usually made when the design $\{e_k\}$ generating a scale $\{W^r\{e_k\}\}$ consists of singular vectors of the operator A (see, e.g., [16]).

It is known [13] that for numerical differentiation reduced to equation (1.3) such design is formed by the functions

$$e_k(t) = e_k^I(t) = 2^{-1/4}\sin\frac{\pi}{2}\left(k+\frac{1}{2}\right)(t+1), \quad k = 1,2,\ldots,$$

and

$$e_k(t) = \bar{e}_k^I(t) = 2^{-1/4}\cos\frac{\pi}{2}\left(k+\frac{1}{2}\right)(t+1), \quad k = 1,2,\ldots.$$

Then the regularization of the problem (1.3) by means of the well-known spectral cut-off scheme leads to an approximation of the derivative y' by the derivative of the n-th partial sum

$$S_n^I y(t) = \sum_{k=1}^{n} e_k^I(t)\langle e_k^I, y\rangle$$

of Fourier series of y with noisy coefficients (1.6), i.e.,

$$y'(t) \approx \frac{d}{dt}(S_n^I y_\delta(t)) := \sum_{k=1}^{n} \frac{\pi(2k+1)}{4}\bar{e}_k^I(t)\langle e_k^I, y_\delta\rangle.$$

From [24] it follows that for a function $y \in W^r\{e_k^I\}$ the error and the expected risk of approximating y' by $\frac{d}{dt}(S_n^I y_\delta(t))$ can be estimated as follows: for deterministic noise (1.1), (1.2) and $n = c\delta^{-1/r}$ we have

$$\left\| y' - \frac{d}{dt}(S_n^I y_\delta(t))\right\|_{L_2} \le c\|y\|\delta^{\frac{r-1}{r}}, \tag{1.7}$$

while for stochastic white noise (1.1), (1.4), (1.5) and $n = c\delta^{-\frac{2}{2r+1}}$ it holds

$$\left(\mathbb{E}\left\| y' - \frac{d}{dt}(S_n^I y_\delta(t))\right\|_{L_2}^2\right)^{1/2} \le c\|y\|\delta^{\frac{r-1}{r+1/2}}, \tag{1.8}$$

where here and below we follow the convention that the symbol c denotes an absolute constant, which may not be the same at different occurrences.

For our further discussion it is important to note two things. First, for a function $y \in W^r\{e_k^I\}$ the orders of the bounds (1.7), (1.8) with respect to δ cannot be, in general, improved by using any other approximations based on noisy data y_δ. Second, in (1.7), (1.8) the symbol $\|y\|$ can mean both the norm of y in the space $W^r\{e_k^I\}$ and the norm of $y \in W^r\{e_k^I\}$ in the Sobolev space $H_2^r = H_2^r(-1,1)$, where it is defined as usual

$$\|y\|_{H_2^r} := \left(\sum_{i=0}^{r} \|y^{(i)}\|_{L_2}^2\right)^{1/2},$$

and the derivatives $y^{(i)}$ are taken in the weak sense.

The bounds (1.7), (1.8) allow us to quantify the ill-posedness of numerical differentiation of functions from $W^r\{e_k^I\}$. Our quantification of ill-posedness is based on the observation that a problem is well posed if and only if its solution can be potentially approximated with an accuracy of order $O(\delta)$ from data blurred by additive noise of intensity δ. The higher the order of the best guaranteed accuracy deviates from $O(\delta)$ the more ill-posed a problem is, and the value of such a deviation can be called the order of ill-posedness. Note that this is a bit different from the notion of degree of ill-posedness coined by Wahba [34], since the latter one may take values from $[0,\infty)$, while the order of ill-posedness varies from 0 (for a well-posed problem) to 1 (for a severe ill-posed problem).

From (1.7), (1.8) it follows that the problem of approximating the first derivative of a function $y \in W^r\{e_k^I\}$ from data y_δ blurred by deterministic noise (1.1), (1.2) has the order of ill-posedness equal to $\frac{1}{r}$, while in the case of stochastic white noise (1.1), (1.4), (1.5) this order is equal to $\frac{3}{2r+1}$. This means that in the Hilbert scale of spaces $W^r\{e_k^I\}$ the stochastic white noise increases the order of ill-posedness of numerical differentiation compared to the deterministic noise of the same intensity δ. We will show that in general such an increase is not inevitable.

Recall that the scale $\{W^r\{e_k^I\}\}$ naturally appears when the problem of numerical differentiation is reduced to the equation (1.3). Along this scale the regularity index r coincides with the Sobolev differentiability, since

$$W^r\{e_k^I\} \subset H_2^r.$$

However, as we know from [23], this scale, as well as other Hilbert scales, does not exhaust the whole Sobolev scale $\{H_2^r\}$ naturally originated from the space H_2^1, that is, the domain of the differentiation operator $\frac{d}{dt}$, in which we are interested to approximate.

For instance, as it follows from [5] (see Example 5 there), if the simple analytic function $y(t) = at + b$ is considered in a scale of spaces $W^\psi\{e_k^I\}$, then its derivative $y'(t)$ cannot be approximated from (1.1)–(1.3) in the space L_2 with the accuracy better than $O(\delta^{1/3})$. In view of (1.7), such an accuracy corresponds to the regularity index $r = \frac{2}{3}$ in the scale $\{W^r\{e_k^I\}\}$, while the considered function has arbitrary high regularity (differentiability) in the Sobolev scale $\{H_2^r\}$.

To the best of our knowledge, in the case of L_2-valued noise an error bound of numerical differentiation in the Sobolev scale $\{H_2^r\}$ has been obtained only recently. Namely, in [36] the design $\{e_k\}$ consisting of Legendre polynomials

$$e_k(t) = P_k(t) = \frac{\sqrt{k+1/2}}{2^k k!} \frac{d^k}{dt^k}[(t^2-1)^k], \quad k = 1, 2, \ldots, \tag{1.9}$$

has been considered in approximating the derivatives of functions by the derivatives of their partial sums of Fourier–Legendre series with noisy coefficients, i.e.

$$y'(t) \approx D_n y_\delta(t) := \frac{d}{dt}\left(\sum_{k=1}^{n} y_k^\delta P_k(t)\right), \tag{1.10}$$

where y_k^δ are defined by (1.6) with $e_k = P_k$, $k = 1, 2, \ldots$.

