Supplementary Matrial of Identification of Partial-Differential-Equations-Based Models from Noisy Data via Splines

[1] *Biostatistics and Research Decision Sciences Department, Merck & Co., Inc*
[2] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Tech*

Supplementary Material

This is the online supplementary material of paper *Identification of Underlying Partial Differential Equations from Noisy Data via Spines*. It includes (1) lemmas to derive the main theory; (2) numerical details of the figures in the simulation; (3) proofs and other technical details which is not covered in the main body of the paper due to the page limitation.

S1 Overview of Our Proposed Algorithm

We give an overview our SAPDEMI method in Algorithm 1.

Algorithm 1: Overview of our proposed SAPDEMI method

Input: Data from the unknown PDE model as in (1.1); penalty parameter used in the Lasso identify model: $\lambda > 0$; smoothing parameter used in the cubic spline: $\alpha, \bar{\alpha} \in (0, 1]$.

Output: The identified/recovered PDE model.

Functional estimation stage:

Estimate $\mathbf{X}, \nabla_t \mathbf{u}$ by cubic spline with $\alpha, \bar{\alpha} \in (0, 1]$.

Model identification stage:

The unknown PDE system is recovered as: $\frac{\partial}{\partial t} u(x,t) = \mathbf{x}^\top \widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2MN} \|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1$ and

$$\mathbf{x} = \left(1, \quad u(x,t), \quad \frac{\partial u(x,t)}{\partial x}, \quad \frac{\partial^2 u(x,t)}{\partial x^2}, \quad (u(x,t))^2, \quad u(x,t)\frac{\partial u(x,t)}{\partial x}, \quad \ldots, \quad \right)^\top.$$

S2 Derivation of the 0-th, First, Second Derivative of the Cubic Spline in Section 2.1

In this section, we focus on solving the derivatives of $u(x, t_n)$ with respective to x, i.e., $\left\{u(x_i, t_n), \frac{\partial}{\partial x} u(x_i, t_n), \frac{\partial^2}{\partial x^2} u(x_i, t_n)\right\}_{i=0,1\ldots,M-1}$, for any $n = 0, 1, \ldots, N-1$. To realize this objective, we first fix t as t_n for a general $n \in \{0, 1, \ldots, N-1\}$. Then we use cubic spline to fit data $\{(x_i, u_n^i)\}_{i=0,1,\ldots,M-1}$.

Suppose the cubic polynomial spline over the knots $\{(x_i, u_n^i)\}_{i=0,1,\ldots,M-1}$ is $s(x)$. So under good approximation, we can regard $s(x), s'(x), s''(x)$ as the estimators of $u(x_i, t_n)$, $\frac{\partial}{\partial x} u(x, t_n)$, $\frac{\partial^2}{\partial x^2} u(x, t_n)$, where $s'(x), s''(x)$ is the first and second derivatives of $s(x)$, respectively.

Let's first take a look at the zero-order derivatives of $s(x)$. By introducing matrix algebra,

the objective function in equation (2.4) can be rewritten as

$$J_\alpha(s) = \alpha(\mathbf{u}_:^n - \mathbf{f})^\top \mathbf{W}(\mathbf{u}_:^n - \mathbf{f}) + (1-\alpha)\mathbf{f}^\top \mathbf{A}^\top \mathbf{M}^{-1}\mathbf{A}\mathbf{f} \tag{S2.1}$$

where vector

$$\mathbf{f} = \begin{pmatrix} s(x_0) \\ s(x_1) \\ \vdots \\ s(x_{M-1}) \end{pmatrix} \triangleq \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{M-1} \end{pmatrix}, \mathbf{u}_:^n = \begin{pmatrix} u_0^n \\ u_1^n \\ \vdots \\ u_{M-1}^n \end{pmatrix}$$

and matrix $\mathbf{W} = \mathrm{diag}(w_0, w_1, \ldots, w_{M-1})$ and matrix $\mathbf{A}$ is defined in (2.6). By taking the derivative of (S2.1) with respective to $\mathbf{f}$ and set it as zero, we have

$$\widehat{\mathbf{f}} = [\alpha\mathbf{W} + (1-\alpha)\mathbf{A}^\top\mathbf{M}\mathbf{A}]^{-1}\alpha\mathbf{W}\mathbf{u}_:^n. \tag{S2.2}$$

Then we solve the second-order derivative with respective to x. Let us first suppose that the cubic spline $s(x)$ in $[x_i, x_{i+1}]$ is denoted $s_i(x)$, and we denote $s_i''(x_i) = \sigma_i, s_i''(x_{i+1}) = \sigma_{i+1}$. Then we have $\forall x \in [x_i, x_{i+1}]$ $(0 \le i \le M-2)$,

$$s_i''(x) = \sigma_i \frac{x_{i+1} - x}{h_i} + \sigma_{i+1}\frac{x - x_i}{h_i},$$

where matrix $\mathbf{M}$ is defined in (2.7). This is because $s_i''(x)$ with $x \in [x_i, x_{i+1}]$ is a linear function. By taking a double integral of the above equation, we have

$$s_i(x) = \frac{\sigma_i}{6h_i}(x_{i+1} - x)^3 + \frac{\sigma_{i+1}}{6h_i}(x - x_i)^3 + c_1(x - x_1) + c_2(x_{i+1} - x), \tag{S2.3}$$

where c_1, c_2 is the unknown parameters to be estimated. Because $s_i(x)$ interpolates two endpoints (x_i, f_i) and (x_{i+1}, f_{i+1}) , if we plug x_i, x_{i+1} into the above $s_i(x)$, we have

$$\begin{cases} f_i &= s_i(x_i) = \frac{\sigma_i}{6}h_i^2 + c_2 h_i \\ f_{i+1} &= s_i(x_{i+1}) = \frac{\sigma_{i+1}}{6}h_i^2 + c_1 h_i, \end{cases}$$

where we can solve c_1, c_2 as

$$\begin{cases} c_1 &= (f_{i+1} - \frac{\sigma_{i+1}}{6}h_i^2)/h_i, \\ c_2 &= (f_i - \frac{\sigma_i}{6}h_i^2)/h_i. \end{cases}$$

By plugging in the value of c_1, c_2 into equation (S2.3), we have ($0 \le i \le M-2$)

$$s_i(x) = \frac{\sigma_i}{6h_i}(x_{i+1} - x)^3 + \frac{\sigma_{i+1}}{6h_i}(x - x_i)^3 + \left(\frac{f_{i+1}}{h_i} - \frac{\sigma_{i+1}h_i}{6}\right)(x - x_i) + \left(\frac{f_i}{h_i} - \frac{\sigma_i h_i}{6}\right)(x_{i+1} - x),$$

with its first derivative as

$$s_i'(x) = -\frac{\sigma_i}{2h_i}(x_{i+1} - x)^2 + \frac{\sigma_{i+1}}{2h_i}(x - x_i)^2 + \frac{f_{i+1} - f_i}{h_i} - \frac{h_i}{6}(\sigma_{i+1} - \sigma_i). \tag{S2.4}$$

Because $s_{i-1}'(x_i) = s_i'(x_i)$, for $1 \le i \le M-2$, we have

$$\frac{1}{6}h_{i-1}\sigma_{i-1} + \frac{1}{3}(h_{i-1} + h_i)\sigma_i + \frac{1}{6}h_i\sigma_{i+1} = \frac{f_{i+1} - f_i}{h_i} - \frac{f_i - f_{i-1}}{h_{i-1}}. \tag{S2.5}$$

Equation (S2.5) gives $M-2$ equations. Recall $\sigma_0 = \sigma_{M-1} = 0$, so totally we get M equations, which is enough to solve M parameters, i.e., $\sigma_0, \sigma_1, \ldots, \sigma_{M-1}$. We write out the above system of linear equations, where we hope to identify a fast numerical approach to solve it. The system of linear equations is:

$$\left\{\begin{array}{ccccccc} & & \frac{1}{3}(h_0+h_1)\sigma_1 & + & \frac{1}{6}h_1\sigma_2 & = & \frac{u_2^n-u_1^n}{h_1} - \frac{f_1-u_0}{h_0} \\ \frac{1}{6}h_1\sigma_1 & + & \frac{1}{3}(h_1+h_2)\sigma_2 & + & \frac{1}{6}h_2\sigma_3 & = & \frac{f_3-f_2}{h_1} - \frac{f_2-f_1}{h_0} \\ & & & \vdots & & & \\ \frac{1}{6}h_{M-4}\sigma_{M-4} & + & \frac{1}{3}(h_{M-4}+h_{M-3})\sigma_{M-3} & + & \frac{1}{6}h_{M-3}\sigma_{M-2} & = & \frac{f_{M-2}-f_{M-3}}{h_{M-3}} - \frac{f_{M-3}-f_{M-4}}{h_{M-4}} \\ \frac{1}{6}h_{M-3}\sigma_{M-3} & + & \frac{1}{3}(h_{M-3}+h_{M-2})\sigma_{M-2} & & & = & \frac{f_{M-1}-f_{M-2}}{h_{M-2}} - \frac{f_{M-2}-f_{M-3}}{h_{M-3}} \end{array}\right. .$$

From the above system of equation, we can see that the second derivative of cubic spline $s(x)$ can be solved by the above system of linear equation, i.e.,

$$\widehat{\boldsymbol{\sigma}} = \mathbf{M}^{-1}\mathbf{A}\widehat{\mathbf{f}} \tag{S2.6}$$

where vector $\widehat{\mathbf{f}}$ is defined in (S2.2), matrix $\mathbf{A} \in \mathbb{R}^{(M-2)\times M}$ is defined in (2.6), and matrix $\mathbf{M} \in \mathbb{R}^{(M-2)\times(M-2)}$ is defined as (2.7).

Finally, we focus on solving the first derivative of cubic spline $s(x)$. Let $\theta_i = s'(x_i)$ for $i = 0, 1, \ldots, M-1$, then we have

$$\begin{aligned} s_i(x) &= \theta_i \tfrac{(x_{i+1}-x)^2(x-x_i)}{h_i^2} - \theta_{i+1}\tfrac{(x-x_i)^2(x_{i+1}-x)}{h_i^2} + f_i\tfrac{(x_{i+1}-x)^2[2(x-x_i)+h_i]}{h_i^3} + \\ &\quad f_{i+1}\tfrac{(x-x_i)^2[2(x_{i+1}-x)+h_i]}{h_i^3}, \\ s_i'(x) &= \theta_i\tfrac{(x_{i+1}-x)(2x_i+x_{i+1}-3x)}{h_i^2} - \theta_{i+1}\tfrac{(x-x_i)(2x_{i+1}+x_i-3x)}{h_i^2} + 6\tfrac{u_{i+1}^n-u_i^n}{h_i^3}(x_{i+1}-x)(x-x_i), \\ s_i''(x) &= -2\theta_i\tfrac{2x_{i+1}+x_i-3x}{h_i^2} - 2\theta_{i+1}\tfrac{2x_i+x_{i+1}-3x}{h_i^2} + 6\tfrac{u_{i+1}^n-u_i^n}{h_i^3}(x_{i+1}+x_i-2x). \end{aligned}$$

By plugging x_i into $s_i''(x)$ and $s_{i-1}''(x)$, we have

$$\left\{\begin{array}{rcl} s_i''(x) &=& -2\theta_i\frac{2x_{i+1}+x_i-3x}{h_i^2} - 2\theta_{i+1}\frac{2x_i+x_{i+1}-3x}{h_i^2} + 6\frac{f_{i+1}-f_i}{h_i^3}(x_{i+1}+x_i-2x) \\ s_{i-1}''(x) &=& -2\theta_{i-1}\frac{2x_i+x_{i-1}-3x}{h_{i-1}^2} - 2\theta_i\frac{2x_{i-1}+x_i-3x}{h_{i-1}^2} + 6\frac{f_i-f_{i-1}}{h_{i-1}^3}(x_i+x_{i-1}-2x) \end{array}\right.$$

which gives

$$\left\{\begin{array}{rcl} s_i''(x) &=& \frac{-4}{h_i}\theta_i + \frac{-2}{h_i}\theta_{i+1} + 6\frac{f_{i+1}-f_i}{h_i^2} \\ s_{i-1}''(x) &=& \frac{2}{h_{i-1}}\theta_{i-1} + \frac{4}{h_{i-1}}\theta_i - 6\frac{f_i-f_{i-1}}{h_{i-1}^2}. \end{array}\right.$$

Because $s_i''(x_i) = s_{i-1}''(x_i)$, we have $(\forall i = 1, 2, \ldots, M-2)$

$$\begin{aligned} & \tfrac{-4}{h_i}\theta_i + \tfrac{-2}{h_i}\theta_{i+1} + 6\tfrac{f_{i+1}-f_i}{h_i^2} = \tfrac{2}{h_{i-1}}\theta_{i-1} + \tfrac{4}{h_{i-1}}\theta_i - 6\tfrac{f_i-f_{i-1}}{h_{i-1}^2} \\ \Leftrightarrow\quad & \tfrac{2}{h_{i-1}}\theta_{i-1} + (\tfrac{4}{h_{i-1}} + \tfrac{4}{h_i})\theta_i + \tfrac{2}{h_i}\theta_{i+1} = 6\tfrac{f_{i+1}-f_i}{h_i^2} + 6\tfrac{f_i-f_{i-1}}{h_{i-1}^2} \\ \Leftrightarrow\quad & \tfrac{1}{h_{i-1}}\theta_{i-1} + (\tfrac{2}{h_{i-1}} + \tfrac{2}{h_i})\theta_i + \tfrac{1}{h_i}\theta_{i+1} = 3\tfrac{f_{i+1}-f_i}{h_i^2} + 3\tfrac{f_i-f_{i-1}}{h_{i-1}^2}. \end{aligned}$$

By organizing the above system of equation into matrix algebra, we have

$$\begin{pmatrix} \frac{1}{h_0} & \frac{2}{h_0}+\frac{2}{h_1} & \frac{1}{h_1} & 0 & \dots & 0 & 0 & 0 \\ 0 & \frac{1}{h_1} & \frac{2}{h_1}+\frac{2}{h_2} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{h_{M-3}} & \frac{2}{h_{M-3}}+\frac{2}{h_{M-2}} & \frac{1}{h_{M-2}} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{M-1} \end{pmatrix}$$
$$= \begin{pmatrix} 3\frac{f_2-f_1}{h_1^2}+3\frac{f_1-f_0}{h_0^2} \\ 3\frac{f_3-f_2^n}{h_2^2}+3\frac{f_2-f_1}{h_1^2} \\ \vdots \\ 3\frac{f_{M-1}-f_{M-2}}{h_{M-2}^2}+3\frac{f_{M-2}^n-f_{M-3}}{h_{M-3}^2} \end{pmatrix}.$$

For the endpoint θ_0, because $s_0''(x_0)=0$, we have

$$s_0''(x) = -2\theta_0\frac{2x_1+x_0-3x}{h_0^2} - 2\theta_1\frac{2x_0+x_1-3x}{h_0^2} + 6\frac{f_1-f_0}{h_0^3}(x_1+x_0-2x).$$

When we take the value of x as x_0, we have

$$\begin{aligned} s_0''(x_0) &= -2\theta_0\tfrac{2x_1+x_0-3x_0}{h_0^2} - 2\theta_1\tfrac{2x_0+x_1-3x_0}{h_0^2} + 6\tfrac{f_1-f_0}{h_0^3}(x_1+x_0-2x_0) \\ &= \tfrac{-4}{h_0}\theta_0 + \tfrac{-2}{h_0}\theta_1 + 6\tfrac{f_1-f_0}{h_0^2} \\ &= 0. \end{aligned}$$

For the two endpoint θ_{M-1}, because $s_{M-2}''(x_{M-1})=0$, we have

$$\begin{aligned} s_{M-2}''(x) &= -2\theta_{M-2}\tfrac{2x_{M-1}+x_{M-2}-3x}{h_{M-2}^2} - 2\theta_{M-1}\tfrac{2x_{M-2}+x_{M-1}-3x}{h_{M-2}^2} + \\ &\quad 6\tfrac{f_{M-1}-f_{M-2}}{h_{M-2}^3}(x_{M-1}+x_{M-2}-2x). \end{aligned}$$

When we take the value of x as x_{M-1}, we have

$$\begin{aligned} s_{M-2}''(x_{M-1}) &= -2\theta_{M-2}\tfrac{2x_{M-1}+x_{M-2}-3x_{M-1}}{h_{M-2}^2} - 2\theta_{M-1}\tfrac{2x_{M-2}+x_{M-1}-3x_{M-1}}{h_{M-2}^2} + \\ &\quad 6\tfrac{f_{M-1}-f_{M-2}}{h_{M-2}^3}(x_{M-1}+x_{M-2}-2x_{M-1}) \\ &= \tfrac{2}{h_{M-2}}\theta_{M-2} + \tfrac{4}{h_{M-2}}\theta_{M-1} - 6\tfrac{f_{M-1}-f_{M-2}}{h_{M-2}^2} \\ &= 0. \end{aligned}$$

So the first order derivative $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{M-1})^\top$ can be solved by

$$\underbrace{\begin{pmatrix} \frac{2}{h_0} & \frac{1}{h_0} & 0 & 0 & \ldots & 0 & 0 & 0 \\ \frac{1}{h_0} & \frac{2}{h_0}+\frac{2}{h_1} & \frac{1}{h_1} & 0 & \ldots & 0 & 0 & 0 \\ 0 & \frac{1}{h_1} & \frac{2}{h_1}+\frac{2}{h_2} & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \frac{1}{h_{M-3}} & \frac{2}{h_{M-3}}+\frac{2}{h_{M-2}} & \frac{1}{h_{M-2}} \\ 0 & 0 & 0 & 0 & \ldots & 0 & \frac{1}{h_{M-2}} & \frac{2}{h_{M-2}} \end{pmatrix}}_{\mathbf{Q}\in\mathbb{R}^{M\times M}} \underbrace{\begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \vdots \\ \theta_{M-1} \end{pmatrix}}_{\boldsymbol{\theta}}$$

$$= \underbrace{\begin{pmatrix} 3\frac{f_1-f_0}{h_0^2} \\ 3\frac{f_2-f_1}{h_1^2}+3\frac{f_1-f_0}{h_0^2} \\ 3\frac{f_3-f_2}{h_2^2}+3\frac{f_2-f_1}{h_1^2} \\ \vdots \\ 3\frac{f_{M-1}-f_{M-2}}{h_{M-2}^2}+3\frac{f_{M-2}-f_{M-3}}{h_{M-3}^2} \\ 3\frac{f_{M-1}-f_{M-2}}{h_{M-2}^2} \end{pmatrix}}_{\mathbf{q}}.$$

In matrix notation, the first order derivative $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{M-1})^\top$ can be solved by

$$\widehat{\boldsymbol{\theta}} = \mathbf{Q}^{-1}\widehat{\mathbf{q}} = \mathbf{Q}^{-1}\mathbf{B}\widehat{\mathbf{f}}, \tag{S2.7}$$

where $\widehat{\mathbf{f}}$ is defined in (S2.2), and matrix $\mathbf{B} \in \mathbb{R}^{M\times M}$ is defined as

$$\mathbf{B} = \begin{pmatrix} \frac{-3}{h_0^2} & \frac{3}{h_0^2} & 0 & 0 & \ldots & 0 & 0 & 0 \\ \frac{-3}{h_0^2} & \frac{3}{h_0^2}-\frac{3}{h_1^2} & \frac{3}{h_1^2} & 0 & \ldots & 0 & 0 & 0 \\ 0 & \frac{-3}{h_1^2} & \frac{3}{h_1^2}-\frac{3}{h_2^2} & \frac{3}{h_2^2} & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ & 0 & 0 & 0 & 0 & \frac{-3}{h_{M-3}^2} & \frac{3}{h_{M-3}^2}-\frac{3}{h_{M-2}^2} & \frac{3}{h_{M-2}^2} \\ & 0 & 0 & 0 & 0 & 0 & \frac{-3}{h_{M-2}^2} & \frac{3}{h_{M-2}^2} \end{pmatrix}.$$

S3 Computational Complexity of Local Polynomial Regression Method

In the functional estimation stage, the computational complexity of the local polynomial regression method is stated in the following proportion.

