Karlsruhe Institute of Technology

Wave phenomena
analysis and numerics

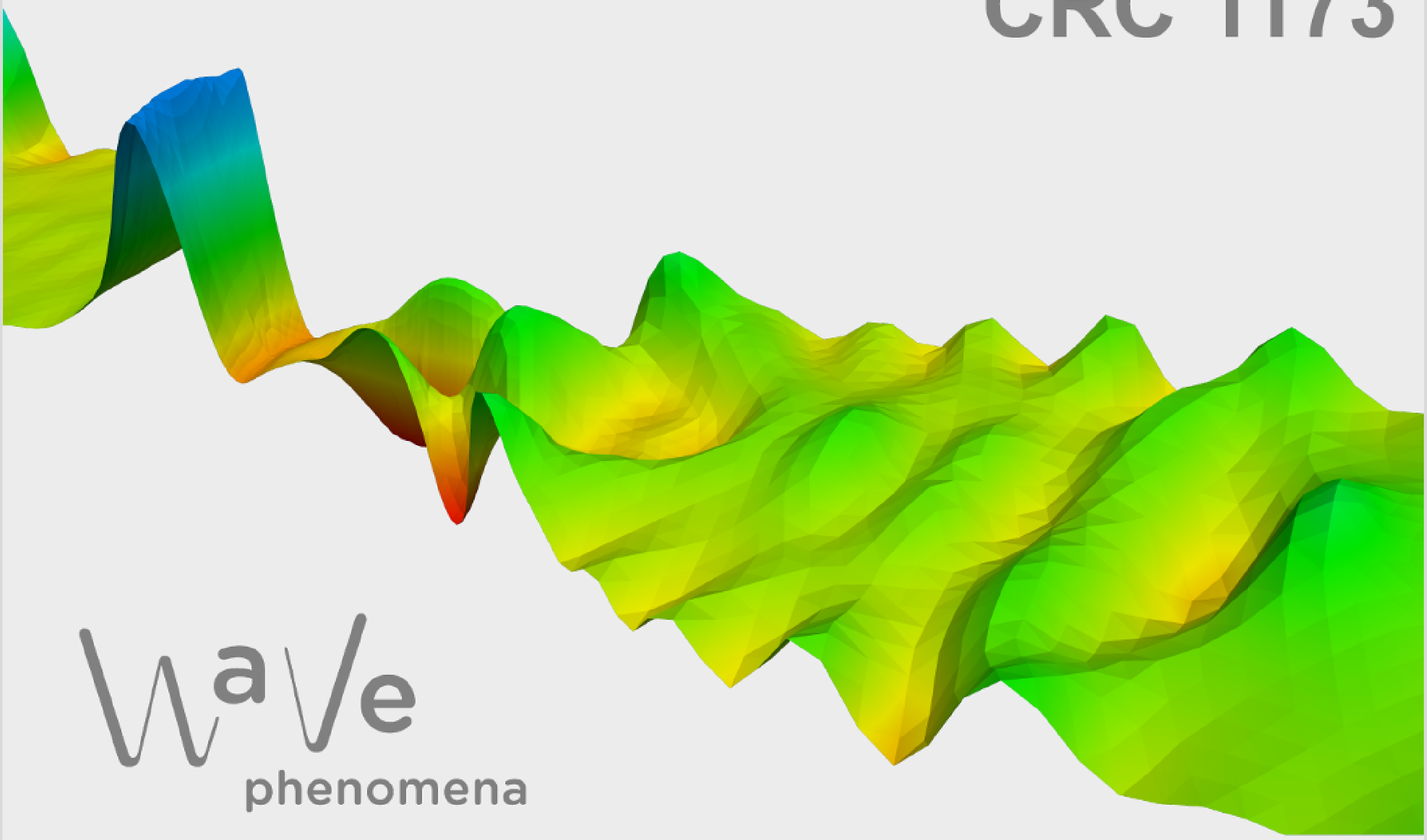
The taut string approach to statistical inverse problems: theory and applications

Sangjin Kim, Michael Pokojovy, Xiang Wan

CRC Preprint 2020/19, July 2020

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Wave phenomena

Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

TU WIEN

Funded by

DFG

ISSN 2365-662X

The Taut String Approach to Statistical Inverse Problems: Theory and Applications

Sangjin Kim[*], Michael Pokojovy[†], Xiang Wan[‡]

July 10, 2020

Abstract

A novel solution approach to a class of nonlinear statistical inverse problems with finitely many observations collected over a compact interval on the real line blurred by Gaussian white noise of arbitrary intensity is presented. Exploiting the nonparametric taut string estimator, we prove the state recovery strategy is convergent to a solution of the unnoisy problem at the rate of $n^{-1/2}$ as the number of observations n grows to infinity. Illustrations of the method's application to real-world examples from hydrology, civil & electrical engineering are given and an empirical study on the robustness of our approach is presented.

Key words: inverse problems; signal processing; nonparametric statistics; taut string estimator; white noise

MSC (2010): 60G35, 60G15, 62G08, 62G20, 62G35, 93E10, 93E11, 93E14

1 Introduction

Let $\mathscr{X}, \mathscr{Y}$ be Banach spaces and let $\mathcal{A}\colon \mathfrak{D}(\mathcal{A}) \subset \mathscr{X} \to \mathscr{Y}$ be an operator with a domain $\mathfrak{D}(\mathcal{A})$ and a range $\mathfrak{R}(\mathcal{A})$. With x denoting the unknown state variable and y being a measurement or an observation, many nonlinear inverse problems can be written in the form

$$y = \mathcal{A}(x). \tag{1.1}$$

Equation (1.1) with an appropriately chosen $\mathcal{A}$, dictated by each particular application and typically referred to as a forward operator, arises in control theory when observing dynamical systems given by ordinary, partial or stochastic differential equations (see, e.g., the monographs [18, 28]). Equation (1.1) also occurs in optimization theory and statistics when reconstructing or estimating system's parameters (cf. [5, 10, 26]), etc. Applications range from physics (continuum mechanics, geophysics & geology, astronomy, optics) and engineering (nondestructive testing, material science, remote sensing) to medicine & biology (medical imaging, noninvasive diagnostics, population modeling) and computer science (image processing, signal filtering and decoding), etc.

Different versions and variants of Equation (1.1) are known in the literature. In most studies, $\mathscr{X}$ and $\mathscr{Y}$ are assumed to be Hilbert spaces (often even $\mathscr{X} = \mathscr{Y}$) and the operator $\mathcal{A}$ is chosen

[*]Dept. of Management Information Systems, Dong-A University, 225 Gudeok-ro, Seo-gu, Busan, Republic of Korea (49236) skim10@donga.ac.kr

[†]Dept. of Mathematical Sciences, The University of Texas at El Paso, 500 W University Ave, El Paso, TX 79968 mpokojovy@utep.edu

[‡]Dept. of Mathematics, The George Washington University, 801 22nd St. NW, Washington, D.C., 20052 xiangwan@gwu.edu

to be linear. Depending on the particular framework, both the operator $\mathcal{A}$ and the observation y can be deterministic, stochastic or fuzzy. Additionally, either a single or multiple batches of full or partial observations can be available. Usually, the operator $\mathcal{A}$ is invertible. If this is not the case, various generalized solution notions (strong/mild/weak/extrapolated solution, quasi-solution, ε-solution, etc.) can sometimes be employed. Nonetheless, even if $y \in \mathfrak{R}(\mathcal{A})$ and $\mathcal{A}$ is invertible, Equation (1.1) is typically ill-posed in the sense of Hadamard (cf. [26, Definition 3.1]) due to the lack of continuity of $\mathcal{A}^{-1}$.

In [26, Chapters 3 to 9], Schuster et al. give an overview of recent literature on solution approaches to Equation (1.1), in particular, they outline the Tikhonov regularization method, iterative regularization techniques, the method of approximate inverse, etc. In contrast to nonlinear versions of Equation (1.1), many studies exist in the linear situation. We refer the reader to [3, 14, 20] for details.

Nonlinear inverse problems with stochastic noise are particularly challenging. Given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and assuming Σ to be a σ-algebra on $\mathscr{Y}$, a general inverse problem with additive random noise reads as

$$y = \mathcal{A}(x) + \sigma\xi, \tag{1.2}$$

where ξ is assumed to be an $\mathcal{F}$-Σ-measurable $\mathscr{Y}$-valued random element (typically, satisfying $\mathbf{E}[\xi] = 0$) and $\sigma > 0$ is the (constant) noise intensity. A well-known consequence of the Hahn-Banach theorem (cf. [6, p. 199]) permits Equation (1.2) to be equivalently rewritten as

$$y_\varphi = \langle \varphi, \mathcal{A}(x) \rangle_{\mathscr{Y}';\mathscr{Y}} + \sigma \langle \varphi, \xi \rangle_{\mathscr{Y}';\mathscr{Y}} \text{ for any } \varphi \in \mathscr{Y}' \tag{1.3}$$

with $\mathscr{Y}'$ denoting the (Euclidean) topological dual of $\mathscr{Y}$ and $\langle \cdot, \cdot \rangle_{\mathscr{Y}';\mathscr{Y}}$ standing for the dual pairing between $\mathscr{Y}'$ and $\mathscr{Y}$. For given $\varphi \in \mathscr{Y}'$, Equation (1.3) is often referred to as an indirect observation. If $\mathcal{A}$ is linear, densely defined and invertible (at least on the essential range of ξ), Equation (1.3) can also be interpreted as a direct observation

$$y_\psi = \langle \psi, x \rangle_{\mathscr{X}';\mathscr{X}} + \sigma \langle \psi, \zeta \rangle_{\mathscr{X}';\mathscr{X}} \text{ for } \psi \in \mathfrak{R}(\mathcal{A}') = \mathscr{X}'.$$

with $\zeta = \mathcal{A}^{-1}\eta$. Such transformations are typically not desirable since they introduce unnatural error correlations.

In contrast to Equation (1.3), only a finite number of indirect observations are typically available in practice. Hence, Equation (1.3) can be rewritten as

$$y_k = \big\langle \varphi_k, \mathcal{A}(x) \big\rangle_{\mathscr{Y}';\mathscr{Y}} + \sigma \langle \varphi_k, \xi \rangle_{\mathscr{Y}';\mathscr{Y}} \text{ for any } \varphi_k \in \Phi_n, \tag{1.4}$$

where $\Phi_n = \{\varphi_1, \varphi_2, \dots, \varphi_n\} \subset \mathscr{Y}'$ is a finite set referred to as a design. If Φ_n can be chosen freely, say, to optimize some quality criterion or to achieve the best asymptotic approximation rate, the framework of Equation (1.4) is sometimes called an optimal design. Though being very desirable, the optimal design can rarely be implemented in practice. Instead of being arbitrarily selectable, the design is typically given beforehand. In this case, one speaks of a fixed design. The problem may become, for example, to assemble an estimation procedure based on this fixed design mapping the observations $\{y_1, y_2, \dots, y_n\}$ to a state estimate $\hat{x}_n$ to assure the best estimation accuracy as $n \to \infty$ for x or a derived quantity such as $f(x)$ for some linear or nonlinear f, where x is a solution to the unnoisy problem in Equation (1.1). Alternatively, the estimator could be designed to achieve the best order of accuracy for recovering x or $f(x)$ as $\sigma \to 0$ (cf. [21]).

In the present paper, we propose a two-step "plug-in" recovery strategy for Equation (1.2):

I) Using nonparametric statistics, compute an estimator $\hat{y}_n$ for $\mathbf{E}[y]$.

II) Plug $\hat{y}_n$ into Equation (1.1) and solve the resulting inverse problem $\hat{y}_n = \mathcal{A}(x)$ for x as if $\hat{y}_n$ was nonrandom.

Obviously, Step II) is especially simple if $\mathcal{A}$ possesses a Lipschitzian inverse.

To illustrate the power of our new approach, we apply it to the following specific version of Equation (1.4) and assess its performance. Choosing the reflexive Lebesgue space (defined in Section 2.1 below)

$$\mathscr{Y} := L^p\big(0,1;\mathbb{R}^d\big) \text{ for } d \in \mathbb{N} \text{ and some } p \in (1,\infty),$$

we consider the operator

$$\mathcal{A} \colon \mathfrak{D}(\mathcal{A}) \subset \mathscr{X} \to \mathscr{Y} \text{ for some reflexive Banach space } \mathscr{X},$$

which will later be extended to a larger Banach space. The rationale behind the choice of spaces will be explained in the sequel.

The noise ξ is selected to be Gaussian white noise given as the distributional derivative of a standard d-variate Brownian motion $\big(W(t)\big)_{t\geq 0}$ being a random element of $C^0\big([0,1],\mathbb{R}^d\big)$ with its paths almost surely contained in the space $C_0^\alpha\big([0,1],\mathbb{R}^d\big) \cap W_0^{s,p}\big(0,1;\mathbb{R}^d\big)$ for any $\alpha \in (0,\frac{1}{2})$, $s \in (0,\frac{1}{2})$, $p \in [1,\infty)$ (cf. [7, p. 13]). See Section 2.1 below for the definition of these and other spaces used in this paper.

With $T_n := \big\{\frac{1}{n},\frac{2}{n},\ldots,1\big\}$ for $n \in \mathbb{N}$ denoting an equidistant lattice on $[0,1]$, we let

$$\varphi_k := n\int_{\frac{k-1}{n}}^{\frac{k}{n}} (\,\cdot\,)(t)\mathrm{d}t \text{ for } 1 \leq k \leq n, \tag{1.5}$$

where the integration is interpreted in terms of the antiderivative operator, which, in turn, is a bijective mapping from $C^{\alpha-1}\big([0,1],\mathbb{R}^d\big)$ to $C_0^\alpha\big([0,1],\mathbb{R}^d\big)$ for $\alpha \in (0,\frac{1}{2})$ (cf. [12, p. 209]) and isomorphism from $L^p\big(0,1;\mathbb{R}^d\big)$ to $W_0^{1,p}\big(0,1;\mathbb{R}^d\big) \hookrightarrow C_0^{(p-1)/p}\big(0,1;\mathbb{R}^d\big)$ for $p \in (1,\infty)$. This is elaborated upon in Section 2.3. The design Φ_n is then chosen as $\Phi_n := \big\{\varphi_1,\ldots,\varphi_n\big\} \subset \mathscr{L}(\mathscr{Y},\mathbb{R}^d)$. Strictly speaking, the φ_k's are bounded linear operators and not functionals. For the sake of simplicity, we forgo the projection of φ_k's on each of their components since their range is finitely dimensional.

