

NUMERICAL DIFFERENTIATION FROM A VIEWPOINT OF REGULARIZATION THEORY

SHUAI LU AND SERGEI V. PEREVERZEV

ABSTRACT. In this paper, we discuss the classical ill-posed problem of numerical differentiation, assuming that the smoothness of the function to be differentiated is unknown. Using recent results on adaptive regularization of general ill-posed problems, we propose new rules for the choice of the stepsize in the finite-difference methods, and for the regularization parameter choice in numerical differentiation regularized by the iterated Tikhonov method. These methods are shown to be effective for the differentiation of noisy functions, and the order-optimal convergence results for them are proved.

1. INTRODUCTION

How do we approximate a derivative $y'(t)$ of a smooth function $y(t)$, $t \in [0, 1]$? This question is discussed extensively in computational mathematics, and there are many different formulae for numerical differentiation. At the same time, the problem of numerical differentiation is known to be ill-posed [4] in the sense that a small perturbation in the values of $y(t)$ may lead to large errors in the computed derivative. This fact has the unpleasant consequence that the function $y_\delta(t)$, which differs imperceptibly from $y(t)$, has a derivative which differs vastly from $y'(t)$. Since in practice data will almost never be exactly available, one has to be aware of numerical instabilities when a noisy observation y_δ instead of y is known. In this case the perturbation $\delta(t) = y(t) - y_\delta(t)$ represents measurement error or rounding error and can be a nondifferentiable function having a quite erratic nature. Hence, in order to approximate $y'(t)$ in a stable way, regularization methods should be applied.

These methods fall into two categories: methods which use noisy data $y_\delta(t)$ in a nondiscretized form, and methods based only on a finite amount of discrete information regarding $y_\delta(t)$. The well-known forward difference approximation

$$(1.1) \quad y'(t) \approx y_\delta^h(t) = \frac{y_\delta(t+h) - y_\delta(t)}{h}$$

is the simplest example of the method from the first category. Using (1.1) for reconstruction $y'(t)$ one presupposes that the value $y_\delta(t+h)$, for example, is available for any sufficiently small h . Because as it has been observed in [6], [22], difference schemes may construct stable regularizing algorithms only if a stepsize h is chosen properly. An example of the method from the second category can be found in [7],

Received by the editor November 3, 2004 and, in revised form, April 19, 2005.

2000 *Mathematics Subject Classification*. Primary 65D25; Secondary 65J20.

Key words and phrases. Numerical differentiation, adaptive regularization, unknown smoothness, finite-difference methods, Tikhonov regularization.

©2006 American Mathematical Society
Reverts to public domain 28 years from publication

where a derivative $S'_n(t)$ of a natural cubic spline $S_n(t)$ solving the minimization problem

$$(1.2) \quad \frac{1}{n-1} \sum_{i=1}^{n-1} (y_\delta(t_i) - S_n(t_i))^2 + \alpha \|S''_n\|^2 \rightarrow \min$$

is taken as an approximation for $y'(t)$. To construct $S'_n(t)$ one should know only the values $y_\delta(t_i)$ at the points $\{t_i\}_{i=0}^n \subset [0, 1]$, $t_0 = 0$, $t_n = 1$, and choose the regularization parameter α in (1.2).

These examples allow us to see a common distinguishing feature of numerical differentiation methods from both above-mentioned categories. Namely, these methods always have some parameter that should be used for problem regularization. For example, a stepsize h plays the role of regularization parameter for the method (1.1). A number of regularization parameter choice techniques have been developed for numerical differentiation. They yield satisfactory results when the smoothness of the function to be differentiated is given very precisely [1], [2], [22], [26]. However, in applications this smoothness is usually unknown, as one can see it from the following example.

Example 1.1. In the simplest one-dimensional case the cooling of hot glass is modelled by the parabolic system of the form

$$(1.3) \quad \begin{aligned} \frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2}, & u(0, x) &= u_{in}(x), & \frac{\partial u}{\partial x}(t, 0) &= \beta_0(t), & \frac{\partial u}{\partial x}(t, 1) &= \beta_1(t), \\ (t, x) &\in (0, T] \times [0, 1], \end{aligned}$$

and one is interested in determining the heat exchange coefficients β_0, β_1 by means of measurements of the boundary temperature. Assume that we have only a noisy measurement $u_0^\delta(t), u_1^\delta(t)$ of the boundary temperature $u(t, 0), u(t, 1)$ on the whole time interval $[0, T]$, where δ is a small parameter used for measuring the noise level. Such data allow us to determine an approximate distribution $u = u_\delta(t, x)$ of the temperature $u = u(t, x)$ in the whole interval $[0, 1]$ as a solution of the initial boundary-value problem with a noisy Dirichlet condition:

$$(1.4) \quad \begin{aligned} \frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2}, & u(0, x) &= u_{in}(x), & u(t, 0) &= u_0^\delta(t), & u(t, 1) &= u_1^\delta(t), \\ (t, x) &\in (0, T] \times [0, 1]. \end{aligned}$$

From the unique solvability of (1.3), (1.4) it follows that the solution $u(t, x)$ of (1.3) corresponding to the “true” coefficients β_0, β_1 is the same as the solution of (1.4) with “pure” boundary data $u(t, 0), u(t, 1)$ instead of u_0^δ, u_1^δ . In view of the well-posedness of (1.4) the deviation of $u(t, x)$ from $u_\delta(t, x)$ is of the same order as the noise level. Then without loss of generality we can assume that $\|u - u_\delta\| \leq \delta$.

As soon as u_δ has been determined from (1.4), the heat exchange coefficients can be approximated as follows:

$$(1.5) \quad \beta_0(t) \approx \frac{u_\delta(t, h) - u_0^\delta(t)}{h}, \quad \beta_1(t) \approx \frac{u_1^\delta(t) - u_\delta(t, 1-h)}{h}.$$

Keeping in mind that the values $u_\delta(t, h), u_\delta(t, 1-h)$ are available for any $h \in (0, 1)$, approximation (1.5) can be considered as a numerical differentiation method of the first category, because it uses the noisy data $u_\delta(t, x)$ in a nondiscretized form.

At the same time, one does not know a priori the smoothness of the function $u(t, x)$ to be differentiated. This smoothness depends on the so-called compatibility conditions

$$(1.6) \quad \frac{du_{in}}{dx}(0) = \beta_0(0), \quad \frac{du_{in}}{dx}(1) = \beta_1(0).$$

If they are not satisfied, then $\frac{\partial u}{\partial x}(t, \cdot)$ may be discontinuous for $t = 0$. On the other hand, one cannot check (1.6), because $\beta_0(t)$ and $\beta_1(t)$ are the only functions that should be recovered. Thus, the regularization parameter h in (1.5) should be chosen without knowledge of the smoothness of function $u(t, x)$ to be differentiated.

Surprisingly enough, in the case of unknown smoothness we cannot find any results providing us with the recipe for regularization parameter choice for numerical differentiation. Only in the paper [7] the discrepancy principle for the choice of the regularization parameter α in (1.2) has been discussed. This principle is usually applied in the case of unknown smoothness, but its efficiency for numerical differentiation has been proved in [7] under the assumption that the function to be differentiated is two times differentiable.

Thus, there is a gap in the analysis of such an important computational procedure as numerical differentiation, and the goal of the present paper is to fill it out. It will be done on the basis of recent results of regularization theory [15]–[17], [20].

In Section 2 we discuss the case of nondiscretized noisy data and suggest an a posteriori rule for the choice of the stepsize h in the finite-difference methods. Numerical differentiation with discrete noisy information is discussed in Section 3. In this section we make use of the adaptive regularization strategy from [16] for solving the corresponding Volterra equation, which is an alternative to the finite-difference methods. Numerical experiments supporting theoretical results are presented in both sections. Section 4 contains some concluding remarks.