Proposition S3.1. Given data $\mathcal{D}$ in (1.1), if we use the local polynomial regression in the functional estimation stage, i.e., estimate $\mathbf{X} \in \mathbb{R}^{MN\times K}, \nabla_t \mathbf{u} \in \mathbb{R}^{MN}$ via the local polynomial regression described as in this online supplementary material, then the computation complexity of this stage is of order

$$\max\{O(q_{\max}^2 M^2 N), O(MN^2), O(q_{\max}^3 MN), O(p_{\max} MN), O(K^3)\},$$

where $p_{\max}$ is the highest polynomial order in (1.3), $q_{\max}$ is the highest order of derivatives in (1.3), M is the spatial resolution, N is the temporal resolution, and K is the number of columns of $\mathbf{X}$.

If we set $q_{\max} = 2$ to match the derivative order of the local polynomial regression to the cubic spline, then the computation complexity is of order

$$\max\{O(M^2N), O(MN^2), O(p_{\max}MN), O(K^3)\}.$$

See a proof in Section S10.2.

As suggested by Proposition S3.1, the computational complexity of local polynomial regression is much higher than that in the cubic spline. But the advantage of local polynomial regression is that it can derive any order of derivatives, i.e., $q_{\max} \geq 0$ in (1.3), while for the cubic spline, $q_{\max} = 2$. In applications, $q_{\max} = 2$ should be sufficient because most of the PDE models are governed by derivatives up to the second derivative, for instance, heat equation, wave equation, Laplace's equation, Helmholtz equation, Poisson's equation, and so on. In our paper, we mainly use cubic spline as an illustration example due to its simplification and computational efficiency. Readers can extend our proposed SAPDEMI method to the higher-order spline with $q_{\max} > 2$ if they are interested in higher-order derivatives.

S4 Coordinate Gradient Descent to Solve the Optimization problem in Section 2.2.

In this section, we briefly review the implement of the coordinate descent algorithm in Friedman et al. (2010) to solve (2.10). The main idea of the coordinate descent is to update the estimator in a coordinate-wise fashion, which is the main difference between the coordinate descent and regular gradient descent. For instance, in the k-th iteration, the coordinate descent updates the iterative estimator $\boldsymbol{\beta}^{(k)}$ by using partial of the gradient information, instead of the whole gradient information. Mathematically speaking, in the k-th iteration, the coordinate descent optimizes $F(\boldsymbol{\beta}) = \frac{1}{2MN}\|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1$ with respective to $\boldsymbol{\beta}$ by

$$\beta_j^{(k+1)} = \arg\min_{\beta_j} F((\beta_1^{(k)}, \beta_2^{(k)}, \ldots, \beta_{j-1}^{(k)}, \beta_j, \beta_{j+1}^{(k)}, \ldots, \beta_K^{(k)}))$$

for all $j = 1, 2, \ldots, K$. To minimize the above optimization problem, we can derive the first derivative and set it as 0:

$$\frac{\partial}{\partial \beta_j} F(\boldsymbol{\beta}^{(k)}) = \frac{1}{MN}\left(\mathbf{e}_j^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}^{(k)} - \nabla_t \mathbf{u}^\top \mathbf{X}\mathbf{e}_j\right) + \lambda \text{sign}(\boldsymbol{\beta}_j) = 0,$$

where $\mathbf{e}_j$ is a vector of length K whose entries are all zero expect the j-th entry is 1. By solving the above equation, we can solve $\beta_j^{(k+1)}$ by

$$\beta_j^{(k+1)} = S\left(\nabla_t \mathbf{u}^\top \mathbf{X}\mathbf{e}_j - \sum_{l \neq j} (\mathbf{X}^\top \mathbf{X})_{jl} \boldsymbol{\beta}_l^{(k)}, MN\lambda\right) \Big/ (\mathbf{X}^\top \mathbf{X})_{jj},$$

where $S(\cdot)$ is the soft-thresholding function defined as

$$S(x, \alpha) = \begin{cases} x - \alpha & \text{if } x \geq \alpha \\ x + \alpha & \text{if } x \leq -\alpha \\ 0 & \text{otherwise} \end{cases}.$$

The detailed procedure of this algorithm is summarized in Algorithm 2.

Algorithm 2: Algorithm for the coordinate descent to minimize $F(\boldsymbol{\beta})$

Input: response vector $\nabla_t \mathbf{u}$, design matrix $\mathbf{X}$, and number of iterations M
Output: coefficient estimation $\widehat{\boldsymbol{\beta}}$
Initialize $\boldsymbol{\beta}^{(0)}$
for $\ell = 1, \ldots, \mathcal{L}$ **do**
  for $j = 1, \ldots, K$ **do**
    $\boldsymbol{\beta}_j^{(\ell)} = S\left(\nabla_t \mathbf{u}^\top \mathbf{X}\mathbf{e}_j - \sum_{l \neq j} (\mathbf{X}^\top \mathbf{X})_{jl} \boldsymbol{\beta}_l^{(\ell-1)}, MN\lambda\right) \Big/ (\mathbf{X}^\top \mathbf{X})_{jj}$
$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(\mathcal{L})}$

S5 A review of methods to select the smoothing parameter in the cubic spline literature

We consider a noisy data $\{(x_i, y_i)\}_{i=1,\ldots,n}$, where

$$y_i = g(x_i) + \epsilon_i,$$

with $\epsilon_i \sim N(0, \sigma^2)$. To fit this noisy data, the cubic spline use a spline function $s(x)$ to approximate $g(x)$. And the function $s(x)$ can be solved as the minimizer of the following optimization problem:

$$J_\lambda(s) = \frac{1}{n} \sum_{i=1}^{n} [y_i - s(x_i)]^2 + \lambda \int_{x_1}^{x_n} s''(x)^2 dx,$$

where the first term $\sum_{i=1}^{n} [y_i - s(x_i)]^2$ is the sum of squares for residuals. And this term is commonly called *infidelity* of the data. In the second term $\lambda \int_{x_1}^{x_n} s''(x)^2 dx$, the function $s''(x)$ is the second derivative of $s(x)$, and this term is the penalty of the smoothness. In the above optimization problem, the parameter $\lambda > 0$ controls the trade off between the goodness of fit and the smoothness of the cubic spline.

We will discuss the selection of λ under two scenarios: σ is known and σ is unknown.

- **Scenario 1: σ is known.** As suggested by Reinsch (1967), a good value of λ should be the one make the infidelity ($\frac{1}{n} \sum_{i=1}^{n} [y_i - s(x_i)]^2$) equals to σ^2, i.e.,

  $$\lambda^* = \left\{ \lambda : \frac{1}{n} \sum_{i=1}^{n} [y_i - s(x_i)]^2 = \sigma^2 \right\}.$$

  An alternative way to select λ is to choose the optimal λ which minimizes the true mean square error averaged over the data points (Wahba, 1975; Craven and Wahba, 1978). And the true mean square error $R(\lambda)$ is defined as

  $$R(\lambda) = \frac{1}{n} \sum_{i=1}^{n} [g(x_i) - s(x_i)]^2.$$

  So the optimal λ is

  $$\lambda^* = \arg\min_{\lambda} R(\lambda).$$

In practice, the above two approaches from Reinsch (1967); Wahba (1975); Craven and Wahba (1978) are not feasible, because σ is commonly unknown.

- **Scenario 2: σ is unknown.**

  The first representative method is from Mallows (2000); Hudson (1974), where the optimal λ^* is

$$\begin{aligned}\lambda^* &= \arg\min_{\lambda} E\left(\frac{1}{n}A(\lambda)y - g\right)\\ &= \arg\min_{\lambda} \frac{1}{n}\|[I - A(\lambda)]g\|^2 + \frac{\sigma^2}{n}\mathrm{tr}(A^2(\lambda)).\end{aligned}$$

  Here the matrix $A(\lambda) \in \mathbb{R}^{n\times n}$ depends on λ and is defined by the following equation:

$$\begin{pmatrix} s(x_1)\\ s(x_2)\\ \vdots\\ s(x_n)\end{pmatrix} = A(\lambda)\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}.$$

  In the above equation, the vectors y and g are defined $y = (y_1, \ldots, y_n)^\top$ and $g = (g(x_1), \ldots, g(x_n))^\top$. And the norm $\|\cdot\|^2$ is the Euclidean norm.

  The second representative method is generalized cross-validation (GCV) (Craven and Wahba, 1978; Aydin et al., 2013). Mathematically, it takes the optimal λ as the minimizer of $V(\lambda)$, i.e.,

$$\begin{aligned}\lambda^* &= \arg\min_{\lambda} V(\lambda)\\ &= \arg\min_{\lambda} \frac{1}{n}\|[I - A(\lambda)]y\|^2 \Big/ \left[\frac{1}{n}\mathrm{tr}(I - A(\lambda))\right]^2.\end{aligned}$$

S6 Some Important Lemmas

In this section, we present some important preliminaries, which are important blocks for the proofs of the main theories. To begin with, we first give the upper bound of $\widehat{u(x,t_n)} - u(x,t_n)$ for $x \in \{x_0, x_1, \ldots, x_{M-1}\}$, which is distance between the ground truth $u(x,t_n)$ and the estimated zero-order derivatives by cubic spline $\widehat{u(x,t_n)}$.

Lemma S6.1. Assume that

1. for any fixed $n = 0, 1, \ldots, N-1$, we have the spatial variable x is sorted in nondecreasing order, i.e., $x_0 < x_1 \ldots < x_{M-1}$;
2. for any fixed $n = 0, 1, \ldots, N-1$, we have the ground truth function $f^*(x) := u(x,t_n) \in C^4$, where C^4 refers to the set of functions that is forth-time differentiable;
3. for any fixed $n = 0, 1, \ldots, N-1$, we have $\frac{\partial^2}{\partial x^2}u(x_0,t_n) = \frac{\partial^2}{\partial x^2}u(x_{M-1},t_n) = 0$, and $\frac{\partial^3}{\partial x^3}u(x_0,t_n) \neq 0$, $\frac{\partial^3}{\partial x^3}u(x_{M-1},t_n) = 0$;
4. for any fixed $n = 0, 1, \ldots, N-1$, the value of third order derivative of function $f^*(x) := u(x,t_n)$ at point $x = 0$ is bounded, i.e., $\frac{d^3}{dx^3}f^*(0) < +\infty$;

5. for any U_i^n generated by the underlying PDE system $U_i^n = u(x_i, t_n) + w_i^n$ with $w_i^n \overset{i.i.d}{\sim} N(0, \sigma^2)$, we have $\eta^2 := \max_{i=0,\ldots,M-1, n=0,\ldots,N-1} E(U_i^n)^2$ is bounded;

6. for function $K(x) = \frac{1}{2} e^{-|x|/\sqrt{2}} \left[\sin(|x|\sqrt{2}) + \pi/4\right]$, we assume that it is uniformly continuous with modulus of continuity w_K and of bounded variation $V(K)$ and we also assume that $\int |K(x)|dx$, $\int |x|^{1/2}|dK(x)|$, $\int |x \log|x||^{1/2}|dK(x)|$ are bounded and denote

$$K_{\max} := \max_{x \in \max_{x \in [0, X_{\max}] \cup [0, T_{\max}]}} K(x);$$

7. the smoothing parameter in (2.4) is set as $\alpha = \left(1 + M^{-4/7}\right)^{-1}$;

8. the Condition 3.3 - Condition 3.4 hold.

Then there exist finite positive constant $\mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, C_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, \widetilde{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, Q_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, \gamma_{(M)} > 0, \omega_{(M)} > 1$, such that for any ϵ satisfying

$$\begin{aligned} \epsilon \quad > \quad \mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} \max \quad & \Bigg\{ 4K_{\max} M^{-3/7}, 4AM^{-3/7}, 4\sqrt{2} \tfrac{d^3}{dx^3} f^*(0) M^{-3/7}, \\ & \frac{16\left[C_{(\sigma, \|u\|_{L^\infty(\Omega)})} \log(M) + \gamma_{(M)}\right] \log(M)}{M^{3/7}}, \\ & 16\sqrt{\tfrac{\omega_{(M)}}{7}} \widetilde{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} \frac{\sqrt{\log(M)}}{M^{3/7}} \Bigg\}, \end{aligned}$$

there exist a $\dot{M} > 0$, such that when $M > \dot{M}$, we have

$$\begin{aligned} & P\left[\sup_{x \in [0, X_{\max}]} \left| \frac{\partial^k}{\partial x^k} \widehat{u(x, t_n)} - \frac{\partial^k}{\partial x^k} u(x, t_n) \right| > \epsilon \right] \\ < \quad & 2Me^{-\frac{(M^{3/7} - \|u\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma, \|u\|_{L^\infty(\Omega)})} e^{-L\gamma_{(M)}} + 4\sqrt{2}\eta^4 M^{-\omega_{(M)}/7} \end{aligned}$$

for $k = 0, 1, 2$. Here $A = \sup_\alpha \int |u|^s f_M(\alpha, u) du \times \int_{x \in [0, X_{\max}]} |K(x)| dx$.

Proof. See in Section S10.3 in this file. □

In the above lemma, we add $(\sigma, \|u\|_{L^\infty(\Omega)})$ as the subscript of constants $\mathscr{C}, C, \widetilde{C}, Q$ to emphasize that these constant are independent of the temporal resolution N and spatial resolution M, and only depends on the noisy data $\mathcal{D}$ in (1.1) itself. We add M as the subscript of constants γ, ω to emphasize that γ, ω are function of the spatial resolution M, and we will discuss the value of γ, ω in Lemma S6.2.

The above lemma show the closeness between $\frac{\partial^k}{\partial x^k} \widehat{u(x, t_n)}$ and $\frac{\partial^k}{\partial x^k} u(x, t_n)$ for $k = 0, 1, 2$. This results can be easily extend of the closeness between $\frac{\partial}{\partial t} \widehat{u(x_i, t)}$ and $\frac{\partial}{\partial t} u(x_i, t)$, which is shown in the following corollary.

Corollary S6.1. Assume that

1. for any fixed $i = 0, 1, \ldots, M-1$, we have the spatial variable t is sorted in nondecreasing order, i.e., $t_0 < t_1 \ldots < t_{N-1}$;

2. for any fixed $i = 0, 1, \ldots, M-1$, we have the ground truth function $f^*(t) := u(x_i, t) \in C^4$, where C^4 refers to the set of functions that is forth-time differentiable;

3. for any fixed $i = 0, 1, \ldots, M-1$, we have $\frac{\partial^2}{\partial t^2} u(x_i, t_0) = \frac{\partial^2}{\partial t^2} u(x_i, t_{N-1}) = 0$, and $\frac{\partial^3}{\partial t^3} u(x_i, t_0) \neq 0$, $\frac{\partial^3}{\partial t^3} u(x_i, t_{N-1}) = 0$;
4. for any fixed $i = 0, 1, \ldots, M-1$, the value of third order derivative of function $\bar{f}^*(x) := u(x_i, t)$ at point $t = 0$ is bounded, i.e., $\frac{d^3}{dt^3} \bar{f}^*(0) < +\infty$;
5. for any U_i^n generated by the underlying PDE system $U_i^n = u(x_i, t_n) + w_i^n$ with $w_i^n \overset{i.i.d}{\sim} N(0, \sigma^2)$, we have $\max_{i=0,\ldots,M-1, n=0,\ldots,N-1} E(U_i^n)^2$ is bounded;
6. for function $K(x) = \frac{1}{2} e^{-|x|/\sqrt{2}} \left[\sin(|x|\sqrt{2}) + \pi/4\right]$, we have $K(x)$ is uniformly continuous with modulus of continuity w_K and of bounded variation $V(K)$, and we also assume that $\int_{x \in [0, X_{\max}]} |K(x)| dx$, $\int |x|^{1/2} |dK(x)|$, $\int |x \log |x||^{1/2} |dK(x)|$ are bounded and denote $K_{\max} := \max_{x \in [0, X_{\max}] \cup [0, T_{\max}]} K(x)$;
7. the smoothing parameter in (2.4) is set as $\bar{\alpha} = O\left(\left(1 + N^{-4/7}\right)^{-1}\right)$;
8. the Condition 3.3 - Condition 3.4 hold.

then there exist finite positive constant $\mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, C_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, \widetilde{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, Q_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0, \gamma_{(N)} > 0, \omega_{(N)} > 1$, such that for any ϵ satisfying

$$\epsilon > \mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} \max \left\{ 4K_{\max} N^{-3/7}, 4\bar{A} N^{-3/7}, 4\sqrt{2} \frac{d^3}{dx^3} f^*(0) N^{-3/7}, \frac{16\left[C_{(\sigma, \|u\|_{L^\infty(\Omega)})} \log(N) + \gamma_{(N)}\right] \log(N)}{N^{3/7}}, 16\sqrt{\frac{\omega_{(N)}}{7}} \widetilde{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} \frac{\sqrt{\log(N)}}{N^{3/7}} \right\},$$

there exist a $\dot{N} > 0$, such that when $N > \dot{N}$, we have

$$P\left[\sup_{t \in [0, T_{\max}]} \left| \frac{\partial}{\partial t} \widehat{u(x_i, t)} - \frac{\partial}{\partial t} u(x_i, t) \right| > \epsilon \right] < 2Ne^{-\frac{(N^{3/7} - \|u\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma, \|u\|_{L^\infty(\Omega)})} e^{-L\gamma_{(N)}} + 4\sqrt{2}\eta^4 N^{-\omega_{(N)}/7}.$$

Here $\bar{A} = \sup_\alpha \int |u|^s \bar{f}_N(\alpha, u) du \times \int_{t \in [0, T_{\max}]} |K(x)| dx$.