Then under condition (1.2) for $n = c\delta^{-1/r}$ and any $y \in H_2^r$, $r > 2$, the following bound has been proven in [36]:

$$\|y' - D_n y_\delta\|_{L_2} \le c\|y\|_{H_2^r} \delta^{\frac{r-2}{r}}. \tag{1.11}$$

At the same time, as we mentioned above, when dealing with the design (1.9) it is natural to consider a variable Hilbert scale of spaces

$$W_2^\psi := W^\psi\{P_k\} := \left\{ y \in L_2 : \|y\|_\psi^2 := \sum_{k=0}^{\infty} \psi^2(k)|\langle P_k, y\rangle|^2 < \infty \right\}.$$

Note that if we denote by $\{W_2^\mu\}$ the corresponding scale of spaces W_2^ψ with $\psi(k) = k^\mu$, then the regularity index μ along this scale will not coincide any more with the guaranteed Sobolev differentiability of functions from H_2^μ, since it is known from [26, 33] (see also [17]) that a space W_2^μ can be identified with weighted Sobolev space of functions y whose derivatives $y^{(\mu)}$ may not belong to L_2, but

$$\int_{-1}^{1} |y^{(\mu)}(t)|^2 (1-t^2)^\mu dt < \infty.$$

So, a space W_2^μ is essentially wider than a Sobolev space H_2^μ, but nevertheless it can be proven [14] that the derivatives of functions $y \in W_2^\mu$ can be approximated from data y_δ satisfying (1.1) and (1.2) with the same order of accuracy as that obtained in [36] for H_2^μ. More precisely, for $n = c\delta^{-1/\mu}$ and any $y \in W_2^\mu$ it holds

$$\|y' - D_n y_\delta\|_{L_2} \le c\|y\|_\mu \delta^{\frac{\mu-2}{\mu}}, \tag{1.12}$$

where $\|\cdot\|_\mu$ is the norm in W_2^ψ with $\psi(k) = k^\mu$.

From (1.12) it follows that the order of ill-posedness of the problem of approximating the first derivative of a function $y \in W_2^\mu$ from noisy data (1.1), (1.2) can be bounded by $\frac{2}{\mu}$. It is instructive to compare this bound with that previously obtained for $W^r\{e_k^I\}$. But since the scales $\{W_2^\mu\}$ and $\{W^r\{e_k^I\}\}$ are different, such a comparison should be made provided that the functions from W_2^μ have the same Sobolev differentiability as those from $W^r\{e_k^I\}$. In [14] it has been observed that $W_2^\mu \subset H_2^r$ for $r < \frac{\mu}{2}$, which can be seen as an estimation of the Sobolev differentiability r guaranteed for all functions $y \in W_2^\mu$.

Then we have $\frac{2}{\mu} < \frac{1}{r}$, and this may be interpreted as that under the same differentiability r for functions $y \in W^r\{e_k^I\}$ the order of ill-posedness of approximating the derivative y' from noisy data blurred by the deterministic noise is, in general, larger than for functions $y \in W_2^\mu$.

In the present paper we observe another interesting feature distinguishing the scale $\{W_2^\mu\}$ from the traditional scale $\{W^r\{e_k^I\}\}$. Namely, in the scale $\{W_2^\mu\}$ the stochastic white noise does not increase the loss of accuracy compared to the deterministic noise of the same intensity δ, while for the scale of spaces $W^r\{e_k^I\}$ such a loss is well known (see (1.7) and (1.8)). This feature follows from an estimation of the expected risk of numerical differentiation in the scales $\{W_2^\psi\}$ obtained in the next section. Moreover, in Section 3 we discuss a posteriori parameter choice rule that automatically adjusts the value of the truncation level n to unknown index function ψ, such that the chosen level $n = n_+$ gives a risk bound, which only by a log-factor worse than the one obtained with the knowledge of ψ. In Section 4 we consider an approximation of the derivative y' from random noisy data in the space of continuous functions and prove the corresponding risk bounds. The results of Section 3 and 4 are based on the additional assumption that random white noise is Gaussian. In the last section, we discuss the relation of the numerical differentiation scheme (1.10) with the well-known Savitzky–Golay derivative filters [29] and present numerical experiments with simulated data from diabetes technology.

2 Error bounds in L_2

With the notions above we bound the expected risk of approximating the derivative y' by $D_n y_\delta(t)$, see (1.10), by the sum of approximation and noise propagation errors:

$$\mathbb{E}\|y' - D_n y_\delta\|_{L_2}^2 = \|y' - D_n y\|_{L_2}^2 + \mathbb{E}\|D_n y - D_n y_\delta\|_{L_2}^2. \tag{2.1}$$

In this section we proceed similarly as in [14], providing first the estimates for each of the error terms and then stating and proving our main theorem.

Since the approximation error $\|y' - D_n y\|_{L_2}$ does not depend on the noise model, we refer to [14] for the proof of the following estimate.

Lemma 2.1. *For $y \in W_2^\psi$ the approximation error has the following bound, provided the integral below exists:*

$$\|y' - D_n y\|_{L_2} \le C \left(\iint_{\Omega_n} \frac{t\tau}{\psi^2(\tau)} dt\, d\tau \right)^{1/2} \|y\|_\psi, \tag{2.2}$$

where $\Omega_n = [0,n] \times [n,\infty] \cup \{(t,\tau) : n \le t \le \tau < \infty\}$.

In cases $\psi(k) = k^\mu$ and $\psi(k) = e^{kh}$, $h > 0$, the bound (2.2) *reduces to the following ones respectively:*

$$\|y' - D_n y\|_{L_2} \le c n^{2-\mu} (\mu - 2)^{-1/2} \|y\|_\mu, \quad \mu > 2, \tag{2.3}$$

and

$$\|y' - D_n y\|_{L_2} \le c \left(\frac{n^3}{h} + \frac{1}{h^4} \right)^{1/2} e^{-nh} \|y\|_\psi. \tag{2.4}$$

Lemma 2.2. *Under assumptions* (1.1), (1.4) *and* (1.5) *the following bound holds true:*

$$(\mathbb{E}\|D_n y - D_n y_\delta\|_{L_2}^2)^{1/2} \le \frac{\delta}{2} n (n^2 + 6n + 5)^{1/2}. \tag{2.5}$$

Proof. We will use the following representation of the derivatives of the Legendre polynomials [20],

$$P_k'(t) = 2\sqrt{k + 1/2} \sum_{i=0}^{(k-q_k-1)/2} \sqrt{2i + q_k + 1/2}\, P_{2i+q_k}(t), \tag{2.6}$$

where

$$q_k = \begin{cases} 0 & \text{if } k = 2\nu + 1, \\ 1 & \text{if } k = 2\nu, \end{cases} \qquad \nu = 0, 1, 2, \dots.$$

This representation leads to the following identities:

$$\begin{aligned} &D_n y(t) - D_n y_\delta(t) \\ &\quad = \delta \sum_{k=1}^{n} \langle \xi, P_k \rangle \sum_{i=0}^{(k-q_k-1)/2} 2\sqrt{k + 1/2}\sqrt{2i + q_k + 1/2}\, P_{2i+q_k}(t) \\ &\quad = \delta \sum_{j=0}^{n-1} P_j(t) \sum_{i=0}^{\lfloor (n-j-1)/2 \rfloor} 2\sqrt{j + 1/2}\sqrt{j + 2i + 3/2} \langle \xi, P_{j+2i+1} \rangle, \end{aligned} \tag{2.7}$$

where we use the notation

$$\lfloor a \rfloor = \max\{n \in \mathbb{Z} : n \le a\}.$$

Then, keeping in mind the orthonormality of the system $\{P_j\}$, and the fact that $\mathbb{E}|\langle \xi, P_j \rangle|^2 = 1$, while $\mathbb{E}(\langle \xi, P_j \rangle \langle \xi, P_k \rangle) = 0$ for $k \ne j$, we can derive the following bound for the noise propagation error:

$$\begin{aligned}\mathbb{E}\|D_n y - D_n y_\delta\|_{L_2}^2 &= 4\delta^2 \sum_{j=0}^{n-1} \sum_{i=0}^{\lfloor (n-j-1)/2 \rfloor} (j+1/2)(j+2i+3/2) \\ &= 2\delta^2 \sum_{j=0}^{n-1} (j+1/2)(n+j+2)(\lfloor (n-j-1)/2 \rfloor + 1) \\ &\le \delta^2 \sum_{j=0}^{n-1} (j+1/2)(n+j+2)(n-j+1) \\ &= \frac{1}{4}\delta^2 n^2 (n^2+6n+5). \qquad \square\end{aligned}$$

Now we will formulate the main result of this section that follows directly from Lemmas 2.1, 2.2 and (2.1).