With this notation, Equation (1.4) can be written as

$$\mathbf{y}_k = \varphi_k\big(\mathcal{A}(x)\big) + \sigma\boldsymbol{\xi}_k \text{ for } 1 \leq k \leq n, \tag{1.6}$$

where $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n$ are independent identically distributed (iid) standard normal d-variate random vectors. Here and in the sequel, we imploy the standard convention used in statistics and bold vector- and tensor-valued objects. First, a nonparametric estimation procedure needs to be applied to estimate $\mathbf{E}\big[y(\cdot)\big]$. Here, we decided to use the taut string estimator recently introduced by Davies and Kovac in [9]. Our choice will be justified by two important features of the estimation procedure. On one hand, we can achieve the convergence rate of $n^{-1/2}$ in some weak metric and, on the other hand, in contrast to linear estimation strategies, our nonlinear procedure is quite robust to corruption and dilution of a certain data fraction even if the corruption magnitude is significant. Note that the convergence rate of $n^{-1/2}$ is typical – and optimal – for parametric problems and can be attained, e.g., with asymptotically efficient maximum likelihood estimators or their robust analogs. Due to the weak topology employed in this article, we can retain this optimal parametric rate even in the nonparametric context. This rate would have impossible if a stronger topological framework was employed (cf. [19, 30]). Further, under appropriate assumptions on $\mathcal{A}$, we can reconstruct x such that our estimate converges in an appropriate norm to a solution of Equation

(1.2) for $\sigma = 0$ obtained with the Tikhonov regularization approach at the rate of $n^{-1/2}$ as $n \to \infty$. Since our problem is nonlinear, a nonlinear estimation technique will also be adopted.

The outline of our paper is the following. In Section 2, we briefly summarize some important information on probability in Banach spaces. Further, we present a short discourse on nonparametric statistics as well as introduce the taut string estimator and prove a novel weak convergence result for the latter at the "parametric" rate of $n^{-1/2}$. In Section 3, we apply the taut string methodology to design an estimator $\hat{x}$ for x in Equation (1.6). Further, we give two applications and present a numerical robustness study. In the Appendix, some well-known results on the double obstacle problem of differential geometry, used to prove the weak convergence in Section 3, are summarized.

2 Nonparametric Regression and Taut Strings

In this section, we briefly discuss the central topic of nonparametric regression theory, present the 1D taut string estimator introduced in [9] and prove its convergence in the topology of $W^{-1,p}(0,1)$ referred to in this paper as weak convergence. For a more detailed treatise on nonparametric regression, we refer the reader to [11, Chapter 2], [17, Chapter 4]) or [29, Chapter 5], etc.

2.1 Probability in Functional Spaces

Thoughout the paper, we employ standard notation (cf. [1]). Let $G := (a, b)$ be a bounded open interval of $\mathbb{R}$ and $\bar{G} = [a, b]$ denote its topological closure.

2.1.1 Basic Functional Spaces

Consider the following spaces.

- For Banach spaces $\mathscr{X}, \mathscr{Y}$, let $\mathscr{L}(\mathscr{X}, \mathscr{Y})$ denote the (Banach) space of bounded linear operators from $\mathscr{X}$ to $\mathscr{Y}$.

- For a Banach space $\mathscr{X}$, let $\mathscr{X}'$ denote its (real) topological dual, i.e., $\mathscr{X}' := \mathscr{L}(\mathscr{X}, \mathbb{R})$.

- Let $C^0(\bar{G})$ denotes the space of continuous functions on $\bar{G}$ endowed with the maximum norm. Note that $C^0(\bar{G})$ is then a separable Banach space (in particular, a Polish space).

- For $\alpha \in (0, 1]$, define the Hölder space

$$C^\alpha(\bar{G}) := \big\{u \in C^0(\bar{G}) \,|\, \|u\|_{C^\alpha(\bar{G})} < \infty\big\} \quad \text{with} \quad \|u\|_{C^\alpha(\bar{G})} := \|u\|_{C^0(\bar{G})} + \sup_{\substack{x,y \in \bar{G} \\ x \neq y}} \frac{\big|u(x) - u(y)\big|}{|x - y|^\alpha}.$$

  Similar to [7], introduce

$$C_0^\alpha(\bar{G}) := \big\{u \in C^\alpha(\bar{G}) \,|\, u(a) = 0\big\}.$$

- For $\alpha \in (0, 1)$, the negative Hölder space $C^{-\alpha}(\bar{G})$ of Schwartz' distributions u is defined as in [12, Section 13.3] with an additional assumption that $a \equiv \min(\bar{G}) \notin \operatorname{supp}(u)$, where the latter stands for the support of u. Endowed with appropriate topology, $C^{-\alpha}(\bar{G})$ is a Fréchet space.

- For $p \in [1, \infty]$, let $L^p(G)$ and $L^p_{\mathrm{loc}}(G)$ denote the standard Lebesgue space or its local version, respectively.

- For $u \in L^1_{\mathrm{loc}}(G)$ and $s \in \mathbb{N}$, let $u^{(s)}$ denote the distributional derivative of u of order s.

- For $s \in \mathbb{N}$ and $p \in [1,\infty]$, let $W^{s,p}(G)$ and $W^{s,p}_{\mathrm{loc}}(G)$ denote the standard Sobolev space on G or its local version, respectively. Let $W^{0,p}(G) \equiv W_0^{0,p}(G) := L^p(G)$. For $s \in (0,\infty)\backslash\mathbb{N}$, complex interpolation yields the space

$$W^{s,p}(G) := \big[L^p(G), W^{m,p}(G)\big]_{s/m}$$

equipped with an (equivalent) Sobolev-Slobodeckij norm

$$\|u\|_{W^{s,p}(G)} = \begin{cases} \left(\|u\|^p_{W^{\lfloor s\rfloor,p}(G)} + \int_G\int_G \left(\frac{|\partial_x^{\lfloor s\rfloor} u(x) - \partial_y^{\lfloor s\rfloor} u(y)|^p}{|x-y|^{s-\lfloor s\rfloor}}\right)\frac{\mathrm{d}x\,\mathrm{d}y}{|x-y|}\right)^{1/p}, & p<\infty, \\ \max\left\{\|u\|_{W^{\lfloor s\rfloor,\infty}(G)}, \operatorname*{ess\,sup}_{x,y\in G} \frac{|u(x)-u(y)|}{|x-y|^{s-\lfloor s\rfloor}}\right\}, & p=\infty. \end{cases}$$

- For $s>0$ and $p \in [1,\infty]$, let $W_0^{s,p}(G)$ denote the closure of $C^\infty(G)$-functions vanishing in a neighborhood of the left boundary a of G with respect to the norm of $W^{s,p}(G)$. Further, let $W^{-s,p}(G) := \big(W_0^{s,p'}(G)\big)'$ with $\frac{1}{p}+\frac{1}{p'}=1$.

From [1, Chapter 7.3] and [27, Section 15], we know

$$W_0^{s,p}(G) \hookrightarrow \big\{u \in C^0(\bar{G}) \,|\, u(a)=0\big\} \text{ for } p\in(1,\infty], s\in(\tfrac{1}{p},\infty). \tag{2.1}$$

Remark 2.1. *It should be pointed out that (similar to [7]) our definition of $W_0^{s,p}(G)$ and $C_0^\alpha(\bar{G})$ (and, therefore, of $W^{-s,p}(G)$ and $C^{-\alpha}(\bar{G})$) slightly differs from the usual one as we do not require its elements to vanish at the right boundary of G. Still, all essential properties typical for the classical $W_0^{s,p}(G)$ and $C_0^\alpha(\bar{G})$ spaces are preserved.*

2.1.2 Levý Construction of Standard Brownian Motion

Let $(\Omega,\mathcal{F},\mathbf{P})$ be a complete probability space and let $\xi_1,\xi_2,\ldots$ be iid random variables with mean 0 and standard deviation $\sigma>0$. For $n\in\mathbb{N}$, we define the cumulative sum $S_n := \sum_{k=1}^n \xi_k$ and consider the process

$$X_n(t) := \sigma n^{-1/2}\big(S_{\lfloor nt\rfloor} + (nt-\lfloor nt\rfloor)\xi_{\lfloor nt\rfloor+1} \text{ for } t\in[0,1]. \tag{2.2}$$

By construction, for any $n\in\mathbb{N}$, $X_n \in C^0\big([0,1]\big)$.

Let $\mathbf{P}_n$ denote the probability measure of S_n on $C^0\big([0,1]\big)$ and $\mathcal{B}$ be the Borel σ-algebra on $C^0\big([0,1]\big)$. By virtue of Donsker's invariance principle, there exists a probability measure $\tilde{\mathbf{P}}$ on $\big(C^0\big([0,1]\big),\mathcal{B}\big)$ and a $\tilde{\mathbf{P}}$-measurable standard Brownian motion $W := (W(t))_{t\in[0,1]}$ such that, as $n\to\infty$, $\mathbf{P}_n$ weakly converges to $\tilde{\mathbf{P}}$ and X_n converges in distribution to W, i.e.,

$$\mathbf{P}_n \xrightarrow{w} \tilde{\mathbf{P}} \text{ and } X_n \xrightarrow{d} W \text{ as } n\to\infty. \tag{2.3}$$

From [7, p. 13], we know

$$W(\cdot) \in C_0^\alpha\big([0,1],\mathbb{R}^d\big)\cap W^{s,p}\big(0,1;\mathbb{R}^d\big) \text{ for } \alpha\in(0,\tfrac{1}{2}), s\in(0,\tfrac{1}{2}), p\in[1,\infty)\ \mathbf{P}\text{–a.s.} \tag{2.4}$$

Moreover, on the strength of [7, Theorem 3.6],

$$W \in L^p\big(\Omega, H_0^{s,p}\big(0,1;\mathbb{R}^d\big),\mathcal{F},\mathbf{P}\big) \text{ for } p\in(2,\infty), s\in(\tfrac{1}{p},\tfrac{1}{2})$$

with respect to some complete probability space denoted again (for the sake of simplicity) as $(\Omega,\mathcal{F},\mathbf{P})$. Here and in the sequel, the (separable) "image space" is equipped with the standard Borel σ-algebra.

2.2 Nonparametric Regression

Before we introduce a rigorous probabilistic framework for nonparametric regression in Section 2.3, we start with an informal discussion by first ignoring topology and measurability aspects. In this article, we follow the probabilistic approach to nonparametric regression. We refer the reader to a recent work [8] by Davies for a description of the alternative data-driven approach.

Consider a process $\big(Y(t)\big)_{t\in[0,1]}$ possessing a Tukey decomposition

$$Y(t) = \mu(t) + \varepsilon(t) \text{ for } t \in [0,1] \tag{2.5}$$

where $\big(\varepsilon(t)\big)_{t\in[0,1]}$ models the noise and $\mu\colon [0,1] \to \mathbb{R}$ is the signal we aim to reconstruct. Again, as mentioned, evaluating Y and μ pointwise may not be meaningful if Y and μ are irregular. Nevertheless, since the vast majority of statistical literature ignores the (ir)regularity aspects, we assume for a moment evaluating Y and μ pointwise is meaningful.

For $n \in \mathbb{N}$, we define a finite index set $T_n = \big\{t_k = \frac{k}{n} \,|\, 1 \le k \le n\big\}$. For Equation (2.5), consider a regression problem with fixed design given by

$$Y(t_k) = \mu(t_k) + \varepsilon(t_k) \text{ for } t_k \in T_n. \tag{2.6}$$

Our thrust is to obtain an estimate $\hat{\mu}_n$ of the function μ based on a realization

$$\big(y_n\left(\tfrac{1}{n}\right), y_n\left(\tfrac{2}{n}\right), \ldots, y_n(1)\big)$$

of the process $(Y_t)_{t\in T_n}$. That is, given a set $\mathscr{R}$ of regression functions defined on $[0,1]$ and indexed by a parameter set Θ, one needs to find an element of $\mathscr{R}$ which is the "best" approximation for μ.

Within the framework of parametric regression, the set $\mathscr{R}$ is finitely dimensional, e.g., the space of affine linear functions on $[0,1]$ in case of linear regression or polynomials up to certain degree in case of polynomial regression, etc. In this case, one often employs the so-called least squares estimator $\hat{\mu}_n$ of μ obtained as a solution to the minimization problem

$$\hat{\mu}_n = \operatorname*{arg\,min}_{\theta\in\Theta} \sum_{k=1}^{n} \left|y_n\left(\tfrac{k}{n}\right) - f\left(\tfrac{k}{n};\theta\right)\right|^2. \tag{2.7}$$

In contrast to parametric regression, the set $\mathscr{R}$ is infinitely dimensional when a nonparametric or semiparametric framework is used. In this case, the minimization problem in Equation (2.7) is usually ill-posed and a regularization technique is needed. A typical regularization approach is to replace Equation (2.7) with

$$\hat{f}_n = \operatorname*{arg\,min}_{\theta\in\Theta} \Big(\sum_{k=1}^{n} \left|y_n\left(\tfrac{k}{n}\right) - f\left(\tfrac{k}{n};\theta\right)\right|^2 + \mathcal{P}_n\big(f(\cdot;\theta)\big)\Big) \tag{2.8}$$

where $\mathcal{P}_n\colon \mathscr{R} \to [0,\infty)$ is a Tikhonov-type penalization functional. For example, letting $\mathscr{R} = W^{2,2}\big((0,1)\big)$ and

$$\mathcal{P}_n(f) = \lambda_n \int_0^1 \left|f^{(2)}(t)\right|^2 \mathrm{d}t \text{ for } f \in \mathscr{R} \text{ and appropriate } \lambda_n > 0,$$

the regression model from Equation (2.8) reduces to the well-known cubic P-spline regression (see, e.g., [17, Section 4.2.3]). Here, $f^{(s)}$ denotes the s-th order distributional/weak derivative of f (cf. [4, Chapter 1]). Due to the continuity and strict convexity of $\mathcal{P}_n$ as well as the convexity and compactness of $\mathscr{R}$, the unique minimizer in Equation (2.8) exists in $\mathscr{R}$ and is a smooth function by virtue of the Sobolev's imbedding theorem (see [1, Chapter 4]). Hence, the estimation quality can be rather poor if the true mean μ exhibits non-smoothness or discontinuity properties.

