AIMS *Mathematics*

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(11): 26991–27008.
DOI:10.3934/math.20231381
Received: 21 July 2023
Revised: 20 August 2023
Accepted: 28 August 2023
Published: 22 September 2023

Research article

A weighted online regularization for a fully nonparametric model with heteroscedasticity

Lei Hu*

Shenzhen Institute of Information Technology, Shenzhen 518172, China

* **Correspondence:** Email: luoshengying@163.sufe.edu.cn.

Abstract: In this paper, combining B-spline function and Tikhonov regularization, we propose an online identification approach for reconstructing a smooth function and its derivative from scattered data with heteroscedasticity. Our methodology offers the unique advantage of enabling real-time updates based on new input data, eliminating the reliance on historical information. First, to address the challenge of heteroscedasticity and computation cost, we employ weight coefficients along with a judiciously chosen set of knots for interpolation. Second, a reasonable approach is provided to select weight coefficients and the regularization parameter in objective functional. Finally, We substantiate the efficacy of our approach through a numerical example and demonstrate its applicability in solving inverse problems. It is worth mentioning that the algorithm not only ensures the calculation efficiency, but also trades the data accuracy through the data volume.

Keywords: B-spline function; heteroscedasticity; nonparametric identification; weight coefficients
Mathematics Subject Classification: 41A15, 62G05, 93E24

1. Introduction

The underlying model of this article can be formulated quite easily: given some real valued variables x and y fulfill the heteroscedastic model:

$$y = f(x) + \sigma(x)\varepsilon, \tag{1.1}$$

where $f(x)$ belongs to Sobolev space $W^{2,2}(0, 1)$, variance function $\sigma(x)$ is a positive function defined on $(0, +\infty)$ and the random error term ε is independent of x and satisfies $E(\varepsilon) = 0$ and $Var(\varepsilon) = 1$.

Such a problem is widely applied in various fields such as Computerized Tomography (CT) and the inverse problem of option pricing (IPOP) in [1–3]. Moreover, this is a classical ill-posed problem. There are plenty of regularization methods for this ill-posed problem in one dimension or higher dimensions [4, 5]. Researchers [6–9] have proposed some statistical methods to solve this ill-posed problem. Zhang [10] employs a relatively small number of knots for interpolation to reduce computation cost. These classical approaches are based on the equal variance of the error term and, when N is large enough, needs large amount of calculating. However, the variance $\sigma(x)$ in (1.1) is typically unequal, a phenomenon known as heteroscedasticity [11, 12].

Heteroscedasticity, which is an econometric term, is the variances of perturbations $\sigma(x)$ in the model which are not exactly equal [13]. Heteroscedasticity will lead to large errors in the model. In order to reduce the error caused by heteroscedasticity, the principal methods are mainly divided into three parts. The most common method is to take the logarithm of the data [14]. However, not all data must be logarithmic. The second method is Robust Standard Error Regression, which is the most popular and effective treatment method at present [15, 16]. The main idea is to modify the objective function in the classical least squares regression, which is very sensitive to outliers. Robust method is only applicable to heteroscedasticity and independent observation. Another method is FGLS regression [17]. For points with larger residual value, the smaller weight is given to solve the heteroscedasticity problem. It is still the least square method in essence. However, the convergence speed of this method is slow, that is, the deviation of limited samples will be large.

B-spline functions are crucial elements in various fields, especially in computer graphics, computer-aided design (CAD), and numerical analysis [18–20]. They play a significant role in curve and surface representation, interpolation, approximation, and modeling [21–23]. B-spline functions offer a versatile mathematical framework for representing complex shapes and data, enabling efficient and accurate solutions in various industries and scientific disciplines. Their ability to balance smoothness, flexibility, and local control makes them a cornerstone of computational modeling and design processes. Among them, cubic splines are the most representative. Cubic B-spline functions strike a balance between simplicity and expressiveness. They are more flexible than linear or quadratic B-splines, allowing for smoother curves and surfaces, while still being relatively straightforward to manipulate.

In this paper, based on B-splines function and FGLS regression, we propose a weighted online regularization method which apply weight coefficients and a relatively small number of knots for interpolation to reduce the effect of heteroscedasticity and computation cost, and exchange the amount of data for the accuracy of reconstruction. Moreover, this algorithm has the characteristics of smaller memory occupation and can process online data, which means that when new data is added, there is no need to re-process the processed data.

The paper is organized as follows. In Section 2, we introduce the issue which is to be studied and propose an online reconstruction algorithm. Error estimation is carried out in Section 3. Section 4 provides a principle of optimal selection for weight coefficients and the regularization parameter in objective function. The performance of the proposed algorithm is illustrated in Section 5. In Section 6, we present two applications of our reconstruction algorithm in inverse problems. Section 7 concludes main results.

2. Problem statement and algorithm design

In this section, we consider the following problem: for a lager positive integer N, we try to reconstruct function $f(x)$ and its derivative $f'(x)$ in the model (1.1) from observation data $\{(x_i, y_i)\}_{i=1}^N$

and function $\sigma(x)$.

2.1. Regularized solution

First, we define the Tikhonov objective functional as follows:

$$J(g) = \frac{1}{N}\sum_{i=1}^{N} w_i(g(x_i) - y_i)^2 + \alpha\|g''(x)\|^2_{L^2(0,1)}, \tag{2.1}$$

where $g(x) \in W^{2,2}(0,1)$, $\{w_i\}_{i=1}^N$ are weight coefficients and α is a regularization parameter. The minimization problem is reformulated as follows:

$$f_N = \arg\min_g J(g) = \arg\min_g \left(\frac{1}{N}\sum_{i=1}^{N} w_i(g(x_i) - y_i)^2 + \alpha\|g''(x)\|^2_{L^2(0,1)}\right). \tag{2.2}$$

f_N is defined as the approximated solution of $f(x)$ in (1.1).

In this paper, we reconstruct not only function $f(x)$ but also its derivative $f'(x)$. This is a classical ill-posed problem. It is computationally costly to solve the problem with local regression methods such as kernel regression and local polynomial regression. This problem is usually solved by regularization method and spline functions. Moreover, B-spline basis functions have compact support which makes it possible to speed up calculations. Thus, we choose B-spline function.

Then we solve suitable $g(x)$ in the finite dimensional function space of cubic B-spline function instead of the infinite dimensional space. Inspired by the method in [25], we just select some equidistant points not sample points as interpolation node.

Let M be a positive integer and mesh size $d = M^{-1}$. Equidistant knots $\{p_j\}_{j=0}^M$ are defined as

$$p_j = jd, \quad j = 1, 2, \cdots, M.$$

The vector space of all cubic B-spline functions with knots $\{p_j\}_{j=0}^M$ is called space $\mathcal{V}_m$. Assume function $g(x)$ belongs to space $\mathcal{V}_m$, $g(x)$ can be written as

$$g(x) = \sum_{j=-1}^{M+1} \lambda_j \phi_j(x), \tag{2.3}$$

where $\{\lambda_j\}_{j=-1}^{M+1}$ are constants and $\phi_j(x) = \phi(\frac{x-p_j}{d})$. $\phi(x)$ is standard cubic B-spline function defined in [24].