After bounding the error of all the derivatives, we then aim to bound $\|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty$. It is important to bound $\|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty$, with the reason described as follows in Lemma S6.2.

Lemma S6.2. Suppose the conditions in Lemma S6.1 and Corollary S6.1 hold and we set $M = O(N)$, then there exist finite positive constant $\mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} > 0$ such that for any ϵ satisfying

$$\epsilon > \mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})} \frac{\log(N)}{N^{3/7 - r}},$$

and any $r \in \left(0, \frac{3}{7}\right)$, there exist $\dot{N} > 0$, such that when $N > \dot{N}$, we have

$$P\left(\|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty > \epsilon\right) < Ne^{-N^r},$$

where $\mathscr{C}_{(\sigma, \|u\|_{L^\infty(\Omega)})}$ is a constant which do not depend on the temporal resolution M and spatial resolution N.

Proof. See Section S10. □

S7 Justification of $\alpha, \bar{\alpha}$ in Theorem 3.1

We acknowledge that our way to select the smoothing parameters $\alpha, \bar{\alpha}$ is different from that in the cubic spline literature (see a detailed literature review in the supplementary material). The root cause of the difference lies in the different objectives in theory. For the existing methods in the cubic spline literature, the objective is to minimize the fitting error when one fits the data (similar to *single objective optimization*). However, for our proposed SAPDEMI method, the objective is to maximize the accuracy when one identifies the underlying PDE models. To build a path to this objective, we apply the cubic spline as an important block. And the selection of the smoothing parameter is required to, on the one hand, have a relatively small fitting error; on the other hand, leads to a high identification accuracy (similar to *multiple objective optimizations*).

S8 Tables to draw the curve in Fig. 2 and Fig. 8

In this section, we present the table to draw the curves in Fig. 2,8 in Table 1, 2, respectively.

Table 1: Computational complexity of the functional estimation by cubic spline and local polynomial regression in transport equation

	M = 20						
	N=200	N=400	N=800	N=1000	N=1200	N=1600	N=2000
cubic spline	374,389	748,589	1,496,989	1,871,189	2,245,389	2,993,789	3,742,189
local poly	14,136,936	45,854,336	162,089,136	246,606,536	348,723,936	605,758,736	933,193,536
	N = 20						
	M=200	M=400	M=800	M=1000	M=1200	M=1600	M=2000
cubic spline	398,573	875,773	207,0173	2,787,373	3,584,573	5,418,973	7,573,373
local poly	33,046,336	125,596,136	489,255,736	760,365,536	1,090,995,336	1,930,814,936	3,008,714,536

Table 2: Correct identification probability of transport equation, inviscid Burgers equation and viscous Burgers's equation

	σ											
	0.01	0.05	0.1	0.25	0.3	0.4	0.5	0.7	0.75	0.8	0.9	1
	transport equation											
$M = N = 100$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
$M = N = 150$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
$M = N = 200$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	inviscid Burgers equation											
$M = N = 100$	100%	100%	100%	100%	100%	100%	100%	99.9%	99.8%	99.8%	99.8%	99.1%
$M = N = 150$	100%	100%	100%	100%	100%	100%	100%	100%	99.8%	99.7%	99.7%	99.7%
$M = N = 200$	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	viscous Burgers equation											
$M = N = 100$	100%	99.4%	89.8%	78.0%	71.4%	82.0%	91.6%	72.8%	79.0%	72.9%	57.9%	51.3%
$M = N = 150$	100%	100%	100%	97.3%	96.5%	96.2%	97.6%	95.6%	93.3%	86.6%	79.9%	73.6%
$M = N = 200$	100%	100%	100%	100%	99.6%	99.6%	98.2%	98.8%	98.2%	97.0%	94.3%	91.3%

[1] The simulation results are based on 1000 times of simulations.

S9 The reasons why the RK4 is not feasible.

In this section, we discuss the reasons why the RK4 is not feasible. Generally speaking, RK4 is used to approximate solutions of ordinary differential equations. In our content, it aims at

solveing the solution of the following differential equation with fixed $i \in \{0, \ldots, M-1\}$:

$$\begin{cases} \frac{\partial}{\partial t} u(x_i, t) = p(u, t) \\ u(x_i, t_0) = u_i^0 \end{cases}, \tag{S9.8}$$

where $p(u, t)$ is the function interpolated through data set

$$\left\{ t_n, u_i^n, \frac{u(x_i, t_n + \Delta t) - u(x_i, t_n)}{\Delta t} \right\}_{n=0,\ldots,N-2}.$$

Then as shown by Chapter 5 in Lambert et al. (1991), the solution can be approximate by

$$u(x_i, t_{n+1}) = u(x_i, t_n) + \frac{\Delta t}{6}(k_1 + k_2 + k_3 + k_4),$$

where $\Delta t = t_{n+1} - t_n$ and k_1, k_2, k_3, k_4 are defined as

$$\begin{cases} k_1 = p(t_n, u_i^n) \\ k_2 = p(t_n + \Delta t/2, u_i^n + k_1 \Delta t/2) \\ k_3 = p(t_n + \Delta t/2, u_i^n + k_2 \Delta t/2) \\ k_4 = p(t_n + \Delta t, u_i^n + k_3 \Delta t). \end{cases} \tag{S9.9}$$

Given the above implementation of RK4, we find it is infeasible to be used in our case study due to the following two reasons. First, it is infeasible to obtain $p(u, t)$ in (S9.8), even though the interpolation methods. Second, it is infeasible to get the value of k_3 in (S9.9). Because the calculation of k_3 depends on the value of k_2 and k_2 needs to be obtained (at least) by interpolation, it is complicated to calculate k_1, k_2, k_3, k_4 through one-time operation. Given the complicated implementation of RK4, we use the explicit Euler method in our case study.

S10 Proofs

S10.1 Proof of Proposition 2.1

Proof. The computational complexity in the functional estimation stage lies in calculating all elements in matrix $\mathbf{X}$ and vector $\nabla_t \mathbf{u}$, including

$$\left\{ \widehat{u(x_i, t_n)}, \frac{\partial}{\partial x} \widehat{u(x_i, t_n)}, \frac{\partial^2}{\partial x^2} \widehat{u(x_i, t_n)}, \frac{\partial}{\partial t} \widehat{u(x_i, t_n)} \right\}_{i=0,\ldots,M-1, n=0,\ldots,N-1}.$$

by cubic spline in (2.4).

We divide our proof into two scenarios: (1) $\alpha = 1$ and (2) $\alpha \in (0, 1)$.

- First of all, we discuss a very simple case, i.e., $\alpha = 1$. When $\alpha = 1$, we call the cubic spline as *interpolating cubic spline* since there is no penalty on the smoothness.

  For the zero-order derivative, i.e., $\left\{ \widehat{u(x_i, t_n)} \right\}_{i=0,\ldots,M-1, n=0,\ldots,N-1}$, it can be estimated as $\widehat{u(x_i, t_n)} = u_i^n$ for $i = 0, 1, \ldots, M-1, n = 0, 1, \ldots, N-1$. So there is no computational complexity involved.

  For the second order derivatives, i.e., $\left\{ \frac{\partial^2}{\partial x^2} \widehat{u(x_i, t_n)} \right\}_{i=0,\ldots,M-1}$, with $n \in \{0, \ldots, N-1\}$ fixed, it can be solved in a closed-form, i.e.,

  $$\widehat{\boldsymbol{\sigma}} = \mathbf{M}^{-1} \mathbf{A} \mathbf{u}_{:}^n$$

where $\widehat{\boldsymbol{\sigma}} = \left(\frac{\partial^2}{\partial x^2}\widehat{u(x_0,t_n)}, \frac{\partial^2}{\partial x^2}\widehat{u(x_1,t_n)}, \ldots, \frac{\partial^2}{\partial x^2}\widehat{u(x_{M-1},t_n)}\right)^\top$. So the main computational load lies in the calculation of $\mathbf{M}^{-1}$. Recall $\mathbf{M} \in \mathbb{R}^{(M-2)\times(M-2)}$ is a tri-diagonal matrix:

$$\mathbf{M} = \begin{pmatrix} \frac{h_0+h_1}{3} & \frac{h_1}{6} & 0 & \ldots & 0 & 0 \\ \frac{h_1}{6} & \frac{h_1+h_2}{3} & \frac{h_2}{6} & \ldots & 0 & 0 \\ 0 & \frac{h_2}{6} & \frac{h_2+h_3}{3} & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & \frac{h_{M-4}+h_{M-3}}{3} & \frac{h_{M-3}}{6} \\ 0 & 0 & 0 & \ldots & \frac{h_{M-3}}{6} & \frac{h_{M-3}+h_{M-2}}{3} \end{pmatrix}.$$

For this type of tri-diagonal matrix, there exist a fast algorithm to calculate its inverse. The main idea of this fast algorithm is to decompose $\mathbf{M}$ through Cholesky decomposition as

$$\mathbf{M} = \mathbf{L}\mathbf{D}\mathbf{L}^\top,$$

where $\mathbf{L} \in \mathbb{R}^{(M-2)\times(M-2)}, \mathbf{D} \in \mathbb{R}^{(M-2)\times(M-2)}$ has the form of

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ l_1 & 1 & 0 & \ldots & 0 \\ 0 & l_2 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & l_{M-3} & 1 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{M-2} \end{pmatrix}.$$

After decomposing matrix $\mathbf{M}$ into $\mathbf{L}\mathbf{D}\mathbf{L}^\top$, the second derivatives $\widehat{\boldsymbol{\sigma}}$ can be solved as

$$\widehat{\boldsymbol{\sigma}} = (\mathbf{L}^\top)^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1}\underbrace{\mathbf{A}\mathbf{u}^n_{\vdots}}_{\boldsymbol{\xi}}.$$

In the remaining of the proof in this scenario, we will verify the following two issues:

1. the computational complexity to decompose $\mathbf{M}$ into $\mathbf{L}\mathbf{D}\mathbf{L}^\top$ is $O(M)$ with $n \in \{0, \ldots, N-1\}$ fixed;

2. the computational complexity to compute $\widehat{\boldsymbol{\sigma}} = (\mathbf{L}^\top)^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1}\boldsymbol{\xi}$ is $O(M)$ with $n \in \{0, \ldots, N-1\}$ fixed and $\mathbf{L}, \mathbf{D}$ available.

For the decomposition of $\mathbf{M} = \mathbf{L}\mathbf{D}\mathbf{L}^\top$, its essence is to derive $l_1, \ldots l_{M-3}$ in matrix $\mathbf{L}$ and $d_1, \ldots, d_{M-2}$ in matrix $\mathbf{D}$. By utilizing the method of undetermined coefficients to

inequality $\mathbf{M} = \mathbf{L}\mathbf{D}\mathbf{L}^\top$, we have:

$$
\begin{bmatrix}
d_1 & d_1 l_1 & 0 & \dots & 0 & 0 \\
d_1 l_1 & d_2 & d_2 l_2 & \dots & 0 & 0 \\
0 & d_2 l_2 & d_3 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \dots & d_{M-3} l_{M-3} & d_{M-3} l_{M-3}^2 + d_{M-2}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
M_{11} & M_{12} & \dots & 0 \\
M_{21} & M_{22} & \dots & 0 \\
0 & M_{32} & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \dots & M_{M-2,M-2}
\end{bmatrix},
$$

where $M_{i,j}$ is the (i,j)th entry in matrix $\mathbf{M}$. Through the above method of undetermined coefficients, we can solve the exact value of the entries in matrix $\mathbf{L}, \mathbf{D}$, which is summarized in Algorithm 3. It can be seen from Algorithm 3 that, the computational complexity of solve $\mathbf{L}, \mathbf{D}$ is of order $O(M)$.

For the calculation of $\widehat{\boldsymbol{\sigma}} = (\mathbf{L}^\top)^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1}\boldsymbol{\xi}$ with matrix $\mathbf{L}, \mathbf{D}$ available, we will first verify that the computational complexity to solve $\bar{\boldsymbol{\xi}} = \mathbf{L}^{-1}\boldsymbol{\xi}$ is $O(M)$. Then, we will verify that the computational complexity to solve $\bar{\bar{\boldsymbol{\xi}}} = \mathbf{D}^{-1}\bar{\boldsymbol{\xi}}$ is $O(M)$. Finally, we will verify that the computational complexity to solve $\bar{\bar{\bar{\boldsymbol{\xi}}}} = (\mathbf{L}^\top)^{-1}\bar{\bar{\boldsymbol{\xi}}}$ is $O(M)$. First, the computational complexity of calculating $\bar{\boldsymbol{\xi}} = \mathbf{L}^{-1}\boldsymbol{\xi}$ is $O(M)$, this is because by $\mathbf{L}\bar{\boldsymbol{\xi}} = \boldsymbol{\xi}$, we have the following system of equations:

$$
\begin{cases}
\xi_1 = \bar{\xi}_1 \\
\xi_2 = \bar{\xi}_2 + l_1 \bar{\xi}_1 \\
\vdots \\
\xi_{M-2} = \bar{\xi}_{M-2} + l_{M-3}\bar{\xi}_{M-3},
\end{cases}
$$

where $\xi_i, \bar{\xi}_i$ is the i-th entry in $\boldsymbol{\xi}, \bar{\boldsymbol{\xi}}$,respectively. Through the above system of equations, we can solve the values of all entries in $\bar{\boldsymbol{\xi}}$, which is summarized in Algorithm 4. From Algorithm 4, we know that the computational complexity of solving $\mathbf{L}^{-1}\boldsymbol{\xi}$ is $O(M)$. Next, it is obvious that the computational complexity of $\bar{\bar{\boldsymbol{\xi}}} = \mathbf{D}^{-1}\bar{\boldsymbol{\xi}}$ is $O(M)$, because $\mathbf{D}$ is a diagonal matrix. Finally, with the similar logic flow, we can verify that the computational complexity of $\bar{\bar{\bar{\boldsymbol{\xi}}}} = (\mathbf{L}^\top)^{-1}\bar{\bar{\boldsymbol{\xi}}}$ is still $O(M)$. So, the computational complexity is to calculate $\widehat{\boldsymbol{\sigma}} = (\mathbf{L}^\top)^{-1}\mathbf{D}^{-1}\mathbf{L}^{-1}\boldsymbol{\xi}$, with known $\mathbf{L}, \mathbf{D}$ is $O(M)$.

As a summary, the computational complexity is to calculate $\left\{\widehat{\frac{\partial^2}{\partial x^2} u(x_i, t_n)}\right\}_{i=0,\dots,M-1}$ with a fixed $n \in \{0, 1, \dots, N-1\}$ is $O(M)$. Accordingly, the computational complexity to solve $\left\{\widehat{\frac{\partial^2}{\partial x^2} u(x_i, t_n)}\right\}_{i=0,\dots,M-1, n=0,\dots,N-1}$ is $O(MN)$.

For the first order derivatives, i.e., $\left\{\frac{\partial}{\partial x} u(x_i, t_n), \frac{\partial}{\partial x} u(x_i, t_n)\right\}_{i=0,\dots,M-1, n=0,\dots,N-1}$, we can verify the computational complexity to solve them is also $O(MN)$ with the similar logic as that in the second order derivatives.

Algorithm 3: Pseudo code to solve $\mathbf{L}, \mathbf{D}$

Input: matrix $\mathbf{M}$
Output: matrix $\mathbf{L}, \mathbf{D}$
Initialize $d_1 = M_{1,1}$
for $i = 1, 2, \ldots, M-3$ **do**
  $l_i = M_{i,i+1}/d_i$
  $d_{i+1} = M_{i+1,i+1} - d_i l_i^2$

Algorithm 4: Pseudo code to solve $\mathbf{L}^{-1}\boldsymbol{\xi}$

Input: matrix $\mathbf{L}, \boldsymbol{\xi}$
Output: matrix $\bar{\boldsymbol{\xi}}$
Initialize $\bar{\xi}_1 = \xi_1$
for $i = 2, \ldots, M-2$ **do**
  $\bar{\xi}_i = \xi_i - l_{i-1}\bar{\xi}_{i-1}$

- Next, we discuss the scenario when $\alpha \in (0,1)$.

  Since all the derivatives has similar closed-form formulation as shown in (2.5), (S2.7), (S2.6), we take the zero-order derivative $\{u(x_i, t_n)\}_{i=0,\ldots,M-1,n=0,\ldots,N-1}$ as an illustration example, and other derivatives can be derived similarly.