2.3 The Taut String Estimator

In the following, we outline the taut string regression technique proposed by [9]. For the process $\big(Y(t)\big)_{t\in[0,1]}$ given in Equation (2.5), the key idea of their approach is to apply nonparametric regression to the antiderivative of $\big(Y(t)\big)_{t\in[0,1]}$ rather then applying it directly to $\big(Y(t)\big)_{t\in[0,1]}$ since the antiderivative has better regularity properties than the original process. As proved by Grasmair [15], the taut string regression can be substituted into the framework of Equation (2.8) by selecting

$$\mathcal{P}_n(f) = \lambda_n \mathrm{TV}(f),$$

where $\mathrm{TV}(f)$ stands for the total variation of f. Correspondingly, $\mathscr{R}$ is selected as the space $\mathrm{BV}\big([0,1]\big)$ of bounded variation functions. With $\mathscr{R}$ including functions having discontinuities at the node points T_n, this approach provides a good estimation for discontinuous μ.

Before we proceed with the estimator description, we put Equation (2.5) into a rigorous form. Let $\big(\varepsilon(t)\big)_{t\in[0,1]}$ be the white noise given as the weak/distributional derivative of a standard univariate Wiener process $\big(W(t)\big)_{t\in[0,1]}$ with respect to a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Using the results of Section 2.1, we have

$$\begin{aligned} &\varepsilon(\cdot) \in W_0^{s-1,p}\big((0,1)\big) \cap C_0^{\alpha-1}\big([0,1]\big) \text{ for } \alpha, s \in (0, \tfrac{1}{2}), p \in [1,\infty) \ \mathbf{P}\text{–a.s. and} \\ &\varepsilon \in L^p\big(\Omega, W_0^{s-1,p}\big((0,1)\big), \mathcal{F}, \mathbf{P}\big) \text{ for any } s \in (\tfrac{1}{p}, \tfrac{1}{2}), p \in (2,\infty). \end{aligned} \tag{2.9}$$

Further, let $\mu \in L^p\big((0,1)\big)$ for some $p \in (1,\infty)$. With the grid T_n defined in Section 2.2, we consider the regression problem

$$n\int_{t_{k-1}}^{t_k} Y(s)\mathrm{d}s = n\int_{t_{k-1}}^{t_k} \mu(s)\mathrm{d}s + \sigma n \int_{t_{k-1}}^{t_k} \varepsilon(s)\mathrm{d}s \text{ for } 1 \le k \le n \tag{2.10}$$

with $t_k := \frac{k}{n}$ for $0 \le k \le n$ and a given $\sigma > 0$. Here, the integration operator is formally interpreted in the sense

$$\int_a^b f(s)\mathrm{d}s := \big((\partial_t)^{-1} f\big)(b) - \big((\partial_t)^{-1} f\big)(a),$$

where the derivative operator $\partial_t := (\cdot)^{(1)}$ is an isomorphism between $W_0^{1,p}\big((0,1)\big)$ and $L^p\big((0,1)\big)$, which, using standard interpolation and duality techniques, can be extended to an isomorphism between the spaces $W_0^{s,p}\big((0,1)\big)$ and $W^{s-1,p}\big((0,1)\big)$ for $s \in (0,1)$. Further, ∂_t can be viewed as a bijective mapping between the (non-separable) Banach space $C_0^\alpha\big([0,1]\big)$ and the Fréchet space $C^{\alpha-1}\big([0,1]\big)$ (cf. [12, p. 209]). In view of these facts, $(\partial_t)^{-1}\varepsilon(\cdot) \in C_0^\alpha\big([0,1]\big)$ for $\alpha \in (0, \frac{1}{2})$ $\mathbf{P}$–a.s. and $(\partial_t)^{-1}\mu \in W_0^{1,p}\big((0,1)\big) \hookrightarrow C_0^{p/(p-1)}\big([0,1]\big)$. Thus, Equation (2.10) can be interpreted in the strong sense as both $(\partial_t)^{-1}\mu$ and $(\partial_t)^{-1}\varepsilon(\cdot)$ can be evaluated pointwise in $[0,1]$.

Letting

$$Y_n(t_k) = n\int_{t_{k-1}}^{t_k} Y(s)\mathrm{d}s, \quad \mu_n(t_k) = n\int_{t_{k-1}}^{t_k} \mu(s)\mathrm{d}s, \quad \varepsilon_n(t_k) = n\int_{t_{k-1}}^{t_k} \varepsilon(s)\mathrm{d}s$$

and extending Y_n, μ_n and ε_n to cádlág step functions with jumps at t_k's, Equation (2.10) can be written as

$$Y_n(t_k) = \mu_n(t_k) + \sigma\varepsilon_n(t_k) \text{ for } 1 \le k \le n, \tag{2.11}$$

where $\varepsilon_n(t_k)$'s are iid standard Gaussian random variables. In fact, the convergence results to follow will remain valid also for non-Gaussian iid $\varepsilon_n(t_k)$'s with bounded moments up to a certain order.

It should also be pointed out that Equation (2.11) is now rigorously defined but still resembles the original naive formulation in Equation (2.6).

Given a realization $\left(y_n\left(\frac{1}{n}\right), y_n\left(\frac{2}{n}\right), \ldots, y_n(1)\right)$ of the locally time-averaged process $\left(Y_n(t)\right)_{t\in T_n}$, we define the cumulative process

$$y_n^\circ\left(\tfrac{k}{n}\right) = \frac{1}{n}\sum_{j=1}^{k} y_n\left(\tfrac{j}{n}\right) \text{ for } 0 \le k \le n.$$

Hence, y_n° is an approximation of its continuous counterpart $\left(Y_n^\circ(t)\right)_{t\in[0,1]}$ with the latter given by

$$Y_n^*(t) = \int_0^t Y_n(s)\mathrm{d}s \text{ for } t \in [0,1].$$

For $C > 0$, we define the lower and upper bounds for x_n° via

$$\begin{aligned} \varphi_n(0) = u_n(0) = 0, \quad \varphi_n\left(\tfrac{k}{n}\right) = y_n^\circ\left(\tfrac{k}{n}\right) - C\sigma n^{-1/2} \text{ for } 1 \le k \le n \text{ and} \\ \psi_n(0) = u_n(0) = 0, \quad \psi_n\left(\tfrac{k}{n}\right) = y_n^\circ\left(\tfrac{k}{n}\right) + C\sigma n^{-1/2} \text{ for } 1 \le k \le n, \end{aligned} \tag{2.12}$$

respectively, and extend them onto $[0,1]$ by piecewise linear functions (cf. Equation (2.2) and [9, pp. 3]). In practice, since σ is not known, it can roughly be estimated by

$$\hat{\sigma}_n = \tfrac{1.48}{\sqrt{2}} \operatorname{Median}\left\{|y_n(t_2) - y_n(t_1)|, \ldots, |y_n(t_n) - y_n(t_{n-1})|\right\}. \tag{2.13}$$

As for the constant C, Davies and Kovac recommended in [9, pp. 32] taking

$$C = \operatorname{Median}\left(\max_{t\in[0,1]} |W(t)|\right) \approx 1.149. \tag{2.14}$$

Next, we define

$$\theta_n(t) = t y_n^\circ(0) \text{ for } t \in [0,1].$$

Obviously, $\varphi_n, \psi_n \in W^{1,\infty}\left((0,1)\right)$ and θ_n is a straight line. Using Equation (A.1), we can thus define the functional tube

$$\mathcal{K}_{\varphi_n,\psi_n}^{\theta_n} = \left\{s \in W^{1,\infty}\left((0,1)\right) \,\middle|\, s(0) = 0, s(1) = y_n^\circ(1), \varphi_n \le s_n \le \psi_n \text{ in } [0,1]\right\}.$$

Consider the length functional $\mathcal{L}$ defined in Equation (A.3). From Theorem A.3, we know $\mathcal{L}$ attains its unique minimum over the set $\mathcal{K}_{\varphi_n,\psi_n}^{\theta_n}$ at some $s_n^* \in W^{1,\infty}\left((0,1)\right)$. Further, the minimizer s_n^* is exactly the taut string running through the tube $\mathcal{K}_{\varphi_n,\psi_n}^{\theta_n}$ and satisfying the "boundary conditions"

$$s_n^*(0) = 0 \text{ and } s_n^*(1) = y_n^\circ(0).$$

Moreover, s_n^* is a piecewise linear function uniquely determined by its values at the node points $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$. The taut string estimator $\hat{\mu}_n$ for μ is then given as the cádlág realization of $s_n^{*(1)}$, i.e.,

$$\hat{\mu}_n(t) = s_n^{*(1)}(t) \text{ for } t \in [0,1].$$

For the sake of convenience, we modify $\hat{\mu}_n$ on $\left[0, \frac{1}{n}\right)$ according to

$$\hat{\mu}_n(t) := \hat{\mu}_n\left(\tfrac{1}{n}\right) \text{ for } t \in \left[0, \tfrac{1}{n}\right)$$

such that $\hat{\mu}_n$ is uniquely determined by its values on T_n. This way, we do not have to distinguish between $\hat{\mu}_n$ and $\hat{\mu}_n|_{T_n}$. Hence, the taut string estimator $\mathcal{T}_n$ of μ is a nonlinear mapping defined as

$$\mathcal{T}_n(\cdot; C\sigma)\colon \mathbb{R}^n \to \mathbb{R}^n, \quad x_n \mapsto \hat{\mu}_n.$$

Figure 1 shows the result of application of the taut string estimator to a dataset of size $n = 200$ selected from a distribution modelled by Equation (2.11) with $\mu(t) = 2\exp(t)\operatorname{sign}\left(1 - 2t - \cos(3.1\pi t)\right)$, $t \in [0,1]$, and $\sigma = 1$.

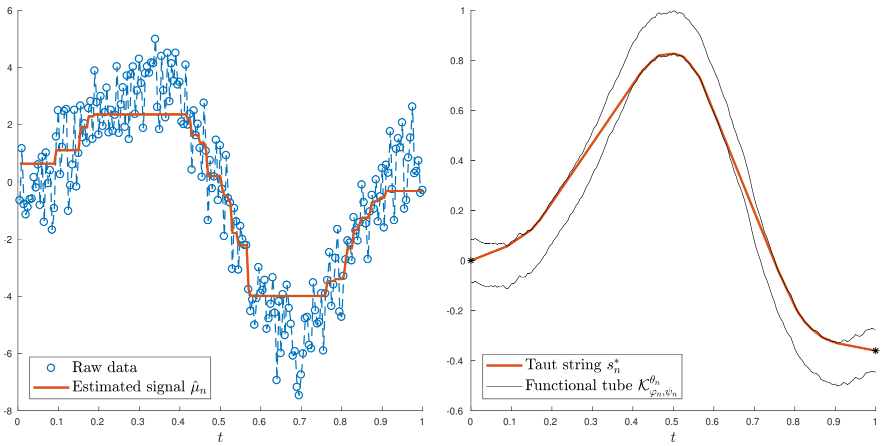


Figure 1: Illustration of the taut string methodology.

2.4 Asymptotic Consistency of the Taut String Estimator

Before presenting our new convergence study, we briefly summarize some important results on the asymptotic behavior of the taut string estimator from [9].

With $K_n = K_n^{C\sigma}$ denoting the number of local extremes of $\hat{\mu}_n$, we trivially observe $0 \le K_n \le n-1$. In [9, Theorem 4], the following result was proved.

Theorem 2.2. *Let μ have k local extrema. If $\varepsilon_n(t_k)$'s are iid random variables with zero mean and unit variance, then*

$$\lim_{n\to\infty} \mathrm{P}\big(\{k \ge K_n^{C\sigma}\}\big) = H(C), \tag{2.15}$$

where H denotes the cdf of the random variable $\max_{t\in[0,1]} |W(t)|$ with $\big(W(t)\big)_{t\in[0,1]}$ standing for the standard univariate Wiener process.

Among other assertions, [9, Theorem 6] gives the following consistency result for the taut string estimator.

Theorem 2.3. *Let $\mu \in C^2\big([0,1]\big)$ possess exactly k local extreme values $\{t_1^e, \dots, t_k^e\}$ such that $\mu^{(1)}$ vanishes only at $\{t_1^e, \dots, t_k^e\}$ whereas $\mu^{(2)}$ does not vanish at these points. Further, let $\varepsilon_n(t_k)$'s be iid sub-Gaussian random variables. Thus, there exists a number $A > 0$ such that*

$$\lim_{C\to\infty} \lim_{n\to\infty} \mathbf{P}\Big(\Big\{ \operatorname*{ess\,sup}_{t\in A_n} \frac{|\mu(t) - \hat{\mu}_n(t)|}{|\mu^{(1)}(t)|^{1/3}} \le AC^{2/3} n^{-1/3} \Big\}\Big) = 1$$

with $A_n = \big[A\big(\frac{\log n}{n}\big)^{1/3}, 1 - \big(\frac{\log n}{n}\big)^{1/3}\big]$ and

$$\lim_{C\to\infty} \lim_{n\to\infty} \mathbf{P}\Big(\Big\{ \operatorname*{ess\,sup}_{t\in[0,1]} \frac{|\mu(t) - \hat{\mu}_n(t)|}{|\mu^{(2)}(t)|^{1/3}} \le AC^{2/3} n^{-1/3} \Big\}\Big) = 1.$$

Recall that a random variable ζ is called sub-Gaussian if there exists a number $c > 0$ such that $\mathrm{E}\big[\exp(\lambda\zeta)\big] < \exp(c\lambda^2)$ *for any $\lambda \in \mathbb{R}$.*

Theorem 2.3, however, has both attractive and unattractive aspects. On one hand, it yields the convergence in probability of the taut string estimator in the rather strong uniform norm. On the other hand, the convergence rate of $n^{-1/3}$ is suboptimal compared to $n^{-2/5}$ predicted in [19] and the limiting procedure assumes the constant C goes to infinity whereas in practice, it is small and fixed. Further, the proof relies on rather restrictive smoothness as well as local monotonicity and convexity assumptions. Another downside is that no information is provided on the constant A.

