23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

Weighted pseudoinverse of compact operators and differentiation of signals with random noise

Sergiy Zhuk

***Abstract*— The paper presents a new algorithm which "differentiates" noisy signals with values in a Hilbert space. The algorithm provides an iterative estimation procedure which converges to the weighted pseudoinverse of a compact operator with (possibly unbounded) linear weighting operator for a wide class of signals. The case of differentiation of 1D signals corrupted by a random noise with uncertain but bounded second moments is studied in details.**

I. Introduction

The problem of computing derivatives of a given signal is fundamental in diverse fields including control engineering (e.g. design of observers requiring to compute numerical derivatives of the input [1]), image processing (e.g. edge detection and motion estimation by computing numerical gradients of the image brightness function [2]), and signal processing. The latter provides a very simple example: given a signal y represented by the sum of an absolutely continuous function $x \mapsto \psi(x) \in R$ and a bounded measurable function $x \mapsto \eta(x) \in R$ on an interval (x_1, x_2), i.e.

$$y(x) = \psi(x)+\eta(x), \quad t \in (x_0, x_1), \quad -\infty < x_0 < x_1 < +\infty$$

one needs to compute $\frac{dy}{dx}$, i.e. to estimate $\frac{d\psi}{dx}$ from the noisy data y. Noting that $\psi(x) = \psi(x_0) + \int_{x_0}^{x} \varphi(z)dz$, *the aforementioned problem can be reformulated as follows:* **given**

- the signal $y(t,x) = [C(t)\varphi](x) + \eta(t,x)$ on $[x_0, t]$, and operator

$$[C(t)\varphi](x) = \int_{x_0}^{x} k(t,x)\varphi(x)dx$$

$$k(t,x) = \begin{cases} 1, & x_0 \le x \le t \\ 0, & x > t \end{cases}$$

- the class of admissible derivatives φ of ψ in the form of solutions of a linear operator equation $\{\varphi : N\varphi = f, \delta\varphi = f_0, [f_0, f] \in G_0\}$ for a given bounding set G_0 and operators N, δ

one needs to estimate $\varphi = \frac{d\psi}{dx}$ **from** y.

Intuitively, $t \mapsto C(t)\varphi$ can be interpreted as a "window function", which "reveals" all the information about the function φ up until the current time instant t, but no information is provided about φ in the "future", for $x > t$ as $k(t,x) = 0$ for $x > t$ so that $y(t,x) = \eta(t,x)$ if $t < x \le x_1$. In this interpretation, the estimator should gradually accumulate the information about φ over time, i.e. when $t \to x_1$, and this information is then used to improve the estimate $\hat{\varphi}$ of φ.

The aforementioned reformulation allows us to relate differentiation to a more general problem of state estimation for linear operator equations in a Hilbert space. As a result, one can use powerful abstract methods developed for state estimation [3], [4], [5], [6], [7], [8], [9], [10], and in this paper we specifically rely upon results of [4] where optimal estimates of φ were derived assuming that $C(t)$ is a continuous operator-valued function with values in the space of linear bounded operators, η is a random process with values in a Hilbert space, and N, δ correspond to an abstract Neumann problem. It turns out that this abstract setting is closely related to the weighted pseudoinverse of a compact operator $\varphi \mapsto \int_{x_0}^{T} C^\star(t)C(t)\varphi dt$ with the weighting operator/regularizer defined by N: if, in particular, C corresponds to a Green function of a differential operator L then the resulting estimate of φ will "mimic" the effect of applying L to y provided $\eta = 0$, and, more importantly, the impact of the noise η will be minimized if $\eta \neq 0$. Thus, this abstract setting provides a generic template for (i) the design of optimal differentiating algorithms for multidimensional signals corrupted by random noise, and (ii) the optimal estimation of differentiating errors. The proposed abstract setting represents the main theoretical contribution of this paper.

The aforementioned template is then "implemented" to design a differentiator for signals in $L^2(x_0, x_1)$ subject to a random noise with unknown but bounded correlation operator. The resulting differentiation algorithm is given in the form of a well-known two-point boundary value problem for a linear quadratic optimal control problem provided $N = \frac{d^2}{dx^2} + \mu I$ and C is defined as suggested above. The algorithm allows to estimate first, second and third derivatives of the signal y, and is confirmed numerically on a simple example. The higher-order L^2-differentiator for signals subject to random measurement noise represent the main practical contribution of this paper.

The literature on estimating the derivative of an L^2-signal from noisy measurements is rich enough, and many authors transform this problem into a state estimation problem in order to apply different types of observers/filters to estimate both the derivative and the differentiation error. Popular techniques rely upon sliding mode differentiators [11] which are exact in a finite time for continuously differentiable Lipschitz signals and noise free measurements, and otherwise the corresponding derivative estimates converge into a zone

Sergiy Zhuk is with IBM Research, IBM Tech. campus, Damastown, Dublin, D15 HN66, Ireland sergiy.zhuk@ie.ibm.com

provided the measurement noise is uniformly bounded. A related family of differentiators is based upon high-gain observers [12] and the corresponding derivative estimates are asymptotically exact for noise free measurements. To achieve the convergence the observer's gain must grow unbounded which amplifies the noise in observations, and also leads to stiff time discretizations. This technique also assumes uniformly bounded measurement's noise. For the case of a stochastic noise one can construct a differentiator assuming stochastic models of the signal to be differentiated and of the measurement noise [13], and then apply a linear filter. It is well-known that the estimate obtained by the filter/observer is optimal only at the current time instant t in contrast to the estimate provided by an optimal smoother which "fits" the entire function to the data and thus improves the estimate of the entire function when new observations arrive. Of couse this comes at price: the smoother is not suitable for online estimation, but provides better estimates over the entire interval. The algorithms developed here have a similar property as they estimate the entire function on $[x_0, t]$. Further, we work with a stochastic noise with uncertain but bounded second moments - a very practical assumption especially when the statistical characteristics of the noise process are obtained experimentrally, and so the moments are not given exactly. In addition, we provide an optimal worst-case differentiation error which depends only on N and the bounding set for noise correlation operators, and hence is robust w.r.t. errors of the statistical/empirical estimates of the moments of the noise process. Finally, in the proposed framework the equations $N\varphi = f$ and $\delta\varphi = f_0$ do not play a role of the standard state equations, and f_0, f are not considered as uncertain parameters, as this is the case in the aforementioned works. Instead, we consider N and δ as regularizers/weighting operators as it was mentioned above, and f_0 and f control the signal's energy. In other words, the estimate of the derivative is sought in a special class defined by N, δ and the bounding set for f_0 and f, and this estimate is, in fact a solution of a certain ill-posed least-squares problem with minimal energy norm induced by N. This, in particular, allows us to achieve asymptotic exactness of the estimate (under some assumptions on the operator C) without assuming a specific stochastic model for the noise, or introducing strong observability assumptions.