  Recall that in Section 2.1, the zero-order derivative $\{u(x_i, t_n)\}_{i=0,\ldots,M-1}$ with $n \in \{0, 1, \ldots, N-1\}$ fixed can be estimated through cubic spline as in equation (2.5):

$$\widehat{\mathbf{f}} = [\underbrace{\alpha\mathbf{W} + (1-\alpha)\mathbf{A}^\top\mathbf{M}\mathbf{A}}_{\mathbf{Z}}]^{-1}\underbrace{\alpha\mathbf{W}\mathbf{u}_{:}^{n}}_{\mathbf{y}},$$

  where $\alpha \in (0,1)$ trades off the fitness of the cubic spline and the smoothness of the cubic spline, vector $\widehat{\mathbf{f}} = \left(\widehat{u(x_0, t_n)}, \widehat{u(x_1, t_n)}, \ldots, \widehat{u(x_{M-1}, t_n)}\right)^\top$, vector $\mathbf{u}_{:}^{n} = (u_0^n, \ldots, u_{M-1}^n)^\top$, matrix $\mathbf{W} = \mathrm{diag}(w_0, w_1, \ldots, w_{M-1})$, matrix $\mathbf{A} \in \mathbb{R}^{(M-2)\times M}, \mathbf{M} \in \mathbb{R}^{(M-2)\times(M-2)}$ are defined as

$$\mathbf{A} = \begin{pmatrix} \frac{1}{h_0} & -\frac{1}{h_0} - \frac{1}{h_1} & \frac{1}{h_1} & 0 & \ldots & 0 & 0 & 0 \\ 0 & \frac{1}{h_1} & -\frac{1}{h_1} - \frac{1}{h_2} & \frac{1}{h_2} & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \frac{1}{h_{M-3}} & -\frac{1}{h_{M-3}} - \frac{1}{h_{M-2}} & \frac{1}{h_{M-2}} \end{pmatrix},$$

$$\mathbf{M} = \begin{pmatrix} \frac{h_0+h_1}{3} & \frac{h_1}{6} & 0 & \ldots & 0 & 0 \\ \frac{h_1}{6} & \frac{h_1+h_2}{3} & \frac{h_2}{6} & \ldots & 0 & 0 \\ 0 & \frac{h_2}{6} & \frac{h_2+h_3}{3} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \frac{h_{M-4}+h_{M-3}}{3} & \frac{h_{M-3}}{6} \\ 0 & 0 & 0 & \ldots & \frac{h_{M-3}}{6} & \frac{h_{M-3}+h_{M-2}}{3} \end{pmatrix}$$

  with $h_i = x_{i+1} - x_i$ for $i = 0, 1, \ldots, M-2$.

By simple calculation, we know that matrix $\mathbf{Z} = \alpha\mathbf{W} + (1-\alpha)\mathbf{A}^\top\mathbf{M}\mathbf{A} \in \mathbb{R}^{M\times M}$ is a symmetric seventh-diagonal matrix:

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & z_{13} & z_{14} & 0 & \dots \\ z_{21} & z_{22} & z_{23} & z_{24} & z_{25} & \dots \\ z_{31} & z_{32} & z_{33} & z_{34} & z_{35} & \ddots \\ z_{41} & z_{42} & z_{43} & z_{44} & z_{45} & \ddots \\ 0 & z_{52} & z_{53} & z_{54} & z_{55} & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

By applying Cholesky decomposition to matrix $\mathbf{Z}$ as $\mathbf{Z} = \mathbf{P}\mathbf{\Sigma}\mathbf{P}^\top$, we can calculate $\widehat{\mathbf{f}}$ as

$$\widehat{\mathbf{f}} = \mathbf{Z}^{-1}\mathbf{y} = (\mathbf{P}^\top)^{-1}\mathbf{\Sigma}^{-1}\mathbf{P}^{-1}\mathbf{y},$$

where $\mathbf{P} \in \mathbb{R}^{M\times M}, \mathbf{\Sigma} \in \mathbb{R}^{M\times M}$ has the form of

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ \ell_1 & 1 & 0 & 0 & \dots & 0 & 0 \\ \gamma_1 & \ell_2 & 1 & 0 & \dots & 0 & 0 \\ \eta_1 & \gamma_2 & \ell_3 & 1 & \dots & 0 & 0 \\ 0 & \eta_2 & \gamma_3 & \ell_4 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & 1 & 0 \\ 0 & 0 & \dots & \eta_{M-3} & \gamma_{M-2} & \ell_{M-1} & 1 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} s_1 & 0 & 0 & \dots & 0 \\ 0 & s_2 & 0 & \dots & 0 \\ 0 & 0 & s_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \dots & s_M \end{pmatrix}.$$

In the remaining of the proof in this scenario, we will verify the following two issues:

1. the computational complexity to decompose $\mathbf{Z}$ into $\mathbf{P}\mathbf{\Sigma}\mathbf{P}^\top$ is $O(M)$ with $n \in \{0, \dots, N-1\}$ fixed;
2. the computational complexity to compute $(\mathbf{P}^\top)^{-1}\mathbf{\Sigma}^{-1}\mathbf{P}^{-1}\mathbf{y}$ is $O(M)$ with $n \in \{0, \dots, N-1\}$ fixed.

First of all, we verify that the computational complexity to decompose $\mathbf{Z}$ into $\mathbf{P}\mathbf{\Sigma}\mathbf{P}^\top$ is $O(M)$ when $n \in \{0, \dots, N-1\}$ fixed By applying method of undetermined coefficients to equality $\mathbf{Z} = \mathbf{P}\mathbf{\Sigma}\mathbf{P}^\top$, we have

$$\begin{bmatrix} s_1 & s_1\ell_1 & s_1\gamma_1 & \dots & 0 \\ s_1\ell_1 & s_1\ell_1^2 + s_2 & s_1\ell_1\gamma_1 + s_2\ell_2 & \dots & 0 \\ s_1\gamma_1 & s_1\ell_1\gamma_1 + s_2\ell_2 & s_1\gamma_1^2 + s_2\ell_2^2 + s_3 & \dots & 0 \\ s_1\eta_1 & s_1\eta_1\ell_1 + s_2\gamma_2 & s_1\eta_1\gamma_1 + s_2\gamma_2\ell_2 + s_3\ell_3 & \dots & 0 \\ 0 & s_2\eta_2 & s_2\eta_2\ell_2 + s_3\gamma_3 & \dots & 0 \\ 0 & 0 & s_3\eta_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \begin{matrix} s_{M-3}\eta_{M-3}^2 + s_{M-2}\gamma_{M-2}^2 \\ +s_{M-1}\gamma_{M-1}\ell_{M-1}^2 + s_M \end{matrix} \end{bmatrix} = [z_{i,j}],$$

where $[z_{ij}]$ denotes matrix $\mathbf{Z}$ with its (i,j)th entry as $z_{i,j}$. Through the above method of undetermined coefficients, we can solve the explicit value of all entries in matrix $\mathbf{P}, \mathbf{\Sigma}$, i.e.,

$\ell_1, \ldots, \ell_{M-1}, \gamma_1, \ldots, \gamma_{M-2}, \eta_1, \ldots, \eta_{M-3}$ in matrix $\mathbf{P}$ and $s_1, \ldots, s_M$ in matrix $\mathbf{\Sigma}$, which is summarized in Algorithm 5. From Algorithm 5, we can see that the computational complexity to decompose $\mathbf{Z}$ into $\mathbf{P\Sigma P}^\top$ is $O(M)$ with $n \in \{0, \ldots, N-1\}$ fixed.

Second, we verify the computational complexity to compute $(\mathbf{P}^\top)^{-1}\mathbf{\Sigma}^{-1}\mathbf{P}^{-1}\mathbf{y}$ is $O(M)$ with $n \in \{0, \ldots, N-1\}$ fixed and matrix $\mathbf{P}, \mathbf{\Sigma}$ available. To realize this objective, we will first verify that the computational complexity to calculate $\bar{\mathbf{y}} = \mathbf{P}^{-1}\mathbf{y}$ is $O(M)$. Then, we will first verify that the computational complexity to calculate $\bar{\bar{\mathbf{y}}} = \mathbf{\Sigma}^{-1}\bar{\mathbf{y}}$ is $O(M)$. Finally, we will first verify that the computational complexity to calculate $\bar{\bar{\bar{\mathbf{y}}}} = (\mathbf{P}^\top)^{-1}\bar{\bar{\mathbf{y}}}$ is $O(M)$. First of all, let us verify the computational complexity to compute $\bar{\mathbf{y}} = \mathbf{P}^{-1}\mathbf{y}$ is $O(M)$ with $n \in \{0, \ldots, N-1\}$ fixed. Because we have a system of equations derived from $\mathbf{P}\bar{\mathbf{y}} = \mathbf{y}$:

$$\begin{cases} \bar{y}_1 &= y_1 \\ \bar{y}_2 &= y_2 - \ell_1\bar{y}_1 \\ \bar{y}_3 &= y_3 - \gamma_1\bar{y}_1 - \ell_2\bar{y}_2 \\ \bar{y}_4 &= y_4 - \eta_1\bar{y}_1 - \gamma_2\bar{y}_2 - \ell_3\bar{y}_3 \\ \bar{y}_5 &= y_5 - \eta_2\bar{y}_3 - \gamma_3\bar{y}_3 - \ell_4\bar{y}_4 \\ \vdots & \\ \bar{y}_M &= y_M - \eta_{M-3}\bar{y}_{M-3} - \gamma_{M-2}\bar{y}_{M-2} - \ell_{M-1}\bar{y}_{M-1}, \end{cases}$$

we can solve vector $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_M)^\top$ explicitly through Algorithm 6, which only requires $O(M)$ computational complexity. After deriving $\bar{\mathbf{y}} = \mathbf{P}^{-1}\mathbf{y}$, we can easily verify that the computational complexity to derive $\bar{\bar{\mathbf{y}}} = \mathbf{\Sigma}^{-1}\bar{\mathbf{y}}$ is still $O(M)$ because $\boldsymbol{\sigma}$ is a diagonal matrix. Finally, after deriving $\bar{\bar{\mathbf{y}}} = \mathbf{\Sigma}^{-1}\bar{\mathbf{y}}$, we can verify that the computational complexity to derive $\bar{\bar{\bar{\mathbf{y}}}} = (\mathbf{P}^\top)^{-1}\bar{\bar{\mathbf{y}}}$ is still $O(M)$ with the similar logic as that in $\bar{\mathbf{y}} = \mathbf{P}^{-1}\mathbf{y}$.

From the above discussion, we know that the computational complexity to calculate $\widehat{\mathbf{f}} = (\widehat{u(x_0, t_n)}, \widehat{u(x_1, t_n)}, \ldots, \widehat{u(x_{M-1}, t_n)})^\top$, is $O(M)$ with $n \in \{0, 1, \ldots, N-1\}$ fixed. In other words, the computational complexity to derive $\left\{\widehat{u(x_i, t_n)}\right\}_{i=0,\ldots,M-1}$ is $O(M)$. According, the computational complexity to derive $\left\{\widehat{u(x_i, t_n)}\right\}_{i=0,\ldots,M-1, n=0,\ldots,N-1}$ is $O(MN)$.

Algorithm 5: Pseudo code to solve $\mathbf{P}, \mathbf{\Sigma}$

Input: matrix $\mathbf{Z}$
Output: matrix $\mathbf{P}, \mathbf{\Sigma}$
Initialize $s_j = \eta_j = \gamma_j = \ell_j = 0 \; \forall j \leq 0$
for $i = 1, 2, \ldots, M$ **do**
  $s_i = z_{ii} - s_{i-3}\eta_{i-3}^2 - s_{i-2}\gamma_{i-2}^2 - s_{i-1}\ell_{i-1}^2$
  $\ell_i = (z_{i,i+1} - s_{i-2}\gamma_{i-2}\eta_{i-2} - s_{i-1}\gamma_{i-1}\ell_{i-1})/s_i$
  $\eta_i = a_{i,i+3}/s_i$

Algorithm 6: Pseudo code to solve $\mathbf{P}^{-1}\mathbf{y}$

Input: matrix $\mathbf{P}, \mathbf{y}$
Output: vector $\bar{\mathbf{y}}$
Initialize $\eta_i = \gamma_i = \ell_i = 0 \; \forall i \leq 0$
for $i = 1, \ldots, M$ **do**
  $\bar{y}_i = y_i - \eta_{i-3}\bar{y}_{i-3} - \gamma_{i-2}\bar{y}_{i-2} - \ell_{i-1}\bar{y}_{i-1}$

□

S10.2 Proof of Proposition S3.1

Proof. We have discussed how to use cubic spline to derive derivatives of $u(x,t)$. In this section, we discuss how to use local polynomial regression to derive derivatives, as a benchmark method.

Recall that the derivatives can be estimated by local polynomial regression includes $u(x_i,t_n)$, $\frac{\partial}{\partial x}u(x_i,t_n)$, $\frac{\partial^2}{\partial x^2}u(x_i,t_n), \ldots$. And here we take the derivation $\frac{\partial^l}{\partial x^l}u(x,t_n)$ as an example ($l = 0,1,2,\ldots$), and the other derivatives can be derived with the same logic flow. To derive the estimation of $\frac{\partial^l}{\partial x^l}u(x,t_n)$, we fix the temporal variable t_n for a general $n \in \{0,1,\ldots,N-1\}$. Then we locally fit a degree $\check{p}$ polynomial over the data $\{(x_i,u_i^n)\}_{i=0,\ldots,M-1}$, i.e.,

$$\begin{cases} u(x_0,t_n) &= u(x,t_n) + \frac{\partial}{\partial x}u(x,t_n)(x_0-x) + \ldots + \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)(x_0-x)^{\check{p}} \\ u(x_1,t_n) &= u(x,t_n) + \frac{\partial}{\partial x}u(x,t_n)(x_1-x) + \ldots + \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)(x_1-x)^{\check{p}} \\ \vdots & \vdots \\ u(x_{M-1},t_n) &= u(x,t_n) + \frac{\partial}{\partial x}u(x,t_n)(x_{M-1}-x) + \ldots + \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)(x_{M-1}-x)^{\check{p}} \end{cases}.$$

For the choice of $\check{p}$, we choose $\check{p} = l+3$ to realize minmax efficiency (see Fan et al., 1997). If we denote $\mathbf{b}(x) = \left(u(x,t_n), \frac{\partial}{\partial x}u(x,t_n), \ldots, \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)\right)^\top$, then $\frac{\partial^l}{\partial x^l}u(x,t_n)$ can be obtained as the $(l+1)$-th entry of the vector $\widehat{\mathbf{b}(x)}$, and $\widehat{\mathbf{b}(x)}$ is obtained by the following optimization problem:

$$\widehat{\mathbf{b}(x)} = \arg\min_{\mathbf{b}(x)} \sum_{i=0}^{M-1} \left[u_i^n - \sum_{j=0}^{\check{p}} \frac{\partial^j}{\partial x^j}u(x,t_n)(x_i-x)^j \right]^2 \mathcal{K}\left(\frac{x_i-x}{h}\right), \tag{S10.10}$$

where h is the bandwidth parameter, and $\mathcal{K}$ is a kernel function, and in our paper, we use the Epanechikov kernel $\mathcal{K}(x) = \frac{3}{4}\max\{0, 1-x^2\}$ for $x \in \mathbb{R}$. Essentially, the optimization problem in equation (S10.10) is a weighted least squares model, where $\mathbf{b}(x)$ can be solved in a close form:

$$\mathbf{b}(x) = \left(\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{X}_{\text{spa}}\right)^{-1} \mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{u}_{:}^n, \tag{S10.11}$$

where

$$\mathbf{X}_{\text{spa}} = \begin{bmatrix} 1 & (x_0-x) & \ldots & (x_0-x)^{\check{p}} \\ 1 & (x_1-x) & \ldots & (x_1-x)^{\check{p}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_{M-1}-x) & \ldots & (x_{M-1}-x)^{\check{p}} \end{bmatrix}, \mathbf{u}_{:}^n = \begin{bmatrix} u_0^n \\ u_1^n \\ \vdots \\ u_{M-1}^n \end{bmatrix}$$

and $\mathbf{W}_{\text{spa}} = \text{diag}\left(\mathcal{K}\left(\frac{x_0-x}{h}\right), \ldots, \mathcal{K}\left(\frac{x_{M-1}-x}{h}\right)\right)$.

By implementing the local polynomial in this way, the computational complexity is much higher than our method, and we summarize its computational complexity in the following proposition.

Following please find the proof.