Denote the weight coefficient $W = diag(\sqrt{w_1}, \sqrt{w_2}, \cdots, \sqrt{w_N})$, the noisy sample $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)^T$, the function parameter $\boldsymbol{\lambda} = (\lambda_{-1}, \lambda_0, \cdots, \lambda_{M+1})^T$ and the row vector $H_x = (\phi_{-1}(x), \phi_0(x), \cdots, \phi_{M+1}(x))$. Through the Eq (2.3) and the definition above, the Eq (2.2) can be rewritten as

$$\boldsymbol{\lambda}_N = \arg\min_{\boldsymbol{\lambda}} J_1(\boldsymbol{\lambda}) := \arg\min_{\boldsymbol{\lambda}} \left(\frac{1}{N}(WH\boldsymbol{\lambda} - W\boldsymbol{y})^T(WH\boldsymbol{\lambda} - W\boldsymbol{y}) + \alpha\boldsymbol{\lambda}^T P\boldsymbol{\lambda}\right), \tag{2.4}$$

where $H := (H_{x_1}, H_{x_2}, \cdots, H_{x_N})^T$ and $P \in \mathbb{R}^{(M+3)\times(M+3)}$ is given by

$$P = (p_{ij})_{i,j=-1}^{M+1}, \quad p_{ij} = \int_0^1 \phi_i''(x)\phi_j''(x)dx.$$

When $\boldsymbol{\lambda}_N$ is known, the approximated solution $f_N(x) = H_x\boldsymbol{\lambda}_N$.

Theorem 1. *Suppose $N > 2$, and the observation points $\{x_i\}_{i=1}^N$ are not identical. Then, the minimization problem* (2.4) *has a unique minimizer $\boldsymbol{\lambda}_N$ which satisfies:*

$$\left(\frac{1}{N}H^T W^T W H + \alpha P\right)\boldsymbol{\lambda}_N = \frac{1}{N}H^T W^T W\boldsymbol{y}. \tag{2.5}$$

Proof. Since $J_1(\boldsymbol{\lambda})$ in (2.4) is a quadratic form with respect to $\boldsymbol{\lambda}$, the derivative of $J_1(\boldsymbol{\lambda})$ takes zero only at $\boldsymbol{\lambda} = \boldsymbol{\lambda}_N$, that is

$$\left(\frac{1}{N}H^T W^T W H + \alpha P\right)\boldsymbol{\lambda}_N = \frac{1}{N}H^T W^T W\boldsymbol{y}.$$

Moreover, P is positive definite and $H^T W^T W H$ is positive semidefinite. Thus, $\boldsymbol{\lambda}_N$ is unique. □

2.2. Algorithm

Through the theorem above, we propose a online Tikhonov regularization algorithm in Algorithm 1. In "Require" sentence, we input the observation data $\{(x_i, y_i)\}_{i=1}^N$ and the necessary parameters–the number of knots M, weight coefficients W in (4.6) and regularization parameter α in Table 1. The algorithm proceeds from Line 1 to Line 8. In Line 1–2, we initialize matrix $A_0 = 0 \in \mathbb{R}^{(M+3)\times(M+3)}$ and $b_0 = 0 \in \mathbb{R}$ and generate the matrix $P \in \mathbb{R}^{(M+3)\times(M+3)}$. Line 3–6 provides a method to update A and b in (2.5) according to the newly entered data. Through solving linear system in Line 7, we have $\boldsymbol{\lambda}_N$ and the approximated solution f_N satisfies $f_N(x) = H_x\boldsymbol{\lambda}_N$.

Algorithm 1 The Online Tikhonov regularization.

Require: The number of knots M, mesh size $d = 1/M$, the number of sample N, the observation data $\{(x_i, y_i)\}_{i=1}^N$, weight coefficients W and regularization parameter α.
Ensure: The approximated solution $f_N \in \mathcal{V}_m$.

1: Initialize $A_0 = 0 \in \mathbb{R}^{(M+3)\times(M+3)}$ and $b_0 = 0 \in \mathbb{R}$;
2: Generate the matrix $P \in \mathbb{R}^{(M+3)\times(M+3)}$

$$P = (p_{ij})_{i,j=-1}^{M+1}, \quad p_{ij} = \int_0^1 \phi_i''(x)\phi_j''(x)dx;$$

3: **for** $i \leftarrow 1, 2, \cdots, N$ **do**
4: Generate row vector $H_{x_i} = (\phi_{-1}(x_i), \phi_0(x_i), \cdots, \phi_{M+1}(x_i))$;
5: Upgrade A_i by $A_i \leftarrow \frac{i-1}{i}A_{i-1} + \frac{w_i^2}{i}H_{x_i}^T H_{x_i}$;
6: Upgrade b_i by $b_i \leftarrow \frac{i-1}{i}b_{i-1} + \frac{w_i^2}{i}H_{x_i}^T y_i$;
7: Solve linear system $(\alpha P + A_N)\boldsymbol{\lambda}_N = b_N$ for $\boldsymbol{\lambda}_N$;
8: **return** $f_N(x) = H_x\boldsymbol{\lambda}_N$.

Table 1. Parameter selection.

Parameter	M	α	w_i
Value	$N^{\frac{1}{5}}$	$\sigma_*^2 N^{-\frac{4}{5}}$	$1/(k + \sigma(x_i))^2$

Remark 1. *When a new data is inputed, one just run the algorithm from Line 4 to Line 6. The computational complexity at Line 4, 5 and 6 are* $\mathcal{O}(1)$ *and the computational complexity at Line 7 is* $\mathcal{O}(M)$. *Moreover, the total data storage of Algorithm 1 is* $\mathcal{O}(M)$.

Weight coefficients W are introduced to reduce the effect of heteroscedasticity. The selection of weight coefficients W will be thoroughly discussed in Section 3.

Note that, N continues to increase with the influx of new data, when N becomes so large that $M \ll N^{\frac{1}{5}}$, we need to increase M to $M \approx N^{\frac{1}{5}}$ and restart the algorithm.

3. Error estimation

In this chapter, we analyze the reconstruction error of function $f(x)$. Some assumptions are necessary before proving. Suppose that the matrix P is positive definited. Without losing generality, it is advisable to set the domain of variable x to $[0, 1]$. The weighting coefficient matrix W can be redefined as

$$W = diag(\sqrt{w_1}, \sqrt{w_2}, \cdots, \sqrt{w_M})^T.$$

Make w_{min} and w_{max} are the minimum and maximum values of parameters $\{w_j\}_{j=1}^N$ respectively. Then for the observation data $\{(x_i, y_i)\}_{i=1}^N$, y_i can be expressed as

$$y_j = f(x_j) + \sigma(x_j)\varepsilon_j, \quad j = 1, 2, \cdots, N,$$

where ε_j are independent with expectation 0 and variance 1.

Recording error function as

$$e_N(x) = f_N(x) - f(x).$$

Next, consider the influence of regularization function on function itself and random noise.