Notation. Given an abstract Hilbert space H we denote by $(\cdot,\cdot)_H$ its canonical inner product with values in R, and set $\|x\|_H^2 := (x, x)$ for any $x \in H$. $[x, y]$ denotes an element of $X \times Y$, the Cartesian product of two Hilbert spaces X and Y. We also define a space of all linear continuous operators from a Hilbert space H_1 to Hilbert space H_2 by $\mathscr{L}(H_1, H_2)$, $H_1^\star$ denotes the adjoint space of H_1, Λ_{H_1} denotes the canonical isomorphism of the Hilbert space H_1 onto $H_1^\star$, I denotes the identity operator. $\langle\cdot,\cdot\rangle$ denotes the duality pairing between H and its adjoint space $H^\star$.

Outline. Section II provides a brief overview of abstract Neumann problems, and recalls the basics of singular value decomposition for compact operators which is closely related to the construction of the pseudoinverse of a compact operator. Section III states the main results: a generic template for the design of optimal differentiating algorithms, and its implementation in $L^2(x_0, x_1)$. Section III-B provides a numerical example. Appendix contains all the proofs.

II. Mathematical Preliminaries

In this section we collect all the required mathematical notions and results on state estimation for abstract Neumann problems, which define the class the derivatives of the admissible signals. We also review the pseudoinversion of compact operators by means of singular value decomposition. These tools are then used to design the estimates of the derivatives of the observed signal subject to a stochastic noise.

A. Abstract Neumann problems

Assume that H_0, H_- and H_+ are given Hilbert spaces such that

$$H_+ \subseteq H_0 \subseteq H_- .$$

In what follows, we identify H_0 with its adjoint space $H_0^\star$. Let $(\cdot,\cdot)_{0,-,+}$ denote inner products in $H_{0,-,+}$ respectively, and let a be a continuous bilinear form on $H_+ \times H_+$ such that

$$\exists\alpha > 0 : \quad a(\phi,\phi) \ge \alpha^2(\phi,\phi)_+ , \quad \forall\phi \in H_+ , \alpha \neq 0 . \quad (1)$$

Let H_∂ denote a Hilbert space and consider a linear operator $\gamma \in \mathscr{L}(H_+, H_\partial)$ such that

- $\gamma(H_+) = H_\partial$
- $\mathring{H}_+ := \ker(\gamma) = \{\phi \in H_+ : \gamma\phi = 0\}$ is dense in H_+

Define $\mathring{H}_- := \mathring{H}_+^\star$. Then, clearly,

$$\mathring{H}_+ \subseteq H_0 \subseteq \mathring{H}_- .$$

Define a linear operator

$$\forall\phi \in H_+ : \quad N\phi := a(\phi,\cdot) \in \mathring{H}_- ,$$

i.e. N maps a vector $\phi \in \mathring{H}_+$ to a linear continuous functional $\psi \mapsto \ell_\phi = a(\phi,\psi)$ over $\mathring{H}_+$, so that $\ell_\phi \in \mathring{H}_- = \mathring{H}_+^\star$. The operator N is a bounded linear operator in the Hilbert space $H_+(N) := \{\phi \in H_+ : N\phi \in H_0\}$ equipped with the graph norm $\|\phi\|_{H_+(N)} := (\|\phi\|_+^2 + \|N\phi\|_0^2)^{\frac{1}{2}}$, and, by definition of N, it follows that

$$N \in \mathscr{L}(H_+, \mathring{H}_-) \cap \mathscr{L}(H_+(N), H_0) .$$

Define N^+, the formal adjoint of N as follows:

$$\forall\psi \in H_+ : \quad N^+\psi := a(\cdot,\psi) \in \mathring{H}_- .$$

Clearly, $N = N^+$ provided $a(\phi,\psi) = a(\psi,\phi)$, and $N^+ \in \mathscr{L}(H_+, \mathring{H}_-) \cap \mathscr{L}(H_+(N^+), H_0)$ where $H_+(N^+) := \{\phi \in H_+ : N^+\phi \in H_0\}$.

There exist a linear bounded operator $\delta \in \mathscr{L}(H_+(N), H_\partial^\star)$ such that the following Green formula holds true:

$$a(\phi,\psi) = (N\phi,\psi)_0 + \langle\delta\phi,\gamma\psi\rangle , \quad \phi \in H_+(N) , \psi \in H_+ ,$$

and the formal adjoint of N, N^+ verifies the following equality: $\forall\phi \in H_+(N) , \psi \in H_+(N^+)$

$$(N^+\psi,\phi)_0 - (\psi, N\phi)_0 = \langle\gamma\psi,\delta\phi\rangle - \langle\delta^+\psi,\gamma\phi\rangle,$$

where $\delta^+ \in \mathscr{L}(H_+, H_\partial^\star)$ is such that

$$a(\psi, \phi) = (N^+\phi, \psi)_0 + \langle \delta^+\phi, \gamma\psi \rangle, \quad \phi \in H_+, \psi \in H_+(N^+).$$

The abstract Neumann problem associated with the form a is to find $\varphi \in H_+(N)$ such that:

$$N\varphi = f, \quad \delta\varphi = f_0, \quad f_0 \in H_\partial^\star, f \in H_0. \tag{2}$$

This latter problem can be equivalently reformulated in a variational form [14], namely $\varphi \in H_+(N)$ solves eq. (2) if and only if

$$a(\varphi, \psi) = (f, \psi)_0 + \langle f_0, \gamma\psi \rangle, \quad \forall \psi \in H_+. \tag{3}$$

We stress that eq. (2) has the unique solution provided eq. (1) holds true [14].

B. Pseudoinverse of a compact operator

Let K be a compact linear operator in H_0. Then, the compact self-adjoint operator $K^\star K$ possesses total orthonormal system of eigenvectors $\{\varphi_n\}$ with eigenvalues $\lambda_n \geq 0$: $K^\star K\varphi_n = \lambda_n\varphi_n$. Note that the numerical range of $K^\star K$, defined by $\{(Kx, Kx)_0, \|x\|_0 = 1\}$ contains the continuous spectrum of $K^\star K$, and so $\inf_{\psi:\|\psi\|_0=1}(K\psi, K\psi)_0 = 0$ if 0 is in the spectrum of $K^\star K$. The latter is the case, for instance, for the Volterra operator: $K\varphi = \int_{x_0}^x \varphi(z)dz$. In what follows we provide formulas for pseudoinversion of $K^\star K$ and K.