Now, we present an alternative convergence study of the taut string estimator. Our approach requires less regularity assumptions on μ and provides a convergence rate of $n^{-1/2}$ even if μ has finitely many points of discontinuity. The convergence, however, holds in a rather week topology. Instead of the $L^\infty\big((0,1)\big)$-norm used by Davies and Kovac in [9, Theorem 6], we employ the topology of $W^{-1,p}\big((0,1)\big)$.

For $p \in [1,\infty]$, consider the pathwise L^p-norm of the standard Brownian motion on $(0,1)$. This yields a random variable $\vartheta_p := \|W\|_{L^p((0,1))}$. Using Hölder's inequality to estimate

$$|\vartheta_p| = \Big(\int_0^1 |W(t)|^p \mathrm{d}t\Big)^{1/p} \leq \max_{0\leq t\leq 1} |W(t)| \leq \vartheta_\infty$$

and taking into account that all moments of ϑ_∞ are finite, we conclude that the moments of ϑ_p for any $p \in [1,\infty]$ are also finite.

Theorem 2.4. *For some $\alpha \in (1/2, 1]$ and $p \in [1,\infty)$, assuming*

- *$\mu \in C^\alpha(\bar{I}_j)$, I_j open for $1 \leq j \leq m$ such that $\bigcup_{j=1}^{m} \bar{I}_j = [0,1]$. (The "signal" function μ may be discontinuous at ∂I_j.)*
- *$\varepsilon_n(t_k)$'s are iid random variables with zero mean and unit variance,*

we then have

$$\|\hat{\mu}_n - \mu\|_{W^{-1,p}((0,1))} \leq n^{-1/2}\sigma(C + \vartheta_p) + o_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty.$$

Proof. For $n \in \mathbb{N}$, let $S_n^* \in W^{1,\infty}((0,1))$ denote the taut string through the function tube

$$\mathcal{K}^{\theta_n}_{\varphi_n,\psi_n} = \big\{S_n \in W^{1,\infty}((0,1)) \,|\, \varphi_n \leq S_n \leq \psi_n\big\}$$

centered around the cumulative process $\big(Y_n^\circ(t)\big)_{t\in[0,1]}$, where the bounds φ_n and ψ_n are defined as in Equation (2.12). Recall that the taut string estimator $\hat{\mu}_n$ is given as a cádlág realization of the weak derivative $(S_n^*)^{(1)}$ of S_n^*.

Using the definition of φ_n, ψ_n and exploiting Equation (2.11),

$$\varphi_n(t) = \frac{1}{n}\sum_{k=0}^{\lfloor nt\rfloor}\big(\mu_n(\tfrac{k}{n}) + \varepsilon_n(\tfrac{k}{n})\big) + \{nt\}\Big(\mu_n\big(\tfrac{\lceil nt\rceil}{n}\big) + \varphi_n\big(\tfrac{\lceil nt\rceil}{n}\big)\Big) - C\sigma n^{-1/2},$$

$$\psi_n(t) = \frac{1}{n}\sum_{k=0}^{\lfloor nt\rfloor}\big(\mu_n(\tfrac{k}{n}) + \varepsilon_n(\tfrac{k}{n})\big) + \{nt\}\Big(\mu_n\big(\tfrac{\lceil nt\rceil}{n}\big) + \varphi_n\big(\tfrac{\lceil nt\rceil}{n}\big)\Big) + C\sigma n^{-1/2}$$

with $\{\cdot\} := (\cdot) - \lfloor\cdot\rfloor$ denoting the fractional part of a real number.

Consider the set

$$A_n := \bigcup_{k=1}^{n}\big\{(t_{k-1},t_k) \,\big|\, (t_{k-1},t_k)\cap\partial I_j = \varnothing \text{ for } 1\leq j\leq m\big\}.$$

Obviously, as $n \to \infty$, the Lebesgue measure of $[0,1]\backslash A_n$ goes to zero as $\frac{1}{n}$. For any $t \in A_n$, we estimate

$$|\mu_n(t) - \mu(t)| \le \frac{1}{n} \sum_{j=1}^{m} \int_{t_{k-1}}^{t_k} |\mu(s) - \mu(t)| \mathrm{d}s \le L n^{-\alpha} \tag{2.16}$$

with $L := \max_{1 \le j \le m} \|\mu\|_{C^\alpha(I_j)}$. Letting $M(t) := \int_0^t \mu(s)\mathrm{d}s$ for $t \in [0,1]$ and recalling Equation (2.16), we find for any $t \in [0,1]$

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{k=0}^{\lfloor nt \rfloor} \mu_n(\tfrac{k}{n}) + (nt - \lfloor nt \rfloor)\mu_n(\tfrac{\lceil nt \rceil}{n}) - M(t)\Big| \\
&\quad\le \Big|\frac{1}{n}\sum_{k=0}^{\lfloor nt \rfloor} \mu_n(\tfrac{k}{n}) - M(t)\Big| + n^{-1}\|\mu\|_{L^\infty((0,1))} \\
&\quad\le \Big|\frac{1}{n}\sum_{\substack{0 \le k \le \lfloor nt \rfloor \\ t_k \notin A_n}} \mu_n(\tfrac{k}{n}) - M(t)\Big| + \tfrac{m+1}{n}\|\mu\|_{L^\infty((0,1))} + \tfrac{m}{n}\|M\|_{L^\infty((0,1))} \\
&\quad\le \Big|\frac{1}{n}\sum_{\substack{0 \le k \le \lfloor nt \rfloor \\ t_k \notin A_n}} \mu(\tfrac{k}{n}) - M(t)\Big| + L n^{-\alpha} + o(n^{-1}) \\
&\quad\le \sum_{k=0}^{\lceil nt \rceil} \int_{t_k}^{t_{k+1}} |\mu(s) - \mu(k/n)|\mathrm{d}s + O(n^{-\alpha}) = O(n^{-\alpha}).
\end{aligned}$$

Combining the latter inequality with Equation (2.3),

$$\begin{aligned}
\varphi_n(t) &= M(t) + n^{-1/2}\sigma W(t) - C\sigma n^{-1/2} + o_{\mathbf{P}}(n^{-1/2}) \text{ and} \\
\psi_n(t) &= M(t) + n^{-1/2}\sigma W(t) + C\sigma n^{-1/2} + o_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty
\end{aligned} \tag{2.17}$$

uniformly with respect to $t \in [0,1]$. Therefore, as $n \to \infty$,

$$\begin{aligned}
\|\varphi_n - M\|_{L^p((0,1))} &= n^{-1/2}\sigma\big(C + \vartheta_p\big) + o_{\mathbf{P}}(n^{-1/2}), \\
\|\psi_n - M\|_{L^p((0,1))} &= n^{-1/2}\sigma\big(C + \vartheta_p\big) + o_{\mathbf{P}}(n^{-1/2}).
\end{aligned} \tag{2.18}$$

Theorem A.3 implies

$$\|S_n^* - M\|_{L^p((0,1))} = n^{-1/2}\sigma\big(C + \vartheta_p\big) + o_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty.$$

Using the fact that the weak derivative operator $(\cdot)^{(1)}$ is a contraction from the Banach space $L^p((0,1))$ into $W^{-1,p}((0,1))$, i.e.,

$$\begin{aligned}
\|f^{(1)}\|_{W^{-1,p}((0,1))} &= \sup_{\|\varphi\|_{W^{1,p'}((0,1))}=1} \big|\langle f^{(1)}, \varphi\rangle_{W^{-1,p}((0,1));W_0^{1,p'}((0,1))}\big| \\
&= \sup_{\|\varphi\|_{W^{1,p'}((0,1))}=1} \Big|\int_0^t f(t)\varphi^{(1)}(t)\mathrm{d}t\Big| \\
&\le \sup_{\|\varphi\|_{W^{1,p'}((0,1))}=1} \|f\|_{L^p((0,1))}\|\varphi\|_{W^{1,p'}((0,1))} \\
&= \|f\|_{L^p((0,1))},
\end{aligned} \tag{2.19}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$, we finally get

$$\begin{aligned}\|\mu_n - \mu\|_{W^{-1,p}((0,1))} &= \big\|(S_n^*)^{(1)} - M^{(1)}\|_{W^{-1,p}((0,1))} \leq \|S_n^* - M\|_{L^p((0,1))} \\ &\leq n^{-1/2}\sigma(C + \vartheta_p) + o_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty\end{aligned}$$

as claimed. □

Remark 2.5. *The weak convergence result in Theorem 2.4 can be interpreted as follows. For an arbitrary $f \in W_0^{1,p}(G)$, there holds*

$$\Big|\int_0^1 \big(\mu_n(t) - \mu(t)\big) f(t)\mathrm{d}t\Big| \leq n^{-1/2}\sigma(C + \theta_p)\|f\|_{W^{1,p}(G)} + o_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty,$$

i.e., the difference $(\hat{\mu}_n - \mu)$ being multiplied in $L^2(G)$ with any test function $f \in W_0^{1,p}(G)$ vanishes at the rate of $n^{-1/2}$.

Theorem 2.4 can be used to obtain confidence balls for μ in the norm of $W^{-1,p}\big((0,1)\big)$. The following corollary is straightforward.

Corollary 2.6. *Let the assumptions of Theorem 2.4 be satisfied. For a significance level $\alpha \in (0,1)$, let $\theta_{p,1-\alpha}$ denote the $(1-\alpha)$-th quantile of θ_p. Consider a random functional ball*

$$B_{n,1-\alpha} := \big\{u \in W^{-1,p}\big((0,1)\big) \mid \|\hat{\mu}_n - u\|_{W^{-1,p}((0,1))} \leq n^{-1/2}\sigma(C + \vartheta_p)\big\},$$

then

$$\limsup_{n\to\infty} \mathbf{P}\big\{\mu \in B_{n,1-\alpha}\big\} \geq 1 - \alpha.$$

Remark 2.7. *In practice, σ is not known. With an estimate $\hat{\mu}_n$ of μ being available, $\hat{\sigma}_n$ can be estimated using Equation* (2.13)*. If $\hat{\mu}_n$ can be verified to converge to μ in $C^0\big([0,1]\big)$ in probability* $\mathbf{P}$*, Equation* (2.13) *can be refined via*

$$\hat{\sigma}_n = \tfrac{1.48}{\sqrt{2}} \,\text{Median}\,\big\{|y_n(t_1) - \hat{\mu}_n(t_1)|, \ldots, |y_n(t_n) - \hat{\mu}_n(t_n)|\big\}.$$

The topology in Theorem 2.4 can be strengthened at the price of lowering the convergence rate.

Theorem 2.8. *Let $s \in [0,1)$. In addition to the assumptions of Theorem 2.4, let*

$$\mathbf{E}\big[|\varepsilon_n(t_k)|^{p+2}\big] < \infty,$$

then there exist constant $C_p, C > 0$ such that

$$\|\hat{\mu}_n - \mu\|_{W^{s-1,p}((0,1))} \leq 2CC_p n^{-\frac{1-s}{2}} \|\mu\|_{L^p((0,1))}^{1-s} \Big(\|\mu\|_{L^p((0,1))} + \Big(\mathbf{E}\big[|\varepsilon_n(0)|^p\big]\Big)^{1/p}\Big)^s + o_{\mathbf{P}}(1) \text{ as } n \to \infty.$$

Proof. First, we show an L^p-bound for $\hat{\mu}_n$. Being piecewise linear functions, φ_n and ψ_n are weakly differentiable with piecewise constant derivatives $\varphi_n^{(1)}, \psi_n^{(1)} \in L^\infty\big((0,1)\big)$ expressed as

$$\varphi_n^{(1)}(t) = \psi_n^{(1)}(t) = \mu_n\big(\lceil \tfrac{t}{n} \rceil\big) + \varepsilon_n\big(\lceil \tfrac{t}{n} \rceil\big) \text{ for } t \in [0,1]. \tag{2.20}$$

Next, using Equation (2.20) and exploiting Minkowski inequality, by virtue of the weak law of large numbers, we obtain

$$\|\varphi_n^{(1)}\|_{L^p((0,1))} = \|\psi_n^{(1)}\|_{L^p((0,1))}$$

$$
\begin{aligned}
&= \Big(\sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} \big|\mu_n\big(\tfrac{k}{n}\big) + \varepsilon_n\big(\tfrac{k}{n}\big)\big|^p \mathrm{d}t \Big)^{1/p} \\
&= \Big(\sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} \big|\mu_n\big(\tfrac{k}{n}\big)\big|^p \mathrm{d}t + \int_{(k-1)/n}^{k/n} \big|\varepsilon_n\big(\tfrac{k}{n}\big)\big|^p \mathrm{d}t \Big)^{1/p} \qquad (2.21)\\
&\leq \|\mu\|_{L^p((0,1))} + \Big(n^{-1} \sum_{k=1}^{n} \big|\varepsilon_n\big(\tfrac{k}{n}\big)\big|^p \Big)^{1/p} + o(1) \\
&= \|\mu\|_{L^p((0,1))} + \Big(\mathbf{E}\big[|\varepsilon_n(0)|^p\big] \Big)^{1/p} + o_{\mathbf{P}}(1)
\end{aligned}
$$

as $n \to \infty$ based on Equation (2.16).