When the data is divided into deterministic part and random part, the reconstruction result can also be divided into two parts

$$f_N = f_{abs} + f_{noise},$$

where the deterministic part f_{abs} satisfies

$$f_{abs} = \arg\min_{g \in \mathcal{V}_m} J(g; \alpha, \boldsymbol{y} - \boldsymbol{\varepsilon}),$$

here $\boldsymbol{\varepsilon}$ is a column vector of the observed noise

$$\boldsymbol{\varepsilon} = (\sigma_1\varepsilon_1, \sigma_2\varepsilon_2, \cdots, \sigma_N\varepsilon_N)^T \in \mathbb{R}^N,$$

let $\sigma_*^2 = \max_{x_j} \sigma^2(x_j)$. And the random part f_{noise} satisfies

$$f_{noise} = \arg\min_{g \in \mathcal{V}_m} J(g; \alpha, \boldsymbol{\varepsilon}).$$

Then, the error function can be rewritten as

$$e_N(x) = (f_{abs}(x) - f(x)) + f_{noise}(x) := e_{abs}(x) + f_{noise}(x).$$

Two important lemmas are introduced before error estimation.

Firstly, the indicator function is defined to describe the distribution of observation points.

Definition 1. *Divide the interval* $[0,1]$ *into M disjoint subspaces* I_i, $1 \leq i \leq M$,

$$I_1 = [0, d], \quad I_i = (id - d, id],$$

where $d = 1/M$ is the grid spacing. There are N_i observation data on each subspace I_i, $1 \leq i \leq M$. Define the indicator function $\rho_M(x)$ as

$$\rho_N(x) = \begin{cases} \rho_{N,i} = N_i(Md)^{-1}, & z \in I_i, \\ 0, & x \notin [0,1]. \end{cases} \tag{3.1}$$

Lemma 1. *([10], Lemma 3.3) Assume that the indicator function* $\rho_M(x)$ *has an upper bound* ρ_u *in the interval* $[0,1]$,

$$\sup_{x\in[0,1]} \rho_N(x) \leq \rho_u.$$

Then, for the first order continuous differentiable function $f(x)$, $x \in [0,1]$, the mean square sum of values at all observation points satisfies:

$$\frac{1}{N}\sum_{j=1}^{N} f^2(x_j) \leq 2\rho_u \left(\|f\|^2_{L(0,1)} + d^2 \|f'\|^2_{L^2(0,1)} \right).$$

Next, the error estimation of cubic spline interpolation is introduced.

Lemma 2. *([24], Theorem 1.56) Assume the objective function* $f(x) \in H^2(a,b)$, $x_0 = a$, $x_M = b$, *the grid spacing is d. Let* s_f *be a cubic spline interpolation function of f with natural or fixed boundary conditions, then the interpolation error can be estimated as*

$$\|s_f - f\|_{L^2(a,b)} \leq \frac{d^2}{4}\|s''_f - f''\|_{L^2(a,b)} \leq \frac{d^2}{4}\|f''\|_{L^2(a,b)},$$

$$\|s'_f - f'\|_{L^2(a,b)} \leq \frac{d}{\sqrt{2}}\|s''_f - f''\|_{L^2(a,b)} \leq \frac{d}{\sqrt{2}}\|f''\|_{L^2(a,b)}.$$

In addition, if s_f has natural boundary conditions, then the relation between the second derivatives of f and s_f is

$$\|s''_f\|^2_{L^2(a,b)} + \|s''_f - f''\|^2_{L^2(a,b)} = \|f''\|^2_{L^2(a,b)}.$$

Now, the error estimate of the deterministic part is analyzed.

Theorem 2. *Assume that* ρ_M *is an indicator function on the interval* $[0,1]$ *and has an upper bound* ρ_u. $f(x) \in H^2(0,1)$. *Then, the mean square sum of* $e_{abs}(x)$ *at the observation points is estimated as follows:*

$$\|e_{abs}\|^2_{L^2(0,1)} = \frac{1}{N}\sum_{j=1}^{N}(f_{abs}(x_j) - f(x_j))^2 \leq \frac{9}{8}\frac{w_{max}}{w_{min}}\rho_u d^4 \|f''\|^2_{L^2(0,1)} + \frac{\alpha}{w_{min}}\|f''\|^2_{L^2(0,1)}.$$

Proof. The deterministic part. Note

$$E_N = \frac{1}{N}\sum_{j=1}^{N}(s_f(x_j) - f(x_j))^2,$$

where $s_f(x)$ is a cubic spline interpolation function satisfies natural or fixed boundary conditions on the interval $[0, 1]$. By Lemmas 1 and 2, we have

$$\begin{aligned} E_N &\le 2\rho_u\left(\|f - s_f\|^2_{L^2(0,1)} + d^2\|f' - s'_f\|^2_{L^2(0,1)}\right) \\ &\le 2\rho_u\left(\frac{d^4}{16}\|f''\|^2_{L^2(0,1)} + \frac{d^4}{2}\|f''\|^2_{L^2(0,1)}\right) \\ &\le \frac{9}{8}\rho_u d^4\|f''\|^2_{L^2(0,1)}. \end{aligned} \tag{3.2}$$

And

$$\frac{1}{N}\sum_{j=1}^{N}(f_{abs}(x_j) - f(x_j))^2 \le \frac{1}{w_{min}}\left(\frac{1}{N}\sum_{j=1}^{N}w_j(f_{abs}(x_j) - f(x_j))^2 + \alpha\|f''_{abs}\|^2_{L^2(0,1)}\right). \tag{3.3}$$

For f_{abs} is the smallest element of the weighted regular functional $J(\cdot;\alpha,\boldsymbol{y}-\boldsymbol{\varepsilon})$,

$$\frac{1}{N}\sum_{j=1}^{N}w_j(f_{abs}(x_j) - f(x_j))^2 + \alpha\|f''_{abs}\|^2_{L^2(0,1)} \le \frac{1}{N}\sum_{j=1}^{N}w_j(s_f(x_j) - f(x_j))^2 + \alpha\|s''_f\|^2_{L^2(0,1)}. \tag{3.4}$$

Then,

$$\frac{1}{N}\sum_{j=1}^{N}w_j(f_{abs} - f(x_j))^2 + \alpha\|f''_{abs}\|^2_{L^2(0,1)} \le w_{max}\left(E_N + \frac{\alpha}{w_{max}}\|s''_f\|^2_{L^2(0,1)}\right). \tag{3.5}$$

From Eqs (3.3)–(3.5), we have

$$\frac{1}{N}\sum_{j=1}^{N}(f_{abs}(x_j) - f(x_j))^2 \le \frac{w_{max}}{w_{min}}E_N + \frac{\alpha}{w_{min}}\|s''_f\|^2_{L^2(0,1)} \tag{3.6}$$

On the other hand, by Lemma 2,

$$\|s''_f\|^2_{L^2(0,1)} \le \|f''\|^2_{L^2(0,1)}.$$

Combining (3.2) and (3.6),

$$\frac{1}{N}\sum_{j=1}^{N}(f_{abs}(x_j) - f(x_j))^2 \le \frac{9}{8}\frac{w_{max}}{w_{min}}\rho_u d^4\|f''\|^2_{L^2(0,1)} + \frac{\alpha}{w_{min}}\|f''\|^2_{L^2(0,1)}. \tag{3.7}$$

□

Remark 2. *According to formula (3.7), in order to control the error effectively, it is necessary to select the same order of regularization parameter α as d^4 to control the interpolation error.*

Next, consider the error estimation of the random part. The reconstruction result of the random part is

$$f_{noise} = \arg\min_{g\in\mathcal{V}_m} J(g;\alpha,\varepsilon).$$

By the conclusion in chapter 4, we have

$$f_{noise} = H\boldsymbol{\lambda}^{\varepsilon},$$

where $\boldsymbol{\lambda}^{\varepsilon}$ satisfies

$$\boldsymbol{\lambda}^{\varepsilon} = \frac{1}{N}\left(\frac{1}{N}H^T W^T W H + \alpha P\right)^{-1} H^T W^T W\varepsilon.$$

Theorem 3. *Under the above assumptions, the mean square sum of the random part f_{noise} on the observation points is satisfied with the probability of* $1-\beta$

$$\|f_{noise}\|^2_{L^2(0,1)} = \frac{1}{N}\sum_{j=1}^{N} f^2_{noise}(x_j) \le \frac{4\sigma_*^2 M}{w_{min}\beta N}.$$

Proof. The proof is divided into three steps.