2. REGULARIZED FINITE-DIFFERENCE METHODS

2.1. Adaptive choice of the stepsize. Within the framework of the finite-difference method D_h^l , $l \in N$, the approximate value of the derivative y' at the point $t \in (0, 1)$ is calculated as

$$y'(t) \approx D_h^l y(t) = h^{-1} \sum_{j=-l}^l a_j^l y(t + jh),$$

where a_j^l are some fixed real numbers, and a stepsize h is so small that $t + jh \in (0, 1)$ for $j = -l, -l + 1, \dots, 0, \dots, l$. The last restriction means that the distance between point t and the boundary points of the interval should be sufficiently large. If instead of y only $y_\delta \in C[0, 1]$ is available such that

$$(2.1) \quad \|y - y_\delta\|_{C[0,1]} \leq \delta,$$

then the method D_h^l produces the approximation

$$D_h^l y_\delta(t) = h^{-1} \sum_{j=-l}^l a_j^l y_\delta(t + jh),$$

and its error can be estimated as

$$(2.2) \quad |y'(t) - D_h^l y_\delta(t)| \leq |y'(t) - D_h^l y(t)| + |D_h^l y(t) - D_h^l y_\delta(t)|,$$

where the first term on the right-hand side is the consistency error of D_h^l , whereas the second term is a propagation error. It can be bounded as

$$(2.3) \quad |D_h^l y(t) - D_h^l y_\delta(t)| \leq \frac{\delta}{h} \sum_{j=-l}^l |a_j^l|,$$

and under the assumption (2.1) this bound is the best possible one. Moreover, it does not depend on the smoothness of the function y , which is assumed to be unknown.

On the other hand, the consistency error crucially depends on the smoothness of the function to be differentiated. Usually, for properly chosen coefficients a_j^l one has the bound

$$(2.4) \quad |y'(t) - D_h^l y(t)| \leq c_r h^{r-1} \|y^{(r)}\|_{C[0,1]},$$

for the consistency error provided that $y \in C^r[0, 1]$.

At the same time method D_h^l should be robust in the sense that its consistency error should converge to zero with $h \rightarrow 0$ even for function y having very modest smoothness properties. It can be seen from the following example.

Example 2.1. For $l = 1$, $a_{-1}^1 = 0$, $a_0^1 = -1$, $a_1^1 = 1$ the finite-difference method D_h^1 gives us the well-known forward difference approximation (1.1). It satisfies (2.4) with $r = 2$ provided that $y \in C^2[0, 1]$, but for $y \in C^r[0, 1]$, $r > 2$, the order $O(h)$ cannot be improved in general. On the other hand, for $y \in C^1[0, 1]$ the consistency error of the forward difference approximation can be estimated as follows:

$$\begin{aligned} \left| y'(t) - \frac{y(t+h) - y(t)}{h} \right| &= \frac{1}{h} \left| \int_t^{t+h} \{y'(t) - y'(\tau)\} d\tau \right| \\ &\leq \frac{1}{h} \int_t^{t+h} |y'(t) - y'(\tau)| d\tau \leq \omega(y'; h), \end{aligned}$$

where the quantity

$$\omega(f; h) := \sup_{\substack{t, \tau: \\ |t-\tau| \leq h}} |f(t) - f(\tau)|$$

is known in function theory as a modulus of continuity of a real-valued function f . It is well known that for $y' \in C[0, 1]$ (i.e., $y \in C^1[0, 1]$) $\omega(y'; h) \rightarrow 0$ as $h \rightarrow 0$. It means that the consistency error of the forward difference approximation converges to zero for any continuously differentiable function.

As Example 2.1 shows, the standard finite-difference methods have in common that their consistency error is decreasing for decreasing h . Therefore, in general it is natural to assume that there exists a nondecreasing function $\psi(h) = \psi(y; D_h^l; h)$ such that $0 = \psi(0) \leq \psi(h)$ and

$$|y'(t) - D_h^l y(t)| \leq \psi(h).$$

Combining it with (2.3) one arrives at the bound

$$(2.5) \quad |y'(t) - D_h^l y_\delta(t)| \leq \psi(h) + d_l \frac{\delta}{h},$$

where

$$d_i = \sum_{j=-l}^l |a_j^l|.$$

Suppose we are given a finite set H_N of possible stepsizes $h = h_i, i = 1, 2, \dots, N$,

$$\delta = h_1 < h_2 < \dots < h_N < 1,$$

and the corresponding set of the approximate values $D_{h_i}^l y_\delta(t)$ of the derivative produced by the finite-difference method D_h^l for $h = h_i, i = 1, 2, \dots, N$.

As in [17], a nondecreasing function $\psi : [0, 1] \rightarrow [0, 1]$ will be called admissible for $t \in (0, 1), y \in C^1[0, 1], H_N, D_h^l$ if it satisfies (2.5) for any $h \in H_N$ and $\psi(\delta) < d_l$.

Let $\Psi_t(y) = \Psi_t(y, H_N, D_h^l)$ be the set of all such admissible functions. In view of (2.5) the quantity

$$e_\delta(y, H_N, D_h^l) = \inf_{\psi \in \Psi_t(y)} \min_{h \in H_N} \left\{ \psi(h) + d_l \frac{\delta}{h} \right\}$$

is the best possible accuracy that can be guaranteed for approximation $y'(t)$ within the framework of the method D_h^l under assumption (2.1). We will now present a principle for the adaptive choice of the stepsize $h_+ \in H_N$ that allows us to reach this best possible accuracy up to multiplier 6ρ , where

$$\rho = \rho(H_N) = \max_{i=1,2,\dots,N} \frac{h_{i+1}}{h_i}.$$

As we will see, such h_+ can be chosen without any a priori information concerning smoothness $y \in C^1[0, 1]$. The idea of our adaptive principle has its origin in the paper [11], devoted to statistical estimation of function $y(t)$ with unknown Hölder smoothness from direct observation blurred by Gaussian white noise. In the context of general ill-posed problems this idea has been realized in [5], [25], [8], [15]–[17], [20]. We use this idea for the adaptive choice of the stepsize in finite-difference methods because the structure of the error estimate (2.5) is very similar to the loss function of statistical estimation, where some parameter always controls the trade-off between the bias and the variance of the risk. If, as is usual for statisticians, we will treat the terms in the right-hand side of (2.5) as bias and variance, respectively, then the idea is to choose the maximal h for which the “bias” $\psi(h)$ is still dominated by the “variance” $d_l \frac{\delta}{h}$.

Let $H_N^\delta(D_h^l)$ be the set of all $h_i \in H_N$ such that for any $h_j \leq h_i, h_j \in H_N$,

$$|D_{h_i}^l y_\delta(t) - D_{h_j}^l y_\delta(t)| \leq 4d_l \frac{\delta}{h_j}, \quad j = 1, 2, \dots, i.$$

The stepsize h_+ we are interested in is now defined as

$$h_+ = \max\{h_i \in H_N^\delta(D_h^l)\}.$$

We stress that the admissible functions $\psi \in \Psi_t(y)$, as well as any other information concerning the smoothness of the function to be differentiated, are not involved in the process of the choice of h_+ . We can now formulate the main result of this section.