Theorem 2.3. *Let assumptions* (1.1), (1.4) *and* (1.5) *be satisfied. Assume that* $y \in W_2^\psi$ *with* $\psi(k) = k^\mu$. *Then for* $\mu > 2$ *and* $n = c\delta^{-1/\mu}$ *we have*

$$(\mathbb{E}\|y' - D_n y_\delta\|_{L_2}^2)^{1/2} = O(\delta^{\frac{\mu-2}{\mu}}). \tag{2.8}$$

If $y \in W_2^\psi$ *with* $\psi(k) = e^{kh}$, $h > 0$, *then for* $n = \frac{c}{h}\log(\frac{1}{\delta})$ *we obtain*

$$(\mathbb{E}\|y' - D_n y_\delta\|_{L_2}^2)^{1/2} = O(\delta \log^2 \delta). \tag{2.9}$$

By comparing the present results (2.8), (2.9) with those for the deterministic noise model [14] (see also (1.12)) we are able to conclude that in the scale of function spaces generated by the system of Legendre polynomials the stochastic white noise does not increase the loss of accuracy compared to the deterministic noise of the same intensity δ, as it was the case for the scale of spaces $W^r\{e_k^I\}$. This effect deserves further study. Its ad hoc explanation is that this is because the Fourier–Legendre coefficients of the noise propagation error (2.7) are correlated. Note that in the case of orthonormal system $\{e_k^I\}$ of the singular functions of the integration operator (1.3), which are used for generating scale $\{W^r\{e_k^I\}\}$, the corresponding Fourier coefficients would be uncorrelated.

3 Adaptation to the unknown bound of the approximation error

It is obvious that the truncation level n in (1.10) plays the role of the regularization parameter and, thus, need to be properly chosen. Note that the optimal choice of n indicated in Theorem 2.3 can be realized only when the form of the smoothness index function $\psi(k)$ and the noise intensity δ are known a priori. Since in reality the form of the smoothness index function can be rather complex and, indeed, a priori unknown, the truncation level n indicated in Theorem 2.3 is mainly of a theoretical nature.

In this section we present a posteriori parameter choice rule that automatically adjusts the value of the truncation level n to unknown smoothness index function $\psi(k)$. Moreover, we are going to show and prove that the automatically chosen truncation level $n = n_+$ provides a risk bound which is only by a log-factor worse that the one obtained with the knowledge of the index function $\psi(k)$.

In the following, without loss of generality, we assume that the noise propagation error is controlled by some known increasing continuous function λ such that

$$\mathbb{E}\|D_n y - D_n y_\delta\|_{L_2}^2 \le \lambda^2(n)\delta^2. \tag{3.1}$$

From Lemma 2.2 it follows that one may use $\lambda(n) = \frac{1}{2}n(n^2 + 6n + 5)^{1/2}$. With the above observation, we restrict our attention to the finite set

$$\mathcal{N} = \{n : n = 1, 2, \dots, N,\ N = \lfloor(\lambda^{-1}(1/\delta)\rfloor\}$$

of possible truncation levels. Such choice is natural, since for $n > \lfloor(\lambda^{-1}(1/\delta))\rfloor$ the estimation for the noise propagation error (3.1) becomes trivial, i.e. $O(1)$.

At this point it is worth mentioning that due to the random nature of noise one cannot conclude from (3.1) that the following estimation holds for all $n \in \mathcal{N}$:

$$\|D_n y - D_n y_\delta\|_{L_2} \le \lambda(n)\delta. \tag{3.2}$$

Thus, one needs to use an additional assumption on the distribution of random noise to be able to control noise propagation. One of the accepted assumptions is that the stochastic white noise ξ is Gaussian. Under this assumption, we will see in the sequel that for sufficiently large κ, which does not depend on the truncation level n, the estimation

$$\|D_n y - D_n y_\delta\|_{L_2} \le \kappa\lambda(n)\delta \tag{3.3}$$

holds with a large probability. From this representation it follows that κ is a design parameter and, moreover, is one of the key ingredients in the process of choosing the optimal truncation level n_+. Concerning the appropriate choice of this parameter, we refer to the work [1].

Note that if ξ is a Gaussian white noise, then in view of (2.7), (3.1) a normalized difference

$$h_n = \frac{D_n y - D_n y_\delta}{\lambda(n)\delta} \tag{3.4}$$

is a Gaussian random variable in L_2 with $\mathbb{E}h_n = 0$ and $\mathbb{E}\|h_n\|_{L_2}^2 \le 1$. Therefore, the following lemma is a direct consequence of the concentration inequality, see [12, p. 59].

Lemma 3.1. *The following probability estimate holds for all* $\tau > 0$*:*

$$\mathbb{P}\left\{\frac{\|D_n y - D_n y_\delta\|_{L_2}}{\lambda(n)\delta} > \tau\right\} \leq 4\exp\left(-\frac{\tau^2}{8}\right). \tag{3.5}$$

As in [19] we define the notion of an admissible function that is though used to bound the approximation error, but, as we will see later, is not involved in the a posteriori choice of n_+.

Definition 3.2. A function $\varphi(n) = \varphi(n; \lambda, y, \delta)$ is said to be admissible for given y, λ and δ if the following holds:

(i) $\varphi(n)$ is a non-increasing function on $[1, N]$,

(ii) $\varphi(N) < \lambda(N)\delta$,

(iii) for all $n \in \mathcal{N}$,

$$\|y' - D_n y\|_{L_2} \leq \varphi(n). \tag{3.6}$$

For given λ, y, δ the set of admissible functions is denoted by $\Phi(\lambda, y, \delta)$. In the sequel, we always assume that the set $\Phi(\lambda, y, \delta)$ is not empty. From Lemma 2.1 it follows that this is the case when δ is large enough (for negligible noise levels no adaptation is required) and $y \in W_2^\psi$, where $\psi(k)$ tends to infinity faster than k^μ, $\mu > 2$. The latter assumption is not restrictive, since as we mentioned above, the inclusion $W_2^\mu \subset H_2^1$, which is necessary for considering y' as an element of L_2, can be guaranteed for $\mu > 2$.

From (2.1), (3.1) and (3.6) it follows that for any $n \in \mathcal{N}$ the expected risk of approximating y' by $D_n y_\delta(t)$, see (1.10), has the following bound:

$$\mathbb{E}\|y' - D_n y_\delta\|_{L_2}^2 \leq \varphi^2(n) + \lambda^2(n)\delta^2. \tag{3.7}$$

In view of (3.7) the quantity

$$e(\lambda, y, \delta) = \inf_{\varphi \in \Phi(\lambda, y, \delta)} \min_{n \in \mathcal{N}} \{(\varphi^2(n) + \lambda^2(n)\delta^2)^{1/2}\}$$

can be seen as the best error bound that can be guaranteed for approximation of $y'(t)$ by $D_n y_\delta$ under the assumptions (1.1), (1.4), (1.5), and (3.1).