Step 1: Rewrite $\boldsymbol{\lambda}^{\varepsilon}$ as

$$\begin{aligned}\boldsymbol{\lambda}^{\varepsilon} &= \left(H_*^T H_* + \alpha N P\right)^{-1} H_*^T\varepsilon_*, \quad H_* := WH, \quad \varepsilon_* := W\varepsilon \\ &= P^{-1}H_*^T(H_*P^{-1}H_*^T + \alpha N I)^{-1}\varepsilon_* \\ &= P^{-1}H_*^T(S + \alpha N I)^{-1}\varepsilon_*, \quad S := H_*P^{-1}H_*^T.\end{aligned} \tag{3.8}$$

where I is an identity matrix of $N \times N$.

Step 2: Feature decomposition S. For S satisfying:

$$S := H_*P^{-1}H_*^T.$$

Obviously, it is semi-positive definite, thus S can be rewritten as

$$S = UTU^T,$$

where U is an orthogonal matrix and T is a diagonal matrix composed of the eigenvalues of S. For

$$rank(S) \le rank(P^{-1}) = M + 3.$$

T can be rewritten as

$$T = diag(t_1, t_2, \cdots, t_{M+3}, 0, \cdots, 0).$$

where $\{t_j\}_{j=1}^{M+3}$ is a monotonically decreasing sequence,

$$t_1 \ge t_2 \ge \cdots \ge t_{M+3} \ge 0.$$

Step 3: Obtaining estimates in the form of confidence intervals.

For the mean square sum of f_{noise} on observation points satisfies:

$$\frac{1}{N}\sum_{j=1}^{N} f_{noise}^2 = \frac{1}{N}\|H\boldsymbol{\lambda}\|_2^2.$$

And

$$\begin{aligned}\mathbb{E}\|H\boldsymbol{\lambda}\|_2^2 &= \mathbb{E}\left[\varepsilon_*^T(S+\alpha NI)^{-1}H_*P^{-1}H^THP^{-1}H_*^T(S+\alpha NI)^{-1}\varepsilon_*\right]\\ &= \mathbb{E}\left[tr((W^TW)^{-1}S^2(S+\alpha NI)^{-2}\varepsilon_*\varepsilon_*^T)\right]\\ &= tr((W^TW)^{-1}S^2(S+\alpha NI)^{-2}\mathbb{E}[\varepsilon_*\varepsilon_*^T])\\ &= \sigma_*^2 tr((W^TW)^{-1}S^2(S+\alpha NI)^{-2})\\ &\le \frac{\sigma_*^2}{w_{min}}\sum_{j=1}^{N}\frac{t_j^2}{(t_j+\alpha N)^2}\\ &\le \frac{\sigma_*^2}{w_{min}}(M+3).\end{aligned} \tag{3.9}$$

Combined with Markov inequality: for non-negative variable X, for any given $a > 0$, the following formula holds

$$P(X \ge a) \le \frac{\mathbb{E}[|X|]}{a}.$$

Therefore, the estimation of noise mean square sum error of noise under the probability of $1-\beta$ satisfies:

$$\frac{1}{N}\sum_{j=1}^{N} f_{noise}^2(x_j) \le \frac{\mathbb{E}\|H\boldsymbol{\lambda}^{\varepsilon}\|_2^2}{\beta N} \le \frac{\sigma_*^2(M+3)}{w_{min}\beta N}.$$

□

Finally, we estimate the error $\|e_N\|_{L^2(0,1)}^2$. Through Theorems 2 and 3, the errors of the deterministic and random parts of the reconstruction function are analyzed. Moreover, we can obtain the mean square error of $f(x)$ under the probability of $1-\beta$.

Theorem 4. *Assume that ρ_M is an indicator function on the interval $[0,1]$ and has an upper bound ρ_u. Let $f(x) \in H^2(0,1)$, for any $\beta \in [0,1]$, such that the error e_N under the probability of $1-\beta$ satisfies:*

$$\|e_N\|_{L^2(0,1)}^2 \le \frac{9}{8}\frac{w_{max}}{w_{min}}\rho_u d^4\|f''\|_{L^2(0,1)}^2 + \frac{\alpha}{w_{min}}\|f''\|_{L^2(0,1)}^2 + \frac{\sigma_*^2(M+3)}{w_{min}\beta N}. \tag{3.10}$$

Proof. By Theorems 2 and 3, the conclusion is easy to get. □

4. Parameters selection

From inequality (3.10), the estimation error is determined by the model parameters. Thus, it is a very important problem to select appropriate model parameters. In this chapter, we are going to determine the regularization parameter α, the number of knots M, and weight coefficients W in Algorithm 1.

First, we consider the regularization parameter α and the number of knots M. Through Theorem 4, the three formulas

$$\frac{9}{8}\frac{w_{max}}{w_{min}}\rho_u d^4\|f''\|^2_{L^2(0,1)},\ \frac{\alpha}{w_{min}}\|f''\|^2_{L^2(0,1)}\ \ and\ \ \frac{\sigma_*^2(M+3)}{w_{min}\beta N}$$

must be in the same order of magnitude. Otherwise, it will result in oversize error $\|e_N\|^2_{L^2(0,1)}$ or overfitting. Assume the value of ρ_u and $\|f''\|^2_{L^2(0,1)}$ do not effect the order of magnitude. Thus, we have

$$w_{max}d^4 \approx \alpha \approx \sigma_*^2 M/N,$$

where $d = 1/M$. Through the equation above, the regularization parameter α and the number of knots M satisfy:

$$M \approx (Nw_{max}/\sigma_*^2)^{\frac{1}{5}}, \quad \alpha \approx w_{max}^{\frac{1}{5}}(\sigma_*^2/N)^{\frac{4}{5}}. \tag{4.1}$$

Next, we consider the optimal choice of weight cofficients W. For weighted linear regression, researchers provide a method to choose weight cofficients. Thus, based on the method of weighted linear regression and the linear form of the reconstruction results, we provide Theorem 3.1 to choose weight coefficients.