Theorem 2.1. *For any $y \in C^1[0, 1]$*

$$|y'(t) - D_{h_+}^l y_\delta(t)| \leq 6\rho e_\delta(y, H_N, D_h^l).$$

Proof. Let $\psi \in \Psi_t(y, H_N, D_h^l)$ be any admissible function and let us temporarily introduce the stepsizes

$$\begin{aligned}
 h_{j_0} &= h_{j_0}(\psi) = \max\{h_j \in H_N : \psi(h_j) \leq d_l \frac{\delta}{h_j}\}, \\
 h_{j_1} &= h_{j_1}(\psi) = \operatorname{argmin}\{\psi(h_j) + d_l \frac{\delta}{h_j}, h_j \in H_N, j = 1, 2, \dots, N\}.
 \end{aligned}$$

Observe that

$$(2.6) \quad \frac{d_l \delta}{h_{j_0}} \leq \rho(\psi(h_{j_1}) + \frac{d_l \delta}{h_{j_1}}),$$

because either $h_{j_1} \leq h_{j_0}$ in which case

$$\frac{d_l \delta}{h_{j_0}} \leq \frac{d_l \delta}{h_{j_1}} < \rho(\psi(h_{j_1}) + \frac{d_l \delta}{h_{j_1}}), \quad \rho > 1,$$

or $h_{j_0} < h_{j_0+1} \leq h_{j_1}$. But then, by the definition of h_{j_0} , it holds true that $\frac{d_l \delta}{h_{j_0+1}} < \psi(h_{j_0+1})$ and

$$\frac{d_l \delta}{h_{j_0}} = \frac{h_{j_0+1}}{h_{j_0}} \frac{d_l \delta}{h_{j_0+1}} \leq \rho \psi(h_{j_0+1}) \leq \rho \psi(h_{j_1}) < \rho(\psi(h_{j_1}) + \frac{d_l \delta}{h_{j_1}}).$$

We now show that $h_{j_0} \leq h_+$. Indeed, for any $h_j \leq h_{j_0}$, $h_j \in H_N$,

$$\begin{aligned}
 (2.7) \quad |D_{h_{j_0}}^l y_\delta(t) - D_{h_j}^l y_\delta(t)| &\leq |y'(t) - D_{h_{j_0}}^l y_\delta(t)| + |y'(t) - D_{h_j}^l y_\delta(t)| \\
 &\leq \psi(h_{j_0}) + d_l \frac{\delta}{h_{j_0}} + \psi(h_j) + d_l \frac{\delta}{h_j} \\
 &\leq 2\psi(h_{j_0}) + d_l \frac{\delta}{h_{j_0}} + d_l \frac{\delta}{h_j} \\
 &\leq 3d_l \frac{\delta}{h_{j_0}} + d_l \frac{\delta}{h_j} \leq 4d_l \frac{\delta}{h_j}.
 \end{aligned}$$

It means that $h_{j_0} \in H_N^\delta(D_h^l)$ and

$$h_{j_0} \leq h_+ = \max\{h_i \in H_N^\delta(D_h^l)\}.$$

Using this and (2.6), one can continue as follows:

$$\begin{aligned}
 |y'(t) - D_{h_+}^l y_\delta(t)| &\leq |y'(t) - D_{h_{j_0}}^l y_\delta(t)| + |D_{h_{j_0}}^l y_\delta(t) - D_{h_+}^l y_\delta(t)| \\
 &\leq \psi(h_{j_0}) + d_l \frac{\delta}{h_{j_0}} + 4d_l \frac{\delta}{h_{j_0}} \leq 6d_l \frac{\delta}{h_{j_0}} \leq 6\rho(\psi(h_{j_1}) + d_l \frac{\delta}{h_{j_1}}) \\
 &\leq 6\rho \min_{h \in H_N} \{\psi(h) + d_l \frac{\delta}{h}\}.
 \end{aligned}$$

This estimation holds true for the arbitrary admissible function $\psi \in \Psi_t(y)$. Therefore, we conclude that

$$|y'(t) - D_{h_+}^l y_\delta(t)| \leq 6\rho \inf_{\psi \in \Psi_t(y)} \min_{h \in H_N} \{\psi(h) + d_l \frac{\delta}{h}\}.$$

The proof is complete. □

Remark 2.1. In view of (2.7) the rule for the choice h_+ can also be formulated in one of the following forms:

$$h_+ = \max\{h_j \in H_N : |D_{h_j}^l y_\delta(t) - D_{h_i}^l y_\delta(t)| \leq d_l \delta \left(\frac{3}{h_j} + \frac{1}{h_i}\right), i = 1, 2, \dots, j\},$$

$$h_+ = \max\{h_j \in H_N : |D_{h_j}^l y_\delta(t) - D_{h_i}^l y_\delta(t)| \leq 2d_l \delta \left(\frac{1}{h_j} + \frac{1}{h_i}\right), i = 1, 2, \dots, j\}.$$

It is easy to check that for these rules Theorem 2.1 is still valid. On the other hand, in a practical test, these rules did sometimes produce more accurate results.

2.2. Numerical test. We test our rule for the adaptive choice of the stepsize on the function

$$(2.8) \quad y(t) = |t|^7 + |t - 0.25|^7 + |t - 0.5|^7 + |t - 0.75|^7 + |t - 0.85|^7 \in C^6[0, 1].$$

The numerical values $y_\delta(t + jh)$ used in the test are the results of a simple program in which the perturbation $y(t + jh) - y_\delta(t + jh) \in [-\delta, \delta]$, $\delta = 0.01 \cdot y(0.5) \sim 10^{-5}$, is produced by a uniform random number generator.

The function (2.8) was used for numerical experiments in the paper [21], where several new and rather sophisticated finite-difference methods were proposed. We borrow two of them. Namely, D_h^2 with the coefficients

$$a_0^2 = 0, \quad a_1^2 = a_{-1}^2 = \frac{2}{3}, \quad a_2^2 = -a_{-2}^2 = -\frac{1}{12},$$

and D_h^4 with the coefficients

$$a_0^4 = 0, a_1^4 = a_{-1}^4 = \frac{6528}{8760}, \quad a_2^4 = -a_{-2}^4 = -\frac{1272}{8760},$$

$$a_3^4 = -a_{-3}^4 = \frac{128}{8760}, \quad a_4^4 = -a_{-4}^4 = \frac{3}{8760}.$$

Moreover, we also use well-known centered difference approximation D_h^1 with the coefficients $a_0^1 = 0, a_1^1 = -a_{-1}^1 = \frac{1}{2}$. The above-mentioned formulae meet (2.4) with a different value r . Namely, for D_h^1 one has

$$D_h^1 y(t) = y'(t) + \frac{y^{(3)}(\theta)}{2} h^2,$$

and in [21] it has been shown that

$$D_h^2 y(t) = y'(t) - \frac{y^{(6)}(\theta)}{30} h^5, \quad D_h^4 y(t) = y'(t) + \frac{4}{511} y^{(8)}(\theta) h^7,$$

for some $\theta \in (0, 1)$. As to the stability problem due to noise error propagation, this was not discussed in [21]. As the same time, in Remark 5.1 of [21] it has been suggested that according to the regularity of the function, the corresponding formula should be used for the estimate of its derivative, because for nonsmooth functions like (2.8) all the higher order formulae produce even worse results. In accordance with this suggestion the method D_h^2 with $h = 0.1$ has been used in [21] for estimating $y'(0.5) = 0.09650714062500$. Obtained accuracy is $0.001071 \sim 10^{-3}$.

In our test we applied the above-mentioned formulae to noisy data $y_\delta(t + jh)$ with the stepsizes $h = h_i = 0.02 \cdot i, i = 1, 2, \dots$, and the first rule from Remark 2.1 was used for determining h_+ . For the method D_h^2 it gave $h_+ = 0.12$, and the value $y'(0.5)$ was estimated with error $0.001885 \sim 10^{-3}$, i.e., for noisy data the same order of accuracy as in [21] was obtained. It is perhaps also instructive to see that

for the stepsize $h = 0.2$, for example, the error is 0.016934, i.e., almost ten times larger.