Now we are going to present an adaptive choice of the truncation level $n = n_+$, which is based on the so-called balancing principle that has been extensively studied recently (see, e.g., [7, 19] and references therein). This choice relies on the most probable bound of the noise propagation error (3.3) and is read as follows:

$$n_+ = \min\{n \in \mathcal{N} : \|D_n y_\delta - D_m y_\delta\|_{L_2} \leq 4\kappa\lambda(m)\delta,\ m = n, \dots, N\}. \tag{3.8}$$

Note that such adaptive choice of the truncation level n_+ does not depend on the unknown $\varphi(n)$, it actually depends only on y_δ, λ and δ, which are assumed to be known. Moreover, n_+ is well defined as the minimum over a finite set of numbers, which is not empty since it contains at least $n = N$.

Hence, the main result of this section which is given below shows that the adaptively chosen truncation level n_+ provides a risk bound that only by a log-factor worse than the one obtained with the knowledge of $\psi(k)$.

Theorem 3.3. *Let the truncation level n_+ be chosen in accordance with* (3.8) *and a design parameter κ is chosen as $\kappa = 4\sqrt{p \ln \lambda^{-1}(1/\delta)} \asymp \sqrt{\ln(1/\delta)}$ with a constant p chosen in such a way that the following equality holds,*

$$2^{11/2}(\lambda^{-1}(1/\delta))^{-p+1} = \delta^2,$$

then

$$\mathbb{E}\|y' - D_{n_+} y_\delta\|_{L_2}^2 \le c\eta\sqrt{\ln(1/\delta)} \inf_{\varphi\in\Phi(\lambda,y,\delta)} \min_{n\in\mathcal{N}}\{(\varphi^2(n)+\lambda^2(n)\delta^2)^{1/2}\}, \quad (3.9)$$

where $\eta = \max\{\frac{\lambda(n+1)}{\lambda(n)} : n \in \mathcal{N}\}$.

Remark 3.4. The constant p, used for determining the design parameter κ, can be numerically calculated for given noise level, e.g., for $\delta = 10^{-6}$, $p = 5.34$, for $\delta = 10^{-4}$, $p = 5.50$, and for $\delta = 10^{-2}$, $p = 6.11$.

Remark 3.5. The factor η in the error bound (3.9) can be easily found using the bound of Lemma 2.2, for example, for our choice of $\lambda(n) = \frac{1}{2}n(n^2 + 6n + 5)^{1/2}$ we have $\eta = \sqrt{7}$.

Note that Theorem 3.3 can be proven similar to one in [2], but since our definition of the admissible function is different from that of [2], we present the proof of the theorem for the sake of self-containedness.

Proof. First of all, recall that $D_{n_+} y_\delta = D_{n_+} y_\delta(\xi(\omega))$ is a random element defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and

$$\mathbb{E}\|y' - D_{n_+} y_\delta\|_{L_2}^2 = \int_\Omega \|y' - D_{n_+} y_\delta\|_{L_2}^2 \, d\mathbb{P}(\omega). \quad (3.10)$$

Let $\varphi \in \Phi(\lambda, y, \delta)$ be any admissible function and let us temporarily introduce the values

$$n_0 = \min\{n \in \mathcal{N} : \varphi(n) \le \lambda(n)\delta\}, \quad (3.11)$$

$$n_1 = \arg\min\{\varphi(n) + \lambda(n)\delta : n \in \mathcal{N}\}, \quad (3.12)$$

and the random variable

$$\Xi_{\lambda,\delta}(\omega) = \max_{n\in\mathcal{N}} \|h_n\|_{L_2},$$

where h_n is defined by (3.4).

We divide the probability space Ω into two subspaces

$$\Omega_\kappa = \{\omega \in \Omega : \Xi_{\lambda,\delta} \le \kappa\}, \quad \bar{\Omega}_\kappa = \Omega \setminus \Omega_\kappa = \{\omega \in \Omega : \Xi_{\lambda,\delta} > \kappa\},$$

and decompose the expected risk (3.10) as follows:

$$\mathbb{E}\|y' - D_{n_+} y_\delta\|_{L_2}^2 = \int_{\Omega_\kappa} \|y' - D_{n_+} y_\delta\|_{L_2}^2 d\mathbb{P}(\omega) + \int_{\bar{\Omega}_\kappa} \|y' - D_{n_+} y_\delta\|_{L_2}^2 d\mathbb{P}(\omega). \tag{3.13}$$

Similar to [2], we divide the proof of the theorem into two parts. The first one estimates the behavior of the $\|y' - D_{n_+} y_\delta\|_{L_2}$ for "good" event $\omega \in \Omega_\kappa$ when the behavior of the random component $\|D_m y - D_{n_+} y_\delta\|_{L_2}$ can be controlled by (3.3).

The second part of the proof deals with the "bad" events when the stochastic noise property produces results far away from the average.

Part 1 ("good" event $\omega \in \Omega_\kappa$). First of all, we observe that

$$\lambda(n_0)\delta \le \eta(\varphi(n_1) + \lambda(n_1)\delta),$$

because either $n_1 \le n_0 - 1 \le n_0$, so that by definition (3.11) we have

$$\lambda(n_0 - 1)\delta \le \varphi(n_0 - 1)$$

and

$$\lambda(n_0)\delta = \frac{\lambda(n_0)}{\lambda(n_0 - 1)} \lambda(n_0 - 1)\delta \le \eta\varphi(n_0 - 1) \le \eta(\varphi(n_1) + \lambda(n_1)\delta),$$

or $n_0 \le n_1$ in which case

$$\lambda(n_0)\delta \le \lambda(n_1)\delta \le \eta(\varphi(n_1) + \lambda(n_1)\delta).$$

Using the definition of n_0 (3.11) we want to show that $n_0 \ge n_+$. Indeed, for any $m > n_0$ we obtain the following bound for the behavior of the random variable:

$$\begin{aligned}
\|D_m y_\delta - D_{n_0} y_\delta\|_{L_2} &\le \|y' - D_m y_\delta\|_{L_2} + \|y' - D_{n_0} y_\delta\|_{L_2} \\
&\le \varphi(m) + \|D_m y - D_m y_\delta\|_{L_2} + \varphi(n_0) \\
&\qquad + \|D_{n_0} y - D_{n_0} y_\delta\|_{L_2} \\
&\le 2\kappa\lambda(m)\delta + 2\kappa\lambda(n_0)\delta \le 4\kappa\lambda(m)\delta.
\end{aligned}$$

This means that

$$\begin{aligned}
n_0 \ge n_+ = \min\{n \in \mathcal{N} : \|D_n y_\delta - D_m y_\delta\|_{L_2} \le 4\kappa\lambda(m)\delta, \\
m = N, N-1, \dots, n+1\}.
\end{aligned}$$