Let $e_N = f_N - f$ be the error function of the proposed regularization algorithm. From error function e_N and Eq (2.5), the error comes from two aspects: the interpolation error caused by pre selected interpolation nodes and the random error caused by observation noise. The error e_N can not be reduced easily. But we can select suitable weight coefficients W to reduce the variance of the error e_N. Denote $f^*(x)$ as a best approximated solution of $f(x)$ in space V_m as follows:

$$f^* = \arg\min_{g\in V_m}\left(\frac{1}{N}\sum_{i=1}^{N}(g(x_i) - f(x_i))^2 + \alpha\|g''(x)\|^2_{L^2(0,1)}\right), \tag{4.2}$$

where the error $e^* = f^* - f$ just comes from the interpolation error. Similar the process of Theorem 1, f^* can also be written as $f^*(x) = H_x\boldsymbol{\lambda}_N^*$ and $\boldsymbol{\lambda}_N^*$ is given by

$$\left(\alpha P + \frac{1}{N}H^T H\right)\boldsymbol{\lambda}_N^* = \frac{1}{N}H^T\boldsymbol{f}, \tag{4.3}$$

where $\boldsymbol{f} := (f(x_1), f(x_2), \cdots, f(x_N))^T$.

Theorem 5. *Assume regularization parameter $\alpha \ll 1$, M is large enough and variance function $\sigma(x)$ satisfies $\sigma : [0,1] \to (0,+\infty)$. If W^* is given by*

$$W^* = diag(\frac{1}{\sigma(x_1)}, \frac{1}{\sigma(x_2)}, \cdots, \frac{1}{\sigma(x_N)}),$$

then, $\boldsymbol{\lambda}_N(W^*)$ is the approximated minimum empirical variance unbiased estimation of $\boldsymbol{\lambda}_N^*$, that is, $\boldsymbol{\lambda}_N^* \approx \mathbb{E}[\boldsymbol{\lambda}_N(W^*)]$ and, for any $W \in \mathbb{R}^{N\times N}$,

$$Var(\boldsymbol{\lambda}_N(W^*)) \approx \min_W Var(\boldsymbol{\lambda}_N(W)).$$

Proof. $\boldsymbol{f}$ can be written as

$$\boldsymbol{f} = H\boldsymbol{\lambda}_N^* + \boldsymbol{\varepsilon_1}, \quad \boldsymbol{\varepsilon_1} \in \mathbb{R}^N, \tag{4.4}$$

where $\boldsymbol{\varepsilon_1}$ is the interpolation error which tends to zero, when M large enough. From the equation above, Eqs (2.5) and (1.1), we have

$$\begin{aligned}\mathbb{E}[\boldsymbol{\lambda}_N(W)] &= \left(\alpha P + \frac{1}{N}H^T W^T WH\right)^{-1} \frac{1}{N}H^T W^T W(H\boldsymbol{\lambda}_N^* + \boldsymbol{\varepsilon_1})\\ &\approx \left(\alpha P + \frac{1}{N}H^T W^T WH\right)^{-1} \frac{1}{N}H^T W^T WH\boldsymbol{\lambda}_N^*, \boldsymbol{\varepsilon_1} \to 0\\ &\approx \left(\alpha N(H^T W^T WH)^{-1}P + I_{(M+3)\times(M+3)}\right)^{-1} \boldsymbol{\lambda}_N^*, \alpha \to 0\\ &\approx \boldsymbol{\lambda}_N^*.\end{aligned} \tag{4.5}$$

Denote $\boldsymbol{\varepsilon} = (\sigma(x_1)\varepsilon, \sigma(x_2)\varepsilon, \cdots, \sigma(x_N)\varepsilon)$, diagonal matrix $D := W^T W$ and diagonal matrix $V := diag(\sigma^2(x_1), \sigma^2(x_2), \cdots, \sigma^2(x_N))$. Through the definition above and Eqs (2.5), (1.1) and (4.5), for any $W \in \mathbb{R}_+^{N\times N}$, if $\varepsilon_1 \to 0$ and $\alpha \to 0$, we have

$$\begin{aligned}Var[\boldsymbol{\lambda}_N(W)] &= \mathbb{E}(\boldsymbol{\lambda}_N(W) - \boldsymbol{\lambda}^*)(\boldsymbol{\lambda}_N(W) - \boldsymbol{\lambda}^*)^T\\ &\approx \mathbb{E}((\alpha P + \frac{1}{N}H^T DH)^{-1}\frac{1}{N}H^T D\boldsymbol{\varepsilon})((\alpha P + \frac{1}{N}H^T DH)^{-1}\frac{1}{N}H^T D\boldsymbol{\varepsilon})^T\\ &\approx (H^T DH)^{-1}H^T D\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]DH(H^T DH)^{-1}\\ &= (H^T DH)^{-1}H^T DVDH(H^T DH)^{-1}\\ &\geq (H^T VH)^{-1}.\end{aligned}$$

Moreover, when $W = W^*$,

$$Var[\boldsymbol{\lambda}_N(W^*)] \approx (H^T VH)^{-1} \approx \min_W Var(\boldsymbol{\lambda}_N(W)).$$

□

From Theorem 2, when $\alpha \to 0$ and M is large enough, we can choose W^* to reduce the variance of estimator $\boldsymbol{\lambda}_N(W)$. Moreover, the value $\{1/\sigma(x_i)\}_{i=1}^N$ in W^* can not be infinity. When some weight coefficients $1/\sigma(x_i)$ are too large, the reconstruction results are only affected by the data corresponding to these large weight coefficients. Thus, we suggest to choose the suitable weight coefficients W as follows:

$$W_{suit} = diag(\frac{1}{k + \sigma(x_1)}, \frac{1}{k + \sigma(x_2)}, \cdots, \frac{1}{k + \sigma(x_N)}), \tag{4.6}$$

where k is a small positive constant such as 0.1 and 0.01 and w_i satisfies:

$$w_i = 1/(k + \sigma(x_i))^2, \quad i = 1, 2, \cdots, N. \tag{4.7}$$

Combining Eqs (4.1) and (4.7), we have the parameter selection Table 1.

5. An illustrative example

In this section, we provide a illustrative example to show the feasibility and advantages of the weighted regularization method. Assume that the heteroscedastic model satisfies:

$$y = f(x) + \sigma(x)\varepsilon,$$

where $f(x) := 40(x - 0.5)^2$ and $\sigma(x) := 3\sin(150\pi x) + 3.5$. We consider following problem: for a lager positive integer $N = 300$, given observation data $\{(x_i, y_i)\}_{i=1}^N$ and function $\sigma(x)$, we try to reconstruct function $f(x)$ and its derivative $f'(x)$. We select the number of knots $M = 10$, mesh size $d = 1/M = 0.1$, regularization parameter $\alpha = 10^{-4}$ and weight coefficients W satisfying (4.6) with $k = 0.1$.

Figure 1 compares the reconstruction results-“improve” in this paper with the results-“tradition” in [10]. When function $\sigma(x)$ changes sharply, our results are closer to exact function than that those in [10].