For the methods D_h^1 and D_h^4 the results are even better. Namely, D_h^1 , $h = h_+ = 0.06$, and D_h^4 , $h = h_+ = 0.2$, give the value $y'(0.5)$ with errors $-5.3209 \cdot 10^{-4}$ and $-2.44106 \cdot 10^{-4}$, respectively. These tests do not support the suggestion from Remark 5.4 of [21] that in practice only the lower order formulae should be used so that no unexpected errors could occur. But from a viewpoint of regularization theory the results of the tests are not so surprising, because the order of some finite-difference method D_h^l (i.e., the highest possible r in (2.4)) can be interpreted as a qualification of the regularization method based on D_h^l , and, as it is well known (see, for example, [15]), if the regularization parameter (i.e., stepsize h) is properly chosen, then the higher the qualification of the method is, the better the results that can be obtained.

Another interesting observation is that in the considered case $D_{h_+}^1$ and $D_{h_+}^4$ give the same order of accuracy, 10^{-4} . It also is in good agreement with the theory, because function (2.8) belongs to $C^r[0, 1]$ for $r = 6$, and the best possible order of accuracy that can be guaranteed for such functions under the noise level $\delta \sim 10^{-5}$ is $\delta^{\frac{r-1}{r}} = \delta^{\frac{5}{6}} \sim 10^{-4}$. Thus, in the considered case both methods can realize the best order of accuracy provided that the stepsize is chosen properly.

Note that for each specific y and y_δ the best possible stepsize $h_{\text{ideal}} \in H_N$ could be defined as

$$h_{\text{ideal}} = \arg \min\{|y'(t) - D_{h_i}^l y_\delta(t)|, h_i \in H_N, i = 1, 2, \dots, N\}.$$

Of course, such h_{ideal} is not numerically feasible. Our next test shows how far h_+ can be from this ideal stepsize.

Consider $y(t) = \sin(t - 0.4)/(t - 0.4)$ and simulate noisy data in such a way that $y_\delta(t \pm jh) = y(t \pm jh) \pm (-1)^j \delta$, $\delta = 10^{-5}$. We use a centered difference approximation D_h^1 defined above. Then for $t = 0.5$ and $H_{15} = \{0.02 \cdot i, i = 1, 2, \dots, 15\}$, $h_{\text{ideal}} = 0.16$, $h_+ = 0.28$. On the other hand, the error of the method D_h^1 with $h = h_{\text{ideal}}$ is $1.47654 \cdot 10^{-4}$, while for $h = h_+$ D_h^1 approximate $y'(0.5)$ with the error $2.96007 \cdot 10^{-4}$. As one can see, in considered case h_+ differs from h_{ideal} . Nevertheless, D_h^1 gives the accuracy of the same order as a finite-difference method with the ideal stepsize.

3. NUMERICAL DIFFERENTIATION WITH DISCRETE NOISY INFORMATION

3.1. Setting of the problem. Suppose $y(t)$ is a smooth function that has at least one square integrable derivative on the interval $[0, 1]$, i.e., $y \in W_2^1(0, 1)$. Assume that noisy samples y_i^ϵ of the values $y(t_i)$ are known at the points of the grid $\sigma = \{0 = t_0 < t_1 < \dots < t_n = 1\}$. Let $|\sigma| = \max\{t_i - t_{i-1}, i = 1, 2, \dots, n\}$ be the mesh size of the grid and suppose

$$(3.1) \quad |y(t_i) - y_i^\epsilon| \leq \epsilon, \quad i = 0, 1, 2, \dots, n,$$

where ϵ is a known level of noise in the data. We are interested in finding an approximation of $y'(t)$ from the given data $\{y_i^\epsilon\}$.

At this point we note that in the situation considered the use of finite-difference methods discussed in Section 2 may not be very satisfactory since the values $y_\delta(t + jh)$ may not be available for all desired stepsizes h .

Therefore, one usually takes an approach discussed, for example, in [4, 19, 22], and rewrites the numerical differentiation of a smooth function y as a Volterra problem

$$(3.2) \quad Ax(t) := \int_0^t x(\tau)d\tau = y(t), \quad 0 \leq t \leq 1.$$

Below we assume that

$$(3.3) \quad y(0) = y_0^\epsilon = 0,$$

i.e., the initial data are known exactly. Then it is clear that $x(t) = y'(t)$ is a unique solution of (3.2).

On the other hand, since only the noisy samples $\{y_i^\epsilon\}$ are available, one has an equation

$$(3.4) \quad Ax = y, \quad \|y - y_\delta\|_{L_2} \leq \delta,$$

which is exactly the problem to be solved, where y_δ and A are given and x and y are unknown. If the grid σ is fixed, then, as it is well known (see, for example, [13]), the best possible order of accuracy that can be guaranteed for the recovery of function $y \in W_2^1$ from a noisy sample $\{y_i^\epsilon\}$ is $O(|\sigma| + \epsilon)$. This optimal order can be realized by the piecewise linear interpolation

$$S_\sigma(\{y_i^\epsilon\}; t) = \sum_{i=1}^n b_i^\sigma(t)y_i^\epsilon,$$

where

$$b_i^\sigma(t) = \begin{cases} \frac{t-t_{i-1}}{t_i-t_{i-1}}, & t \in [t_{i-1}, t_i], \\ \frac{t_{i+1}-t}{t_{i+1}-t_i}, & t \in [t_i, t_{i+1}], i = 1, 2, \dots, n-1, \\ 0, & t \notin [t_{i-1}, t_{i+1}], \end{cases}$$

$$b_n^\sigma(t) = \begin{cases} \frac{t-t_{n-1}}{t_n-t_{n-1}}, & t \in [t_{n-1}, t_n], \\ 0, & t \notin [t_{n-1}, t_n]. \end{cases}$$

Indeed,

$$\begin{aligned} \|y - S_\sigma(\{y_i^\epsilon\}; \cdot)\|_{L_2} &\leq \|y - S_\sigma(\{y(t_i)\}; \cdot)\|_{L_2} + \left\| \sum_{i=1}^n b_i^\sigma(\cdot)(y(t_i) - y_i^\epsilon) \right\|_{L_2} \\ &\leq c|\sigma|\|y'\|_{L_2} + \epsilon. \end{aligned}$$

It means that y_δ and δ can be chosen in (3.4) as

$$(3.5) \quad y_\delta(t) = S_\sigma(\{y_i^\epsilon\}; \cdot), \quad \delta = d_0(\epsilon + |\sigma|),$$

where d_0 is some designed parameter (constant) that can be fixed using any a priori estimation for $\|y'\|_{L_2}$, or assuming $|\sigma|$ to be so small that $\|y'\|_{L_2}|\sigma|$ can be neglected compared to ϵ .

Several regularization methods can be applied to numerical differentiation treated as an ill-posed problem (3.4), (3.5). For example, it has been observed in [22] that the operator A from (3.4) acts in the Hilbert space $L_2(0, 1)$ as a monotone operator, i.e., $\forall x \in L_2(0, 1), \langle Ax, x \rangle \geq 0$, where $\langle \cdot, \cdot \rangle$ is the standard inner product in $L_2(0, 1)$.

This observation allows us to use Lavrentiev regularization, as has been suggested in [12, 24] for ill-posed operator equations with monotone operators. Moreover, in [22] a regularization method for solving (3.4) as a special case of a nonlinear ill-posed problem with monotone operators has been proposed, and its convergence for $\delta \rightarrow 0$ has been proved. But these methods as well as any other regularization techniques are only numerically feasible after appropriate discretization. On the other hand, a general discretization strategy for linear ill-posed problems has been recently discussed in [16]. In the next subsection we apply the results from [16] to the numerical differentiation problem (3.4), (3.5). It will allow us to construct a simple numerical scheme which automatically adapts to any unknown smoothness of the function to be differentiated.