Taking into account the one-dimensional Poincaré & Friedrichs inequality and using the interpolation inequality for the fractional Sobolev space

$$
W_0^{s,p}((0,1)) = \big[L^p((0,1)), W_0^{1,p}((0,1))\big]_s
$$

(cf. [1, p. 250]), we obtain a constant $C_p > 0$, which only depends on p, such that

$$
\begin{aligned}
\|S_n^* - M\|_{W_0^{s,p}((0,1))} &\leq C_p \big\|(S_n^*)^{(1)} - M^{(1)}\big\|_{L^p((0,1))}^{1-s} \|S_n^* - M\|_{L^p((0,1))}^{s} \\
&\leq C_p \big\|\hat{\mu}_n - \mu\|_{L^p((0,1))}^{1-s} \|\psi_n - \varphi_n\|_{L^p((0,1))}^{s} \\
&\leq C_p \big(\|\hat{\mu}_n\|_{L^p((0,1))} + \|\mu\|_{L^p((0,1))}\big)^{1-s} \|\psi_n - \varphi_n\|_{L^p((0,1))}^{s}
\end{aligned}
$$

Calculating

$$
\|\psi_n - \varphi_n\|_{L^p((0,1))} = 2Cn^{-1/2}
$$

and using Theorem 2.4, we obtain

$$
\|S_n^* - M\|_{W_0^{s,p}((0,1))} = 2CC_p n^{-\frac{1-s}{2}} \|\mu\|_{L^p((0,1))}^{1-s} \Big(\|\mu\|_{L^p((0,1))} + \Big(\mathbf{E}\big[|\varepsilon_n(0)|^p\big] \Big)^{1/p} \Big)^{s} + o_{\mathbf{P}}(1)
$$

as $n \to \infty$. Using the fact that the derivative operator is a contraction from $W_0^{s,p}((0,1))$ into $W^{s-1,p}((0,1))$ (cf. Equation 2.19), the claim follows. □

3 An $n^{-1/2}$-Rate State Recovery Strategy

In this section, we rigorously define our two-step state recovery strategy based on the taut string estimator and a Tikhonov regularization technique. Under appropriate assumptions, we prove our estimator converges.

For $q \in (1, \infty)$, let $\mathscr{X}$, $\mathscr{X}_-$ with $\mathscr{X} \hookrightarrow \mathscr{X}_-$ be L^q-based (reflexive) Sobolev spaces over an arbitrary domain. For $d \in \mathbb{N}$ and $p \in (1, \infty)$, consider an a (possibly) nonlinear operator

$$
\mathcal{A} \colon \mathfrak{D}(\mathcal{A}) \subset \mathscr{X} \to L^p(0,1;\mathbb{R}^d)
$$

with a domain $\mathfrak{D}(\mathcal{A})$. For $\mathbf{y} \in \mathfrak{R}(\mathcal{A})$, an inverse problem free of noise can be expressed as

$$
\mathcal{A}(x) = \mathbf{y}, \tag{3.1}
$$

where $\mathbf{y} \in L^p(0,1;\mathbb{R}^d)$ play the role of an observation variable. In the following, $\mathcal{A}$ will be extended to an operator

$$
\mathcal{A}_- \colon \mathfrak{D}(\mathcal{A}_-) \subset \mathscr{X}_- \to W^{-s,p}(0,1;\mathbb{R}^d) \quad \text{for some } s \in [1,\infty)
$$

with a domain $\mathfrak{D}(\mathcal{A}_-)$.

For some $\bar{x} \in \mathscr{X}_-$ and $q \in (1,\infty)$, consider a Tikhonov penalization functional

$$\mathcal{P}(x) = \|x - \bar{x}\|^q_{\mathscr{X}_-} \text{ for } x \in \mathscr{X}_-.$$

For Equation (3.1), the following generalized solution notion will be employed (cf. [26, Definition 3.25]).

Definition 3.1. *Let $\mathbf{y} \in \mathfrak{R}(\mathcal{A}) \subset \mathfrak{R}(\mathcal{A}_-)$. An element $x^\dagger \in \mathfrak{D}(\mathcal{A}_-)$ is referred to as a $\mathcal{P}$-minimizing solution to Equation (3.1) if*

$$\mathcal{A}(x^\dagger) = \mathbf{y} \textit{ and } \mathcal{P}(x^\dagger) = \inf\big\{\mathcal{P}(x)\,|\, x \in \mathfrak{D}(\mathcal{A}_-) \textit{ with } \mathcal{A}_-(x) = \mathbf{y}\big\}.$$

(Note that the definition of $x^\dagger$ depends on the choice of $\bar{x}$ fixed above.)

Unless $\mathcal{A}_-$ is linear (cf., e.g., [26, p. 53]), $\mathcal{P}$-minimizing solutions are not uniquely determined. If $\mathcal{A}_-$ is injective, $x^\dagger$ belongs even to $\mathfrak{D}(\mathcal{A})$. Further, if x is a unique solution to operator Equation (3.1), then it is also a $\mathcal{P}$-minimizing solution.

The following existence result in the class of $\mathcal{P}$-minimizing solutions is known from [26, Proposition 3.14].

Theorem 3.2. *Let the following conditions hold true:*

1) *The operator $\mathcal{A}$ is extendable to $\mathcal{A}_- : \mathfrak{D}(\mathcal{A}_-) \subset \mathscr{X}_- \to W^{-s,p}(0,1;\mathbb{R}^d)$ such that the set $\mathfrak{D}(\mathcal{A}_-)$ is convex and closed in $\mathscr{X}_-$.*

2) *The operator $\mathcal{A}_-$ is weak-to-weak sequentially continuous from $\mathscr{X}_-$ to $W^{-s,p}(0,1;\mathbb{R}^d)$, i.e., for any $x \in \mathfrak{D}(\mathcal{A}_-)$, $x_n \rightharpoonup x$ in $\mathscr{X}_-$ for some $(x_n)_{n\in\mathbb{N}} \subset \mathfrak{D}(\mathcal{A}_-)$ implies $\mathcal{A}_-(x_n) \rightharpoonup \mathcal{A}_-(x)$ in $W^{-s,p}(0,1;\mathbb{R}^d)$ as $n \to \infty$.*

Then Equation (3.1) possesse a (not necessarily unique) $\mathcal{P}$-minimizing solution.

Now we turn to a version of Equation (3.1) with noise. Let $\big(\mathbf{W}(t)\big)_{t\in[0,1]}$ be the standard d-variate Wiener process with respect to a probability space $(\Omega,\mathcal{F},\mathbf{P})$ and let $\boldsymbol{\varepsilon}$ denote its first variational derivative. Using the results of Section 2.1, $\varepsilon \in L^p\big(\Omega, W^{s-1,p}\big(0,1;\mathbb{R}^d\big),\mathcal{F},\mathbf{P}\big)$ with $\varepsilon(\cdot) \in C^{\alpha-1}\big([0,1]\big)$ for any $s,\alpha \in (0,\frac{1}{2})$ $\mathbf{P}$–a.s. For $\sigma > 0$, consider the statistical inverse problem with noise

$$\mathcal{A}(x) = \mathbf{y} + \sigma\boldsymbol{\varepsilon} \equiv \mathbf{y}_{\text{obs}}. \tag{3.2}$$

Since $\boldsymbol{\varepsilon}$ neither attains its values in $\mathfrak{R}(\mathcal{A})$ nor even in $L^p\big(0,1;\mathbb{R}^d\big)$, the notion of a $\mathcal{P}$-minimizing solution can not directly be applied to Equation (3.2) as the right-hand side of Equation (3.2) is too irregular.

To overcome this difficulty, we proceed as follows. Assuming the system in Equation (3.2) is indirectly observed over the “functionals” φ_k's from Equation (1.5), we obtain

$$\mathcal{A}(x_n) = \mathbf{y}_n + \sigma\boldsymbol{\varepsilon}_n \equiv \mathbf{y}_{\text{obs},n}, \tag{3.3}$$

where x_n is some (unknown) element of $\mathfrak{D}(\mathcal{A})$, $y_n \equiv y$ remains unchanged and for $t_k = \frac{k}{n}$, $1 \le k \le n$,

$$\mathbf{y}_n(t_k) = n\int_{t_{k-1}}^{t_k} \mathbf{y}(s)\mathrm{d}s, \quad \boldsymbol{\varepsilon}_n(t_k) = n\int_{t_{k-1}}^{t_k} \boldsymbol{\varepsilon}(s)\mathrm{d}s, \quad \mathbf{y}_{\text{obs},n}(t_k) = n\int_{t_{k-1}}^{t_k} \mathbf{y}_{\text{obs}}(s)\mathrm{d}s \tag{3.4}$$

with a unique extension by cádlág functions with jumps at most at t_k, $2 \leq k \leq n-1$ (cf. Section 2.3).

A standard solution approach to Equation (3.3) consists of constructing a regularization procedure in the sense of [26, Definition 3.20] such that, as $\sigma = \sigma(n) \to 0$ and $n \to \infty$, the sequence of solutions to regularized problems converge to an $\bar{x}$-minimum norm solution of the original problem (3.1) without noise. In addition to a uniform $\mathbf{P}$-a.s. boundedness condition on $\boldsymbol{\varepsilon}$ in the norm of $L^p\big(0,1;\mathbb{R}^d\big)$, which is obviously violated in our case, this framework is not practical since σ typically does not vanish in most applications. Hence, we apply our two-step regularization approach briefly outlined in the introduction to obtain a regularization scheme converging to a $\mathcal{P}$-minimizing solution of Equation (3.1) as $n \to \infty$ for *arbitrary*, but fixed $\sigma > 0$.

With the univariate taut string estimator $\mathcal{T}_n$ introduced in Section 2.3, we define its d-variate version

$$\mathcal{T}_n^d \colon (\mathbb{R}^d)^{T_n} \to (\mathbb{R}^d)^{T_n}$$

by applying $\mathcal{T}_n$ to each component of the sample, where $T_n = \big\{t_k = \frac{k}{n} \,|\, 1 \leq k \leq n\big\}$. Further, we define the extension operator

$$\mathcal{E}_n \colon (\mathbb{R}^d)^{T_n} \to L^p(0,1;\mathbb{R}^d)$$

mapping a lattice function to its unique extension to a cádlág step function with (possible) jumps at t_k, $2 \leq k \leq n-1$.

Definition 3.3. *For $\alpha > 0$, consider the following two-step regularization scheme.*

I) Define

$$\hat{\mathbf{y}}_n := \mathcal{E}_n\Big(\mathcal{T}_n^d\big(\mathbf{y}_{\mathrm{obs},n}|_{T_n}\big)\Big),$$

where $\mathbf{y}_{\mathrm{obs},n}|_{T_n}$ is the restriction of $\mathbf{y}_{\mathrm{obs},n}$ onto T_n.

II) Let $\hat{x}_n$ be a minimizer of the functional

$$\mathcal{F}_n(x) = \frac{1}{p}\big\|\mathcal{A}_-(x) - \hat{\mathbf{y}}_n\big\|^p_{W^{-s,p}(0,1;\mathbb{R}^d)} + \frac{\alpha\sigma^p}{q} n^{-p/2} \|x - \bar{x}\|^q_{\mathscr{X}_-} \tag{3.5}$$

over all $x \in \mathfrak{D}(\mathcal{A}_-)$.

We call this scheme a taut-string-based regularization approach.

The well-posedness of our taut-string based regularization technique is presented in Theorem 3.4.

Theorem 3.4. *Let the conditions of Theorem 3.2 be satisfied. The taut-string-based regularization approach yields then $\mathbf{P}$-a.s. for any $n \in \mathbb{N}$ an estimate $\hat{x}_n \in \mathfrak{D}(\mathcal{A}_-)$ as a minimizer of the functional in* (3.5).

Proof. As cádlág functions, both $\mathbf{y}_{\mathrm{obs},n}$ and $\hat{\mathbf{y}}_n$ $\mathbf{P}$-a.s. belong to $L^p\big(0,1;\mathbb{R}^d\big)$. With $\mathfrak{D}(\mathcal{A}_n) = \mathfrak{D}(\mathcal{A}_-)$, the set $\mathfrak{D}(\mathcal{A}_n)$ is convex and closed in $\mathscr{X}_-$. Hence, $\mathcal{A}_n$ satisfies conditions of Theorem 3.2. This fact together with [26, Proposition 4.1] yields the existence of a minimizer $\hat{x}_n$ of $\mathcal{F}_n$. □

Now, we prove a convergence result for our regularization scheme.

Theorem 3.5. *In addition to conditions of Theorem 3.4, suppose $\mathbf{y}$ is a piecewise Hölder-continuous function with a Hölder-exponent strictly bigger than $\frac{1}{2}$. Further, assume for some number $L > 0$*

$$\|x_1 - x_2\|_{\mathscr{X}_-} \leq L\big\|\mathcal{A}_-(x_1) - \mathcal{A}_-(x_2)\big\|_{W^{-s,p}(0,1;\mathbb{R}^d)} \quad \text{for } x_1, x_2 \in \mathfrak{D}(\mathcal{A}_-). \tag{3.6}$$

For $n \in \mathbb{N}$, let $\hat{x}_n \in \mathfrak{D}(\mathcal{A}_-)$ be given by the taut-string-based regularization approach introduced in Definition 3.3. With $x^\dagger \in \mathfrak{D}(\mathcal{A})$ standing for a $\mathcal{P}$-minizing solution to Equation (3.1), we have:

1) $\hat{x}_n$ *is* $\mathbf{P}$*-a.s. uniquely determined.*

2) $\hat{x}_n \in L^p\big(\Omega, \mathscr{X}_-, \mathcal{F}, \mathbf{P}\big)$ *for the complete probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *associated with* $(W_t)_{t\in[0,1]}$.

3) $\|\hat{x}_n - x^\dagger\|_{\mathscr{X}_-} = O_{\mathbf{P}}(n^{-1/2})$ *as* $n \to \infty$.