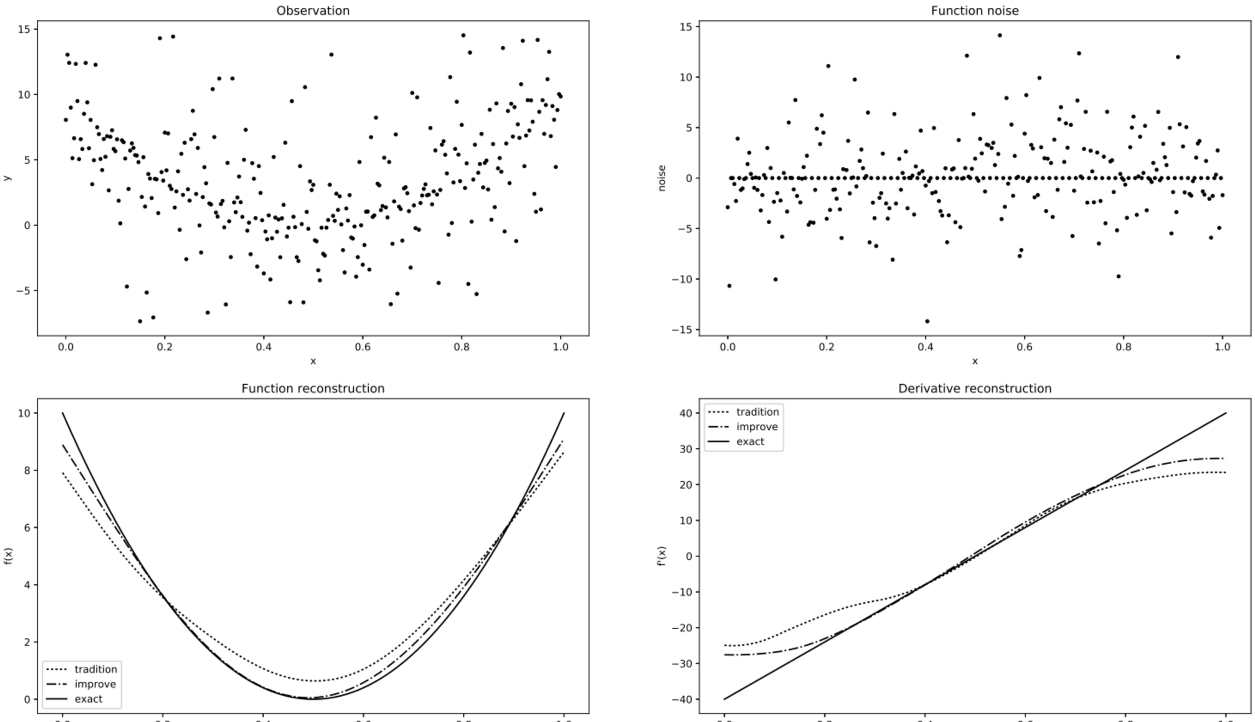

Figure 1. Observation points,noise data and reconstruction results “improve” and “tradition” in this paper and [10], respectively.

6. Applications in inverse problems

In this section, we will give two applications of our numerical differentiation method in two kinds of inverse problems.

6.1. The inverse problems of identifying coefficients in the boundary value problem

We consider a nonlinear ill-posed problem–the identification of the coefficient c in the boundary value problem

$$f'(x) + c(x)f(x) = p(x), \quad f(0) = h_0,$$

from the solution $f(x)$. We can get $c(x)$ directly from

$$c(x) = \frac{p(x) - f(x)}{f'(x)}. \tag{6.1}$$

The example we take here is that $f(x) = 10e^{-\sin(\pi x)}$, $p(x) = 0$ and $h_0 = 10$. In our computation, $\{(x_i, y_i)\}_{i=1}^{N}$ is denoted as the observation data where y satisfies:

$$y = f(x) + \sigma(x)\varepsilon,$$

where $\sigma(x) = 5\sin(150\pi x) + 5.1$. From the observation data above and the improved method in this paper, we reconstruct $f(x)$ and $f'(x)$. The results are present in Figure 2.

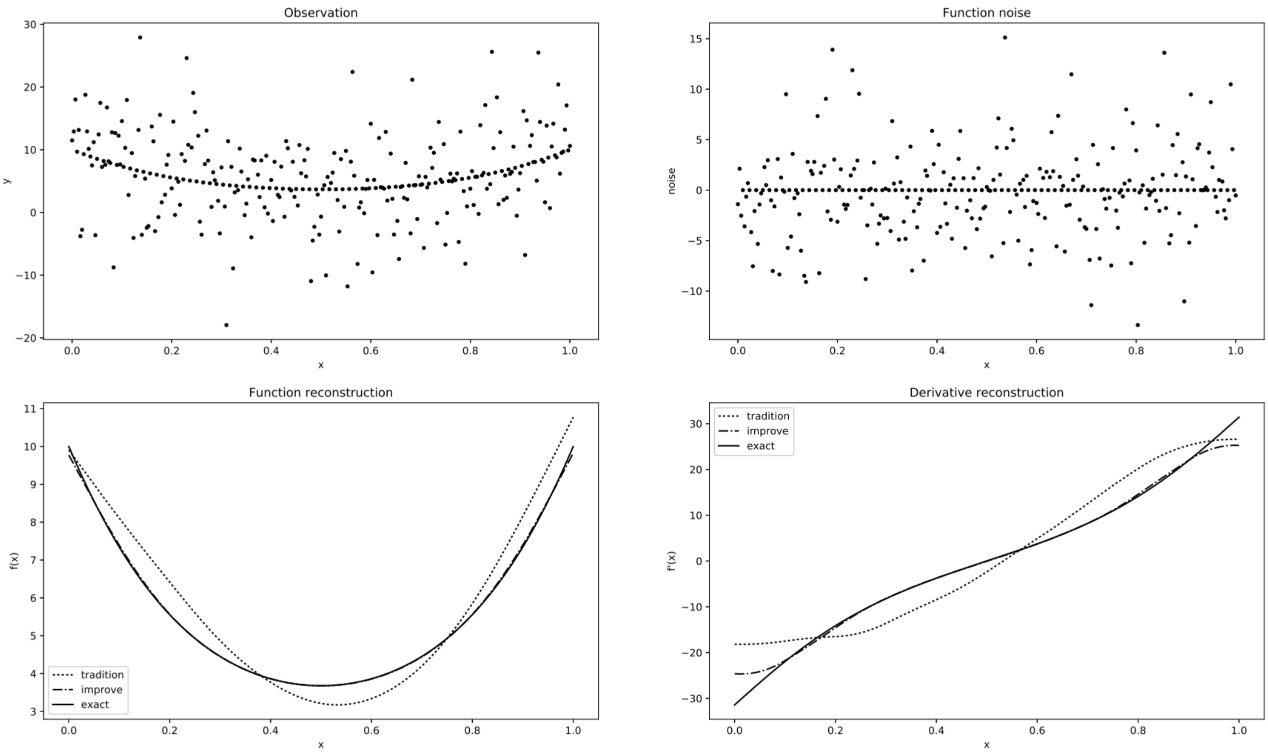

Figure 2. Observation points,noise data and reconstruction results “improve” and “tradition” in this paper and [10], respectively.

Through the reconstruction results of $f(x)$ and $f'(x)$ and Eq (6.1), we have the estimator of $c(x)$ in Figure 3, where our method is better than the traditional.