3.2. Discretized Tikhonov regularization for numerical differentiation.

Tikhonov regularization and its iterated version are probably the most widely known regularization techniques. Therefore, we restrict ourselves to the analysis of these methods.

The Tikhonov method for a noisy linear equation (3.4) consists, it will be recalled, in determining the regularized approximation x_α^δ as a unique solution of the equation

$$(3.6) \quad \alpha x_\alpha^\delta + A^* A x_\alpha^\delta = A^* y_\delta,$$

where α is a positive regularization parameter.

Within the framework of the iterated Tikhonov method of order p the regularized approximation $x_{\alpha,p}^\delta$ is determined by the recursion

$$(3.7) \quad \begin{aligned} \alpha x_{\alpha,l}^\delta + A^* A x_{\alpha,l}^\delta &= \alpha x_{\alpha,l-1}^\delta + A^* y_\delta, \\ l = 1, 2, \dots, p, \quad x_{\alpha,0}^\delta &= 0, \quad x_{\alpha,1}^\delta = x_\alpha^\delta, \end{aligned}$$

i.e., the equation of the form (3.6) should be solved p times.

To discretize the equations (3.6), (3.7) one can use a Galerkin method based, for example, on the trial space of piecewise linear functions $V_{m+1} = \text{span}\{b_i^m\}_{i=0}^m$, where

$$b_i^m(t) = b_i^\sigma(t), \quad i = 1, 2, \dots, m, \quad \sigma_m = \left\{ \frac{i}{m} \right\}_{i=0}^m, \quad b_0^m(t) = b_m^m(1-t),$$

and b_i^σ are defined above. Then the Galerkin approximation $x_{\alpha,l,m}^\delta$ of $x_{\alpha,l}^\delta$ has the form

$$(3.8) \quad x_{\alpha,l,m}^\delta(t) = \sum_{i=0}^m z_i^l b_i^m(t),$$

and should solve the variational problem

$$(3.9) \quad \langle v, \alpha x_{\alpha,l,m}^\delta + A^* A x_{\alpha,l,m}^\delta - \alpha x_{\alpha,l-1,m}^\delta - A^* y_\delta \rangle = 0$$

for all $v \in V_{m+1}$. It is convenient to rewrite (3.9) as the following system of linear algebraic equations with respect to unknown coefficients z_i^l from (3.8):

$$\begin{aligned}
 (3.10) \quad & \alpha \sum_{i=0}^m z_i^l \langle b_i^m, b_j^m \rangle + \sum_{i=0}^m z_i^l \langle Ab_i^m, Ab_j^m \rangle \\
 & = \alpha \sum_{i=0}^m z_i^{l-1} \langle b_i^m, b_j^m \rangle + \langle Ab_j^m, y_\delta \rangle, \\
 & j = 0, 1, \dots, m, \quad l = 1, 2, \dots, p, \quad z_i^0 = 0, \quad i = 0, 1, \dots, m.
 \end{aligned}$$

Keeping in mind that b_i^m and y_δ are piecewise linear functions, the primitives Ab_j^m as well as the entries of the associated stiffness matrix and of the right-hand side of (3.10) can be computed exactly. Moreover, Ab_i^m , $\langle Ab_i^m, Ab_j^m \rangle$, $\langle b_i^m, b_j^m \rangle$ can be precomputed in advance.

Thus, the numerical scheme (3.8), (3.10) can be easily realized, and the main question now is connected with the choice of the discretization parameter m and the regularization parameter α .

In principle, one could choose the regularization parameter α following the approach described in Section 2, provided the error $\|y' - x_{\alpha,l,m}^\delta\|$ has the bias-variance structure given in (2.5). The difference is that $\|y' - x_{\alpha,l,m}^\delta\|$ is driven by two parameters α and m , while $|y' - D_h^l y^\delta|$ depends only on one parameter h . One possibility is to choose m in such a way that the error caused by the discretization will be dominated by the regularization error. To choose such m one should estimate the contribution of the discretization to the total error, and this requires special assumptions concerning the smoothness of $x = y'$. This distinguishes discretized regularization of numerical differentiation (3.2) from finite-difference methods, which can be regularized without any additional smoothness assumptions.

When studying numerical differentiation (3.2), one usually assumes that the function y belongs to an appropriate Sobolev or Besov space. The most convenient way, however, constructs the smoothness class directly from the underlying operator A , and thus measures the smoothness in terms of general source conditions

$$(3.11) \quad x = y' = \varphi(A^*A)v, \quad v \in L_2(0, 1),$$

an approach which recently has become attractive (see [9], [23], [3], [15]-[18]). The function φ here is continuous, increasing and satisfies $\varphi(0) = 0$. It is more flexible to describe smoothness than just the usual scales of Sobolev or Besov spaces. For example, (3.11) with $\varphi(\lambda) = \lambda^k$, $k = 1, 2, \dots$, means that y' has a Sobolev smoothness described as $y' \in W_2^{2k}(0, 1)$.

In [16] the authors argue that in dealing with discretized regularization methods it is convenient to assume that the smoothness index function φ in (3.11) can be represented as a product

$$(3.12) \quad \varphi(\lambda) = \theta(\lambda)\psi(\lambda), \quad \lambda \in [0, 1], \quad \theta(0) = 0,$$

of some nondecreasing Lipschitz continuous function $\psi(\lambda)$ and operator monotone function $\theta(\lambda)$. Recall that the function $\theta(\lambda)$ is operator monotone on $[0, 1]$ if, for any pair of self-adjoint operators U, V with spectra in $[0, 1]$ such that $U \leq V$, we have $\theta(U) \leq \theta(V)$ (i.e., $\forall x, \langle \theta(U)x, x \rangle \leq \langle \theta(V)x, x \rangle$).

For the sake of simplicity we assume, more specifically, either $\theta^2(\lambda)$ to be concave, or $\theta(\lambda) \leq c\sqrt{\lambda}$, where c is some positive constant. The classes of such operator

monotone functions will be denoted by M_+ and M_- , respectively. Observe that up to a certain extent these classes complement each other, because for any $\theta \in M_+$, $\theta(0) = 0$, $\theta^2(\lambda) \geq \theta^2(1)\lambda$, and thus $\theta(\lambda) > c\sqrt{\lambda}$ for $c < \sqrt{\theta(1)}$.

Let $(M_+ \cup M_-) \times \text{Lip}$ be the class of all functions represented in the form (3.12) with $\theta \in M_+ \cup M_-$, $\psi \in \text{Lip}$. Examples 1-4 of [14] show that the general source conditions (3.11) with $\varphi \in (M_+ \cup M_-) \times \text{Lip}$ cover all types of smoothness studied so far in the theory of ill-posed problems.

Let Q_m be an orthogonal projector onto the trial space V_{m+1} defined above. It is easy to see that the numerical scheme (3.8), (3.10) is a combination of iterated Tikhonov method (3.7) using a projection method based on Q_m , V_{m+1} , because $x_{\alpha,l,m}^\delta$ solves an equation

$$\alpha x + Q_m A^* A Q_m x = \alpha Q_m x_{\alpha,l-1,m}^\delta + Q_m A^* y_\delta,$$

which is just a projection of (3.7) onto V_{m+1} .