Proof. The $\mathbf{P}$-a.s. existence of $\hat{x}_n$ follows from Theorem 3.4, while the uniqueness is due to the injectivity of $\mathcal{A}_-$ furnished by Equation (3.6). Using Equations (3.5) and (3.6), we estimate

$$\begin{aligned}\big\|\mathcal{A}_-(\hat{x}_n) - \hat{\mathbf{y}}_n\big\|_{W^{-s,p}(0,1;\mathbb{R}^d)} &\leq \Big(p\mathcal{F}_n(\hat{x}_n)\Big)^{1/p} \leq \Big(p\mathcal{F}_n(x^\dagger)\Big)^{1/p} \\ &= \Big(\big\|\mathcal{A}_-(x^\dagger) - \hat{\mathbf{y}}_n\big\|^p_{W^{-s,p}(0,1;\mathbb{R}^d)} + \frac{p\alpha\sigma^p}{q} n^{-p/2}\|x^\dagger - \bar{x}\|^q_{\mathscr{X}_-}\Big)^{1/p} \\ &= \Big(\big\|\hat{\mathbf{y}}_n - \mathbf{y}\big\|^p_{W^{-s,p}(0,1;\mathbb{R}^d)} + \frac{p\alpha\sigma^p}{q} n^{-p/2}\|x^\dagger - \bar{x}\|^q_{\mathscr{X}_-}\Big)^{1/p}\end{aligned}$$

and

$$\begin{aligned}\|\hat{x}_n - x^\dagger\|_{\mathscr{X}_-} &\leq L\big\|\mathcal{A}_-(\hat{x}_n) - \mathcal{A}_-(x^\dagger)\big\|_{W^{-s,p}(0,1;\mathbb{R}^d)} \\ &\leq L\big\|\mathcal{A}_-(\hat{x}_n) - \hat{\mathbf{y}}_n\big\|_{W^{-s,p}(0,1;\mathbb{R}^d)} + L\|\hat{\mathbf{y}}_n - \mathbf{y}\|_{W^{-s,p}(0,1;\mathbb{R}^d)} \\ &= O_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty.\end{aligned}$$

Using the Lipschitzianity of $\mathcal{A}_-^{-1}$. the Lipschitz-continuous dependence of $\hat{x}_n$ on $\hat{\mathbf{y}}$ and, therefore, ε follows.

Combining the last two estimates, invoking Theorem 2.4 and using the norm continuity, we get

$$\hat{\mathbf{y}}_n \in L^p\big(\Omega, W^{-s,p}\big(0,1;\mathbb{R}^d\big), \mathcal{F}, \mathbf{P}\big) \quad \text{along with} \quad \|\hat{x}_n - x^\dagger\|_{\mathscr{X}_-} = O_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty,$$

thus, finishing the proof. □

Remark 3.6. *The assumption in Equation* (3.6) *is quite restrictive. For deterministic problems, a less restrictive requirement known as the "source condition" (cf. [26, Chapter 3 and 4]) is often employed. For stochastic problems, in the absence of uniqueness, the latter condition makes the analysis quite problematic as the measurability of* $\hat{x}_n$ *becomes an issue. In contrast, for linear problems, lack of uniqueness does not occur and measurability becomes more straightforward.*

4 Examples and Applications

4.1 Groundwater Level Prediction

A mud slide in North Vancouver caused the deaths of two people. An engineering company was commisioned to monitor the groundwater level thereafter. If groundwater rises to levels that reach between stable and unstable soil structures, residents need to be evacuated. In [24], Ramsay et al. discussed a way to predict the groundwater level based on hourly rainfall amount. After a prediction model is established, the rainfall data are not measured a posteriori but predicted beforehand based on weather forecasts.

With $R(t)$ denoting the rainfall amount hourly measured in mm and $G(t)$ the groundwater level in meters at some time $t \in [0, T]$, [24] proposed the following ODE model to establish a relation between $G(t)$ and $R(t)$:

$$\dot{G}(t) = -\beta(t)\big(G(t) - \mu(t)\big) + \alpha(t)R(t - \delta) \text{ for } t \in [0, T], \quad G(0) = G_0, \tag{4.1}$$

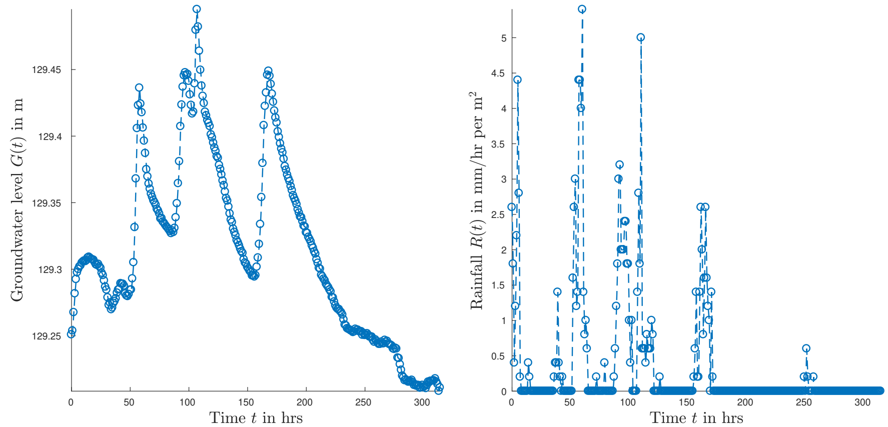


Figure 2: Groundwater data (left) and rainfall data (right)

where $\dot{G}(t) = \frac{\mathrm{d}}{\mathrm{d}t}G(t)$ denotes the derivative of G at time $t \in [0, T]$, $\beta(t)$ is the groundwater discharge rate, $\alpha(t)$ represents the recharge capacity per 1 mm rainfall, δ is a delay parameter set to 3, $\mu(t)$ is the wellhead level and and G_0 stands for the initial groundwater level. First, Ramsay et al. [24] assumed α, β, μ are constant and applied their multi-step smoothing approach to estimate these parameters. Plugging these estimates into Equation (4.1), a rather good agreement between the predicted and the measured groundwater level was obtained. The functional parameters α, β and μ were not explicitly reported, rather described as "slowly varying over time." No rigorous functional-analytic framework was employed.

Since Equation (4.1) is non-autonomous, the model can mainly be used only retrospectively. Unless α, β and μ are constant, the forecasting of $G(t)$ based on the measurements of $R(t)$ is limited to very small time horizons. To overcome this deficiency, we propose the following model of distributed delay-type:

$$\dot{G}(t) = -\gamma\big(G(t) - \mu\big) + \int_{-\delta}^{0} K(s)R(t+s)\mathrm{d}s \text{ for } t \in [0, T], \quad G(0) = G_0. \tag{4.2}$$

where γ is the discharge rate of the initial groundwater level measured with respect to the waterhead level μ, G_0 stands for the initial value of $G(t)$ and $K(s)$ represents the time-delayed recharge coefficient from the rainfall. As in [24], we let $\delta = 3$. For physical reasons, γ, μ, G_0 and $K(s)$ are nonnegative – same as $R(t)$ for all t. Additionally, it is natural to assume $K(s)$ is non-decreasing to give more weight to more recent R-values in the integral. With K being retrospectively estimated over $[0, T]$ with $T \geq \delta$, it can be used to forecast G at over future time horizons.

To put Equation (4.2) into the framework of Section 3, we use the Duhamel's principle (the so-called "variation of constant formula") to get the equivalent formulation

$$G(t) = \mu + e^{-\gamma t}(G_0 - \mu) + \int_0^t \int_{-\delta}^{0} e^{-\gamma(t-s)} K(\xi)R(s+\xi)\mathrm{d}\xi\mathrm{d}s. \tag{4.3}$$

As we will later see, by density and continuity Equation (4.3) will be extendable to a distributional version. Assuming the presence of white noise in the measurements of G, we finally obtain

$$G(t) = \mu + e^{-\gamma t}(G_0 - \mu) + \int_0^t \int_{-\delta}^{0} e^{-\gamma(t-s)} K(\xi)R(s+\xi)\mathrm{d}\xi\mathrm{d}s + \sigma\varepsilon(t), \tag{4.4}$$

where $\varepsilon(t)$ is a 1D white noise process. For the sake of simplicity, the explicit dependence on the functional parameter $R(t)$ will be suppressed in the notation used below.

Define the forward operator

$$\begin{aligned}
&\mathcal{A}\colon \mathfrak{D}(\mathcal{A}) \subset \mathscr{X} \to L^2\big((0,T)\big),\\
&\quad (G_0,\gamma,\mu,K) \mapsto \mu + e^{-\gamma\cdot}(G_0-\mu) + \int_0^{\cdot} e^{-\gamma(\cdot-s)} \int_{-\delta}^{0} K(\xi)R(s+\xi)\mathrm{d}\xi\mathrm{d}s
\end{aligned}$$

with $\mathscr{X} := \mathbb{R}^3 \times W^{1,2}\big((-\delta,0)\big)$ and

$$\mathfrak{D}(\mathcal{A}) := \big([0,\infty)\big)^3 \times \big\{K \in W^{1,2}\big((-\delta,0)\big) \,|\, K \geq 0, K' \geq 0\big\}.$$

Assuming $R \in L^2\big((0,T)\big)$, $\mathcal{A}$ is well-defined and continuous by standard delay ODE results [16].

Next, assuming $R \in W^{-1,2}\big((0,1)\big)$, we canonically extend $\mathcal{A}$ to

$$\mathcal{A}_- : \mathfrak{D}(\mathcal{A}_-) \subset \mathscr{X}_- \to W^{-1,2}\big((0,1)\big)$$

with $\mathscr{X}_- := \mathbb{R}^3 \times W^{1,2}\big((0,T)\big)$ and

$$\mathfrak{D}(\mathcal{A}_-) := \big([0,\infty)\big)^3 \times \big\{K \in W^{1,2}\big((-\delta,0)\big) \,|\, K \geq 0, K' \leq 0\big\}.$$

Using Young's and Sobolev's inequalities to estimate

$$\begin{aligned}
\Big\| \int_0^{\cdot} e^{-\gamma(\cdot-s)} \int_{-\delta}^{0} K(\xi)R(s+\xi)\mathrm{d}\xi\mathrm{d}s \Big\|_{W^{-1,2}((0,1))} &= \Big\| \int_0^{\cdot} e^{-\gamma(\cdot-s)} \int_{-\delta}^{0} K(\xi)\partial_\xi R^\circ(s+\xi)\mathrm{d}\xi\mathrm{d}s \Big\|_{W^{-1,2}((0,1))}\\
&= \Big\| \int_0^{\cdot} e^{-\gamma(\cdot-s)} \Big(-\int_{-\delta}^{0} K'(\xi)R^\circ(s+\xi)\mathrm{d}\xi\mathrm{d}s + \Big(K(0)R^\circ(s) - K(-\delta)R^\circ(s-\delta)\Big) \Big\|_{W^{-1,2}((0,1))}\\
&\leq \|K'\|_{L^2((-\delta,0))}\|R^\circ\|_{L^2((0,T))} + \big|K(-\delta)\big|\|R^\circ\|_{L^2((0,T))}\\
&\leq C\|K\|_{W^{1,2}((-\delta,0))}\|R\|_{W^{-1,2}((0,1))}
\end{aligned}$$

with $K^\circ = (\partial_t)^{-1}K = \int_{-\delta}^{\cdot} K(s)\mathrm{d}s$ and some generic constant $C > 0$, it easily follows $\mathcal{A}_-$ is well-defined and continuous in K (and R). Hence, being a composition of continuous functions, $\mathcal{A}_-$ is strongly continuous in the product topology. The remaining conditions of Theorem 3.4 can easily be verified. This yields the existence of a minimizer $(\hat{G}_0, \hat{\gamma}, \hat{\mu}, \hat{K})$ of the functional (3.5).

Hence, our methodology is applicable to the inverse problem in Equation (4.2). To illustrate its performance, we now apply our parameter estimation procedure to a particular dataset. Unfortunately, only a subset of the dataset used in [24] is available and can be found in the `CollocInfer` package of `R` in the file `NSdata.rda`. This dataset (two univariate time series of size 315) is displayed in Figure 2.

Since the last 100+ time periods did not contain any significant amount of rainfall, we used the first 100 times periods for model training purposes and the next 100 times periods for prediction, while discarding the last 115 time periods. In contrast to [24], our goal was not to "predict" $G(t)$ retrospectively, but rather estimate/recover the parameters and use them for future prediction. This is possible with our model in Equation (4.4).

The taut string estimator was subsequently applied to $\big(G(1), \ldots, G(100)\big)$ with estimated $\hat{\sigma} = 4.6884 \cdot 10^{-4}$. With the amount of additive noise being small, the role of taut string smoothing in this example is less substantial. The integral in Equation (4.3) was approximated by a first-order integral sum. Similarly, the negative Sobolev norms of a lattice function f were approximated by

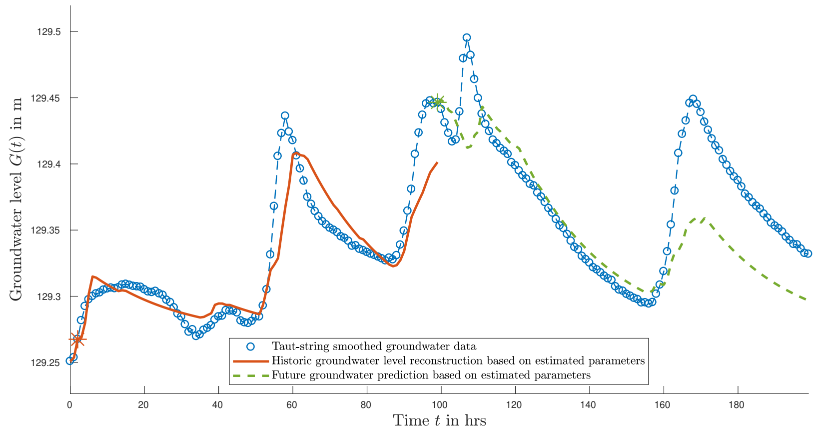


Figure 3: Smoothed groundwater level and its prediction (left) and estimated parameters (right)

first-order integral sums of the squared cumulative sums of f. Finally, the constrained optimization problem for the discretized version of functional $\mathcal{F}_n$ from Equation (3.5) with

$$\bar{x} = \left(G(1), 0, \frac{1}{100}\sum_{i=1}^{100} G(i), (0, 0, \ldots, 0)\right)$$

was solved using the constrainted optimization routine `fmincon` of `Matlab`. A full implementantion of the algorithm is available in the supplement. The blue dashed & circled curve in Figure 3 displays the taut-string smoothed groundwater level, while the orange solid curve displays the fitted groundwater lever for the first 100 times periods. As it is the case with time-dependent predictions, the quality deteriorates with time. The estimated parameters are contained in Table 1 below.