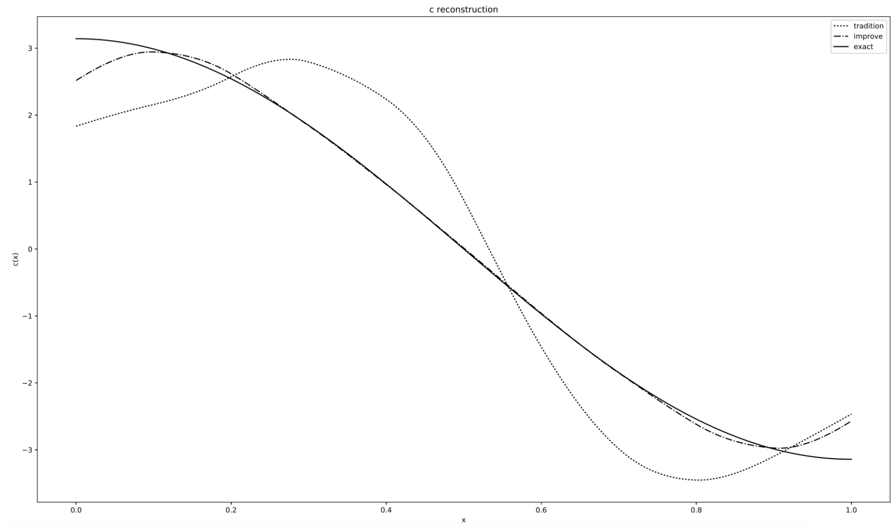

Figure 3. Reconstruction c-“improve” and “tradition” in this paper and [10], respectively.

6.2. *Determination of the interface in computerized tomography*

In this subsection, we apply the method in this paper to a simple but interesting problem in Computerized Tomography (CT).

We assume that the attenuation coefficient of an object with respect to an X-ray at x is $p(x)$. We scan the cross-section by an X-ray beam L of unit intensity. The intensity past the object is $e^{-\int_L p(x)dx}$. We denote function f as follows:

$$f(L) := \int_L p(x)dx. \tag{6.2}$$

The main problem in CT is to recover the function p from f. However, in many cases, it will be enough if one can reconstruct the interface of the different mediums, that is, the discontinuous points of p which is related directly with the nondifferentiable points of f. Thus we just need to determine the nondifferentiable points in $f(x)$.

D is denoted as the triangular cross section of a object. The attenuation coefficient is 0 outside the object and 1 inside the object, that is:

$$p(x) = \begin{cases} 1, & x \in D; \\ 0, & x \notin D, \end{cases} \tag{6.3}$$

see Figure 4.

From Figure 4 and Eq (6.2), the function $f(x)$ can be calculated directly:

$$f(x) = \begin{cases} 10x - 2, & 0.2 \le x < 0.5, \\ 8 - 10x, & 0.5 \le x \le 0.8, \\ 0, & elsewhere. \end{cases} \tag{6.4}$$

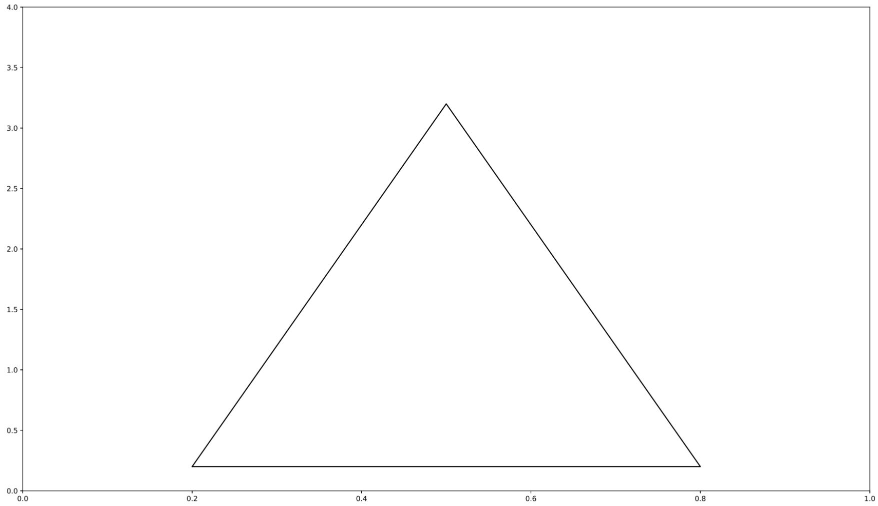

Figure 4. picture of D.

In our computation, $\{(x_i, y_i)\}_{i=1}^{N}$ is denoted as the observation data where y satisfies Eq (1.1) and $\sigma(x) = 5\sin(150\pi x) + 5$. Let the number of data $N = 300$, the number of knots $M = 50$, mesh size $d = 1/M = 0.02$, regularization parameter $\alpha \approx 10^{-4}$ and the parameter $k = 0.01$ in (4.6).

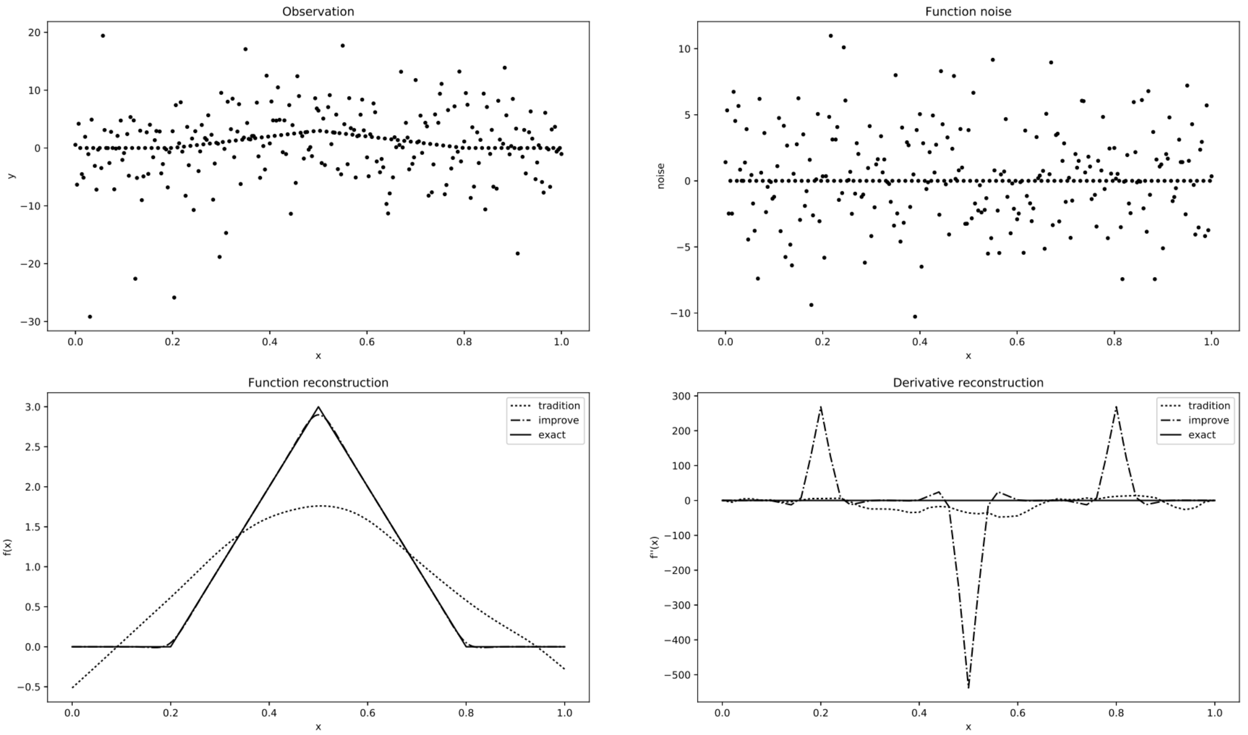

Figure 5. Observation points, noise data and reconstruction results-"improve" and "tradition" in this paper and [10], respectively.