A general discretization strategy for solving linear ill-posed problems by projection methods has been proposed recently in [16]. In the considered case this strategy suggests choosing m to be the smallest integer such that

$$(3.13) \quad \|A - A Q_m\|_{L_2 \rightarrow L_2} \leq \min\{\sqrt{\alpha}, \frac{\delta}{\sqrt{\alpha}}\}.$$

Observe that

$$\|A - A Q_m\|_{L_2 \rightarrow L_2} = \|(I - Q_m)A^*\|_{L_2 \rightarrow L_2},$$

where

$$A^* x(t) = \int_t^1 x(\tau) d\tau, \quad t \in [0, 1],$$

acts as a linear continuous operator from L_2 to the Sobolev space W_2^1 , $\|A^*\|_{L_2 \rightarrow W_2^1} < 2$. Moreover, it is well known that the accuracy of the piecewise linear approximation can be estimated as

$$\|I - Q_m\|_{W_2^1 \rightarrow L_2} = \sup_{g: \|g\|_{W_2^1} \leq 1} \inf_{g_m \in V_{m+1}} \|g - g_m\|_{L_2} \leq c_0 m^{-1},$$

where c_0 is some absolute constant. Then

$$\|A - A Q_m\|_{L_2 \rightarrow L_2} \leq \|I - Q_m\|_{W_2^1 \rightarrow L_2} \|A^*\|_{L_2 \rightarrow W_2^1} < 2c_0 m^{-1},$$

and the discretization parameter m can be chosen as

$$(3.14) \quad m = 2c_0 \max\{\alpha^{-\frac{1}{2}}, \alpha^{\frac{1}{2}} \delta^{-1}\}.$$

It has been observed in [16] that such discretization strategy is the most efficient among all known ones in the sense of the size of the corresponding linear system (3.10). If this size is not a crucial point of the computational procedure, then the discretization parameter can be chosen independently of α in such a way that the error caused by the data noise (inevitable error) would not be dominated by the discretization error. Such a value of m can also be extracted from (3.13), (3.14) using the well-known fact that the value of regularization parameter α should be larger than δ^2 . Then $\max\{\alpha^{-\frac{1}{2}}, \alpha^{\frac{1}{2}} \delta^{-1}\} \leq \delta^{-1}$ and

$$(3.15) \quad m = 2c_0 \delta^{-1}$$

also meet (3.13).

Proposition 3.1. *Assume that y' satisfies (3.11) with $\varphi \in (M_+ \cup M_-) \times \text{Lip}$, and p is such that $\frac{\lambda^p}{\varphi(\lambda)}$ is nondecreasing for $\lambda \in [0, 1]$. Then for m chosen as (3.14) or (3.15)*

$$\|y' - x_{\alpha,p,m}^\delta\|_{L_2} \leq c_1\varphi(\alpha) + \frac{2(2 + \sqrt{p})\delta}{\sqrt{\alpha}},$$

where δ is given by (3.5), and the constant c_1 does not depend on α, δ, m .

This proposition follows immediately from Corollary 1 of [16].

Thus, under the assumptions of Proposition 3.1 the estimation for the error $\|y' - x_{\alpha,p,m}^\delta\|$ has the bias-variance structure just as in (2.5), and the approach described in Section 2 can be applied for choosing the regularization parameter α .

In practical applications the value of the regularization parameter α is often selected from some geometric sequence

$$\Delta_N = \{\alpha_i = \delta^2 q^i, i = 0, 1, \dots, N\}, \quad q > 1, \quad q^N \sim \delta^{-2}.$$

Then by analogy with the adaptive strategy described in Section 2 we choose $\alpha_+ \in \Delta_N$ as

$$(3.16) \quad \alpha_+ = \max\{\alpha_i \in \Delta_N : \|x_{\alpha_i,p,m}^\delta - x_{\alpha_j,p,m}^\delta\|_{L_2} \leq \frac{8(2 + \sqrt{p})\delta}{\sqrt{\alpha_j}}, \\ j = 0, 1, \dots, i\}.$$

Theorem 3.1. *Under the assumption of Proposition 3.1 for m chosen as (3.14) or (3.15)*

$$\|y' - x_{\alpha_+,p,m}^\delta\|_{L_2} \leq c\varphi(\theta_\varphi^{-1}(\delta)),$$

where $\theta_\varphi(\lambda) = \sqrt{\lambda}\varphi(\lambda)$, and the constant c depends only on φ, c_1, q .

Proof (Sketch). We apply the same arguments as in the proof of Theorem 2.1, but this time $h_j = \sqrt{\alpha_j}$, $j = 0, 1, \dots, N$, $\rho = \sqrt{q}$, $l = p$, $d_p = 2(2 + \sqrt{p})$, and we deal only with one admissible function $\psi(h_j) = c_1\varphi(h_j^2)$, where c_1 is the constant from Proposition 3.1. Let \bar{h} be a solution of equation $\psi(h) = d_p \frac{\delta}{h}$, i.e., $\bar{h} = (\theta_\varphi^{-1}(d_p \frac{\delta}{c_1}))^{\frac{1}{2}}$.

In the same way as in the proof of Theorem 2.1 one can show that

$$h_{j_0} \leq h_+ = \sqrt{\alpha_+}, \quad h_{j_0} \leq \bar{h} \leq \rho h_{j_0},$$

and

$$\|y' - x_{\alpha_+,p,m}^\delta\| \leq \frac{6d_p\delta}{h_{j_0}} \leq \frac{6\rho d_p\delta}{\bar{h}} = 6\rho\psi(\bar{h}) = 6\rho\varphi(\theta_\varphi^{-1}(\frac{d_p\delta}{c_1})).$$

To complete the proof we need to show that there exists a constant c_p (depending only on p and c_1) such that

$$\varphi(\theta_\varphi^{-1}(\frac{d_p\delta}{c_1})) \leq c_p\varphi(\theta_\varphi^{-1}(\delta)).$$

If $d_p \leq c_1$, then in view of the monotony of φ and θ_φ^{-1} this inequality holds true with any $c_p \geq 1$.

Consider the case $d_p > c_1$. Since $\frac{\lambda^p}{\varphi(\lambda)}$ is nondecreasing we have

$$\frac{\lambda^p}{\varphi(\lambda)} \leq \frac{(2\lambda)^p}{\varphi(2\lambda)} \Rightarrow \varphi(2\lambda) \leq 2^p\varphi(\lambda).$$

It means that $\varphi(\lambda)$ satisfies the so-called Δ_2 -condition: $\varphi(2\lambda) \leq c\varphi(\lambda)$ with $c = 2^p$.

For any $\gamma > 1$ iterating this Δ_2 -condition one can find the integer number $i \leq \log_2 2\gamma$ such that $2^{i-1} \leq \gamma \leq 2^i$, and for any $\lambda \in [0, 1]$

$$\varphi(\gamma\lambda) \leq 2^p \varphi\left(\frac{\gamma}{2}\lambda\right) \leq 2^{2p} \varphi\left(\frac{\gamma}{2^2}\lambda\right) \leq \dots \leq 2^{pi} \varphi\left(\frac{\gamma}{2^i}\lambda\right) \leq (2\gamma)^p \varphi(\lambda).$$

Moreover, in view of monotony of φ and θ_φ^{-1} for any $\beta > 1$

$$\beta = \frac{\theta_\varphi(\theta_\varphi^{-1}(\beta\lambda))}{\theta_\varphi(\theta_\varphi^{-1}(\lambda))} = \frac{\varphi(\theta_\varphi^{-1}(\beta\lambda))\sqrt{\theta_\varphi^{-1}(\beta\lambda)}}{\varphi(\theta_\varphi^{-1}(\lambda))\sqrt{\theta_\varphi^{-1}(\lambda)}} \geq \sqrt{\frac{\theta_\varphi^{-1}(\beta\lambda)}{\theta_\varphi^{-1}(\lambda)}},$$

i.e.,

$$\theta_\varphi^{-1}(\beta\lambda) \leq \beta^2 \theta_\varphi^{-1}(\lambda).$$

Then,

$$\varphi\left(\theta_\varphi^{-1}\left(\frac{d_p}{c_1}\delta\right)\right) \leq \varphi\left(\frac{d_p^2}{c_1^2}\theta_\varphi^{-1}(\delta)\right) \leq \left(2\frac{d_p^2}{c_1^2}\right)^p \varphi(\theta_\varphi^{-1}(\delta)),$$

which allows us to complete the proof. □

Note that smoothness index function φ from (3.11) is not involved in the process of the choice of α_+ . Moreover, from Corollary 1 of [15] it follows that under the assumptions (3.4) and (3.11) the order of accuracy $\varphi(\theta_\varphi^{-1}(\delta))$ cannot be improved in general.