The unique solution of the optimization problem $\|Kq - K\varphi\|_0^2 + \varepsilon(q, q)_0 \to \min_q$ coincides with the unique solution of the equation $\varepsilon q + K^\star Kq = K^\star K\varphi$, and this solution $q(\varepsilon)$ is given by the following formula:

$$q(\varepsilon) = (\varepsilon I + K^\star K)^{-1} K^\star K\varphi = \sum_{n=1}^{\infty} \frac{\lambda_n}{\varepsilon + \lambda_n}(\varphi, \varphi_n)_0\varphi_n.$$

Clearly, $\frac{\lambda_n}{\varepsilon+\lambda_n} \neq 0$ if $\lambda_n > 0$ and $\frac{\lambda_n}{\varepsilon+\lambda_n} = 0$ otherwise. Hence, for $\varepsilon \to 0$ we get that $q(\varepsilon) \to \varphi$ provided $\min_n \lambda_n > 0$ and $q(\varepsilon) \to \varphi^\perp (:= (K^\star K)^+\varphi)$ otherwise, where $\varphi^\perp = (K^\star K)^+\varphi$ is the solution of the following ill-posed least-squares problem: $\|Kq - K\varphi\|_0^2 \to \min_q$ with the minimal $\|\cdot\|_0$-norm. Note that $\varphi^\perp$ coincides with the projection of φ onto the orthogonal completion of the null-space of $K^\star K$.

The situation changes when one tries to solve $\varepsilon q + K^\star Kq = K^\star\varphi$. Indeed, by using singular value decomposition of K, i.e.

$$Kq = \sum_n \lambda_n^{\frac{1}{2}}(v_n, q)_0 u_n, \quad K^\star\varphi = \sum_n \lambda_n^{\frac{1}{2}}(u_n, \varphi)_0 v_n,$$

for some total orthonormal systems $\{u_n\}$ and $\{v_n\}$ we get that

$$(\varepsilon I + K^\star K)q = \sum_n (\varepsilon + \lambda_n)(v_n, q)_0 v_n$$

hence

$$q(\varepsilon) := (\varepsilon I + K^\star K)^{-1} K^\star\varphi = \sum_n \frac{\lambda_n^{\frac{1}{2}}}{\varepsilon + \lambda_n}(u_n, \varphi)_0 v_n$$

and clearly $q(\varepsilon)$ converges to the solution of $\|Kq - \varphi\|_0^2 \to \min_q$ with the minimal $\|\cdot\|_0$-norm.

III. Main results

a) Problem statement: Assume that we observe a "signal" in the form of a vector-function $y(t)$ with values in a Hilbert space H_0 such that

$$y(t) = C(t)\varphi + \eta(t), \tag{4}$$

where $C(t) \in \mathscr{L}(H_0)$ is a given linear bounded transformation for all $t \in [T_0, T]$, and $t \mapsto C(t)$ is a continuous operator-valued function for all $t \in [T_0, T]$, $T \leq \bar{T}$. Here $\bar{T}$ indicates the upper bound of the "time interval" where the information is available[1]. The noise $\eta(t)$ is modelled as a realisation of a random process with values in H_0 such that:

$$E\eta(t) = 0, \quad \int_{T_0}^T E\|\eta(t)\|_{H_0}^2 dt < +\infty$$

i.e. the process has zero mean and finite second moments on $[T_0, T]$, $T \leq \bar{T}$; moreover, the correlation operator

$$(R_\eta(t, s)x_1, x_2)_{H_0} := E(\eta(t), x_1)_H(\eta(s), x_2)_{H_0}$$

is unknown and belongs to the given bounding set G_1:

$$G_1 = \{R_\eta : \int_{T_0}^T E(Q(t)\eta(t), \eta(t))_{H_0} dt \leq \gamma_T^2\} \tag{5}$$

where $Q(t) \in \mathscr{L}(H_0)$ and $Q(t) = Q^\star(t) \geq \alpha^2 I$, $\alpha \neq 0$, and $t \mapsto \|Q(t)\| \in C(T_0, T)$.

As noted in section I, we will interpret φ as a "derivative" (w.r.t. x!) of a function $\psi(t, x) = [C(t)\varphi](x)$, so that the inverse of $C(t)$, if it does exists, is a differential operator L_x. The most simple example of this abstract setup is considered below in section III-A where $\psi(t, x) = \int_{T_0}^x \varphi(z)dz$ if $x \leq t$ and $\psi(t, x) = 0$ if $x > t$. In this case $C(T)$ has unbounded inverse, $L_x = \dfrac{d}{dx}$ on $T_0 \leq x \leq T$.

Now, given y we want to "differentiate" it, i.e. we need to compute $L_x y$. The latter cannot be done directly as y is subject to the additive random noise. To overcome this we propose to estimate "the derivative", φ instead. The latter would be equivalent to the aforementioned "differentiation" if there was no noise in the signal y. To proceed we further assume a class of admissible "derivatives", namely φ is sought among solutions of the following abstract Neumann problem:

$$N\varphi = B_1 f_1 \quad \delta\varphi = B_0 f_0. \tag{6}$$

Here $B_0 \in \mathscr{L}(F_0, H_\partial^\star)$ and $B_1 \in \mathscr{L}(F_1, H_0)$ are given, and input f_1 and boundary condition f_0 belong to the ellipsoid

$$G_0 = \{[f_0, f_1] \in F_0 \times F_1 : (Q_0 f_0, f_0)_{F_0} + (Q_1 f_1, f_1)_{F_1} \leq 1\}, \tag{7}$$

where F_0 and F_1 are Hilbert spaces, $Q_i \in \mathscr{L}(F_i)$ is a given self-adjoint linear operator such that $Q_i \geq \beta_i I$, $\beta_i > 0$, $i = 1, 2$. Hence, in fact, all the admissible "derivatives" φ belong to a pre-image of the ellipsoid G_0 with respect to the operators N and δ. This representation has three key advantages:

[1] For instance, in the example given in section I $\bar{T}$ would be equal to x_1 – the upper bound of the domain where the observed signal is available.