Similar to the proof of the computational complexity in cubic spline, the proof of the computational complexity of local polynomial regression in the funcational estimation stage lies in calculating all elements in matrix $\mathbf{X}$ and vector $\nabla_t \mathbf{u}$, including

$$\left\{ \widehat{u(x_i,t_n)}, \frac{\partial}{\partial x}\widehat{u(x_i,t_n)}, \frac{\partial^2}{\partial x^2}\widehat{u(x_i,t_n)}, \frac{\partial}{\partial t}\widehat{u(x_i,t_n)} \right\}_{i=0,\ldots,M-1,n=0,\ldots,N-1}.$$

We will take the estimation of $\frac{\partial^p}{\partial x^p}\widehat{u(x_i,t_n)}$ with a general $p \in \mathbb{N}$ as an example. To solve $\left\{\frac{\partial^p}{\partial x^p}\widehat{u(x_i,t_n)}\right\}_{i=0,\ldots,M-1,n=0,\ldots,N-1}$, we first focus on $\left\{\frac{\partial^p}{\partial x^p}\widehat{u(x_i,t_n)}\right\}_{i=0,\ldots,M-1}$, with $n \in \{0,\ldots,N-1\}$ fixed. To solve it, the main idea of local polynomial regression is to do Taylor expansion:

$$\left\{ \begin{array}{ccl} u(x_0,t_n) & = & u(x,t_n) + \frac{\partial}{\partial x}u(x,t_n)(x_0-x) + \ldots + \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)(x_0-x)^{\check{p}} \\ u(x_1,t_n) & = & u(x,t_n) + \frac{\partial}{\partial x}u(x,t_n)(x_1-x) + \ldots + \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)(x_1-x)^{\check{p}} \\ \vdots & & \vdots \\ u(x_{M-1},t_n) & = & u(x,t_n) + \frac{\partial}{\partial x}u(x,t_n)(x_{M-1}-x) + \ldots + \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n)(x_{M-1}-x)^{\check{p}} \end{array} \right. ,$$

where $\check{p}$ is usually set as $\check{p} = p+3$ to obtain asymptotic minimax efficiency (see Fan et al., 1997). In the above system of equations, if we denote

$$\mathbf{b}(x) = \left(u(x,t_n), \frac{\partial}{\partial x}u(x,t_n), \ldots, \frac{\partial^{\check{p}}}{\partial x^{\check{p}}}u(x,t_n) \right)^\top,$$

then we can solve $\mathbf{b}(x)$ through the optimization problem in (S10.10) with a closed-form solution shown in (S10.11):

$$\mathbf{b}(x) = \left(\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{X}_{\text{spa}} \right)^{-1} \mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{u}_{:}^n, \tag{S10.12}$$

where

$$\mathbf{X}_{\text{spa}} = \begin{bmatrix} 1 & (x_0-x) & \ldots & (x_0-x)^{\check{p}} \\ 1 & (x_1-x) & \ldots & (x_1-x)^{\check{p}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_{M-1}-x) & \ldots & (x_{M-1}-x)^{\check{p}} \end{bmatrix}, \mathbf{u}_{:}^n = \begin{bmatrix} u_0^n \\ u_1^n \\ \vdots \\ u_{M-1}^n \end{bmatrix}$$

and $\mathbf{W}_{\text{spa}} = \text{diag}\left(\mathcal{K}\left(\frac{x_0-x}{h}\right), \ldots, \mathcal{K}\left(\frac{x_{M-1}-x}{h}\right) \right)$.

The main computational complexity to derive $\mathbf{b}(x)$ lies in the computation of inverse of matrix $\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{X}_{\text{spa}} \in \mathbb{R}^{(\check{p}+1)\times(\check{p}+1)}$, where

$$\begin{aligned} & \mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{X}_{\text{spa}} \\ = & \begin{bmatrix} \sum_{i=0}^{M-1} w_i & \sum_{i=0}^{M-1} w_i(x_i-x) & \sum_{i=0}^{M-1} w_i(x_i-x)^2 & \ldots & \sum_{i=0}^{M-1} w_i(x_i-x)^{\check{p}} \\ \sum_{i=0}^{M-1} w_i(x_i-x) & \sum_{i=0}^{M-1} w_i(x_i-x)^2 & \sum_{i=0}^{M-1} w_i(x_i-x)^3 & \ldots & \sum_{i=0}^{M-1} w_i(x_i-x)^{\check{p}+1} \\ \sum_{i=0}^{M-1} w_i(x_i-x)^2 & \sum_{i=0}^{M-1} w_i(x_i-x)^3 & \sum_{i=0}^{M-1} w_i(x_i-x)^4 & \ldots & \sum_{i=0}^{M-1} w_i(x_i-x)^{\check{p}+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=0}^{M-1} w_i(x_i-x)^{\check{p}} & \sum_{i=0}^{M-1} w_i(x_i-x)^{\check{p}+1} & \sum_{i=0}^{M-1} w_i(x_i-x)^{\check{p}+2} & \ldots & \sum_{i=0}^{M-1} w_i(x_i-x)^{2\check{p}} \end{bmatrix} \end{aligned}$$

and

$$\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{u}_{:}^n = \begin{pmatrix} \sum_{i=0}^{M-1} w_i u_i^n \\ \sum_{i=0}^{M-1} w_i (x_i - x) u_i^n \\ \sum_{i=0}^{M-1} w_i (x_i - x)^2 u_i^n \\ \sum_{i=0}^{M-1} w_i (x_i - x)^3 u_i^n \\ \sum_{i=0}^{M-1} w_i (x_i - x)^4 u_i^n \end{pmatrix},$$

we know that for a fixed $n \in \{0, \ldots, N-1\}$ and $x \in \{x_0, \ldots, x_{M-1}\}$, the computational complexity of computing $\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{X}_{\text{spa}}$ and $\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{u}_{:}^n$ is $O(\check{p}^2 M)$. Besides, the computational complexity to derive $(\mathbf{X}_{\text{spa}}^\top \mathbf{W}_{\text{spa}} \mathbf{X}_{\text{spa}})^{-1}$ is $O(\check{p}^3)$. So we know that for a fixed $n \in \{0, \ldots, N-1\}$ and $x \in \{x_0, \ldots, x_{M-1}\}$, the computational complexity of computing $\frac{\partial^p}{\partial x^p} u(x_i, t_n)$ is $\max\{O(\check{p}^2 M), O(\check{p}^3)\}$ with $\check{p}$ usually set as $\check{p} = p + 3$. Accordingly, the computational complexity of computing $\left\{ \frac{\partial^p}{\partial x^p} u(x_i, t_n) \right\}_{i=0,\ldots,M-1, n=0,\ldots,N-1}$ is $\max\{O(\check{p}^2 M^2 N), O(\check{p}^3 MN)\}$. Because $p \leq q_{\max}$, we know that the computational complexity of computing all derivatives with respective to x with highest order as $q_{\max}$ is $\max\{O(q_{\max}^2 M^2 N), O(q_{\max}^3 MN)\}$. Similarly, the computational complexity of computing the first order derivatives with respective to t is $\max\{O(MN^2), O(MN)\}$. In conclusion, the computational complexity to derive all elements in matrix $\mathbf{X}$ and vector $\nabla_t \mathbf{u}$, including

$$\left\{ \widehat{u(x_i, t_n)}, \frac{\partial}{\partial x} \widehat{u(x_i, t_n)}, \frac{\partial^2}{\partial x^2} \widehat{u(x_i, t_n)}, \frac{\partial}{\partial t} \widehat{u(x_i, t_n)} \right\}_{i=0,\ldots,M-1, n=0,\ldots,N-1}.$$

by local polynomial regression in (2.4) is $\max\{O(q_{\max}^2 M^2 N), O(MN^2), O(q_{\max}^3 MN)\}$, where $q_{\max}$ is the highest order of derivatives desired in (1.3). □

S10.3 Proof of Lemma S6.1

Proof. In this proof, we take $k = 0$ as an illustration example, i.e., prove that when

$$\epsilon > \mathscr{C}(\sigma, \|u\|_{L^\infty(\Omega)}) \max \left\{ \frac{4K_{\max}}{M^{3/7}}, 4AM^{-3/7}, 4\sqrt{2} \frac{d^3}{dx^3} f^*(0) M^{-3/7}, \frac{16(C \log M + \gamma) \log(M)}{M^{3/7}}, 16\sqrt{\frac{\omega}{7}} \widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)}) \frac{\sqrt{\log(M)}}{M^{3/7}} \right\},$$

we have

$$P\left[\sup_{x \in [0, X_{\max}]} \left| \widehat{u(x, t_n)} - u(x, t_n) \right| > \epsilon \right] < 2Me^{-\frac{M^{2/7}}{2\sigma^2}} + Qe^{-L\gamma} + 4\sqrt{2}\eta^4 M^{-\frac{2}{7}\omega}$$

for a fixed t_n with $n \in \{0, 1, \ldots, N-1\}$. For $k = 1, 2$, it can be derived with the same logic flow.

Recall in Section 2.1, the fitted value of the smoothing cubic spline $s(x)$ is the minimizer of the optimization problem in (2.4). From Theorem A in Silverman (1984) (also mentioned by Messer (1991) in the Section 1, and equation (2.2) in Craven and Wahba (1978)) that when Condition 3.3 - Condition 3.4 hold and for large M and small $\widetilde{\lambda} = \frac{1-\alpha}{\alpha}$, we have

$$\widehat{f}_i = \frac{1}{M\widetilde{\lambda}^{1/4}} \sum_{j=0}^{M-1} K\left(\frac{x_i - x_j}{\widetilde{\lambda}^{1/4}} \right) u_j^n,$$

where $\widehat{f_i} = \widehat{u(x_i, t_n)}$, $\widetilde{\lambda}$ trades off the goodness-of-fit and smoothness of the cubic spline in (2.4) and $K(\cdot)$ is a fixed kernel function defined as

$$K(x) = \frac{1}{2} e^{-|x|/\sqrt{2}} \left[\sin(|x|/\sqrt{2} + \pi/4) \right].$$

For a general spatial variable x and fixed $n \in \{0, 1, \ldots, N-1\}$, we denote

$$f^*(x) = u(x, t_n),$$

which is the ground truth of the underlying dynamic function $u(x, t_n)$ with t_n fixed. Besides, we denote $\widehat{f}(x) = \widehat{u(x, t_n)}$, which is an estimation of the ground truth of $f^*(x) = u(x, t_n)$ with t_n fixed. Accordingly to the above discussion, this estimation of $\widehat{f}(x)$ can be written as

$$\widehat{f}(x) = \frac{1}{M\widetilde{\lambda}^{1/4}} \sum_{j=0}^{M-1} K\left(\frac{x - x_j}{\widetilde{\lambda}^{1/4}} \right) u_j^n,$$

where $\widehat{f_i} = \widehat{f}(x_i)$ for $i \in \{0, 1, \ldots, M-1\}$

In order to bound $P\left(\sup |\widehat{f}(x) - f^*(x)| > \epsilon \right)$ for a general x, we decompose it as follows:

$$\begin{aligned}
& P\left(\sup |\widehat{f}(x) - f^*(x)| > \epsilon \right) \\
= \; & P\left(\sup |\widehat{f}(x) - \widehat{f}^B(x) + \widehat{f}^B(x) - \widehat{f}^*(x)| > \epsilon \right) \\
= \; & P\left(\sup |\widehat{f}(x) - \widehat{f}^B(x) - E(\widehat{f}(x) - \widehat{f}^B(x)) + E(\widehat{f}(x) - \widehat{f}^B(x)) + \widehat{f}^B(x) - \widehat{f}^*(x)| > \epsilon \right) \\
= \; & P\Bigg(\sup | \underbrace{\widehat{f}(x) - \widehat{f}^B(x)}_{\mathcal{A}} - \underbrace{E(\widehat{f}(x) - \widehat{f}^B(x))}_{\mathcal{B}} + \qquad , \\
& \underbrace{E(\widehat{f}(x)) - f^*(x)}_{\mathcal{C}} + \underbrace{\widehat{f}^B(x) - E(\widehat{f}^B(x))}_{\mathcal{D}} | > \epsilon \Bigg) \\
\leq \; & P\left(\sup |\mathcal{A}| > \tfrac{\epsilon}{4} \right) + P\left(\sup |\mathcal{B}| > \tfrac{\epsilon}{4} \right) + P\left(\sup |\mathcal{C}| > \tfrac{\epsilon}{4} \right) + P\left(\sup |\mathcal{D}| > \tfrac{\epsilon}{4} \right)
\end{aligned} \tag{S10.13}$$

where the $\widehat{f^B}(x)$ in (S10.13) the truncated estimator defined as

$$\widehat{f^B}(x) = \frac{1}{M\widetilde{\lambda}^{1/4}} \sum_{j=0}^{M-1} K\left(\frac{x - x_j}{\widetilde{\lambda}^{1/4}} \right) u_j^n \mathbb{1}\{u_j^n < B_M\}.$$

Here $\{B_M\}$ is an increasing sequence and $B_M \to +\infty$ as $M \to +\infty$, i.e., $B_M = M^b$ with constant $b > 0$, and we will discuss the value of b at the end of this proof.

In the remaining of the proof, we work on the upper bound of the four decomposed terms, i.e., $P\left(\sup |\mathcal{A}| > \frac{\epsilon}{4} \right), P\left(\sup |\mathcal{B}| > \frac{\epsilon}{4} \right), P\left(\sup |\mathcal{C}| > \frac{\epsilon}{4} \right), P\left(\sup |\mathcal{D}| > \frac{\epsilon}{4} \right)$.

First, let us discuss the upper bound of $P\left(\sup |\mathcal{A}| > \frac{\epsilon}{4} \right)$.

Because

$$\begin{aligned} P\left(\sup|\mathcal{A}| > \frac{\epsilon}{4}\right) &= P\left(\sup\left|\widehat{f}(x) - \widehat{f}^B(x)\right| > \frac{\epsilon}{4}\right) \\ &= P\left(\sup\left|\frac{1}{M\widetilde{\lambda}^{1/4}}\sum_{j=0}^{M-1} K\left(\frac{x-x_j}{\widetilde{\lambda}^{1/4}}\right) u_j^n \mathbb{1}\{u_j^n \geq B_M\}\right| > \frac{\epsilon}{4}\right) \\ &\leq P\left(\sup\left|\frac{K_{\max}}{M\widetilde{\lambda}^{1/4}}\sum_{j=0}^{M-1} u_j^n \mathbb{1}\{u_j^n \geq B_M\}\right| > \frac{\epsilon}{4}\right), \end{aligned}$$

where $K_{\max} = \max_{x\in[0,X_{\max}]\cup[0,T_{\max}]} K(x)$. If we let $\frac{\epsilon}{4} > \frac{K_{\max}}{M\widetilde{\lambda}^{1/4}} B_M$, then we have

$$\begin{aligned} P\left(\sup|\mathcal{A}| > \frac{\epsilon}{4}\right) &\leq P\left(\exists\, i = 0, \ldots, M-1, s.t.\ |u_i^n| \geq B_M\right) \\ &= P\left(\max_{i=0,\ldots,M-1} |u_i^n| \geq B_M\right). \end{aligned}$$

Let $C_M = B_M - \|U\|_{L^\infty(\Omega)}$, where U is the random variable generated from the unknown dynamic system, i.e., $U = u(x,t) + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$. Then we have

$$\begin{aligned} P\left(\sup|\mathcal{A}| > \frac{\epsilon}{4}\right) &= P\left(\sup\left|\widehat{f}(x) - \widehat{f}^B(x)\right| > \frac{\epsilon}{4}\right) \\ &\leq P\left(\max_{i=0,\ldots,M-1} |U_i^n - u_i^n| \geq C_M\right) \\ &\leq 2Me^{-C_M^2/(2\sigma^2)}. \end{aligned}$$

Next, let us discuss the upper bound of $P\left(\sup|\mathcal{B}| > \frac{\epsilon}{4}\right)$.

$$\begin{aligned} \mathcal{B} &= E\left(|\widehat{f}(x) - \widehat{f}^B(x)|\right) \\ &= E\left(\left|\frac{1}{M\widetilde{\lambda}^{1/4}}\sum_{j=0}^{M-1} K\left(\frac{x-x_j}{\widetilde{\lambda}^{1/4}}\right) u_j^n \mathbb{1}\{u_j^n \geq B_M\}\right|\right) \\ &\leq E\left(\frac{1}{M\widetilde{\lambda}^{1/4}}\sum_{j=0}^{M-1} \left|K\left(\frac{x-x_j}{\widetilde{\lambda}^{1/4}}\right)\right| |u_j^n| \mathbb{1}\{u_j^n \geq B_M\}\right) \\ &= \frac{1}{\widetilde{\lambda}^{1/4}} \int\int_{|u|\geq B_M} \left|K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right)\right| |u| dF_M(a,u) \qquad \text{(S10.14)} \\ &\leq \int |K(\xi)| d\xi \times \underbrace{\sup_\alpha \int_{|u|\geq B_M} |u| f_M(\alpha,u) du}_{\mathcal{V}}. \qquad \text{(S10.15)} \end{aligned}$$

Here in (S10.14), $F_M(\cdot,\cdot)$ is the empirical c.d.f. of (x,u)'s, and in (S10.15), $f_M(\cdot,\cdot)$ is the empirical p.d.f. of (x,u)'s.

Now let us take a look at the upper bound of $\mathcal{V}$. For any $s > 0$, we have

$$\begin{aligned} \sup_\alpha \int_{|u|\geq B_M} \frac{|u|}{B_M} f_M(\alpha,u) du &\leq \sup_\alpha \int_{|u|\geq B_M} \left(\frac{|u|}{B_M}\right)^s f_M(\alpha,u) du \\ &\leq \sup_\alpha \int \left(\frac{|u|}{B_M}\right)^s f_M(\alpha,u) du, \end{aligned}$$

which gives

$$\mathcal{V} := \sup_{\alpha} \int_{|u| \geq B_M} |u| f_M(\alpha, u) du \leq B_M^{1-s} \underbrace{\sup_{\alpha} \int |u|^s f_M(\alpha, u) du}_{\pi_s}.$$

From the lemma statement we know that when $s = 2$, we have $\pi_s := \sup_\alpha \int |u|^s f_M(\alpha, u) du < +\infty$. If we set $A = \pi_s \int |K(\xi)| d\xi$, then we have

$$\mathcal{B} \leq A B_M^{1-s}.$$

So when $\frac{\epsilon}{4} > A B_M^{1-s}$, we have

$$P\left(\sup |\mathcal{B}| > \frac{\epsilon}{4}\right) = P\left(E\left(|\widehat{f}(x) - \widehat{f}^B(x)|\right) \geq \frac{\epsilon}{4}\right) = 0.$$

Then, let us discuss the upper bound of $P\left(\sup |\mathcal{C}| > \frac{\epsilon}{4}\right)$. According to Lemma 5 in Rice and Rosenblatt (1983), when

$f^*(x) \in C^4$, $\frac{d^2}{dx^2} f^*(x_0) = \frac{d^2}{dx^2} f^*(x_{M-1}) = 0$ and $\frac{d^3}{dx^3} f^*(x_0) \neq 0$, $\frac{d^3}{dx^3} f^*(x_{M-1}) = 0$, we have

$$\begin{aligned} & E(\widehat{f}(x)) - f^*(x) \\ = \quad & \sqrt{2} \frac{d^3}{dx^3} f^*(0) \widetilde{\lambda}^{3/4} \exp\left(\frac{-x}{\sqrt{2}} \widetilde{\lambda}^{-1/4}\right) \cos\left(\frac{x}{\sqrt{2}} \widetilde{\lambda}^{-1/4}\right) + \ell(x), \end{aligned}$$

where the error term $\ell(x)$ satisfies

$$\int [\ell(x)]^2 dx = o\left(\int \left[E(\widehat{f}(x)) - f^*(x)\right]^2 dx\right).$$

So when $\frac{\epsilon}{4} > \sqrt{2} \frac{d^3}{dx^3} f^*(0) \widetilde{\lambda}^{3/4}$ and M is sufficiently large then we have

$$P\left(\sup |\mathcal{C}| > \frac{\epsilon}{4}\right) = 0.$$

Finally, let us discuss the upper bound of $P\left(\sup |\mathcal{D}| > \frac{\epsilon}{4}\right)$.