3.3. Numerical examples. A program to test the numerical viability of the schemes (3.8), (3.10) with parameter choice rule (3.16) was written in MATLAB. In the following numerical tests we choose α from $\Delta_{30} = \{\alpha_i = 0.00008 \cdot (1.1)^i, i = 0, 1, \dots, 30\}$. The number of grid points is $m = 200$. The number of observation points is $n = 200$, and these points are randomly distributed in $[0, 1]$. The perturbations $y(t_i) - y_i^\epsilon$ were produced by a uniform random number generator with $\epsilon = 0.001$. The parameter d_0 in (3.5) was taken as $d_0 = 10^{-1}$, i.e., $\delta \sim \epsilon$. Moreover, we used a simple Tikhonov regularization method, i.e., $p = 1$.

For the above-mentioned values of parameters, four tests of the scheme (3.8), (3.10), (3.16) were conducted. The results are shown in Figures 1–4, where the dotted curve represents y' and the solid curve is $x_{\alpha_+, 1, 200}$.

First, we ran the algorithm using (2.8) as the function to be differentiated (see Figure 1). In this case $\alpha_+ = \alpha_3 = 0.0001065$.

Tests 2–4 show that by our adaptive scheme we can find the discontinuous points of $y(t)$ or $y'(t)$. Since the values of $y(t)$ are only given at finite points, we still can run our algorithm. Tests 2–4 are inspired by the paper [26].

We use the functions

$$\begin{aligned} y_2(t) &= \begin{cases} 1 - t, & t \in [0, 0.5], \\ t, & t \in (0.5, 1], \end{cases} & y_3(t) &= \begin{cases} 1 + t, & t \in [0, 0.5], \\ t, & t \in (0.5, 1], \end{cases} \\ y_4(t) &= \begin{cases} -t, & t \in [0, 0.5], \\ t, & t \in (0.5, 1]. \end{cases} \end{aligned}$$

The results are shown in Figures 2 and 4, respectively, where $t = 0.5$ can be easily detected as the discontinuity point for $y_3(t)$, $y_4(t)$, and $y'_2(t)$, $y'_3(t)$, $y'_4(t)$.

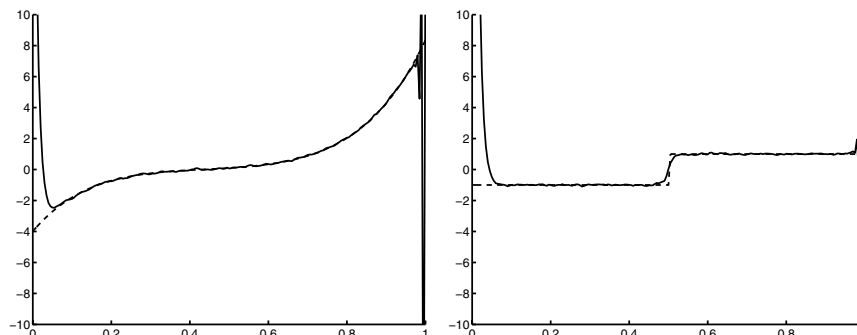


FIGURE 1. Numerical example for (2.8). FIGURE 2. Numerical example for y_2 .

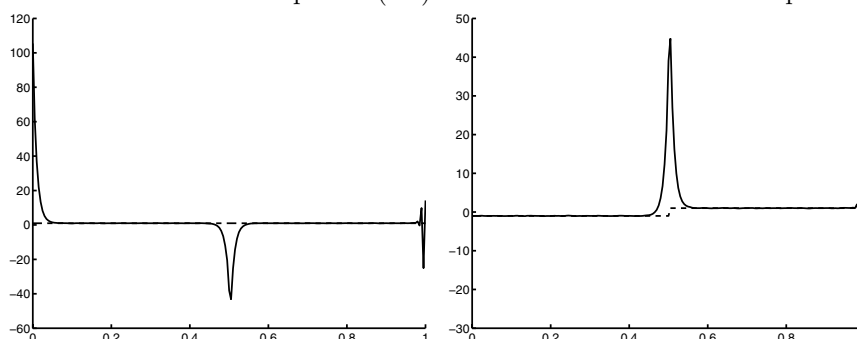


FIGURE 3. Numerical example for y_3 . FIGURE 4. Numerical example for y_4 .

In all these cases $\alpha_+ = \alpha_2 = 0.000088$.

Note that the leap near $t = 0$ in Figures 1–3 is caused by the fact that the functions (2.8), $y_2(t)$, $y_3(t)$ do not satisfy (3.3). So, this circumstance can also be detected by our adaptive scheme.

As to the leap near $t = 1$, in all figures, it is because of our temptation to find the derivative $y'(1)$ using only the values $y(t)$ at the points from one side of $t = 1$. If the values $y(t)$ for $t > 1$ are used, this leap disappears.

It is interesting to note that in the numerical tests from [26] the second derivative $y''(t)$ was used to localize the discontinuity points of $y'(t)$, $y(t)$. Our adaptive regularization procedure allows us to do it with the same accuracy using only current approximation of $y'(t)$.

To show the tightness of the bound from Theorem 3.1, we test the scheme (3.8)–(3.10) in the situation when the index function φ is known.

Consider the function

$$y(t) = \begin{cases} \frac{t^5}{120} - \frac{t^3}{16} + \frac{57}{384}t, & t \in [0, 0.5], \\ \frac{t^4}{48} - \frac{t^3}{12} + \frac{t^2}{96} + \frac{7}{48}t - \frac{1}{3840}, & t \in (0.5, 1], \end{cases}$$

chosen in such a way that the solution of (3.2) is given as

$$x(t) = \begin{cases} \frac{t^4}{24} - \frac{3}{16}t^2 + \frac{57}{384}, & t \in [0, 0.5], \\ \frac{t^3}{12} - \frac{t^2}{4} + \frac{t}{48} + \frac{7}{48}, & t \in (0.5, 1]. \end{cases}$$

It is easy to check that this function can be represented in the form (3.11) with $\varphi(\lambda) = \lambda^2$ and

$$v(t) = \begin{cases} 1, & t \in [0, 0.5], \\ 0, & t \in (0.5, 1], \end{cases}$$

i.e., $x(t) = (A^*A)^2v(t)$, where A is a Volterra integral operator (3.2). Using $y(t) = Ax(t)$ we simulate noisy data $y_\delta(t)$ in the same way as above. But this time $\delta \sim \varepsilon = 10^{-4}$.

For $\varphi(\lambda) = \lambda^2$, Theorem 3.1 guarantees the accuracy of order $\delta^{4/5} \sim 10^{-16/5} \approx 0.6 \cdot 10^{-3}$, and this order can be reached within the framework of scheme (3.8)–(3.10) if $p \geq 2$, since $\lambda^p/\varphi(\lambda)$ should be a nondecreasing function.