- the class $H_+(N)$ is very wide, e.g. H^2-functions with weak derivatives of L^2-class as discussed in section III-A
- for any $\varphi \in H_+(N)$ it holds that: $\|\delta\varphi\|^2_{H^\star_\partial} + \|N\varphi\|^2_{H_0} = C < +\infty$ and hence φ belongs to the preimage of G_0 w.r.t. N and δ provided B_1, B_0 are invertible and Q_0 and Q_1 are chosen appropriately to accomodate the energy of the admissible $N\varphi$ and $\delta\varphi$
- the estimate of φ can be constructed using a suitable state estimation framework for linear operator equations

The next theorem provides an estimate of φ given the signal $y(t)$ on the interval $[T_0, T]$ and relies upon abstract results of [4].

Theorem 1: The boundary value problem

$$\begin{aligned} N^+\hat{p} &= \gamma_T^{-2}\int_{T_0}^T C^\star(t)Q(t)y(t)dt - \gamma_T^{-2}B_T\hat{\varphi}, \\ N\hat{\varphi} &= B_1 Q_1^{-1}\Lambda_1 B_1^\star \hat{p}\,, \quad \delta\hat{\varphi} = B_0 Q_0^{-1}\Lambda_0 B_0^\star \gamma\hat{p}, \\ \delta^+\hat{p} &= 0, \quad B_T := \int_{T_0}^T C^\star(t)Q(t)C(t)dt \end{aligned} \tag{8}$$

has the unique solution $[\hat{p}, \hat{\varphi}]$ such that the mean-squared worst-case distance between φ and $\hat{\varphi}$ in the direction $\ell \in H_0$ is given by:

$$\sup_{[f_0,f_1]\in G_0, R_\eta \in G_1} E[(\ell,\varphi)_0 - (\ell,\hat{\varphi})_0]^2 = (\ell, p)_0 \tag{9}$$

provided p solves

$$\begin{aligned} N^+z &= \ell - \gamma_T^{-2}B_T p\,, \quad \delta^+ z = 0\,, \\ Np &= B_1 Q_1^{-1}\Lambda_1 B_1^\star z\,, \quad \delta p = B_0 Q_0^{-1}\Lambda_0 B_0^\star \gamma z \end{aligned} \tag{10}$$

If $\tilde{B}_1 := B_1 Q_1^{-1}\Lambda_1 B_1^\star$ has bounded inverse, $B_0 = 0$, $Q(t) \equiv q_T$ then $\hat{\varphi}(T)$ defined by eq. (8) solves the following equation:

$$A(T)\hat{\varphi} = Y_T, \delta\hat{\varphi} = \delta^+(\tilde{B}_1)^{-1}N\hat{\varphi} = 0 \tag{11}$$

provided

$$A(T) := \frac{\gamma_T^2}{q_T}N^+(\tilde{B}_1)^{-1}N + \int_{T_0}^T C^\star(t)C(t)dt$$

and

$$Y_T := \int_{T_0}^T C^\star(t)y(t)dt.$$

If, in addition, $\frac{\gamma_T^2}{q_T} \to 0$ when T approaches $\bar{T}$ then

$$\hat{\varphi}(T) = A^{-1}(T)Y_T \to \varphi^\perp \tag{12}$$

where $\varphi^\perp$ denotes least-squares solution of $B_{\bar{T}}\phi = Y_{\bar{T}}$ with minimal norm $\Psi(\phi) := (N\phi, N\phi)_0$.
The proof of this and all the following statements is provided in the appendix.

Remark 1: It should be stressed that $\hat{\varphi}(T)$ obtained in theorem 1 provides an estimate of φ based on the data available on $[T_0, T]$, and eq. (12) suggests a "sliding window" mechanism, namely one can make a number of estimates $\hat{\varphi}(T_1), \hat{\varphi}(T_2)\dots$ for $T_1 < T_2 < \dots \bar{T}$, i.e. to update the estimate when the new information arrives. This process resembles iterative weighted pseudoinverse of a linear operator B_T, and according to the above theorem, it does converge to the least-squares solution of $B_{\bar{T}}\phi = Y_{\bar{T}}$ which at the very least belongs to $H_+(N)$. Now, if $N = I$, $\tilde{B}_1 = I$ and $B_{\bar{T}}$ is compact then $A(T) = \frac{\gamma_T^2}{q_T}I + B_T$ and since

$$Y_{\bar{T}} = B_{\bar{T}}\varphi_{true} + \int_{T_0}^T C^\star(t)\eta(t)dt$$

it follows that (see section II-B)

$$\lim_{T\to\bar{T}} \hat{\varphi}(T) = \varphi^\perp_{true}\,, \text{ provided } \lim_{T\to\bar{T}} \frac{\gamma_T^2}{q_T} = 0,$$

where $\varphi^\perp_{true}$ is the projection of the "true" derivative, φ_{true} onto the orthogonal completion of the null-space of $B_{\bar{T}}$. Hence, if the null-space of $B_{\bar{T}}$ is trivial, we get that $\hat{\varphi}(T) \to \varphi_{true}$.

A. 1D differentiation in L^2

In this section the abstract "differentiation" algorithm of theorem 1 is "implemented" for the case of computing derivatives of noisy signals in one spatial dimension.

Define $\Omega := (x_0, x_1)$ and set $H_+ := H^1(\Omega)$ and $H_0 := L^2(\Omega)$. Define

$$a(\phi,\psi) = \beta\int_{x_0}^{x_1} \frac{d\phi}{dx}\frac{d\psi}{dx}dx + \mu\int_{x_0}^{x_1}\phi\psi dx\,, \beta \geq 0\,, \mu > 0\,.$$

Note that the form a is symmetric so that $N^+ = N$ and $\delta^+ = \delta$. On the other hand,

$$N\varphi = -\beta\frac{d^2\varphi}{dx^2} + \mu\varphi\,, H_+(N) = \{\varphi \in H_+ : \frac{d^2\varphi}{dx^2} \in H_0\}$$

and so, by using standard integration-by-parts formula one can find that, for all $\psi \in H^1(\Omega)$ and $\phi \in H^2(\Omega)$ it holds that:

$$\begin{aligned} \langle\delta\phi, \gamma\psi\rangle &= a(\phi,\psi) - (N\phi,\psi)_0 \\ &= \frac{d\phi}{dx}(x_1)\psi(x_1) - \frac{d\phi}{dx}(x_0)\psi(x_0) \end{aligned}$$