Using the above observations for all $\omega \in \Omega_\kappa$ one can finally obtain

$$\begin{aligned}\|y' - D_{n_+} y_\delta\|_{L_2} &\le \|y' - D_{n_0} y_\delta\|_{L_2} + \|D_{n_0} y_\delta - D_{n_+} y_\delta\|_{L_2}\\ &\le \varphi(n_0) + \kappa\lambda(n_0)\delta + 4\kappa\lambda(n_0)\delta\\ &\le 6\kappa\lambda(n_0)\delta \le 6\kappa\eta(\varphi(n_1) + \lambda(n_1)\delta).\end{aligned}$$

Therefore, we get

$$\begin{aligned}\int_{\Omega_\kappa} \|y' - D_{n_+} y_\delta\|_{L_2}^2 d\,\mathbb{P}(\omega) &\le 36\kappa^2\eta^2(\varphi(n_1) + \lambda(n_1)\delta)^2\\ &\le 36\kappa^2\eta^2(2\varphi^2(n_1) + 2\lambda^2(n_1)\delta^2)\\ &= 72\kappa^2\eta^2(\varphi^2(n_1) + \lambda^2(n_1)\delta^2).\end{aligned}$$

Part 2 ("bad" event $\omega \in \bar{\Omega}_\kappa$). For $\omega \in \bar{\Omega}_\kappa$, using the assumption that

$$N = \lfloor \lambda^{-1}(1/\delta) \rfloor,$$

we have the following bound:

$$\begin{aligned}\|y' - D_{n_+} y_\delta\|_{L_2} &\le \|y' - D_N y_\delta\|_{L_2} + \|D_N y_\delta - D_{n_+} y_\delta\|_{L_2}\\ &\le \varphi(N) + \|D_N y - D_N y_\delta\|_{L_2} + 4\kappa\lambda(N)\delta\\ &< \lambda(N)\delta + \frac{\|D_N y - D_N y_\delta\|_{L_2}}{\lambda(N)\delta}(\lambda(N)\delta) + 4\kappa\lambda(N)\delta\\ &\le \frac{\|D_N y - D_N y_\delta\|_{L_2}}{\lambda(N)\delta} + 5\kappa \le 6\,\Xi_{\lambda,\delta}.\end{aligned}$$

Then using the Cauchy–Schwarz inequality one can estimate the second term in (3.13) as follows:

$$\begin{aligned}\int_{\bar{\Omega}_\kappa} \|y' - D_{n_+} y_\delta\|_{L_2}^2 d\,\mathbb{P}(\omega) &\le 36 \int_{\bar{\Omega}_\kappa} \Xi_{\lambda,\delta}^2 \, d\,\mathbb{P}(\omega)\\ &\le 36 \Big[\int_{\bar{\Omega}_\kappa} \Xi_{\lambda,\delta}^4 \, d\,\mathbb{P}(\omega)\Big]^{1/2} \Big[\int_{\bar{\Omega}_\kappa} 1 d\,\mathbb{P}(\omega)\Big]^{1/2}.\end{aligned}$$

Now we provide separate estimations for each of the above integrals. Lemma 3.1 with $\tau = \kappa$ immediately gives

$$\int_{\bar{\Omega}_\kappa} 1 d\,\mathbb{P}(\omega) = \mathbb{P}\{\omega \in \bar{\Omega}_\kappa\} \le 4N \exp\Big(-\frac{\kappa^2}{8}\Big). \tag{3.14}$$

Using Lemma 3.1 once again, we can estimate the tail distribution

$$G(\tau) = \mathbb{P}\{\Xi_{\lambda,\delta}(\omega) > \tau\} \le \sum_{n=1}^{N} \mathbb{P}\{\|h_n\|_{L_2} > \tau\} \le 4N \exp\left(-\frac{\tau^2}{8}\right),$$

and integrating by parts yields

$$\begin{aligned}
\int_{\bar{\Omega}_\kappa} \Xi^4_{\lambda,\delta} d\,\mathbb{P}(\omega) &\le -\int_0^\infty \tau^4 dG(\tau) \\
&= -\tau^4 G(\tau)|_0^\infty + 4\int_0^\infty \tau^3 G(\tau) d\tau \\
&= 2\int_0^\infty \tau^2 G(\tau) d\tau^2 \\
&\le 8N \int_0^\infty \tau^2 \exp\left(-\frac{\tau^2}{8}\right) d\tau^2 = 2^9 N.
\end{aligned}$$

Then a combination of the above inequalities gives us the following bound:

$$\begin{aligned}
\int_{\bar{\Omega}_\kappa} \Xi^2_{\lambda,\delta} d\,\mathbb{P}(\omega) &\le 2^{11/2} N \exp\left(-\frac{\kappa^2}{16}\right) \\
&\le 2^{11/2} \lambda^{-1}(1/\delta) \exp^{-\frac{16p \ln \lambda^{-1}(1/\delta)}{16}} \\
&\le 2^{11/2} (\lambda^{-1}(1/\delta))^{-p+1} = \delta^2.
\end{aligned}$$

Summing up all the estimations, we get

$$\begin{aligned}
\mathbb{E}\|y' - D_{n_+} y_\delta\|^2_{L_2} &\le 72\kappa^2\eta^2(\varphi^2(n_1) + \lambda^2(n_1)\delta^2) + 36\delta^2 \\
&\le 108\kappa^2\eta^2(\varphi^2(n_1) + \lambda^2(n_1)\delta^2),
\end{aligned}$$

or that is the same,

$$\begin{aligned}
(\mathbb{E}\|y' - D_{n_+} y_\delta\|^2_{L_2})^{1/2} &\le 24\sqrt{3}\eta(\varphi^2(n_1) + \lambda^2(n_1)\delta^2)^{1/2} \sqrt{p \ln \lambda^{-1}(1/\delta)} \\
&\le c\eta\sqrt{\ln(1/\delta)} \min_{n \in \mathcal{N}} \{(\varphi^2(n) + \lambda^2(n)\delta^2)^{1/2}\}.
\end{aligned}$$

This estimation holds true for an arbitrary admissible function $\varphi \in \Phi(\lambda, y, \delta)$. Therefore, we conclude that

$$(\mathbb{E}\|y' - D_{n_+} y_\delta\|^2_{L_2})^{1/2} \le c\eta\sqrt{\ln(1/\delta)} \inf_{\varphi \in \Phi(\lambda,y,\delta)} \min_{n \in \mathcal{N}} \{(\varphi^2(n) + \lambda^2(n)\delta^2)^{1/2}\}.$$

□

4 Error bounds in the space of continuous functions

We first note the following property of the derivatives of the Legendre polynomials [35], which will be useful below:

$$|P_k^{(r)}(t)| \leq |P_k^{(r)}(1)| = \frac{(k+r)!}{2^r(k-r)!r!}\sqrt{k+1/2}. \tag{4.1}$$

At this point, we shall remind that the quality of the reconstruction of the derivative y' from given noisy data depends on the balance between the approximation error related to the smoothness of the function to be differentiated, and the noise propagation error, which is related to the noise nature.

We will first provide the error bounds for the approximation error. These error bounds have been obtained recently in [14] for approximating the derivative of a function $y \in W_2^\psi$ from deterministic noise. We present them here for the sake of completeness.