In order to bound $P\left(\sup |D| > \frac{\epsilon}{4}\right)$, we further decompose $\mathcal{D}$ into two components, i.e.,

$$\mathcal{D} := \widehat{f}^B(x) - E(\widehat{f}^B(x)) = e_M(x, t_n) + \frac{1}{\sqrt{M}} \rho_M(x, t_n).$$

The decomposition procedure and the definition of $e_M(x, t_n), \rho_M(x, t_n)$ are described in the

following system of equations (see Mack and Silverman, 1982, Proposition 2):

$$\begin{aligned}
\mathcal{D} &= \widehat{f}^B(x) - E(\widehat{f}^B(x)) \\
&= \frac{1}{M\widetilde{\lambda}^{1/4}} \sum_{j=0}^{M-1} K\left(\frac{x-x_j}{\widetilde{\lambda}^{1/4}}\right) u_j^n \mathbb{1}\{u_j^n < B_M\} - \\
&\quad E\left(\frac{1}{M\widetilde{\lambda}^{1/4}} \sum_{j=0}^{M-1} K\left(\frac{x-x_j}{\widetilde{\lambda}^{1/4}}\right) u_j^n \mathbb{1}\{u_j^n < B_M\}\right) \\
&= \frac{1}{\sqrt{M}\widetilde{\lambda}^{1/4}} \int_{a\in\mathbb{R}} \int_{|u|<B_M} K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right) u \; d\underbrace{\left(\sqrt{M}(F_M(a,u) - F(a,u))\right)}_{Z_M(a,u)} \qquad \text{(S10.16)} \\
&= \frac{1}{\sqrt{M}\widetilde{\lambda}^{1/4}} \int_{a\in\mathbb{R}} K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right) \int_{|u|<B_M} u \; d(Z_M(a,u)) \\
&= \frac{1}{\sqrt{M}\widetilde{\lambda}^{1/4}} \int_{a\in\mathbb{R}} K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right) \left[\int_{|u|<B_M} u \; d(Z_M(a,u) - B_0(T(a,u))) + \right. \\
&\quad \left. \int_{|u|<B_M} u \; dB_0(T(a,u))\right] \qquad \text{(S10.17)} \\
&= \underbrace{\frac{1}{\sqrt{M}\widetilde{\lambda}^{1/4}} \int_{a\in\mathbb{R}} \int_{|u|<B_M} K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right) u \; d(Z_M(a,u) - B_0(T(a,u)))}_{e_M(x,t_n)} \\
&\quad + \frac{1}{\sqrt{M}} \underbrace{\frac{1}{\widetilde{\lambda}^{1/4}} \int_{a\in\mathbb{R}} \int_{|u|<B_M} K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right) u \; dB_0(T(a,u))}_{\rho_M(x,t_n)}.
\end{aligned}$$

In (S10.16), $F_M(\cdot,\cdot) := F_M(\cdot,\cdot|t_n)$ is the empirical c.d.f of (x,u) with a fixed t_n, and $Z_M(a,u) = \sqrt{M}(F_M(a,u) - F(a,u))$ is a two-dimensional empirical process (see Tusnády, 1977; Mack and Silverman, 1982). In (S10.17), $B_0(T(a,u))$ is a sample path of two-dimensional Brownian bride. And $T(a,u) : \mathbb{R}^2 \to [0,1]^2$ is the transformation defined by Rosenblatt (1952), i.e., $T(a,u) = (F_A(x), F_{U|A}(u|a))$, where F_A is the marginal c.d.f of A and $F_{U|A}$ is the conditional c.d.f of U given A (see Mack and Silverman, 1982, Proposition 2).

Through the above decomposition of $\mathcal{D}$, we have

$$P\left(\sup|D| > \frac{\epsilon}{4}\right) \le P\left(\sup|e_M(x,t_n)| > \frac{\epsilon}{8}\right) + P\left(\sup \frac{1}{\sqrt{M}}|\rho_M(x,t_n)| > \frac{\epsilon}{8}\right).$$

For $e_M(x,t_n)$, we have

$$\begin{aligned}
& P\left(\sup|e_M(x,t_n)| > \frac{\epsilon}{8}\right) \\
= \; & P\left(\sup\left|\frac{1}{\sqrt{M}\widetilde{\lambda}^{1/4}} \int_{a\in\mathbb{R}} \int_{|u|<B_M} K\left(\frac{x-a}{\widetilde{\lambda}^{1/4}}\right) u \; d(Z_M(a,u) - B_0(T(a,u)))\right| > \frac{\epsilon}{8}\right) \\
\le \; & P\left(\frac{2B_M K_{\max}}{\sqrt{M}\widetilde{\lambda}^{1/4}} \sup_{a,u}|Z_M(a,u) - B_0(T(a,u))| > \frac{\epsilon}{8}\right).
\end{aligned}$$

Proved by Theorem 1 in Tusnády (1977), we know that, for any γ, we have

$$P\left(\sup_{a,u}|Z_M(a,u)-B_0(T(a,u))|>\frac{(C\log M+\gamma)\log M}{\sqrt{M}}\right)\leq Qe^{-L\gamma},$$

where C, Q, L are absolute positive constants which is independent of temporal resolution N and spatial resolution M. Thus, when $\frac{\epsilon}{8}\geq\frac{2B_M K_{\max}}{\sqrt{M}\widetilde{\lambda}^{1/4}}\frac{(C\log M+\gamma)\log M}{\sqrt{M}}$, we have

$$P\left(\sup|e_M(x,t_n)|>\frac{\epsilon}{8}\right)<Qe^{-L\gamma}.$$

For $\rho_M(x,t_n)$, by equation (7) in Mack and Silverman (1982), we have

$$\begin{aligned}\frac{\widetilde{\lambda}^{1/8}\sup|\rho_M(x,t_n)|}{\sqrt{\log(1/\widetilde{\lambda}^{1/4})}} \quad\leq\quad & \underbrace{16(\log V)^{1/2}S^{1/2}\left(\log\left(\frac{1}{\widetilde{\lambda}^{1/4}}\right)\right)^{-1/2}\int|\xi|^{1/2}|dK(\xi)|}_{\mathcal{W}_{1,M}}+\\ & \underbrace{16\sqrt{2}\widetilde{\lambda}^{-1/8}\left(\log\left(\frac{1}{\widetilde{\lambda}^{1/4}}\right)\right)^{-1/2}\int q(S\widetilde{\lambda}^{1/4}|\tau|)|d(K(\tau))|}_{\mathcal{W}_{2,M}},\end{aligned}$$

where V is a random variable satisfying $E(V)\leq 4\sqrt{2}\eta^4$ for $\eta^2=\max_{i=0,\ldots,M-1,n=0,\ldots,N-1}E(U_i^n)^2$, $S=\sup_x\int u^2 f(x,u)du$ with $f(\cdot,\cdot)$ as the distribution function of (x_i,u_i^n), and $q(z)=\int_0^z\frac{1}{2}\sqrt{\frac{1}{y}\log\left(\frac{1}{y}\right)}dy$. So we have the following system of equations:

$$\begin{aligned}& P\left(\sup\frac{1}{\sqrt{M}}|\rho_M(x,t_n)|>\frac{\epsilon}{8}\right)\\ =\quad & P\left(\frac{\widetilde{\lambda}^{1/8}\sup|\rho_M(x,t_n)|}{\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}>\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{8\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right)\\ \leq\quad & P\left(\mathcal{W}_{1,M}+\mathcal{W}_{2,M}>\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{8\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right)\\ \leq\quad & P\left(\mathcal{W}_{1,M}\geq\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right)+P\left(\mathcal{W}_{2,M}\geq\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right)\end{aligned}\tag{S10.18}$$

Now let us bound $P\left(\mathcal{W}_{1,M}\geq\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right), P\left(\mathcal{W}_{2,M}\geq\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right)$ in (S10.18) separately.

1. For the first term in (S10.18), we have

$$
\begin{aligned}
& P\left(\mathcal{W}_{1,M} \geq \frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right) \\
= \; & P\left(16(\log V)^{1/2} S^{1/2}\left(\log\left(\frac{1}{\widetilde{\lambda}^{1/4}}\right)\right)^{-1/2}\int |\xi|^{1/2}|dK(\xi)| \geq \frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right) \\
= \; & P\left((\log V)^{1/2} \geq \frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16^2 S^{1/2}\int |\xi|^{1/2}|dK(\xi)|}\right) \\
= \; & P\left(\log V \geq \left(\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16^2 S^{1/2}\int |\xi|^{1/2}|dK(\xi)|}\right)^2\right) \\
= \; & P\left(V \geq \exp\left[\left(\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16^2 S^{1/2}\int |\xi|^{1/2}|dK(\xi)|}\right)^2\right]\right) \\
\leq \; & \frac{E(V)}{\exp\left[\left(\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16^2 S^{1/2}\int |\xi|^{1/2}|dK(\xi)|}\right)^2\right]} \qquad \text{(S10.19)} \\
\leq \; & \frac{4\sqrt{2}\eta^4}{\exp\left[\left(\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16^2 S^{1/2}\int |\xi|^{1/2}|dK(\xi)|}\right)^2\right]} \qquad \text{(S10.20)} \\
= \; & 4\sqrt{2}\eta^4\widetilde{\lambda}^{\omega/4} \qquad \text{(S10.21)}
\end{aligned}
$$

Here inequality (S10.19) is due to Markov's inequality, and inequality (S10.20) is due to the fact that $E(V) \leq 4\sqrt{2}\eta^4$. Equality (S10.21) is because we set $\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}} = \sqrt{\omega}\widetilde{C}(t_n, \sigma, \|u\|_{L^\infty}(\Omega))$, where

$$\widetilde{C}(t_n, \sigma, \|u\|_{L^\infty}(\Omega)) := 16\sqrt{S}\int |\xi|^{1/2}|dK(\xi)|$$

and $\omega > 1$ is an arbitrary scaler.

2. For the second term of (S10.18), it converges to $\widetilde{C}(t_n, \sigma, \|u\|_{L^\infty}(\Omega))$ by using arguments similar to Silverman (1978) (page. 180-181) under the condition in Lemma S6.1 that $\int \sqrt{|x\log(|x|)|}|dK(x)| < +\infty$. Here we add $(t_n, \sigma, \|u\|_{L^\infty(\Omega)})$ after $\bar{C}$ to emphasize that the constant $\bar{C}(t_n, \sigma, \|u\|_{L^\infty(\Omega)})$ is dependent on $t_n, \sigma, \|u\|_{L^\infty(\Omega)}$.

   It should be noted that

$$\widetilde{C}(t_n, \sigma, \|u\|_{L^\infty(\Omega)}) < +\infty, \qquad \text{(S10.22)}$$

   given the reasons listed as follows. First, it can be easily verified that the term $\int |\xi|^{1/2}|dK(\xi)|$ in $\widetilde{C}(t_n, \sigma, \|u\|_{L^\infty}(\Omega))$ is bounded. Second, for $S = \sup_x \int u^2 f(x,u)du$, it is also bounded.

The reasons are described as follows. For a general $\varrho > 0$, we have

$$\begin{aligned}
& \sup_{x\in[0,X_{\max}]} \int |u|^{\varrho} f(x,u)du \\
= & \sup_{x\in[0,X_{\max}]} \int |u|^{\varrho} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(u-u(x,t_n))^2}{2\sigma^2}\right) du \\
= & \sup_{x\in[0,X_{\max}]} \frac{1}{\sqrt{2}}\sigma^2 2^{\varrho/2}\Gamma\left(\frac{1+\varrho}{2}\right) G\left(-\frac{\varrho}{2},\frac{1}{2},-\frac{1}{2}\left(\frac{u(x,t_n)}{\sigma}\right)^2\right),
\end{aligned}$$

where $G(a,b,z)$ is Kummer's confluent hypergeometric function of $z \in \mathbb{C}$ with parameters $a, b \in \mathbb{C}$ (see Winkelbauer, 2012). Because $G\left(-\frac{\varrho}{2},\frac{1}{2},\cdot\right)$ is an entire function for fixed parameters, we have

$$\begin{aligned}
& \sup_{x\in[0,X_{\max}]} \int |u|^{\varrho} f(x,u)du \\
\leq & \sup_{x\in[0,X_{\max}]} \frac{1}{\sqrt{2}}\sigma^2 2^{\varrho/2}\Gamma\left(\frac{1+\varrho}{2}\right) \sup_{z\in\left[-\frac{\max_{t\in\Omega} u^2(x,t)}{2\sigma^2},-\frac{\min_{t\in\Omega} u^2(x,t)}{2\sigma^2}\right]} G\left(-\frac{\varrho}{2},\frac{1}{2},z\right) \\
< & +\infty.
\end{aligned}$$

So we can bound $\sup_{x\in[0,X_{\max}]} \int |u|^{\varrho} f(x,u)du$ by a constant. If we take $\varrho = 2$, we can obtain $S = \sup_x \int u^2 f(x,u)du$ bounded by a constant. So we can declare the statement in (S10.22).

We would also like to declare that there exist a constant $\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)}) > 0$ such that for any $N \geq 1$, we have

$$\max_{n=0,\ldots,N-1} \widetilde{C}(t_n, \sigma, \|u\|_{L^\infty(\Omega)}) \leq \widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)}),$$

where $\bar{C}(\sigma, \|u\|_{L^\infty(\Omega)})$ is independent of t_n, x_i, M, N, and only depends on the noisy data $\mathcal{D}$ itself.

From the above discussion, we learn that $\mathcal{W}_{2,M}$ converges to $\widetilde{C}(t_n, \sigma, \|u\|_{L^\infty(\Omega)})$, which can be bounded by $\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})$. If we set $\frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}} > \sqrt{\omega}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})$ with $\omega > 1$, then there exists a positive integer $\dot{M}(\omega)$ such that as long as $M > \dot{M}(\omega)$, we have $P\left(\mathcal{W}_{2,M} \geq \frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right) = 0$.

For the value of ω, we set it as $\omega = M^{2r}$ with $r > 0$. And we will discuss the value of r later.

By combining $P\left(\mathcal{W}_{1,M} \geq \frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right), P\left(\mathcal{W}_{2,M} \geq \frac{\sqrt{M}\widetilde{\lambda}^{1/8}\epsilon}{16\sqrt{\log(1/\widetilde{\lambda}^{1/4})}}\right)$ together, we have when $\frac{\epsilon}{16} > \sqrt{\omega}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\sqrt{\frac{\log(1/\widetilde{\lambda}^{1/4})}{M\widetilde{\lambda}^{1/4}}}$ and $M > \dot{M}(\omega)$, we have

$$P\left(\sup\left|\frac{1}{\sqrt{M}}\rho_M(x,t_n)\right| > \frac{\epsilon}{8}\right) < 4\sqrt{2}\eta^4\widetilde{\lambda}^{\omega/4}.$$

By combining the discussion on $P\left(\sup|\mathcal{A}| > \frac{\epsilon}{4}\right), P\left(\sup|\mathcal{B}| > \frac{\epsilon}{4}\right), P\left(\sup|\mathcal{C}| > \frac{\epsilon}{4}\right)$, and $P\left(\sup|\mathcal{D}| > \frac{\epsilon}{4}\right)$, we can conclude that when

- $\frac{\epsilon}{4} > \frac{K_{\max}}{M\widetilde{\lambda}^{1/4}} B_M$
- $\frac{\epsilon}{4} > AB_M^{1-s} \quad (s=2)$
- $\frac{\epsilon}{4} > \sqrt{2}\frac{d^3}{dx^3} f^*(0)\widetilde{\lambda}^{3/4}$
- $\frac{\epsilon}{8} > \frac{2B_M K_{\max}(C\log M+\gamma)\log M}{\widetilde{\lambda}^{1/4} M}$
- $\frac{\epsilon}{16} > \sqrt{\omega}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\sqrt{\frac{\log(1/\widetilde{\lambda}^{1/4})}{M\widetilde{\lambda}^{1/4}}}$

we have

$$P\left(\sup|\mathcal{A}+\mathcal{B}+\mathcal{C}+\mathcal{D}|>\epsilon\right) < \underbrace{2Me^{-\frac{C_M^2}{2\sigma^2}}}_{\mathcal{Z}_1} + \underbrace{Qe^{-L\gamma}}_{\mathcal{Z}_2} + \underbrace{4\sqrt{2}\eta^4\widetilde{\lambda}^{\omega/4}}_{\mathcal{Z}_3}. \tag{S10.23}$$

Let

$$\begin{cases} E_1 &= \frac{4K_{\max}}{M\widetilde{\lambda}^{1/4}} B_M \\ E_2 &= 4AB_M^{1-s} \\ E_3 &= 4\sqrt{2}\frac{d^3}{dx^3} f^*(0)\widetilde{\lambda}^{3/4} \\ E_4 &= \frac{16B_M K_{\max}(C\log M+\gamma)\log M}{\widetilde{\lambda}^{1/4} M} \\ E_5 &= 16\sqrt{\omega}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\sqrt{\frac{\log(1/\widetilde{\lambda}^{1/4})}{M\widetilde{\lambda}^{1/4}}} \end{cases},$$

by setting $\widetilde{\lambda} = M^{-a}, B_M = M^b$ with $a, b > 0$, we have

$$\begin{cases} E_1 &= \frac{4K_{\max}}{M\widetilde{\lambda}^{1/4}} B_M = \frac{4K_{\max}}{M^{1-a/4-b}} \\ E_2 &= 4AB_M^{1-s} = 4A\frac{1}{M^{b(s-1)}} \\ E_3 &= 4\sqrt{2}\frac{d}{dx} f^*(0)\widetilde{\lambda}^{3/4} = 4\sqrt{2}\frac{d}{dx} f^*(0)M^{-3a/4} \\ E_4 &= \frac{16B_M K_{\max}(C\log M+\gamma)\log M}{\widetilde{\lambda}^{1/4} M} = \frac{16K_{\max}(C\log M+\gamma)\log(M)}{M^{1-a/4-b}} \\ E_5 &= 16\sqrt{\omega}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\sqrt{\frac{\log(1/\widetilde{\lambda}^{1/4})}{M\widetilde{\lambda}^{1/4}}} = 8\sqrt{a\omega}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\sqrt{\frac{\log(M)}{M^{1-a/4}}}. \end{cases}$$