In other words, in this case the trace operator γ is represented by two Dirac deltas concentrated at the boundary of the interval (x_0, x_1), i.e. $\gamma\phi = (\phi(x_0), \phi(x_1))^\top$, and δ-operator is given by the weak-derivatives of the aforementioned Dirac deltas, i.e. $\delta\phi = (\frac{d\phi}{dx}(x_0), \frac{d\phi}{dx}(x_1))^\top$. Both operators are bounded in $H^2(\Omega)$ as the latter is continuously embedded into $C^1(\Omega)$ (see [15, p.217]). Set $F_1 := H_0$, $B_1 = I$ and $F_0 := \mathbb{R}^2$. We also take $B_0 = \left(\begin{smallmatrix} b & 0 \\ 0 & b\end{smallmatrix}\right)$. In this case, the Neumann problem eq. (6) reads as follows: find $\varphi \in H^2(\Omega)$ such that

$$-\beta\frac{d^2\varphi}{dx^2} + \mu\varphi = f_1\,, \begin{bmatrix} \frac{d\phi}{dx}(x_0) = bf_0^l \\ \frac{d\phi}{dx}(x_1) = bf_0^r \end{bmatrix} \tag{13}$$

We further assume that the set G_0 is given by all f_0^l, f_0^r, f_1 such that

$$(f_0^l)^2 + (f_0^r)^2 + \int_{x_0}^T f_1^2(x)dx \leq 1$$

i.e. we take $Q_0 = I \in \mathbb{R}^{2\times 2}$ and $Q_1 = I$ – the identity mapping in H_0.

To define the observation operator, let us define an indicator function χ as follows:

$$\chi(x,t|x\le t) := \begin{cases} 1\,, & x\le t\,,\\ 0\,, & x>t \end{cases}$$

Set $T_0 := x_0$, $k(x,t) := \chi(x,t|x\le t)$ and define

$$[C(t)\varphi](x) := \int_{x_0}^{x} k(x,t)\varphi(z)dz. \tag{14}$$

Now, let us note that $\sup_{R_\eta\in G_1}$ in eq. (9) is attained at some (deterministic!) element $g\in L^2(x_0,T,H_0)$ such that $\int_{x_0}^T (Q(t)g(t),g(t))_0 = 1$ (see [4, Proof of T.4.4]). Hence, in what follows, without changing the mean-squared worst-case error eq. (9), we can restrict our attention to the deterministic noises. To simplify the presentation we consider the worst-case deterministic noise of the form:

$$\eta(t) = f(t)\xi(x), \quad \int_\Omega \xi^2(x)dx\le 1, \quad q_T\int_{x_0}^T f^2(t)dt\le\gamma_T^2$$

Of course, for this specific type of η, the worst-case error (the r.h.s. of eq. (9)) becomes an upper bound (possibly tight) for the minimal worst-case estimation error associated with the chosen class of noises η. We also set $Q(t)\equiv q_T I$. It then follows that for any $x\in(x_0,x_1)$ we observe

$$y(x,t) = k(x,t)\int_{x_0}^x \varphi_{\text{true}}(z)dz + f(t)\xi(x), \tag{15}$$

for $T_0 = x_0\le t\le T$. The interpretation of this observation equation is as follows: one observes y on $[x_0,T]$, and $T<\bar T$ so that the information about the "true derivative" φ_{true} is available only on $[T_0,T]$, $T<\bar T = x_1$, i.e. we cannot use the "future" $[T,\bar T]$ in our estimates.

The following proposition implements the abstract differentiation procedure from theorem 1 for the specific case of differentiating noisy signals in L^2.

Proposition 1 (1D differentiation): Define

$$\begin{aligned} F(x) &:= \int_{x_0}^T C^\star(t)y(t)dt\\ &= \int_{x_0}^T \frac{(T-z)^2}{2}\varphi_{\text{true}}(z)dz\\ &-\int_{x_0}^x \frac{(x-z)^2}{2}\varphi_{\text{true}}(z)dz\\ &+\int_x^T\left(\int_z^T f(t)dt\right)\eta(z)dz. \end{aligned} \tag{16}$$

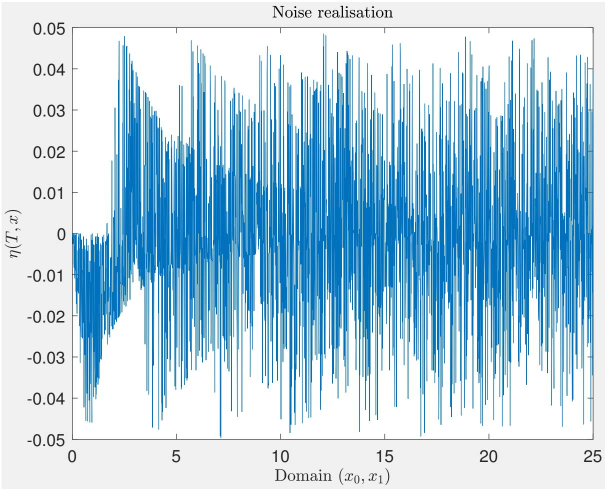
Then $\hat\varphi$, the estimate of φ solves the following boundary-value problem:

$$\begin{aligned} \frac{d\hat p}{dx} &= p_1, \quad \varphi_1(x_0) = b\hat p(x_0),\\ \frac{dp_1}{dx} &= p_2, \quad p_1(x_0) = p_1(x_1) = 0,\\ -\beta\frac{dp_2}{dx}+\mu p_1 &= q_T\gamma_T^{-2}\frac{dF}{dx} + q_T\gamma_T^{-2}q,\\ \frac{d\hat\varphi}{dx} &= \varphi_1, \quad -\beta p_2(x_1)+\mu\hat p_1(x_1) = 0,\\ -\beta\frac{d\varphi_1}{dx}+\mu\hat\varphi &= \hat p, \quad \varphi_1(x_1) = b\hat p(x_1),\\ \frac{dq}{dx} &= q_1, \quad q(x_0) = 0,\\ \frac{dq_1}{dx} &= \hat\varphi, \quad q_1(x_0) = 0. \end{aligned} \tag{17}$$

Hence, to differentiate y given by eq. (15) on $[x_0,T]$ one needs to solve eq. (17). Note that the null-space of $B_{\bar T}$ equals $\{0\}$. Hence, as it was pointed out in remark 1 above, the derivative of y computed from eq. (17), i.e. the estimate $\hat\varphi(T)$ converges to φ_{true} if $\frac{\gamma_T^2}{q_T}\to 0$ when $T\to\bar T = x_1$. In particular, $\hat\varphi(T)$ is "close" to φ_{true} if γ_T^2 is "small enough", and so the worst-case observation error is "small enough". The worst-case mean-squared estimation error in the direction $\ell\in H^2(\Omega)$ given in eq. (9) can be computed by solving eq. (17) with $\dfrac{dF}{dx}$ substituted by $\dfrac{d\ell}{dx}$.
In the next section we illustrate the efficacy of the differentiation algorithm eq. (17) by numerical examples.