Lemma 4.1. *For $y \in W_2^\psi$ the approximation error has the following bound:*

$$\|y' - D_n y\|_C \leq c\left(\int_n^\infty \frac{t^5}{\psi^2(t)}dt\right)^{1/2}\|y\|_\psi. \tag{4.2}$$

In cases $\psi(k) = k^\mu$ and $\psi(k) = e^{kh}$, $h > 0$, from (4.2) *we can derive the following bounds respectively:*

$$\|y' - D_n y\|_C \leq c\frac{n^{3-\mu}}{\sqrt{2\mu-6}}\|y\|_\mu, \quad \mu > 3, \tag{4.3}$$

and

$$\|y' - D_n y\|_C \leq c\left(\frac{n^5}{h} + \frac{n^5}{h^6}\right)^{1/2} e^{-nh}\|y\|_\psi. \tag{4.4}$$

Next we provide and prove an explicit bound for the noise propagation error.

Lemma 4.2. *Under the assumptions* (1.1), (1.4) *and* (1.5) *the following bound holds true:*

$$\mathbb{E}||D_n y - D_n y_\delta||_C \leq c\delta n^3\sqrt{\log n}. \tag{4.5}$$

Proof. Keeping in mind that

$$\langle P_k, y_\delta\rangle = \langle P_k, y\rangle + \delta\langle P_k, \xi\rangle, \quad k = 1, 2, \ldots, n,$$

where $\langle P_k, \xi\rangle = \xi_k$ is a centered Gaussian random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the noise propagation error can be estimated as follows:

$$\mathbb{E}\|D_n y - D_n y_\delta\|_C \leq \delta\mathbb{E}\left\|\sum_{k=1}^n \xi_k P_k'\right\|_C. \tag{4.6}$$

In order to bound the right-hand side of (4.6), we will use Dudley's theorem [12, Theorem 11.17].

Note that

$$X = (X_t)_{t\in T} := \sum_{k=1}^{n} \xi_k P_k'(t), \quad T = [-1,1], \tag{4.7}$$

is a zero mean Gaussian random process because it is a finite linear combination of real-valued elements with Gaussian random coefficients $\xi_k = \langle P_k, \xi \rangle$, $\mathbb{E}\xi_k = 0$ and $\mathbb{E}\xi_k^2 = 1$ for $k = 1, 2, \ldots, n$.

To employ Dudley's theorem we define on T the metric d_X induced by X as follows:

$$d_X(s,t) := (\mathbb{E}|X_s - X_t|^2)^{1/2} = \left(\sum_{k=1}^{n} |P_k'(s) - P_k'(t)|^2 \right)^{1/2}, \quad s,t \in T.$$

In view of (4.1) with $r = 1$ the diameter $D = D(T)$ of $T = [-1,1]$ in this metric admits the estimation

$$D = D(T) \le 2 \left(\sum_{k=1}^{n} |P_k'(1)|^2 \right)^{1/2} \le c \left(\sum_{k=1}^{n} k^5 \right)^{1/2} \le cn^3. \tag{4.8}$$

Moreover, using the Mean Value Theorem and (4.1) with $r = 2$, we can bound the distance $d_X(s,t)$ by a multiple of $|s-t|$:

$$\begin{aligned} d_X(s,t) &\le \left(\sum_{k=1}^{n} \|P_k''\|_C^2 \right)^{1/2} |s-t| \\ &= \left(\sum_{k=1}^{n} \frac{1}{64}(k-1)^2 k^2 (k+1)^2 (k+2)^2 (k+1/2) \right)^{1/2} |s-t| \\ &\le cn^5 |s-t|. \end{aligned} \tag{4.9}$$

Recall that the statement of Dudley's theorem [12, Theorem 11.17] has the form

$$\mathbb{E} \sup_{t\in T} X_t \le 24 \int_0^\infty (\log N(T, d_X; \varepsilon))^{1/2} d\varepsilon,$$

where $N(T, d_X; \varepsilon)$ denotes the minimal number of ε-balls in the metric d_X required to cover T. From (4.9) one can conclude that

$$N(T, d_X; \varepsilon) \le cn^5 \frac{1}{\varepsilon}.$$

Then Dudley's estimate yields

$$\begin{aligned}\mathbb{E}\|\sum_{k=1}^{n}\xi_k P_k'\|_C &\le c\int_0^D\left(\log\frac{n^5}{\varepsilon}\right)^{1/2}d\varepsilon\\ &\le c\sqrt{D}\left(\int_0^D\log\frac{n^5}{\varepsilon}d\varepsilon\right)^{1/2}\\ &= c\sqrt{D}(D\log\frac{n^5}{D}+D)^{1/2}.\end{aligned}$$

Combining this with (4.6)–(4.8) we get the statement of the lemma. □

Now we summarize the above observations from Lemmas 4.1, 4.2 into the following convergence result.

Theorem 4.3. *Let assumptions* (1.1), (1.4) *and* (1.5) *be satisfied. Assume that* $f \in W_2^{\psi}$ *with* $\psi(k) = k^{\mu}$. *Then for* $\mu > 3$ *and* $n = c\delta^{-1/\mu}$ *we have*

$$\mathbb{E}\|y' - D_n y_\delta\|_C = O\big(\delta\log^{1/2}(1/\delta)\big)^{\frac{\mu-3}{\mu}}. \tag{4.10}$$

If $f \in W_2^{\psi}$ *with* $\psi(k) = e^{kh}$, $h > 0$, *then for* $n = \frac{c}{h}\log(1/\delta)$ *we obtain*

$$\mathbb{E}\|y' - D_n y_\delta\|_C = O\big(\delta\log^{7/2}(1/\delta)\big). \tag{4.11}$$

Note that the bounds (4.10), (4.11) are only by a logarithmic factor worse than those obtained in [14] for the deterministic noise model (1.1), (1.2), that can be seen as a reflection of stochastic nature of noise.

Note also that the risk bounds indicated in Theorems 2.3 and 4.3 are achieved for the same order of the truncation level $n = O(\delta^{-1/\mu})$, or $n = O(h^{-1}\log(1/\delta))$. Therefore, one may expect that the truncation level $n = n_+$ chosen in accordance with (3.8) is effective not only in L_2-space, but also in the space of continuous functions.

5 The relation to Savitzky–Golay method. Numerical examples

In 1964 Savitzky and Golay [29] provided a method for smoothing and differentiation of data by least-squares technique. Since then the Savitzky–Golay approach has been widely used, actually, the proposed algorithm is very attractive for its exceptional simplicity and its ability of producing a significant improvement in computational speed. Moreover, the paper [29] is one of the most cited papers in the journal Analytical Chemistry and is considered by that journal as one of its "10 seminal papers" saying "it can be argued that the dawn of the computer-controlled analytical instrument can be traced to this article" [28].

In this section we would like to discuss the relation between the considered approach (1.10) and the well-known filtering technique [29]. As it will be shown the approach (1.10) is very similar to the Savitzky–Golay method.

At the same time, it is worthwhile to mention that the Savitzky–Golay filter produces excellent results provided that the degree of the polynomial n is correctly chosen [25]. However, this issue has not been well studied in the literature and there is no general rule to advise the choice of the polynomial degree n. In this situation and in view of the similarity between the approach (1.10) and the Savitzky–Golay method the adaptive parameter choice rule (3.8) can be used for addressing the above mentioned issue.