To guarantee that $E_1, E_2, E_3, E_4, E_5 \to 0$ as $M \to +\infty$, we can set

$$\begin{cases} 1 - a/4 - b = 3a/4 \\ b(s-1) > 0 \\ \frac{1}{2}(1-a/4) = 3a/4 \\ a, b > 0 \\ s = 2 \end{cases}.$$

then we have

$$\begin{cases} a = 4/7 \\ b = 3/7 \\ s = 2 \end{cases}.$$

Accordingly, we have

$$\begin{cases} E_1 = \frac{4K_{\max}}{M^{3/7}} \\ E_2 = 4AM^{-3/7} \\ E_3 = 4\sqrt{2}\frac{d^3}{dx^3} f^*(0)M^{-3/7} \\ E_4 = \frac{16K_{\max}(C\log M+\gamma)\log(M)}{M^{3/7}} \\ E_5 = 16\sqrt{\frac{\omega}{7}}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\frac{\sqrt{\log(M)}}{M^{3/7}} \end{cases},$$

where

$$E_1, E_2, E_3, E_5 \lesssim E_4$$

as $M \to +\infty$. Here, the operator $\lesssim$ means that when $M \to +\infty$, the order of the left side hand of $\lesssim$ will be much smaller than that on the right side hand. So we can declare that when M is sufficiently large and

$$\begin{aligned}\epsilon \quad > \quad \max \quad & \left\{ \tfrac{4K_{\max}}{M^{3/7}}, 4AM^{-3/7}, 4\sqrt{2}\tfrac{d^3}{dx^3}f^*(0)M^{-3/7}, \tfrac{16K_{\max}(C\log M+\gamma)\log(M)}{M^{3/7}}, \right. \\ & \left. 16\sqrt{\tfrac{\omega}{7}}\widetilde{C}(\sigma, \|u\|_{L^\infty(\Omega)})\tfrac{\sqrt{\log(M)}}{M^{3/7}} \right\},\end{aligned}$$

we have

$$\begin{aligned}P\left(\sup|\mathcal{A}+\mathcal{B}+\mathcal{C}+\mathcal{D}|>\epsilon\right) \quad &\leq \quad 2Me^{-\frac{C_M^2}{2\sigma^2}} + Qe^{-L\gamma} + 4\sqrt{2}\eta^4\widetilde{\lambda}^{\omega/4} \\ &= \quad 2Me^{-\frac{(M^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Qe^{-L\gamma} + 4\sqrt{2}\eta^4 M^{-\omega/7}.\end{aligned}$$

□

S10.4 Proof of Lemma C.2

Proof. For the estimation error $\|\nabla_t\mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty$, we have

$$\begin{aligned}\|\nabla_t\mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty \quad &= \quad \|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^* + \nabla_t\mathbf{u}^* - \mathbf{X}\boldsymbol{\beta}^*\|_\infty \\ &= \quad \|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^* + \mathbf{X}^*\boldsymbol{\beta}^* - \mathbf{X}\boldsymbol{\beta}^*\|_\infty \\ &\leq \quad \|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^*\|_\infty + \|(\mathbf{X}^* - \mathbf{X})\boldsymbol{\beta}^*\|_\infty. \qquad \text{(S10.24)}\end{aligned}$$

So accordingly, we have

$$P\left(\|\nabla_t\mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty > \epsilon\right) \quad \leq \quad P\left(\|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^*\|_\infty > \frac{\epsilon}{2}\right) + P\left(\|(\mathbf{X}^* - \mathbf{X})\boldsymbol{\beta}^*\|_\infty\right).$$

In the remaining of the proof, we will discuss the bound of $P\left(\|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^*\|_\infty > \frac{\epsilon}{2}\right)$ and $P\left(\|(\mathbf{X}^* - \mathbf{X})\boldsymbol{\beta}^*\|_\infty\right)$ separately.

- First let us discuss the bound of $P\left(\|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^*\|_\infty > \frac{\epsilon}{2}\right)$. Because

$$\begin{aligned}P\left(\|\nabla_t\mathbf{u} - \nabla_t\mathbf{u}^*\|_\infty > \frac{\epsilon}{2}\right) \quad &\leq \quad P\left(\max_{i=0,\ldots,M-1}\sup_{t\in[0,T_{\max}]}\left|\widehat{\frac{\partial}{\partial t}u(x_i,t)} - \frac{\partial}{\partial t}u(x_i,t)\right| > \frac{\epsilon}{2}\right) \\ &\leq \quad \sum_{i=0}^{M-1} P\left(\sup_{t\in[0,T_{\max}]}\left|\widehat{\frac{\partial}{\partial t}u(x_i,t)} - \frac{\partial}{\partial t}u(x_i,t)\right| > \frac{\epsilon}{2}\right),\end{aligned}$$

  if we set

$$\begin{aligned}\tfrac{\epsilon}{2} \quad > \quad \mathscr{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\max \quad & \left\{4K_{\max}N^{-3/7}, 4\bar{A}N^{-3/7}, 4\sqrt{2}\tfrac{d^3}{dx^3}\bar{f}^*(0)N^{-3/7}, \right. \\ & \tfrac{16K_{\max}\left[C_{(\sigma,\|u\|_{L^\infty(\Omega)})}\log(N)+\gamma_{(N)}\right]\log(N)}{N^{3/7}}, \qquad \text{(S10.25)} \\ & \left. 16\sqrt{\tfrac{\omega_{(N)}}{7}}\widetilde{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\tfrac{\sqrt{\log(N)}}{N^{3/7}}\right\},\end{aligned}$$

then we have

$$\begin{aligned} & P\left(\|\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^*\|_\infty > \frac{\epsilon}{2}\right) \\ \leq \quad & M\left[2Ne^{-\frac{(N^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-L\gamma_{(N)}} + 4\sqrt{2}\eta^4 N^{-\omega_{(N)}/7}\right], \end{aligned} \tag{S10.26}$$

where inequity (S10.26) is derived according to Corollary S6.1.

- Second, let us discuss the bound of $P\left(\|(\mathbf{X}^* - \mathbf{X})\boldsymbol{\beta}^*\|_\infty\right)$. Because

$$\begin{aligned} & P\left(\|(\mathbf{X}^* - \mathbf{X})\boldsymbol{\beta}^*\|_\infty > \frac{\epsilon}{2}\right) \\ \leq \quad & P\left(\|\boldsymbol{\beta}^*\|_\infty \max_{n=0,\ldots,N-1} \sup_{x\in[0,X_{\max}]} \sum_{k=1}^{K} \|(\mathbf{X}_k^*(x,t_n) - \mathbf{X}_k(x,t_n))\|_\infty > \frac{\epsilon}{2}\right) \\ = \quad & P\left(\max_{n=0,\ldots,N-1} \sup_{x\in[0,X_{\max}]} \sum_{k=1}^{K} \|(\mathbf{X}_k^*(x,t_n) - \mathbf{X}_k(x,t_n))\|_\infty > \frac{\epsilon}{2\|\boldsymbol{\beta}^*\|_\infty}\right) \\ \leq \quad & \sum_{n=0}^{N-1}\sum_{k=1}^{K} P\left(\sup_{x\in[0,X_{\max}]} \|(\mathbf{X}_k^*(x,t_n) - \mathbf{X}_k(x,t_n))\|_\infty > \frac{\epsilon}{2K\|\boldsymbol{\beta}^*\|_\infty}\right), \end{aligned}$$

if we set

$$\begin{aligned} \frac{\epsilon}{2K\|\boldsymbol{\beta}^*\|_\infty} \quad > \quad & \mathscr{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})} \max \quad \Big\{4K_{\max}M^{-3/7}, 4AM^{-3/7}, 4\sqrt{2}\tfrac{d^3}{dx^3}f^*(0)M^{-3/7}, \\ & \frac{16\left[C_{(\sigma,\|u\|_{L^\infty(\Omega)})}\log M + \gamma_{(M)}\right]\log(M)}{M^{3/7}}, \\ & 16\sqrt{\tfrac{\omega_{(M)}}{7}}\widetilde{C}(\sigma,\|u\|_{L^\infty(\Omega_{(M)})})\tfrac{\sqrt{\log(M)}}{M^{3/7}}\Big\}, \end{aligned} \tag{S10.27}$$

then we have

$$\begin{aligned} & P\left(\|(\mathbf{X}^* - \mathbf{X})\boldsymbol{\beta}^*\|_\infty > \frac{\epsilon}{2}\right) \\ \leq \quad & NK\left[2Me^{-\frac{(M^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-L\gamma_{(M)}} + 4\sqrt{2}\eta^4 M^{-\omega_{(M)}/7}\right] \end{aligned} \tag{S10.28}$$

Inequality (S10.28) is derived by Lemma S6.1.

By combining the results in (S10.25), (S10.26), (S10.27), (S10.28), we have that when

$$\begin{aligned} \frac{\epsilon}{2} \quad > \quad & \mathscr{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})} \max \quad \Big\{4K_{\max}M^{-3/7}, 4KK_{\max}\|\boldsymbol{\beta}^*\|_\infty N^{-3/7}, \\ & 4AM^{-3/7}, 4K\|\boldsymbol{\beta}^*\|_\infty \bar{A}N^{-3/7}, \\ & 4\sqrt{2}\tfrac{d^3}{dx^3}f^*(0)M^{-3/7}, 4\sqrt{2}K\|\boldsymbol{\beta}^*\|_\infty \tfrac{d^3}{dx^3}\bar{f}^*(0)N^{-3/7}, \\ & \frac{16KK_{\max}\|\boldsymbol{\beta}^*\|_\infty\left[C_{(\sigma,\|u\|_{L^\infty(\Omega)})}\log(M)+\gamma_{(M)}\right]\log(M)}{M^{3/7}}, \\ & \frac{16K_{\max}\left[C_{(\sigma,\|u\|_{L^\infty(\Omega)})}\log(N)+\gamma_{(N)}\right]\log(N)}{N^{3/7}}, \\ & 16\sqrt{\tfrac{\omega_{(M)}}{7}}\widetilde{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\tfrac{\sqrt{\log(M)}}{M^{3/7}}, \\ & 16K\|\boldsymbol{\beta}^*\|_\infty\sqrt{\tfrac{\omega_{(N)}}{7}}\widetilde{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\tfrac{\sqrt{\log(N)}}{N^{3/7}}\Big\}, \end{aligned}$$

we have

$$
\begin{aligned}
& P\left(\|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\|_\infty > \epsilon\right) \\
\leq \quad & M\left[2Ne^{-\frac{(N^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-L\gamma_{(N)}} + 4\sqrt{2}\eta^4 N^{-\omega_{(N)}/7}\right] + \\
& NK\left[2Me^{-\frac{(M^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-L\gamma_{(M)}} + 4\sqrt{2}\eta^4 M^{-\omega_{(M)}/7}\right]
\end{aligned}
$$

Now, let us do some simplification of the above results. Let $M = N^\kappa, \gamma_{(M)} = \gamma_{(N)} = \frac{1}{L}N^r, \omega_{(M)} = \omega_{(N)} = N^{2r}$, and

$$
\left\{
\begin{array}{lcl}
\mathcal{J}_1 & = & 4KK_{\max}\|\boldsymbol{\beta}^*\|_\infty N^{-3\kappa/7} \\
\mathcal{J}_1' & = & 4K_{\max}N^{-3/7} \\
\mathcal{J}_2 & = & 4AK\|\boldsymbol{\beta}^*\|_\infty N^{-3\kappa/7} \\
\mathcal{J}_2' & = & 4\bar{A}N^{-3/7} \\
\mathcal{J}_3 & = & 4\sqrt{2}K\|\boldsymbol{\beta}^*\|_\infty \frac{d^3}{dx^3}f^*(0)N^{-3\kappa/7} \\
\mathcal{J}_3' & = & 4\sqrt{2}\frac{d^3}{dx^3}\bar{f}^*(0)N^{-3/7} \\
\mathcal{J}_4 & = & \frac{16KK_{\max}\|\boldsymbol{\beta}^*\|_\infty\left[C_{(\sigma,\|u\|_{L^\infty(\Omega)})}(\log(\kappa)+\log(N))+N^r/L\right](\log(\kappa)+\log(N))}{N^{3\kappa/7}} \\
\mathcal{J}_4' & = & \frac{16K_{\max}\left[C_{(\sigma,\|u\|_{L^\infty(\Omega)})}\log(N)+N^r\right]\log(N)}{N^{3/7}} \\
\mathcal{J}_5 & = & 16K\|\boldsymbol{\beta}^*\|_\infty\sqrt{\frac{N^{2r}}{7}}\widetilde{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\frac{\sqrt{\log(\kappa)+\log(N)}}{N^{3\kappa/7}} \\
\mathcal{J}_5' & = & 16\sqrt{\frac{N^{2r}}{7}}\widetilde{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\frac{\sqrt{\log(N)}}{N^{3/7}}
\end{array}
\right. .
$$

To guarantee that $\mathcal{J}_1, \mathcal{J}_1', \mathcal{J}_2, \mathcal{J}_2', \mathcal{J}_3, \mathcal{J}_3', \mathcal{J}_4, \mathcal{J}_4', \mathcal{J}_5, \mathcal{J}_5' \to 0$, as $N \to +\infty$, we need

$$
\left\{
\begin{array}{l}
3\kappa/7 - r > 0 \\
3/7 - r > 0
\end{array}
\right. ,
$$

where the optimal κ is $\kappa = 1$. Accordingly, we have

$$
\mathcal{J}_1, \mathcal{J}_1', \mathcal{J}_2, \mathcal{J}_2', \mathcal{J}_3, \mathcal{J}_3', \mathcal{J}_5, \mathcal{J}_5' \lesssim \mathcal{J}_4, \mathcal{J}_4'.
$$

Based on the above discussion, we can declare that when N is sufficiently large, with

$$
\epsilon > \mathscr{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\frac{\log(N)}{N^{3/7-r}}
$$

for any $r \in \left(0, \frac{3}{7}\right)$ and $M = O(N)$, we have

$$\begin{aligned}
& P\left\|\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}^*\right\|_\infty > \epsilon) \\
\leq\ & M\left[2Ne^{-\frac{(N^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-L\gamma_{(N)}} + 4\sqrt{2}\eta^4 N^{-\omega_{(N)}/7}\right] + \\
& NK\left[2Me^{-\frac{(M^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-L\gamma_{(M)}} + 4\sqrt{2}\eta^4 M^{-\omega_{(M)}/7}\right] \\
=\ & M\left[2Ne^{-\frac{(N^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-N^r} + 4\sqrt{2}\eta^4 N^{-N^{2r}/7}\right] + \\
& NK\left[2Me^{-\frac{(M^{3/7}-\|U\|_{L^\infty(\Omega)})^2}{2\sigma^2}} + Q_{(\sigma,\|u\|_{L^\infty(\Omega)})}e^{-N^r} + 4\sqrt{2}\eta^4 M^{-N^{2r}/7}\right] \\
=\ & O(Ne^{-N^r}).
\end{aligned}$$

Thus, we finish the proof of the theorem. □

S10.5 Proof of Theorem 3.1

Proof. By KKT-condition, any minimizer $\boldsymbol{\beta}$ of (2.10) must satisfies:

$$-\frac{1}{MN}\mathbf{X}^\top(\nabla_t \mathbf{u} - \mathbf{X}\boldsymbol{\beta}) + \lambda \mathbf{z} = 0 \ \text{ for } \ \mathbf{z} \in \partial\|\boldsymbol{\beta}\|_1,$$

where $\partial\|\boldsymbol{\beta}\|_1$ is the sub-differential of $\|\boldsymbol{\beta}\|_1$. The above equation can be equivalently transformed into

$$\mathbf{X}^\top\mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}^*) + \mathbf{X}^\top\left[(\mathbf{X} - \mathbf{X}^*)\boldsymbol{\beta}^* - (\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^*)\right] + \lambda MN\mathbf{z} = 0. \tag{S10.29}$$

Here matrix $\mathbf{X} \in \mathbb{R}^{MN\times K}$ is defined in (2.9), and matrix $\mathbf{X}^* \in \mathbb{R}^{MN\times K}$ is defined as

$$\mathbf{X}^* \ = \ \begin{pmatrix} \mathbf{x}_0^0 & \mathbf{x}_1^0 & \dots & \mathbf{x}_{M-1}^0 & \mathbf{x}_1^0 & \dots & \mathbf{x}_{M-1}^{N-1} \end{pmatrix}^\top,$$

with

$$\mathbf{x}_i^n \ = \ \left(1, \quad u(x_i,t_n), \quad \frac{\partial u(x_i,t_n)}{\partial x}, \quad \frac{\partial^2 u(x_i,t_n)}{\partial x^2}, \quad \left(\widehat{u(x_i,t_n)}\right)^2, \quad \dots, \left(\frac{\partial^2 u(x_i,t_n)}{\partial x^2}\right)^{p_{\max}}\right)^\top \in \mathbb{R}^K.$$

And vector $\boldsymbol{\beta}^* = (\beta_1, \dots, \beta_K) \in \mathbb{R}^K$ is the ground truth coefficients. Besides, vector $\nabla_t \mathbf{u} \in \mathbb{R}^{MN}$ is defined in (2.8), and vector $\nabla_t \mathbf{u}^* \in \mathbb{R}^K$ is the ground truth, i.e.,

$$\nabla_t \mathbf{u}^* = \left(\frac{\partial u(x_0,t_0)}{\partial t}, \quad \frac{\partial u(x_1,t_0)}{\partial t}, \quad \dots, \quad \frac{\partial u(x_{M-1},t_0)}{\partial t}, \quad \frac{\partial u(x_0,t_1)}{\partial t}, \quad \dots, \quad \frac{\partial u(x_{M-1},t_{N-1})}{\partial t}\right)^\top.$$

Let us denote $\mathcal{S} = \{i : \boldsymbol{\beta}_i^* \neq 0 \ \forall\, i = 0, 1, \dots, K\}$, then we can decompose $\mathbf{X}$ into $\mathbf{X}_\mathcal{S}$ and $\mathbf{X}_{\mathcal{S}^c}$, where $\mathbf{X}_\mathcal{S}$ is the columns of $\mathbf{X}$ whose indices are in $\mathcal{S}$ and $\mathbf{X}_{\mathcal{S}^c}$ is the complement of $\mathbf{X}_\mathcal{S}$. And we can also decompose $\boldsymbol{\beta}$ into $\boldsymbol{\beta}_\mathcal{S}$ and $\boldsymbol{\beta}_{\mathcal{S}^c}$, where $\boldsymbol{\beta}_\mathcal{S}$ is the subvector of $\boldsymbol{\beta}$ only contains elements whose indices are in $\mathcal{S}$ and $\boldsymbol{\beta}_{\mathcal{S}^c}$ is the complement of $\boldsymbol{\beta}_\mathcal{S}$.