G_0	γ	μ	$K(-3)$	$K(-2)$	$K(-1)$	$K(0)$
129.2623	0.043336	129.269	$1.3561\,10^{-7}$	$5.2431\,10^{-7}$	$1.4454\,10^{-7}$	0.0053244

Table 1: Estimated parameters

The trained model was subsequently used to predict the groundwater level for the next 100 times periods based on future rainfall data. Again, while the prediction quality is originally fairly good, a significant deterioration can be observed for larger time horizons. A possible way to improve the prediction quality is to account for time-dependent waterhead levels, which, in turn, may require a sophisticated autonomous models for μ. Also, selecting an alternative model that accounts for rainfall at different locations and other factors will most likely produce predictions that better match the observed data. This is not the focus here. Rather, how to implement the taut string methodology in an inverse solution context for a real problem is our concern.

4.2 Euler-Bernoulli Beams

Consider a simply supported elastic Euler-Bernoulli beam of uniform density $\rho > 0$, uniform thickness $h > 0$ and length $L > 0$. In an equilibrium state of the beam, let $w(x)$ denote the vertical displacement of its midline at point $x \in (0, L)$. Further, let $f\colon (0, L) \to \mathbb{R}$ denote the distributed static force acting on the beam often referred to as loads. Using the equations of linear 3D elasticity (cf. [13, Chapter 1]), exploiting Euler-Bernoulli structural assumptions (cf. [2, Equations (7, 8)]) and postulating a linear Hooke's law for homogeneous isotropic media, w can be shown to satisfy the following boundary value problem

$$\begin{aligned} EIw^{(4)} &= f(x) \text{ for } x \in (0, L), \\ w(x) = w^{(2)}(x) &= 0 \text{ for } x \in \{0, L\}, \end{aligned}$$

where E is the Young's modulus, I is the second moment of area of the beam's cross-section and $f(x)$ are the distributed loads. Figure 4 illustrates an elastic beam before and after bending. We refer the reader to [22, Chapter 4] for a detailed discussion on elastic beams.

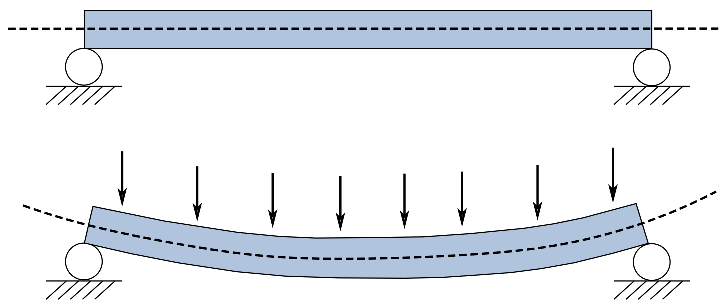


Figure 4: Elastic beam before and after bending

Consider the operator

$$\mathcal{A}\colon \mathfrak{D}(\mathcal{A}) \subset \mathscr{X} \to L^2\big((0, L)\big), \quad w \mapsto EIw^{(4)},$$

where

$$\mathscr{X} := L^2\big((0, L)\big) \quad \text{and} \quad \mathfrak{D}(\mathcal{A}) := \big\{w \in W^{4,2}\big((0, L)\big) \mid w(x) = w^{(2)}(x) = 0 \text{ for } x \in \{0, L\}\big\}.$$

The operator $\mathcal{A}$ can canonically be extended to operator

$$\mathcal{A}_- : \mathfrak{D}(\mathcal{A}_-) \equiv \mathscr{X}_- := \mathscr{V} \to \mathscr{V}', \quad w \mapsto EI\big\langle w^{(2)}, (\cdot)^{(2)}\big\rangle_{L^2((0,L))}$$

with

$$\mathscr{V} := \big\{\varphi \in W^{2,2}\big((0, L)\big) \mid \varphi(x) = 0 \text{ for } x \in \{0, L\}\big\} \text{ and } \mathscr{V}' = W^{-2,2}\big((0, L)\big)$$

equipped with the norm $\|\cdot\|_{\mathscr{V}} = (EI)^{-1/2}\big\|(\cdot)^{(2)}\big\|_{L^2((0,1))}$. With ε denoting the 1D white noise on $[0, L]$, for $\sigma > 0$, we consider the following problem for w

$$\mathcal{A}_- w = f + \sigma\varepsilon. \tag{4.5}$$

In fact, Equation (4.5) is a well-posed direct problem. Nonetheless, the techniques from Section 3 can be employed to estimate w based on observations from Equation (3.3). Indeed, exploiting standard elliptic theory, $\mathcal{A}$ and $\mathcal{A}_-$ can be shown to satisfy the assumptions of Theorems 3.2 and 3.4. Moreover, due to existence and continuity of $\mathcal{A}_-^{-1}$, the conditions of Theorem 3.5 are also satisfied. Hence, with $\hat{w}_n$ denoting the estimate from Definition 3.3, we have

$$\big\|\hat{w}_n - \mathcal{A}_-^{-1}(f)\big\|_{W^{2,2}((0,L))} = O_{\mathbf{P}}(n^{-1/2}) \text{ as } n \to \infty.$$

In particular, the convergence is uniform in $[0, L]$ due to the Sobolev imbedding theorem.

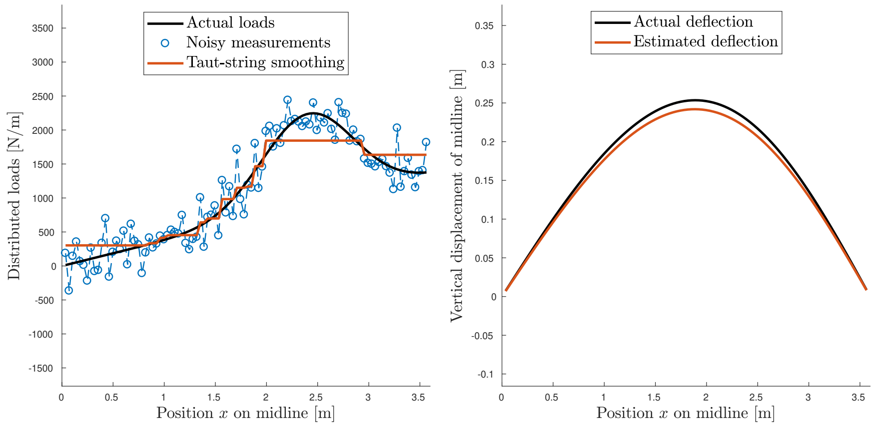


Figure 5: Estimation results for noisy measurements

Next, we present a numerical example. Consider a European wide flange steel I-beam of HE 100 A type of length $L = 3.6\,\mathrm{m}$. For this, $h = 0.096\,\mathrm{m}$, $E = 200\,\mathrm{GPa}$, $I = 5.24 \cdot 10^{-8}\,\mathrm{m}^4$. Assume a total load f of $2 \cdot 9.8 \cdot 10^3\mathrm{N}$ acts on the beam distributed as shown on the left of Figure 5. Assume the distributed loads to be measured at $n = 100$ equidistant points with the measurement errors being iid normal with mean $0\,\mathrm{N}$ and standard deviation $9.8 \cdot 10^3\,\mathrm{N}$. Applying the state recovery procedure from Definition 3.3 with $\bar{x} = 0$ to the noisy measurements, we can estimate deflection of the plate (in the negative vertical direction). Here, the second derivative is discretized using the standard centered three-point finite difference approximation, whereas the $\mathscr{V}$-norm is discretized with an intergral sum for inverse Dirichlet-Laplacian. The right-hand side of Figure 5 illustrates the actual deflection profile $w^\dagger = \mathcal{A}^{-1} f$ and its estimate $\hat{w}_n$. As one can see, $\hat{w}_n$ is in a very good correspondence with w.

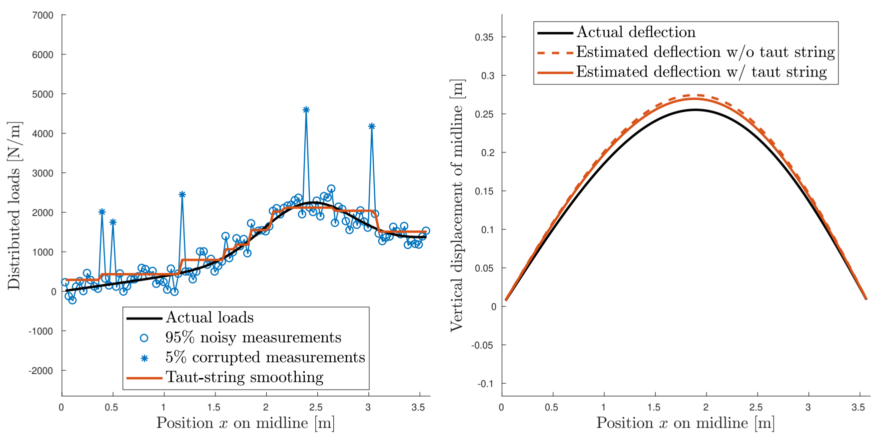


Figure 6: Estimation results for noisy measurements with 5% corruption

Now, we demonstrate the robustness of our approach due to incorporation of the taut-string-based data filtering. In the previous example, 5% of the data were randomly selected and corrupted by adding $483\,\mathrm{N}$. The plot of the corrupted data along with the taut string estimate and actual loads are displayed in Figure 6 on the left. The right-hand side of Figure 6 suggests the deflection estimate obtained using a preliminary filtering with the taut string estimator better matches the actual deflection than the estimate obtained without using the taut string smoothing.

4.3 Electric Circuit

Consider an electric circuit with the diagram displayed in Figure 7 consisting of an AC voltage source, capacitor of 1 F, two resistors of 1 Ω each and an inductor of 1 F. Assuming the voltage

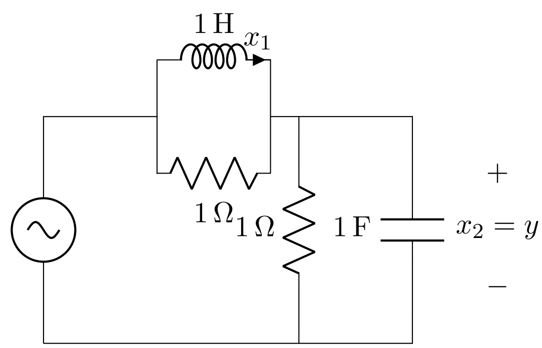


Figure 7: Electric circuit

source is inactive and a measurement at the capacitor can be obtained, the system is governed by the following system of ordinary differential equations over a time horizon $[0, T]$:

$$\begin{aligned} &\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \text{ for } t \in [0,T], \quad \mathbf{x}(0) = \mathbf{x}_0, \\ &y(t) = \mathbf{C}\mathbf{x}(t) \end{aligned} \tag{4.6}$$

with

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0 & 1 \end{pmatrix},$$

where $\mathbf{x}(t) = \big(x_1(t), x_2(t)\big)^{\mathrm{T}}$, $x_1(t)$ is the voltage at the inductor and $x_2(t)$ is the voltage at the capacitor. Given an observation $y(t)$, the goal is to reconstruct the state $\mathbf{x}(t)$ in Equation (4.6). In the absense of noise, this is a standard linear observability problem [31, Chapter 1].

Assuming for a moment $y(t)$ was in $L^2\big((0,T)\big)$ and computing the observability matrix

$$\mathcal{O} = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix},$$

the Kálmán's rank criterion suggests the sytem (4.6) is (continuously) observable. Moreover, the initial state can uniquely and continuously be reconstructed as

$$\mathbf{x}(0) = \Big(\int_0^T e^{\mathbf{A}^{\mathrm{T}}t}\mathbf{C}^{\mathrm{T}}\mathbf{C}e^{t\mathbf{A}}\mathrm{d}t\Big)^{-1}\int_0^T e^{\mathbf{A}^{\mathrm{T}}t}\mathbf{C}^{\mathrm{T}}y(t)\mathrm{d}t, \tag{4.7}$$

where the observability Gramian $\int_0^T e^{t\mathbf{A}^{\mathrm{T}}}\mathbf{C}^{\mathrm{T}}\mathbf{C}e^{t\mathbf{A}}\mathrm{d}t$ is an invertible matrix. Thus,

$$\mathbf{x}(t) = e^{\mathbf{A}t}\Big(\int_0^T e^{\mathbf{A}^{\mathrm{T}}t}\mathbf{C}^{\mathrm{T}}\mathbf{C}e^{\mathbf{A}t}\mathrm{d}t\Big)^{-1}\int_0^T e^{\mathbf{A}^{\mathrm{T}}t}\mathbf{C}^{\mathrm{T}}y(t)\mathrm{d}t \tag{4.8}$$

Eliminating $\mathbf{x}(t)$, Equation (4.6) can be cast into equivalent form:

$$y(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}_0 \text{ for } t \in [0, T]. \tag{4.9}$$

Equation (4.9) can easily be put into the framework of Section 3 by letting

$$\mathcal{A} \colon \mathfrak{D}(\mathcal{A}) \equiv \mathscr{X} \to L^2\big((0, T)\big), \quad \mathbf{x}_0 \mapsto e^{\mathbf{A}\cdot}\mathbf{x}_0$$

with $\mathscr{X} := \mathbb{R}^2$. The operator $\mathcal{A}$ can be trivially extended to

$$\mathcal{A}_- \colon \mathfrak{D}(\mathcal{A}) \equiv \mathscr{X}_- \to W^{-1,2}\big((0, T)\big)$$

with $\mathscr{X}_- = \mathbb{R}^2$.