B. Numerical example

In the setting of section III-A we further assume that $\varphi_{\text{true}}(x) = \cos(3x)\sin(x)$, $x_0 = 0$, $q_T = 1$, $b = 1$ and $x_1 = 75$. We perform a "crash-test", namely we do not force the noise η to be an element of G_1, instead we set $\eta(t,x)\equiv\nu\frac{\sin(2\nu x)}{20}$ where ν is drawn from the uniform distribution supported on $[0,1]$. Fig. 1a displays a realisation of η over the interval $[0,25]$. The L^2-norm of the noise is ≈ 1.8. We also set $\gamma_T^2 := T^{-3}$ so that the noise does not belong to G_1 neither for $T = 25$ ($\gamma_T^{-2} = 6.4e-05$) nor for $T = 75$ ($\gamma_T^{-2} = 2.37e-06$). Nevertheless, the 1st derivative of the noisy signal y, φ_{true} is estimated with very good precision of 11% relative L^2-error. Fig. 2a compares two estimates, $\hat\varphi(25)$ with $\gamma_T^{-2} = 6.4e-05$ and $\hat\varphi(75)$ with $\gamma_T^{-2} = 2.37e-06$ over the interval $[0,25]$. Clearly, $\hat\varphi(75)$ outperforms $\hat\varphi(25)$ (10% relative L^2-error versus 47%) as it is suggested by eq. (12). Fig. 2b compares $\hat\varphi(75)$ and φ_{true} over the larger segment $[0,75]$. We stress that the algorithm provides estimates of the first and second derivatives of φ_{true} (2nd and 3rd derivatives of y!) as well, and these are estimated with a reasonable level of precision which, however, is not as good as that of the first derivative's estimation: see Figs. 3a-3b. Finally we note that setting $\eta(t,x)\equiv\nu\frac{\sin(2\nu x)}{T^3}$ we get that the relative L^2-error for estimating φ_{true}, $\dfrac{d\varphi_{\text{true}}}{dx}$ and $\dfrac{d^2\varphi_{\text{true}}}{dx^2}$ is 3%, 5% and 11% respectively. Setting $\eta(t,x)\equiv\nu\frac{\sin(2\nu x)}{T^6}$ and



(a) Noise: L^2-norm circa 1.8,

Fig. 1: Noise realisation

$T = 175$ we get further reduction down to 1%, 2% and 9% respectively.

APPENDIX

Proof: [Proof of Theorem 1] The solvability of the equations eq. (8)- eq. (10) together with the estimate eq. (9) was demonstrated in [4] in the context of minimax state estimation of solutions of abstract Neumann problems subject to uncertain inputs and boundary conditions which are represented by elements of G_1. By direct substitution it is verified that the unique solution of eq. (8) solves eq. (11) provided $\tilde{B}_1$ has bounded inverse, $B_0 = 0$, and $Q(t) \equiv q_T$. Now, to prove eq. (12) we note that $\hat{\varphi}(T)$ defined by eq. (8) coincides with the unique minimizer of the following convex optimization problem

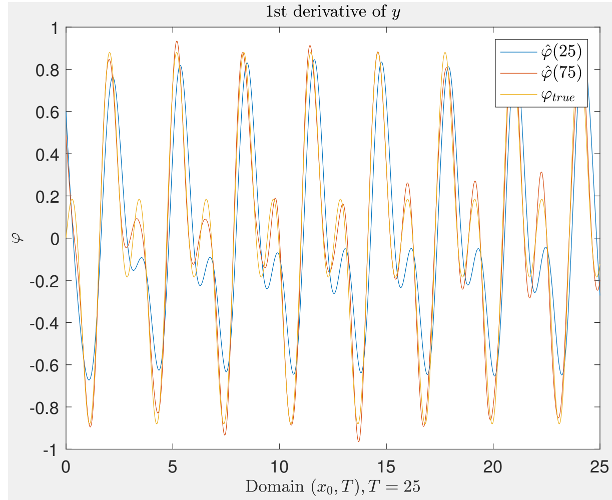
$$J_T(\phi) := \int_{T_0}^{T} \|C(t)\phi - y(t)\|^2_{H_0} dt + \frac{\gamma_T^2}{q_T}(N\phi, N\phi)_0 \to \min_{\phi \in H_+(N)}$$

Since the norm of H_0 is lower semi-continuous, and the graph of N is weakly closed it follows that

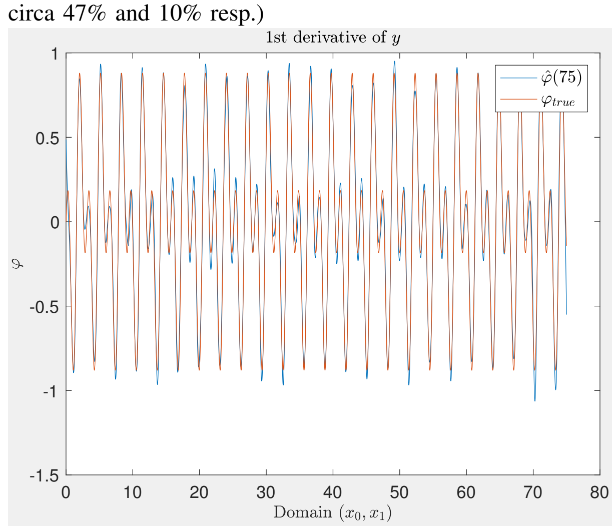
$$\phi \mapsto \Psi(\phi) := (N\phi, N\phi)_0$$

is a lower semi-continuous functional. Moreover, recall that the form a verifies eq. (1). Hence Ψ is a strictly convex lower semi-continuous functional, and so it has the unique minimum point, $\varphi^\perp$ over the set of all minimizers of the quadratic form $\phi \mapsto (B_{\bar{T}}\phi, \phi)_0 - 2(\phi, Y_{\bar{T}})_0$. Finally, analogously to the justification of Tikhonov regularisation method [16], it is easy to demonstrate that $\hat{\varphi}(T)$ converges strongly in H_0 to $\varphi^\perp$ provided $\frac{\gamma_T^2}{q_T} \to 0$ when T approaches $\bar{T}$. Note that in the latter case the second moments of the noise η must go to 0 implying that the observations are exact at time $\bar{T}$. This completes the proof. ∎

Proof: [Proof of Proposition 1] Let us transform eq. (8) according the specific form of operators N, δ and γ given



(a) Estimates for $T = 25, 75$ over $[0, 25]$ (relative L^2-error circa 47% and 10% resp.)