In Figure 5, we present observation data, noise data and the reconstructions of $f(x)$ and $f''(x)$.

Compared with the traditional method in [10], through the improved approach in this paper, we can reconstruct $f(x)$ better and find three nondifferentiable points of $f(x)$: 0.2, 0.5 and 0.8.

7. Conclusions

In this paper, we propose a weighted online regularization method–Algorithm 1, which updates reconstruction results through new input data without old data again. The error estimation in Section 3 illustrates that the algorithm can reach the best convergence order when the relationship between the amount of observed data N and the dimension M of linear space is $M \approx N^{\frac{1}{5}}$. Compared with the classical B-spline method, weight coefficients and a relatively small number of knots for interpolation are used to reduce the effect of heteroscedasticity and computation cost. Moreover, a feasible selection method of the regularization parameter, the number of knots, and weight coefficients are provided in Table 1.

The biggest advantage of this algorithm is to solve the problem of low data accuracy caused by heteroscedasticity by processing a large amount of data. It can significantly reduce the error of observation data while ensuring the calculation efficiency.

In the next work, we will consider the higher dimensional heteroscedasticity problem, which requires higher regularity of the function and analyze other radial basis functions, such as Sobolev radial basis functions, to solve high-dimensional problems.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by Doctoral Research Initiation Fund of Shenzhen Institute of Information Technology with grant (No. SZIIT2023KJ003).

Conflict of interest

The author declare that have no competing interest.

References

1. J. Cheng, X. Z. Jia, Y. B. Wang, Numerical differentiation and its applications, *Inverse Probl. Sci. En.*, **15** (2007), 339–357. https://doi.org/10.1080/17415970600839093
2. M. Hanke, O. Scherzer, Inverse problems light: numerical differentiation, *Am. Math. Mon.*, **108** (2001), 512–521. https://doi.org/10.1080/00029890.2001.11919778
3. D. H. Xu, Y. H. Xu, M. B. Ge, Q. F. Zhang, *Models and numerics for differential equations and inverse problems*, Beijing: Science Press, 2021.
4. P. Craven, G. Wahba, Smoothing noisy data with spline functions, *Numer. Math.*, **31** (1978), 377–403. https://doi.org/10.1007/BF01404567

5. D. L. Ragozin, Error bounds for derivative estimates based on spline smoothing of exact or noisy data, *J. Approx. Theory*, **37** (1983), 335–355. https://doi.org/10.1016/0021-9045(83)90042-4

6. J. P. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, New York: Springer, 2005. https://doi.org/10.1007/b138659

7. L. Wasserman, *All of nonparametric statistics*, New York: Springer, 2006. https://doi.org/10.1007/0-387-30623-4

8. G. Claeskens, T. Krivobokova, J. D. Opsomer, Asymptotic properties of penalized spline estimators, *Biometrika*, **96** (2009), 529–544. https://doi.org/10.1093/biomet/asp035

9. P. H. C. Eilers, B. D. Marx, Flexible smoothing with b-splines and penalties, *Statist. Sci.*, **11** (1996), 89–121. https://doi.org/10.1214/ss/1038425655

10. J. Zhang, J. Cheng, M. Zhong, A tikhonov regularization based algorithm for scattered data with random noise, arXiv:2105.00747. https://doi.org/10.48550/arXiv.2105.00747

11. J. A. Fessler, Penalized weighted least-squares image reconstruction for positron emission tomography, *IEEE T. Med. Imaging*, **13** (1994), 290–300. https://doi.org/10.1109/42.293921

12. K. R. Ridgway, J. R. Dunn, J. L. Wilkin, Ocean interpolation by four dimensional weighted least squares-application to the waters around Australasia, *J. Atmos. Ocean. Tech.*, **19** (2002), 1357–1375. https://doi.org/10.1175/1520-0426(2002)019

13. J. M. Wooldridge, *Introductory econometrics: a modern approach*, Boston: Cengage Learning, 2012.

14. Q. Feng, J. Hannig, J. S. Marron. A note on automatic data transformation, *Stat.*, **5** (2016), 82–87. https://doi.org/10.1002/sta4.104

15. J. Kalina, On heteroscedasticity in robust regression, *International Days of Statistics and Economics*, **41** (2011), 228–237. https://doi.org/10.1111/j.1467-9310.2011.00660.x

16. B. Sun, L. Ma, T. Shen, R. Geng, Y. Zhou, Y. Tian, A robust data-driven method for multiseasonality and heteroscedasticity in time series preprocessing, *Wirel. Commun. Mob. Com.*, **2021** (2021), 6692390. https://doi.org/10.1155/2021/6692390

17. M. Marzjarani, A comparison of a general linear model and the ratio estimator, *International Journal of Statistics and Probability*, **9** (2020), 54–65. https://doi.org/10.5539/ijsp.v9n3p54

18. A. Bashan, N. M. Yagmurlu, Y. Ucar, A. Esen, A new perspective for the numerical solution of the modified equal width wave equation, *Math. Method. Appl. Sci.*, **44** (2021), 8925–8939. https://doi.org/10.1002/mma.7322

19. A. Bashan, N. M. Yagmurlu, Y. Ucar, A. Esen, Finite difference method combined with differential quadrature method for numerical computation of the modified equal width wave equation, *Numer. Meth. Part. D. E.*, **37** (2021), 690–706. https://doi.org/10.1002/num.22547

20. A. Bashan, A. Esen, Single soliton and double soliton solutions of the quadratic-nonlinear Korteweg-de Vries equation for small and long-times, *Numer. Meth. Part. D. E.*, **37** (2021), 1561–1582. https://doi.org/10.1002/num.22597

21. Y. Ucar, N. M. Yagmurlu, A. Bashan, Numerical solutions and stability analysis of modified burgers equation via modified cubic b-spline differential quadrature methods, *Sigma J. Eng. Nat. Sci.*, **37** (2019), 129–142.

22. A. Bashan, N. M. Yagmurlu, A mixed method approach to the solitary wave, undular bore and boundary-forced solutions of the regularized long wave equation, *Comp. Appl. Math.*, **41** (2022), 169. https://doi.org/10.1007/s40314-022-01882-7

23. A. Bashan, N. M. Yagmurlu, Y. Ucar, A. Esen, Numerical approximation to the MEW equation for the single solitary wave and different types of interactions of the solitary waves, *J. Differ. Equ. Appl.*, **28** (2022), 1193–1213. https://doi.org/10.1080/10236198.2022.2132154

24. G. Micula, S. Micula, *Handbook of splines*, Dordrecht: Springer, 1999. https://doi.org/10.1007/978-94-011-5338-6

25. A. Koppel, G. Warnell, E. Stump, A. Ribeiro, Parsimonious online learning with kernels via sparse projections in function space, *J. Mach. Learn. Res.*, **20** (2019), 83–126.