Taking $p = 2$ we ran our algorithm with the same values of other parameters as above. It corresponds to the 2-times iterated Tikhonov method (3.7). In the considered case an adaptive strategy (3.16) gives the value of the regularization parameter $\alpha_+ = 0.013$. The corresponding value of the L_2 -error is $2.2319 \cdot 10^{-3}$, that is in good agreement with the estimation given by Theorem 3.1.

4. CONCLUSION

As has been already mentioned, the representation of a numerical differentiation problem in the form of linear operator equation (3.2) allows us to use a whole arsenal of methods developed in general regularization theory. Discretized Tikhonov regularization discussed in Section 3 is only one of such methods. In principle, any discretized realization of a general regularization scheme (see, for example Ch.4 [4]) can be used as well. It can be analyzed on the basis of the results [15], [16] in the same way as it has been done in Section 3.

At the same time, it should be noted that the general regularization scheme, and the Tikhonov regularization in particular, does not retain the Volterra structure of the numerical differentiation problem (3.2). Typical numerical realizations of the Volterra operator A lead to a lower-triangular matrix, so that the solution of the discretized problem may be handled by efficient, sequential methods. In contrast, the system (3.10), for example, has a full matrix. Several regularization methods which specifically preserve the Volterra structure can be found in a survey [10]. The Lavrentiev regularization is one of them. But the adaptive strategies for choosing regularization and discretization parameters without knowledge of solution smoothness are not developed as much as they are for the general regularization scheme. For example, as far as we know, convergence rate results have been obtained only for the case of a priori known solution smoothness.

From this viewpoint, the results of Sections 2 and 3 can be considered as complementary, because finite-difference methods are not covered by a general regularization scheme. At the same time, the numerical realization of these methods is very simple, and does not require us to solve a system of linear algebraic equations. There is, unfortunately, a price associated with a local nature of the approximation produced by any finite-difference method (it allows us to estimate a point value of the derivative only). Nevertheless these methods are widely used, and to the best of our knowledge, Theorem 2.1, which establishes an order-optimal adaptive strategy for the stepsize choice, is the first such result of its kind.

ACKNOWLEDGMENTS

This research was supported by the Austrian Fonds Zur Förderung der Wissenschaftlichen Forschung (FWF), Grant P17251-N12.

REFERENCES

1. B.Anderssen, F.de Hoog, H.Hegland, *A stable finite difference ansatz for higher order differentiation of non-exact data*, Bull. Austral. Math. Soc., 22 (1998), 223–232. MR1642035 (2000d:65037)
2. J.Cheng, M.Yamamoto, *One new strategy for a priori choice of regularizing parameters in Tikhonov's regularization*, Inverse Problems, 16 (2000), L31–38. MR1776470 (2001g:65073)
3. P.Deuffhard, H.W.Engl, O.Scherzer, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions*, Inverse Problems, 14 (1998), 1081–1106. MR1654603 (99j:65105)
4. H.W.Engl, M.Hanke, A.Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996. MR1408680 (97k:65145)
5. A.Goldenshluger, S.V.Pereverzev, *Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations*, Probab. Theory Relat. Fields, 118 (2000), 169–186. MR1790080 (2001h:62055)
6. C.W.Groetsch, *Differentiation of approximately specified functions*, Am. Math. Mon., 98 (1991), 847–850. MR1133003
7. M.Hanke, O.Scherzer, *Inverse Problems light numerical differentiation*, Am. Math. Mon., 108 (2001), 512–521. MR1840657 (2002e:65089)
8. H.Harbrecht, S.Pereverzev, R.Schneider, *Self-regularization by projection for noisy pseudo-differential equations of negative order*, Numer. Math., 95 (2003), 123–143. MR1993941 (2004j:65075)
9. M.Hegland, *Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization*, Appl. Anal., 59 (1995), 207–223. MR1378036 (97a:65060)
10. P.K.Lamm, *A survey of regularization methods for first-kind Volterra equations (English summary)*, Surveys on solution methods for inverse problems, 53–82, Springer, Vienna, 2000. MR1766739 (2001b:65148)
11. O.V.Lepskii, *A problem of adaptive estimation in Gaussian white noise*, Theory Probab. Appl., 36 (1990), 454–466. MR1091202 (93j:62212)
12. F.Liu, M.Z.Nashed, *Convergence of regularized solutions of nonlinear ill-posed problems with monotone operators*, Partial Differential Equations and Application, Dekker, New York (1996), 353–361. MR1371608 (97m:65112)
13. P.Mathe, S.V.Pereverzev, *Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods*, SIAM J. Numer. Analysis, 38 (2001), 1999–2021. MR1856240 (2002g:62063)
14. P.Mathe, S.V.Pereverzev, *Moduli of continuity for operator valued functions*, Numer. Funct. Anal. Optim., 23 (2002), 623–631. MR1923828 (2003g:47029)
15. P.Mathe, S.V.Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), 789–803. MR1984890 (2004i:47021)
16. P.Mathe, S.V.Pereverzev, *Discretization strategy for linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), 1263–1277. MR2036530 (2004k:65097)
17. P.Mathe, S.V.Pereverzev, *Regularization of some linear ill-posed problems with discretized random noisy data*, Math. Comp. (Accepted).
18. M.T.Nair, E.Schock, U.Tautenhahn, *Morozov's discrepancy principle under general source conditions*, Z. Anal. Anwendungen, 22 (2003), 199–214. MR1962084 (2004a:65069)
19. A.Neumaier, *Solving ill-conditioned and singular linear systems: A tutorial on regularization*, SIAM Rev., 40 (1998), 636–666. MR1642811 (99f:65066)
20. S.Pereverzev, E.Schock, *On the adaptive selection of the parameter in regularization of ill-posed problems*, SIAM J. Numer. Analysis, 43 (2005), 2060–2076. MR2192331
21. R.Qu, *A new approach to numerical differentiation and integration*, Math. Comput. Modelling, 24 (1996), 55–68. MR1426303 (98b:65024)
22. A.G.Ramm, A.B.Smirnova, *On stable numerical differentiation*, Math. Comp., 70 (2001), 1131–1153. MR1826578 (2002a:65046)

23. U.Tautenhahn, *Optimality for ill-posed problems under general source conditions*, Numer. Funct. Anal. Optim., 19 (1998), 377–398. MR1624930 (99g:65073)
24. U.Tautenhahn, *On the method of Lavrentiev regularization for nonlinear ill-posed problems*, Inverse Problems, 18 (2002), 191–207. MR1893590 (2002m:47079)
25. A.Tsybakov, *On the best rate of adaptive estimation in some inverse problems*, C. R. Acad. Sci. Paris Sèr. I Math., 330 (2000), 835–840. MR1769957 (2001c:62058)
26. Y.B.Wang, X.Z.Jia, J.Cheng, *A numerical differentiation method and its application to reconstruction of discontinuity*, Inverse Problems, 18 (2002), 1461–1476. MR1955897 (2004b:65086)

JOHANN RADON INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS (RICAM), AUSTRIAN ACADEMY OF SCIENCE, ALTENBERGERSTRASSE 69, A-4040 LINZ, AUSTRIA

E-mail address: `shuai.lu@oeaw.ac.at`

JOHANN RADON INSTITUTE FOR COMPUTATIONAL AND APPLIED MATHEMATICS (RICAM), AUSTRIAN ACADEMY OF SCIENCE, ALTENBERGERSTRASSE 69, A-4040 LINZ, AUSTRIA

E-mail address: `sergei.pereverzyev@oeaw.ac.at`