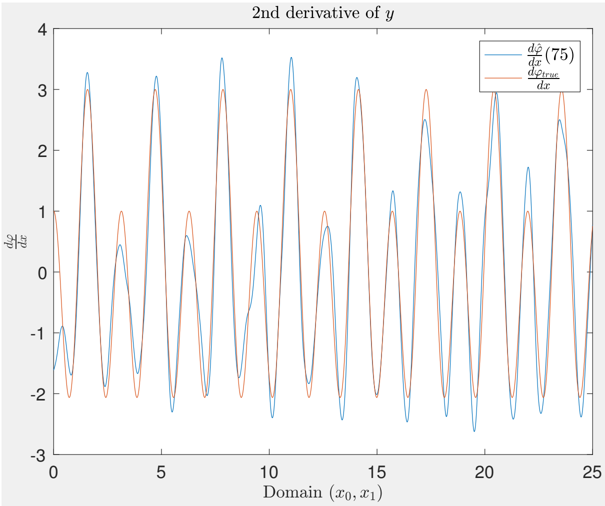


(b) Estimate at $T = 75$ over $[0, 75]$ (relative L^2-error circa 11%)

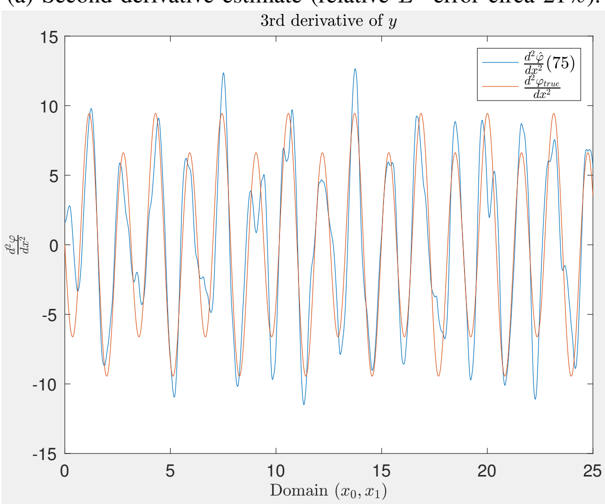
Fig. 2: Estimates of the φ_{true}

in section III-A. We get:

$$\begin{aligned}
&-\beta\frac{d^2\hat{\varphi}}{dx^2} + \mu\hat{\varphi} = \hat{p}, \frac{dp}{dx}(x_0) = \frac{dp}{dx}(x_1) = 0, \\
&-\beta\frac{d^2\hat{p}}{dx^2} + \mu\hat{p} = q_T\gamma_T^{-2}B_T\hat{\varphi} + q_T\gamma_T^{-2}\int_{x_0}^{T} C^\star(t)y(t)dt, \\
&\frac{d\hat{\varphi}}{dx}(x_0) = b\hat{p}(x_0), \frac{d\hat{\varphi}}{dx}(x_1) = b\hat{p}(x_1)
\end{aligned} \tag{18}$$



(a) Second derivative estimate (relative L^2-error circa 21%).



(b) Third derivative estimate (relative L^2-error circa 38%).

Fig. 3: Estimates of the derivatives of φ_{true}

We stress that $t \mapsto C(t)$ is a continuous operator-valued function so that theorem 1 applies. Indeed:

$$\begin{aligned}
&[C(t)\varphi](x) - [C(s)\varphi]^2(x) \\
&\quad = (k(t,x) - k(s,x))^2 \left(\int_{x_0}^{x} \varphi(z) dz \right)^2 \\
&\quad \le (k(t,x) - k(s,x))^2 \left(\int_{x_0}^{x} |\varphi(z)| dz \right)^2 \\
&\quad \le (k(t,x) - k(s,x))^2 \left(\int_{x_0}^{x_1} |\varphi(z)| dz \right)^2 \\
&\quad \le (x_1 - x_0) \|\varphi\|_{H_0}^2
\end{aligned}$$

so that

$$\begin{aligned}
&\|C(t)\varphi - C(s)\varphi\|_{H_0} \\
&\quad \le \left(\int_{x_0}^{x_1} (k(t,x) - k(s,x))^2 dx \right)^{\frac{1}{2}} (x_1 - x_0)^{\frac{1}{2}} \|\varphi\|_H \\
&\quad = (t-s)^{\frac{1}{2}} (x_1 - x_0)^{\frac{1}{2}} \|\varphi\|_H
\end{aligned}$$

We note that:

$$[C^\star(t)\psi](x) := \int_{\min\{t,x\}}^{t} \psi(z) dz \,.$$

Indeed, by definition

$$\begin{aligned}
(C(t)\varphi, \psi)_{H_0} &= \int_{x_0}^{x_1} k(x,t) \int_{x_0}^{x} \varphi(z) dz \psi(x) dx \\
&= \int_{x_0}^{x_1} \int_{x_0}^{x_1} k(x,t) \varphi(z) \psi(x) \chi(z,x|z \le x) dz dx \\
&= \int_{x_0}^{x_1} \varphi(z) \left(\int_{x_0}^{x_1} k(x,t) \psi(x) \chi(z,x|z \le x) dx \right) dz \\
&= \int_{x_0}^{x_1} \varphi(z) \left(\int_{z}^{x_1} k(x,t) \psi(x) dx \right) dz \\
&= \int_{x_0}^{x_1} \varphi(z) \left(\int_{\min\{z,t\}}^{t} \psi(x) dx \right) dz \\
&= (\varphi, C^\star(t)\psi)_{H_0}
\end{aligned}$$

If $x_0 \le T \le x_1$ then

$$\begin{aligned}
&\Big[\int_{x_0}^{T} C^\star(t) C(t) \varphi dt \Big] \\
&\quad = \begin{cases} \int_{x_0}^{T} \frac{(T-z)^2}{2} \varphi(z) dz - \int_{x_0}^{x} \frac{(x-z)^2}{2} \varphi(z) dz \,, & x \le T \,, \\ 0 \,, & x > T \end{cases}
\end{aligned}$$