Moreover, we are going to demonstrate the superiority of the proposed algorithm to finite-difference schemes which are commonly employed in a pointwise estimate of the derivative.

5.1 Formulation of the Savitzky–Golay method

Savitzky–Golay filter approximates a derivative of the function by the derivative of the polynomial of fixed degree n

$$y'(t) \approx \mathrm{SG}_n y_\delta(t) := \frac{d}{dt} \sum_{k=1}^{n} a_k t^k, \tag{5.1}$$

where the coefficients $(a_k)_{k=0}^n$ minimize the sum

$$I_N(y_\delta; (a_k)_{k=0}^n) = \frac{1}{N} \sum_{i=1}^{N} \left(y_\delta(t_i) - \sum_{k=0}^{n} a_k t_i^k \right)^2 \tag{5.2}$$

for given noisy data $(y_\delta(t_i))_{i=1}^N$. For the sake of simplicity we assume for the moment that $(t_i)_{i=1}^N \in [-1, 1]$, but in the numerical tests below the data points (t_i) are taken from other intervals, which are suitable for the considered application and can be transformed to $[-1, 1]$ by an appropriate change of variables.

Keeping in mind that the sum (5.2) is a discrete version of the integral

$$I(y_\delta; (a_k)_{k=0}^n) = \int_{-1}^{1} \left(y_\delta(t) - \sum_{k=0}^{n} a_k t^k \right)^2 dt,$$

and the fact that for the latter one we have

$$\min_{(a_k)} I(y_\delta; (a_k)_{k=0}^n) = \int_{-1}^{1} \left(y_\delta(t) - \sum_{k=0}^{n} y_k^\delta P_k(t) \right)^2 dt, \tag{5.3}$$

the approximation $\mathrm{SG}_n y_\delta(t)$ can be viewed as a discrete version of (1.10). Moreover, in the same spirit as the Savitzky–Golay method (5.1), (5.2), one can consider another discrete version of (1.10) defined as

$$D_{n,N} y_\delta(t) := \frac{d}{dt}\left(\sum_{k=1}^{n} a_k P_k(t)\right), \tag{5.4}$$

where the coefficients $(a_k)_{k=0}^n$ minimize the sum

$$\tilde{I}_N(y_\delta; (a_k)_{k=0}^n) = \frac{1}{N}\sum_{i=1}^{N}\left(y_\delta(t_i) - \sum_{k=0}^{n} a_k P_k(t_i)\right)^2.$$

In view of the similarity of all these three numerical differentiation schemes (1.10), (5.1), (5.4), each of them can be treated as a perturbed version of the others, and if this perturbation is assumed to be within the level of $O(\delta n^2)$, which can be expected since the values of $I_N(y_\delta; (a_k)_{k=0}^n)$ and $\tilde{I}_N(y_\delta; (a_k)_{k=0}^n)$ at minimizers approximate (5.3), then the rule (3.8) for the choice of n can be effectively applied to the Savitzky–Golay method (5.1), as well as to its version (5.4).

5.2 Numerical examples

In the second part of this section we demonstrate how the proposed approach (5.4) together with the adaptive parameter choice rule (3.8) can be effectively used to improve the management of diabetes therapy by providing accurate predictions of the blood glucose (BG) evolution.

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

There are several prediction techniques, and a variety of glucose predictors has been recently proposed, see, for example, [21, 22, 27, 30, 31]. In this section we discuss the predictors based on the numerical differentiation [8]. Such predictors estimate the rate of change of BG-concentration at the prediction moment $t = t_0$ from current and past measurements and a future BG-concentration at any time moment $t \in [t_0, t_k]$ is given as follows:

$$y(t) = y'(t_0)\cdot(t - t_0) + y_\delta(t_0), \tag{5.5}$$

where $t \in [t_0, t_k]$ and $y'(t_0)$ is approximated by (5.4) from the given noisy data $(t_i, y_\delta(t_i))$, $i = 0, \ldots, -m+1$, $SH = 30$ (min), $N = m = 7$ and with the truncation levels $n \in \{1, 2, \ldots, N-1\}$.

To illustrate how these predictors work we use data set of 100 virtual subjects which are obtained from Padova/University of Virginia simulator [10]. 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 [22, 30].

To quantify the clinical accuracy of the considered predictors, we use the Prediction Error-Grid Analysis (PRED-EGA) [30], 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 distinguishes 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 (5.4) with the truncation level chosen in accordance with (3.8), operating on simulated noisy data with $SH = 30$ (min).

We also perform the comparison of the constructed predictors with the predictors based on the linear extrapolation (5.5) considered in [21], where the derivative is estimated by means of the one-sided finite-difference formula [6]

$$y'(t_0) \approx \sum_{i=0}^{-m+1} \frac{a_i}{h} y_\delta(t_i), \tag{5.6}$$

where $h = 5$ (min), $m = 7$,

$$a_0 = \frac{49}{20}, \quad a_{-1} = -6, \quad a_{-2} = \frac{15}{2}, \quad a_{-3} = -\frac{20}{3},$$
$$a_{-4} = \frac{15}{4}, \quad a_{-5} = -\frac{6}{5}, \quad a_{-6} = \frac{1}{6}.$$

The performance of the predictors (5.5), (5.6) is displayed in Table 2 below. The comparison of both tables allows us to conclude that the predictors (5.4), (5.5) with adaptively chosen truncation level (3.8) outperform the predictors (5.5), (5.6) based on the one-sided finite-difference formula of a fixed order.

Patient	BG ≤ 70 (mg/dL) (%)			BG 70-180 (mg/dL) (%)			BG ≥ 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**	–

Table 1. The performance assessment matrix given by the PRED-EGA for the linear extrapolation predictors, where the derivative is found by (5.4) with a truncation level chosen by the balancing principle (3.8), operating on simulated noisy data with $PH = 15$ (min) and $SH = 30$ (min).

Patient	BG ≤ 70 (mg/dL) (%)			BG 70-180 (mg/dL) (%)			BG ≥ 180 (mg/dL) (%)		
Vir. ID	Acc.	Benign	Error	Acc.	Benign	Error	Acc.	Benign	Error
1	–	–	–	52.81	44.44	2.75	47.06	47.06	5.88
2	–	–	–	45.36	51.47	3.17	–	–	–
3	–	–	–	48.48	50.12	1.4	–	–	–
17	62.89	13.52	23.58	48.88	50	1.12	–	–	–
18	54.57	12.98	32.45	54.35	43.91	1.74	–	–	–
24	53.23	14.15	32.62	62.71	36.91	0.38	–	–	–
33	78.17	6.19	15.63	57.28	36.61	6.1	50	12.5	37.5
34	60.96	9.16	29.88	61.87	33.45	4.68	42.86	42.86	14.29
42	65.16	10.66	24.18	59.17	33.39	7.44	28.57	71.43	–
47	67.73	9.6	22.67	52.71	44.17	3.12	–	100	–
Avg.	**63.25**	**10.89**	**25.86**	**54.36**	**42.45**	**3.19**	**33.7**	**54.77**	**11.53**

Table 2. The performance assessment matrix given by the PRED-EGA for the predictors (5.5), (5.6), operating on simulated noisy data with $PH = 15$ (min).