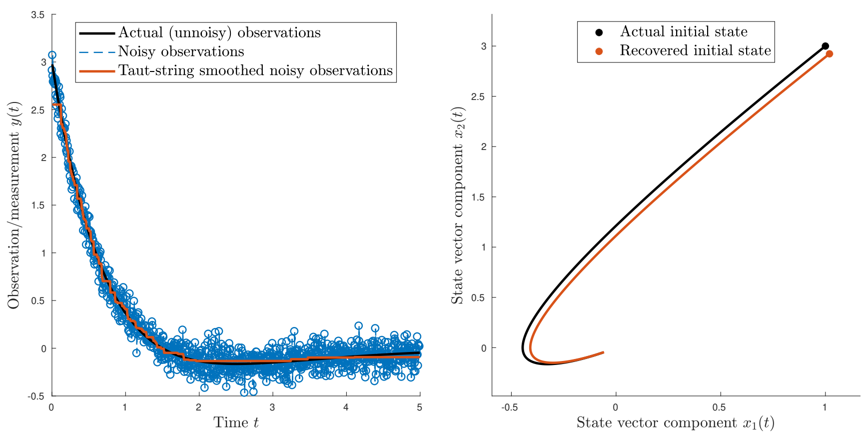


Figure 8: Observability problem for an electric circuit

The noisy version of Equation (4.6) becomes then

$$\mathcal{A}_-(\mathbf{x}^0) = y + \sigma\varepsilon \equiv y_{\text{obs}}, \tag{4.10}$$

where ε is the univariate standard white noise and $\sigma > 0$.

Selecting $\mathbf{x}_0 = (1, 3)^{\mathrm{T}}$ and $y(t) = \big(\mathcal{A}(\mathbf{x}_0)\big)(t) + 0.1\varepsilon(t)$ sampled on an equidistant lattice of size $1,000$ over $[0, 5]$ a numerical implementation of the procedure from Definition 3.3 was obtained by using the matrix exponent to solve for $\mathbf{x}(t)$ and employing a first-order Riemann sum to compute the $W^{-1,2}\big((0, 1)\big)$-norm of $y(t)$. The results are displayed in Figure 8. The estimated value of $\hat{\mathbf{x}}_0$ of $\mathbf{x}_0$ was computed to be $(1.0195, 2.9228)^{\mathrm{T}}$.

Remark 4.1. *The stochastic observation problem* (4.6) *is also solvable via continuous-time Kálmán filter, which is known to have certain L^2-optimality properties. In the presence of outliers (i.e., when the noise process departs from the iid Gaussianity), this optimality can be compromised. Due to a certain degree of robustness originating from the application of the taut string smoothing, we expect our approach to be generally more robust than Kálmán's filter – at the price of reduced statistical efficiency in the absence of outliers.*

5 Conclusions

We presented a Banach space new two-step regularization approach to inverse problems with white noise from Equation (1.2). Our method incorporates a preliminary data smoothing step with the taut string estimator and a subsequent minimization of a discrepancy functional with a Tikhonov regularization. We gave a detailed convergence study for the taut string estimator. In Theorem 2.4, we proved the "parametric" convergence rate of $n^{-1/2}$ in the Sobolev space $W^{-1,p}\big((0,1)\big)$ under rather weak regularity assumptions of the underlying conditional expectation. As a corollary, we obtained an asymptotical confidence ball for the unknown conditional expectation. Further, using interpolation techniques, we extended this result to negative fractional Sobolev spaces at the price of a reduced convergence rate.

Under appropriate assumptions on the forward operator and its continuation, Theorems 3.4 and 3.5 prove the overall regularization procedure is feasible and preserves the convergence rate of $n^{-1/2}$ in the norm of extended parameter/state space where n is the design size. These results are developped in Section 3. In contrast to other solution methodologies, our approach does not require noise to be bounded or vanish asymptotically. Finally, three examples are given that illustrate applicability of the regularization technique.

The set of `Matlab` codes used to produce all examples from Section 4 are available online.

Acknowledgments

Financial support by the Deutsche Forschungsgemeinschaft through the Zukunftskolleg at the University of Konstanz (Germany), ERC-CZ Project LL1202 "MORE" at Charles University in Prague (Czech Republic) and CRC 1173 at Karlsruhe Institute of Technology (Germany) are gratefully acknowledged. Helpful discussions with Professor J. Marcus Jobe (Miami University, Oxford, OH) on practical examples and applications are also greatly appreciated.

A The Double Obstacle Problem in 1D

In this section, we briefly summarize solution theory for the double obstacle problem in 1D. For a detailed treatise, we refer the reader to [23] and [25, Chapter 7].

For functions $\varphi, \psi\colon [0,1] \to [-\infty, \infty]$ with $\varphi \leq \psi$ in $[0,1]$ denoting the lower and the upper obstacle and a linear function $\theta\colon [0,1] \to \mathbb{R}$ satisfying $\varphi(0) \leq \theta(0) \leq \psi(0)$ and $\varphi(1) \leq \theta(1) \leq \psi(1)$, we define the "admissible set"

$$\mathcal{K}^{\theta}_{\varphi,\psi} := \big\{u \in C^0([0,1]) \,|\, u(0) = \theta(0), u(1) = \theta(1), \varphi \leq u \leq \psi \text{ a.e. in } (0,1)\big\}. \tag{A.1}$$

As a result, θ satisfies the differential equations

$$\left(\frac{\theta^{(1)}}{\sqrt{1 + (\theta^{(1)})^2}}\right)^{(1)} = 0 \text{ and } \theta^{(2)} = 0 \text{ in } (0,1).$$

The classical double obstacle problem reads as follows: Find a function u^* with the smallest arc length over all elements of $\mathcal{K}^{\theta}_{\varphi,\psi} \cap W^{1,\infty}((0,1))$, i.e.,

$$u^* = \arg\min\Big\{\mathcal{L}(u) \,|\, u \in \mathcal{K}^{\theta}_{\varphi,\psi} \cap W^{1,\infty}((0,1))\Big\}, \tag{A.2}$$

where the arc length functional is given as

$$\mathcal{L}(u) := \int_0^1 \sqrt{1 + (u^{(1)}(t))^2}\,\mathrm{d}t. \tag{A.3}$$

Under appropriate regularity assumptions, the minimization problem in Equation (A.2) is equivalent with the following variational inequality (cf. [25, Chapter 1.5]):

$$\text{Find } u^* \in \mathcal{K}^\theta_{\varphi,\psi} \text{ such that } \int_0^1 \frac{\big(u^{(1)} - \theta^{(1)}\big)\big(v^{(1)} - u^{(1)}\big)}{\sqrt{1 + (u^{(1)})^2}}\mathrm{d}t \geq 0 \text{ for any } v \in \mathcal{K}^\theta_{\varphi,\psi} \cap W^{1,\infty}((0,1)). \tag{A.4}$$

The condition associated with (A.4) can be viewed as a generalized Karush-Kuhn-Tucker condition for the minimization problem in Equation (A.2). Note that the nonlinear form defined by the integral in Equation (A.4) is the Gâteaux derivative of functional $\mathcal{L}$ at point $u - \theta$ in the direction $v - u$.

In this classical situation, the following well-posedness result of Theorem A.1 below can be shown by adapting techniques from [25, Chapter 7.2] developed in the unilateral case. See also [23] for a study of the linearized double obstacle problem in a Hilbert space situation.

Theorem A.1. *Let the obstacle φ and ψ be such that*

$$\mathcal{K}^\theta_{\varphi,\psi} \cap W^{1,\infty}((0,1)) \neq \emptyset,$$

then the variational inequality in Equation (A.4) possesses a unique solution $u^* \in \mathcal{K}^\theta_{\varphi,\psi} \cap W^{1,\infty}((0,1))$, which is also the unique minimizer of $\mathcal{L}$ in $\mathcal{K}^\theta_{\varphi,\psi} \cap W^{1,\infty}((0,1))$.

In case the obstacles φ, ψ are less regular, e.g., merely continuous and not weakly differentiable with p-integrable derivatives, a weaker solution concept needs to be adopted.

Definition A.2. *Let $\varphi, \psi \in C^0([0,1])$. A function $w \in \mathcal{K}^\theta_{\varphi,\psi}$ is called a generalized solution to Equation (A.4) if it satisfies*

$$-\int_0^1 \left(\frac{v^{(1)}}{\sqrt{1 + (v^{(1)})^2}}\right)^{(1)} (v - u)\,\mathrm{d}t \geq 0 \textit{ for any } v \in \mathcal{K}^\theta_{\varphi,\psi} \cap C^2([0,1]). \tag{A.5}$$

By virtue of Minty's lemma (cf. [25, Chapter 7.3]), any function $u \in \mathcal{K}^\theta_{\varphi,\psi} \cap W^{1,\infty}((0,1))$ satisfying Equation (A.4) also satisfies (A.5).

Let $\mathscr{S}$ denote the solution operator sending each triple (φ, ψ, θ) with $\mathcal{K}^\theta_{\varphi,\psi}$ to the unique generalized solution u.

Theorem A.3. *Let $\varphi, \psi \in C^0([0,1])$ such that $\mathcal{K}^\theta_{\varphi,\psi} \neq \emptyset$, then the following assertions hold true.*

a) There exists a unique generalized solution $u \in \mathcal{K}^\theta_{\varphi,\psi}$ to Equation (A.4).

b) Let $s \in (0,1)$, $p \in [1,\infty]$. If $\varphi, \psi \in W^{s,p}((0,1))$, then $u \in W^{s,p}((0,1))$.

c) The solution operator $\mathscr{S}$ is nonexpansive on L^p for any $p \in [1,\infty]$, i.e.,

$$\big\|\mathscr{S}(\varphi,\psi,\theta) - \mathscr{S}(\tilde\varphi,\tilde\psi,\tilde\theta)\big\|_{L^p((0,1))} \leq \max\big\{\|\varphi - \tilde\varphi\|_{L^p((0,1))}, \|\psi - \tilde\psi\|_{L^p((0,1))}\big\}$$

for all $\varphi, \tilde\varphi, \psi, \tilde\psi \in C^0([0,1])$ with $\mathcal{K}^\theta_{\varphi,\psi} \neq \emptyset$ and $\mathcal{K}^{\tilde\theta}_{\tilde\varphi,\tilde\psi} \neq \emptyset$.

References

[1] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Academic Press, New York-London, 2nd edition, 2003.

[2] S. A. Ambartsumian, M. V. Belubekian, and M. M. Minassian. On the problem of vibrations of non-linear-elastic electroconductive plates in transverse and longitudinal magnetic fields. *International Journal of Non-Linear Mechanics*, 19(2):141–149, 1984.

[3] R. S. Anderssen. *Inverse Problems, Oberwolfach, 1986*, chapter The linear functional strategy for improperly posed problems, pages 11–30. Birkhäuser, Basel, 1986.

[4] W. Arendt, C. J. K. Batty, M. Hieber, and F. Neubrander. *Vector-valued Laplace Transforms and Cauchy Problems*, volume 96 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel – Boston – Berlin, 2001.

[5] R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter Estimation and Inverse Problems*. Academic Press, Amsterdam etc., 2nd edition, 2011.

[6] G. Bachman and L. Narici. *Functional Analysis*. Dover Publications, Mineola, NY, 2000.

[7] Z. Brzeźniak. On Sobolev and Besov spaces regularity of Brownian paths. *Stochastics and Stochastics Reports*, 56:1–15, 1996.

[8] P. L. Davies. Data features. *Statistica Neerlandica*, 49(2):189–245, 1995.

[9] P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Annals of Statistics*, pages 1–48, 2001.

[10] D. L. Donoho. Statistical estimation and optimal recovery. *Annals of Statistics*, 22.

[11] J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. CRC Press, 1996.

[12] P. Fritz and M. Hairer. *A Course on Rough Paths*. Universitext. Springer, Cham etc., 2014.

[13] Y. B. Fu and R. W. Ogden, editors. *Nonlinear Elasticity: Theory and Applications*, volume 283 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, UK, 2001.

[14] M. A. Goldberg. A method of adjoints for solving some ill-posed equations of the first kind. *Applied Mathematics and Computation*, 5.

[15] M. Grasmair. The equivalence of the taut string algorithm and BV-regularization. *Journal of Mathematical Imaging and Vision*, 27:59–66, 2007.

[16] J. K. Hale and S. M. V. Lunel. *Introduction to Functional Differential Equations*. Springer, New York, 1993.

[17] W. K. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer Series in Statistics, 2004.

[18] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Birkhäuser Advanced Texts. Dover Publications, Mineola, New York, 2007.

[19] J.-Y. Koo. Optimal rates of convergence for nonparametric statistical inverse problems. *Annals of Statistics*, 21(2):590–599, 1993.

[20] A. K. Louis. Approximate inverse for linear and some nonlinear problems. *Inverse Problems*, 12.

[21] P. Mathé and S. V. Pereverzev. Direct estimation of linear functionals from indirect noisy obervations. *Journal of Complexity*, 18.

[22] W. M. C. McKenzie. *Examples in Structural Analysis*. Taylor & Francis, London and New York, 2006.

[23] A. Olek and K. Szczepaniak. Continuous dependence on obstacles in double global obstacle problems. *Annales Academiæ Scientarum Fennicæ*, 28:89–97, 2003.

[24] J. Ramsay, D. Campbell, J. Cao, and G. Hooker. Parameter estimation for differential equations: A generalized smoothing approach. In *Parameter Estimation for Dynamical Systems*, Eindhoven, 2009. EURANDOM.

[25] J.-F. Rodrigues. *Obstacle Problems in Mathematical Physics*, volume 134 of *North-Holland Mathematics Studies*. Elsevier Science Publishers B.V., North-Holland – Amsterdam, New York, Oxford, Tokyo, 1987.

[26] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization Methods in Banach Spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter, Berlin/Boston, 2012.

[27] J. Simon. Sobolev, Besov and Nikolskii fractional spaces: Imbeddings and comparisons for vector valued spaces on an interval. *Annali di Matematica Pura ed Applicata*, 157(1):117–148, 1990.

[28] M. Tucsnak and G. Weiss. *Observation and Control for Operator Semigroups*. Birkhäuser Advanced Texts. Birkhäuser, Basel – Boston – Berlin, 2009.

[29] L. Wasserman. *All of Nonparametric Statistics*. Springer Text in Statistics. Springer, 2006.

[30] E. Weinan and Y. Wang. Optimal convergence rate of the universal estimation error. *Research in the Mathematical Sciences*, 4(2):1–14, 2017.

[31] J. Zabczyk. *Mathematical Control Theory: An Introduction*. Systems & Control: Foundations & Applications. Birkhäuser, Boston, 2nd (with corrections) edition, 1995.