Indeed, assume that $z \le T$. We get:

$$\begin{aligned}
&\Big[\int_{x_0}^{T} C^\star(t) C(t) \varphi dt \Big](z) \\
&\quad = \int_{x_0}^{T} \int_{\min\{t,z\}}^{t} \left(\int_{x_0}^{x} k(x,t) \varphi(a) da \right) dx dt \\
&\quad = \int_{x_0}^{T} \int_{x_0}^{t} \chi(z,x|z \le t) \int_{x_0}^{t} \chi(a,x|a \le x) \varphi(a) da dx dt \\
&\quad = \int_{x_0}^{T} \int_{x_0}^{t} \varphi(a) \int_{x_0}^{t} \chi(z,x|z \le t) \chi(a,x|a \le x) dx da dt \\
&\quad = \int_{x_0}^{T} \chi(t,z|z \le t) \int_{x_0}^{t} \varphi(a)(t-a) da dt \\
&\quad = \int_{x_0}^{T} \int_{x_0}^{T} \chi(t,z|z \le t) \chi(a,t|a \le t) \varphi(a)(t-a) da dt \\
&\quad = \int_{x_0}^{T} \varphi(a) \int_{x_0}^{T} \chi(t,z|z \le t) \chi(a,t|a \le t)(t-a) dt da \\
&\quad = \int_{x_0}^{T} \varphi(a) \int_{\max\{a,z\}}^{T} (t-a) dt da \\
&\quad = \int_{x_0}^{z} \varphi(a) \int_{z}^{T} (t-a) dt da \\
&\qquad + \int_{z}^{T} \varphi(a) \int_{a}^{T} (t-a) dt da \\
&\quad = \int_{x_0}^{z} \frac{\varphi(a)}{2} \left((T-a)^2 - (z-a)^2 \right) da \\
&\qquad + \int_{z}^{T} \varphi(a) \frac{(T-a)^2}{2} da \\
&\quad = \int_{x_0}^{T} \varphi(a) \frac{(T-a)^2}{2} da - \int_{x_0}^{z} \varphi(a) \frac{(z-a)^2}{2} da
\end{aligned}$$

If $z > T$ it then follows that

$$\Big[\int_{x_0}^{T} C^\star(t)C(t)\varphi dt\Big](z) = \int_{x_0}^{T}\int_{t}^{t}\left(\int_{x_0}^{x} k(x,t)\varphi(a)da\right)dxdt = 0\,.$$

Hence $\int_{x_0}^{T} C^\star(t)y(t)dt = F(x)$ for F defined as in the proposition's statement. To conclude the proof we note that $F(T) = 0$ and $B_T\hat\varphi(T) = 0$ as it follows from the representation of $\int_{x_0}^{T} C^\star(t)C(t)\varphi dt$ given above, and hence $-\beta\frac{d^2\hat p}{dx^2}(T)+\mu\hat p(T) = 0$. Differentiating the second equation of eq. (18) we get:

$$-\beta\frac{d^3\hat p}{dx^3} + \mu\frac{d\hat p}{dx} = q_T\gamma_T^{-2}\frac{dF}{dx} + q_T\gamma_T^{-2}\,q,$$
$$q(x) := \int_{x_0}^{x}(x-z)\varphi(z)dz$$

Define $p_1 := \frac{d\hat p}{dx}$, $p_2 := \frac{dp_1}{dx}$, $\varphi_1 := \frac{d\hat\varphi}{dx}$, $q_1 := \frac{dq}{dx}$, and recall that $p_2(T) + \mu\hat p(T) = 0$. We deduce that eq. (18) is equivalent to eq. (17). This completes the proof.

■

ACKNOWLEDGMENT

The author is grateful to Dr. A. Polyakov for suggesting to consider a specific form of the observation operator given by eq. (14), and to Professors L. Fridman and A. Levant for fruitful discussions on sliding mode differentiators during IFAC18 congress.

REFERENCES

[1] R. Galván-Guerra, L. Fridman, and J. Dávila, "High-order sliding-mode observer for linear time-varying systems with unknown inputs," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 14, pp. 2338–2356, 2017.

[2] I. Herlin, D. Bereziat, N. Mercier, and S. Zhuk, *Lecture Notes in Computer Science: Computer Vision – ECCV12*. Springer, 2012, ch. Divergence-free motion estimation, pp. 15–27.

[3] S. Zhuk, "Estimation of the states of a dynamical system described by linear equations with unknown parameters," *Ukrainian Mathematical Journal*, vol. 61, no. 2, pp. 214–235, 2009.

[4] A. Nakonechnyi and S. Zhuk, "Minimax state estimates for abstract neumann problems," *Minimax Theory Applications*, 2018.

[5] A. Nakonechny, "On minimax theory for estimating functionals of operator equation solutions," *Dopovidi Akademii Nauk Ukrainskoi RSR Seriya A-Fiziko-Matematichni ta Technichni Nauki*, no. 5, pp. 69–71, 1982.

[6] S. Zhuk, "Minimax state estimation for linear discrete-time differential-algebraic equations," *Automatica*, vol. 46, no. 11, pp. 1785–1789, 2010.

[7] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Trans. Automat. Control*, vol. 30, no. 8, pp. 730–738, 1985.

[8] A. Bensoussan, G. Da Prato, M. Delfour, and S. Mitter, *Representation and Control of Infinite Dimensional Systems*. Birkhauser, 2007, 2nd edition.

[9] A. Nakonechnyi, Y. K. Podlipenko, and Y. A. Zaitsev, "Minimax prediction estimation of solutions of initial-boundary-value problems for parabolic equations with discontinuous coefficients based on imperfect data," *Cybernetics and Systems Analysis*, vol. 36, no. 6, pp. 845–854, 2000.

[10] V. I. Maksimov, *Dynamical inverse problems of distributed systems*. Walter de Gruyter GmbH & Co KG, 2002, vol. 37.

[11] A. Levant, "Robust exact differentiation via sliding mode technique," *automatica*, vol. 34, no. 3, pp. 379–384, 1998.

[12] L. K. Vasiljevic and H. K. Khalil, "Error bounds in differentiation of noisy signals by high-gain observers," *Systems & Control Letters*, vol. 57, no. 10, pp. 856–862, 2008.

[13] B. Carlsson, A. Ahlén, and M. Sternad, "Optimal differentiation based on stochastic signal models," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 341–353, 1991.

[14] J. Aubin, *Approximation of elliptic boundary-value problems*. Wiley, 1972.

[15] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

[16] A. Tikhonov and V. Y. Arsenin, *Methods for solving ill-posed problems*. John Wiley and Sons, Inc, 1977.