Acknowledgments. The authors are grateful to the anonymous referees for valuable suggestions. As it has been commented by one of the referees, the results

achieved in this paper can potentially be extended for solving other similar problems, such as interpolation or data fitting problems, especially in the case that stochastic white noise is considered and noise also propagates with some index in discretization. We plan to study this case in our future publications.

Bibliography

[1] F. Bauer, *An alternate approach to the oblique derivative problem in potential theory*, PhD Thesis, University of Kaiserslautern, 2004.

[2] F. Bauer and S. Pereverzev, Regularization without preliminary knowledge of smoothness and error behaviour, *European J. Appl. Math.* **16** (2005), 303–317.

[3] J. M. Berezanskii, *Expansion in Eigenfunctions of Selfadjoint Operators*, Translation of Mathematical Monographs 17, American Mathematical Society, Providence, 1968.

[4] H. Engl, M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht 1996.

[5] J. Flemming, B. Hofmann and P. Mathé, Sharp converse results for the regularization error using distance functions, *Inverse Problems* **27** (2011), 025006.

[6] B. Fornberg, Generation of finite difference formulas on arbitrarily spaced grids, *Math. Comp.* **51** (1988), 699–706.

[7] U. Hämarik and T. Raus, About the balancing principle for choice of the regularization parameter, *Numer. Funct. Anal. Optim.* **30** (2009), 951–970.

[8] A. Hayes, J. J. Mastrototaro, S. B. Moberg, J. C. Mueller Jr., H. Bud Clark, M. Charles Vallet Tolle, G. L. Williams, B. Wu and G. M. Steil, Algorithm sensor augmented bolus estimator for semi-closed loop infusion system, US Patent Application Publication US2009/0234213 (2009).

[9] M. Hegland, Variable Hilbert scales and their interpolation inequalities with application to Tikhonov regularization, *Appl. Anal.* **59** (1995), 207–223.

[10] B. Kovatchev, M. Breton, C. Dalla Man and C. Cobelli, In silico preclinical trials: A proof of concept in closed-loop control of type 1 diabetes, *J. Diabetes Sci. Technol.* **3** (2009), 44–55.

[11] M. M. Lavrentiev, V. G. Romanov and V. G. Vasiliev, *Multidimensional Inverse Problems for Differential Equations*, Springer-Verlag, Berlin, 1970.

[12] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.

[13] A. K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner Studienbücher Mathematik, B. G. Teubner, Stuttgart, 1989.

[14] S. Lu, V. Naumova and S. V. Pereverzev, Numerical differentiation by means of Legendre polynomials in the presence of square summable noise, RICAM Report No. 2012-15, 2012.

[15] S. Lu and S. V. Pereverzev, Numerical differentiation from a viewpoint of regularization theory, *Math. Comp.* **75** (2006), 1853–1870.

[16] B. A. Mair and F. H. Ruymgaart, Statistical inverse estimation in Hilbert scales, *SIAM J. Appl. Math.* **56** (1996), 1424–1444.

[17] P. Mathé and S. V. Pereverzev, Stable summation of orthogonal series with noisy coefficients, *J. Approx. Theory* **118** (2002), 66–80.

[18] P. Mathé and S. V. Pereverzev, Geometry of linear ill-posed problems in variable Hilbert scales, *Inverse Problems* **19** (2003), 789–803.

[19] P. Mathé and S. V. Pereverzev, Regularization of some linear ill-posed problems with discretized random noisy data, *Math. Comp.* **75** (2006), 1913–1929.

[20] C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves*, Springer-Verlag, Berlin, 1969.

[21] V. Naumova, S. V. Pereverzyev and S. Sivananthan, A meta-learning approach to the regularized learning-case study: Blood glucose prediction, *Neural Networks* **33** (2012), 181–193.

[22] V. Naumova, S. V. Pereverzyev and S. Sivananthan, Adaptive parameter choice for one-sided finite difference schemes and its application in diabetes technology, *J. Complexity* **28** (2012), no. 5–6, 524–538.

[23] A. Neubauer, When do Sobolev spaces form a Hilbert scale?, *Proc. Amer. Math. Soc.* **103** (1988), 557–562.

[24] M. Nussbaum and S. V. Pereverzev, The degree of ill-posedness in stochastic and deterministic noise models, preprint No. 509, WIAS-Berlin, 1999.

[25] P. Persson and G. Strang, Smoothing by Savitzky–Golay and Legendre filters, in: *Mathematical Systems Theory in Biology, Communications, Computation, and Finance* (Notre Dame 2002), Springer-Verlag, New York (2003), 301–315

[26] S. Z. Rafal'son, The approximation of functions by Fourier–Jacobi sums, *Izv. Vysš. Učebn. Zaved. Matematika* **4** (1968), 54–62.

[27] J. Reifman, S. Rajaraman, A. Gribok and W. Kenneth Ward, Predictive monitoring for improved management of glucose levels, *J. Diabetes Sci. Technol.* **1** (2007), 478–486.

[28] J. Riordon, E. Zubritsky and A. Newman, Top 10 Articles. Analytical chemistry looks at 10 seminal papers, *Anal. Chem.* **72** (2000), 324A–329A.

[29] A. Savitzky and M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* **36** (1964), 1627–1639.

[30] S. Sivananthan, V. Naumova, C. Dalla Man, A. Facchinetti, E. Renard, C. Cobelli and S. V. Pereverzyev, Assessment of blood glucose predictors: The prediction-error grid analysis, *Diabetes Technol. Theory* **13** (2011), 787–796.

[31] G. Sparacino, F. Zanderigo, S. Corazza and A. Maran, Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series, *IEEE Trans. Biomed. Eng.* **54** (2007), 931–937.

[32] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Scripta Series in Mathematics, V. H. Winston and Sons, Washington, 1977.

[33] N. G. Tomin, An application of the interpolation of linear operators to questions of the convergence of series of Fourier coefficients with respect to classical orthogonal polynomials, *Dokl. Akad. Nauk SSSR* **212** (1973), 1074–1077.

[34] G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, *SIAM J. Numer. Anal.* **14** (1977), 651–667.

[35] D. E. Winch and P. H. Roberts, Derivatives of addition theorems for Legendre functions, *J. Aust. Math. Soc. Ser. B* **37** (1995), 212–234.

[36] Z. Zhao, A truncated Legendre spectral method for solving numerical differentiation, *Int. J. Comput. Math.* **87** (2010), 3209–3217.

Received July 3, 2012.

Author information

Shuai Lu, School of Mathematical Science, Fudan University, 200433 Shanghai, China.
E-mail: slu@fudan.edu.cn

Valeriya Naumova, Johann Radon Institute for Computational and
Applied Mathematics, Altenbergerstrasse 69, A-4040 Linz, Austria.
E-mail: valeriya.naumova@oeaw.ac.at

Sergei V. Pereverzev, Johann Radon Institute for Computational and
Applied Mathematics, Altenbergerstrasse 69, A-4040 Linz, Austria.
E-mail: sergei.pereverzyev@oeaw.ac.at