By using the decomposition, we can rewrite (S10.29) as

$$\begin{aligned}
\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix} \ = \ & \begin{pmatrix}\mathbf{X}_\mathcal{S}^\top\mathbf{X}_\mathcal{S} & \mathbf{X}_\mathcal{S}^\top\mathbf{X}_{\mathcal{S}^c} \\ \mathbf{X}_{\mathcal{S}^c}^\top\mathbf{X}_\mathcal{S} & \mathbf{X}_{\mathcal{S}^c}^\top\mathbf{X}_{\mathcal{S}^c}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}_\mathcal{S} - \boldsymbol{\beta}_\mathcal{S}^* \\ \boldsymbol{\beta}_{\mathcal{S}^c}\end{pmatrix} + \begin{pmatrix}\mathbf{X}_\mathcal{S}^\top \\ \mathbf{X}_{\mathcal{S}^c}^\top\end{pmatrix}\left[(\mathbf{X} - \mathbf{X}^*)_\mathcal{S}\boldsymbol{\beta}_\mathcal{S}^* - (\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^*)\right] + \\
& \lambda MN\begin{pmatrix}\mathbf{z}_\mathcal{S} \\ \mathbf{z}_{\mathcal{S}^c}\end{pmatrix}
\end{aligned} \tag{S10.30}$$

Suppose the primal-dual witness (PDW) construction gives us an solution $(\check{\boldsymbol{\beta}}, \check{\mathbf{z}}) \in \mathbb{R}^K \times \mathbb{R}^K$, where $\check{\boldsymbol{\beta}}_{\mathcal{S}^c} = 0$ and $\check{\mathbf{z}} \in \partial \|\check{\boldsymbol{\beta}}\|_1$. By plugging $(\check{\boldsymbol{\beta}}, \check{\mathbf{z}})$ into the above equation, we have

$$\begin{aligned}
\check{\mathbf{z}}_{\mathcal{S}^c} &= \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}} (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1} \mathbf{z}_{\mathcal{S}} - \mathbf{X}_{\mathcal{S}^c}^\top \underbrace{(\mathbf{I} - \mathbf{X}_{\mathcal{S}} (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1} \mathbf{X}_{\mathcal{S}}^\top)}_{\mathbf{H}_{X_s}} \frac{[(\mathbf{X} - \mathbf{X}^*)_{\mathcal{S}} \boldsymbol{\beta}_{\mathcal{S}}^* - (\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^*)]}{\lambda MN} \\
&= \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}} (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1} \mathbf{z}_{\mathcal{S}} - \frac{1}{\lambda MN} \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{H}_{X_s} \underbrace{(\mathbf{X}_{\mathcal{S}} \boldsymbol{\beta}_{\mathcal{S}}^* - \nabla_t \mathbf{u})}_{\boldsymbol{\tau}}. \qquad \text{(S10.31)}
\end{aligned}$$

From (S10.31), we have

$$\begin{aligned}
P(\|\check{\mathbf{z}}_{\mathcal{S}^c}\|_\infty \geq 1) &= P\left(\left\| \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}} (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1} \mathbf{z}_{\mathcal{S}} - \frac{1}{\lambda MN} \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{H}_{X_s} \boldsymbol{\tau} \right\|_\infty > 1 \right) \\
&\leq P\left(\left\| \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}} (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1} \mathbf{z}_{\mathcal{S}} \right\|_\infty > 1 - \mu \right) + \\
&\quad P\left(\left\| \frac{1}{\lambda MN} \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{H}_{X_s} \boldsymbol{\tau} \right\|_\infty > \mu \right).
\end{aligned}$$

If we denote $\widetilde{Z_j} = \frac{1}{\lambda MN} (\mathbf{X}_{\mathcal{S}^c})_j^\top \mathbf{H}_{X_s} \boldsymbol{\tau}$, where $(\mathbf{X}_{\mathcal{S}^c})_j$ is the j-th column of $\mathbf{X}_{\mathcal{S}^c}$, then we have

$$P(\|\check{\mathbf{z}}_{\mathcal{S}^c}\|_\infty \geq 1) \leq P\left(\left\| \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}} (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1} \right\|_\infty > 1 - \mu \right) + P\left(\max_{j \in \mathcal{S}^c} |\widetilde{Z_j}| > \mu \right). \qquad \text{(S10.32)}$$

Now let us discuss the upper bound of the second term, i.e., $P\left(\max_{j \in \mathcal{S}^c} |\widetilde{Z_j}| > \mu \right)$. Because

$$\begin{aligned}
P\left(\max_{j \in \mathcal{S}^c} |\widetilde{Z_j}| > \mu \right) &= P\left(\left\| \frac{1}{\lambda MN} \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{H}_{X_s} \boldsymbol{\tau} \right\|_\infty > \mu \right) \\
&\leq P\left(\left\| \frac{1}{\lambda MN} \mathbf{X}_{\mathcal{S}^c}^\top \mathbf{H}_{X_s} \boldsymbol{\tau} \right\|_2 > \mu \right) \\
&\leq P\left(\left\| \frac{1}{\lambda MN} \mathbf{X}^\top \mathbf{H}_{X_s} \boldsymbol{\tau} \right\|_2 > \mu \right) \\
&\leq P\left(\frac{1}{\lambda MN} \|\mathbf{X}\|_2 \|\boldsymbol{\tau}\|_2 > \mu \right) \\
&\leq P\left(\|\boldsymbol{\tau}\|_2 > \lambda \mu \sqrt{\frac{MN}{K}} \right) \\
&\leq P\left(\|\boldsymbol{\tau}\|_\infty > \lambda \mu \frac{1}{\sqrt{K}} \right). \qquad \text{(S10.33)}
\end{aligned}$$

By Lemma S6.2, we know when

$$\lambda \mu \frac{1}{\sqrt{K}} > \mathscr{C}(\sigma, \|u\|_{L^\infty(\Omega)}) \frac{\log(N)}{N^{3/7-r}},$$

we have

$$P\left(\|\nabla_t \mathbf{u} - \mathbf{X} \boldsymbol{\beta}^*\|_\infty > \epsilon \right) < N e^{-N^r}.$$

So we know that

$$\begin{aligned}
P\left(\|\boldsymbol{\tau}\|_\infty > \lambda \mu \frac{1}{\sqrt{K}} \right) &= P\left(\|\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}} \boldsymbol{\beta}_{\mathcal{S}}^*\|_\infty > \lambda \mu \frac{1}{\sqrt{K}} \right) \\
&\leq P\left(\|\nabla_t \mathbf{u} - \mathbf{X} \boldsymbol{\beta}^*\|_\infty > \lambda \mu \frac{1}{\sqrt{K}} \right) \\
&< N e^{-N^r}. \qquad \text{(S10.34)}
\end{aligned}$$

By plugging the results in (S10.33) and (S10.34) into (S10.32), we have

$$\begin{aligned}
P(\|\check{\mathbf{z}}_{\mathcal{S}^c}\|_\infty \geq 1) &\leq P\left(\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty > 1-\mu\right) + P\left(\max_{j\in\mathcal{S}^c} |\widetilde{Z_j}| > \mu\right) \\
&\leq P\left(\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty > 1-\mu\right) + P\left(\|\boldsymbol{\tau}\|_\infty > \lambda\mu\frac{1}{\sqrt{K}}\right) \\
&\leq P\left(\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty > 1-\mu\right) + Ne^{-N^r}.
\end{aligned}$$

The probability for proper support set recovery is

$$\begin{aligned}
P(\|\check{\mathbf{z}}_{\mathcal{S}^c}\|_\infty < 1) &= 1 - P(\|\check{\mathbf{z}}_{\mathcal{S}^c}\|_\infty \geq 1) \\
&\geq 1 - \left[P\left(\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty > 1-\mu\right) + Ne^{-N^r}\right] \\
&= P\left(\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty \leq 1-\mu\right) - Ne^{-N^r} \\
&\leq P_\mu - Ne^{-N^r}.
\end{aligned}$$

Thus, we finish the proof. □

S10.6 Proof of Theorem 3.2

Proof. By equation (S10.30), we can solve $\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}}^*$ as

$$\boldsymbol{\beta}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}}^* = (\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\left[-\mathbf{X}_{\mathcal{S}}^\top (\mathbf{X}_{\mathcal{S}} - \mathbf{X}_{\mathcal{S}}^*)\boldsymbol{\beta}_{\mathcal{S}}^* + \mathbf{X}_{\mathcal{S}}^\top (\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^*) - \lambda MN\mathbf{z}_{\mathcal{S}}\right].$$

Thus, we have the following series of equations:

$$\begin{aligned}
&\max_{k\in\mathcal{S}} |\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^*| \\
\leq\ & \left\|(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty \left\|\mathbf{X}_{\mathcal{S}}^\top [\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^* - (\mathbf{X}_{\mathcal{S}} - \mathbf{X}_{\mathcal{S}}^*)\boldsymbol{\beta}_{\mathcal{S}}^*] - \lambda MN\mathbf{z}_{\mathcal{S}}\right\|_\infty \\
\leq\ & \left\|(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty \left[\left\|\mathbf{X}_{\mathcal{S}}^\top [\nabla_t \mathbf{u} - \nabla_t \mathbf{u}^* - (\mathbf{X}_{\mathcal{S}} - \mathbf{X}_{\mathcal{S}}^*)\boldsymbol{\beta}_{\mathcal{S}}^*]\right\|_\infty + \lambda MN\|\mathbf{z}_{\mathcal{S}}\|_\infty\right] \\
=\ & \left\|(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty \left[\left\|\mathbf{X}_{\mathcal{S}}^\top (\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}}^*)\right\|_\infty + \lambda MN\|\mathbf{z}_{\mathcal{S}}\|_\infty\right] && \text{(S10.35)} \\
\leq\ & \left\|\left(\frac{\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}}}{MN}\right)^{-1}\right\|_\infty \left(\frac{\left\|\mathbf{X}_{\mathcal{S}}^\top (\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}^*)\right\|_\infty}{MN} + \lambda\right) && \text{(S10.36)} \\
\leq\ & \sqrt{K}C_{\min}\left(\frac{\left\|\mathbf{X}_{\mathcal{S}}^\top (\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}^*)\right\|_\infty}{MN} + \lambda\right) && \text{(S10.37)} \\
\leq\ & \sqrt{K}C_{\min}\left(\frac{\|\mathbf{X}_{\mathcal{S}}\|_{\infty,\infty}\|\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}^*\|_\infty}{MN} + \lambda\right) && \text{(S10.38)} \\
\leq\ & \sqrt{K}C_{\min}\left(\frac{\|\mathbf{X}_{\mathcal{S}}\|_F\|\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}^*\|_\infty}{\sqrt{MN}} + \lambda\right) \\
\leq\ & \sqrt{K}C_{\min}\left(\frac{\sqrt{MNK}\|\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}^*\|_\infty}{\sqrt{MN}} + \lambda\right) && \text{(S10.39)} \\
=\ & \sqrt{K}C_{\min}\left(\sqrt{K}\|\nabla_t \mathbf{u} - \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}^*\|_\infty + \lambda\right) \\
\leq\ & \sqrt{K}C_{\min}\left(\sqrt{K}\mathscr{C}_{(\sigma,\|u\|_{L^\infty(\Omega)})}\frac{\log(N)}{N^{3/7-r}} + \lambda\right) && \text{(S10.40)}
\end{aligned}$$

Equation (S10.35) is because $\nabla_t \mathbf{u}^* = \mathbf{X}_{\mathcal{S}}\boldsymbol{\beta}_{\mathcal{S}}$. Inequality (S10.36) is because $\|\mathbf{z}_{\mathcal{S}}\|_\infty = 1$. Inequality (S10.37) is because of Condition 3.2. Inequality (S10.38) is because for a matrix $\mathbf{A}$ and a vector $\mathbf{x}$, we have $\|\mathbf{A}\mathbf{x}\|_q \leq \|\mathbf{A}\|_{p,q}\|\mathbf{x}\|_p$. Here the matrix norm for matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ in $\|\mathbf{A}\|_{\infty,\infty} = \|\text{vector}(\mathbf{A})\|_\infty$. In inequality (S10.39), the norm of matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is that $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2}$, and the norm of vector $\mathbf{a} \in \mathbb{R}^d$ is $\|\mathbf{a}\|_\infty = \max_{1\leq i\leq d} |a_i|$. Inequality (S10.39) is because we normalized columns of $\mathbf{X}$. Inequality (S10.40) is due to Lemma Lemma S6.2 under probability $1 - O(Ne^{-N^r}) \to 1$. □

S11 The full model used in Section 4

The full model used in Section 4 is

$$\begin{aligned}\frac{\partial}{\partial t}u(x,t) \quad = \quad & \beta_1^* + \beta_2^* u(x,t) + \beta_3^* \frac{\partial}{\partial x}u(x,t) + \beta_4^* \frac{\partial^2}{\partial x^2}u(x,t) + \\ & \beta_5^* \left[u(x,t)\right]^2 + \beta_6^* \left[\frac{\partial}{\partial x}u(x,t)\right]^2 + \beta_7^* \left[\frac{\partial}{\partial x}u(x,t)\right]^2 + \\ & \beta_8^* u(x,t)\frac{\partial}{\partial x}u(x,t) + \beta_9^* u(x,t)\frac{\partial^2}{\partial x^2}u(x,t) + \\ & \beta_{10}^* \frac{\partial}{\partial x}u(x,t)\frac{\partial^2}{\partial x^2}u(x,t)\end{aligned} .$$

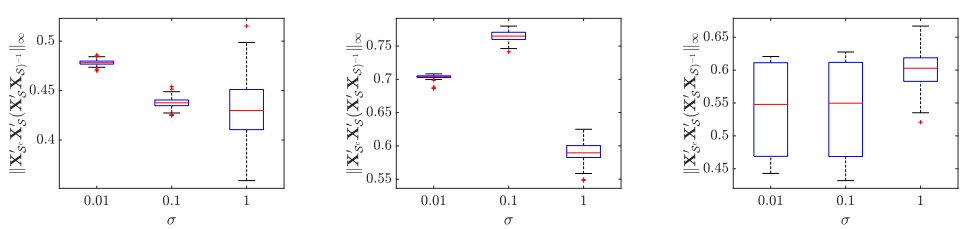
S12 Checking Conditions of Example 1,2,3

In this section, we check Condition 3.1 - Condition 3.5 of the above three examples: (1) example 1 (the transport equation), (2) example 2 (the inviscid Burgers' equation) and (3) example 3 (the viscous Burgers' equation).

S12.1 Verification of Condition 3.1, 3.2

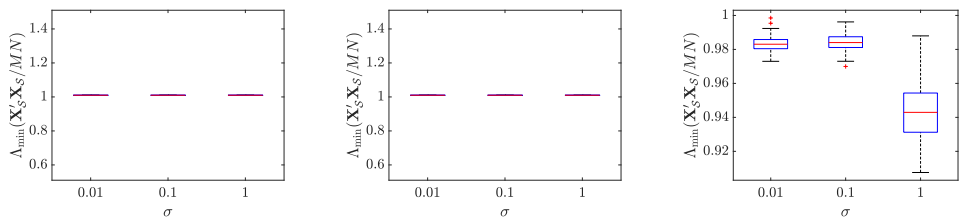
In this section, we check the Condition 3.1 - Condition 3.2 under example 1,2,3, though the applicability of the results is by no means restricted to these.

The verification results can be found in Fig. 1 and Fig. 2, where (a),(b),(c) are the box plot of $\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty$ and the minimal eigenvalue of matrix $\frac{1}{NM}\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}}$ of these three examples under $\sigma = 0.01, 0.1, 1$, respectively. From Fig. 1, we find the value of $\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty$ is smaller than 1, so there exist a $\mu \in (0, 1]$ such that Condition 3.1 is met. From Fig. 2, we find the minimal eigenvalue of matrix $\frac{1}{MN}\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}$ are all strictly larger than 0, so we declare Condition 3.2 is satisfied.



(a) example 1: (b) example 2: (c) example 3:

Figure 1: Box plots of $\left\|\mathbf{X}_{\mathcal{S}^c}^\top \mathbf{X}_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}})^{-1}\right\|_\infty$ under $\sigma = 0.01, 0.1, 1$ when $M = N = 100$.



(a) example 1: (b) example 2: (c) example 3:

Figure 2: Box plots of the minimal eigenvalue of matrix $\frac{1}{NM}\mathbf{X}_{\mathcal{S}}^{\top}\mathbf{X}_{\mathcal{S}}$ under $\sigma = 0.01, 0.1, 1$ when $M = N = 100$.

S12.2 Verification of Condition 3.3 and Condition 3.4

In example 1,2,3, the design points $x_0, x_1, \ldots, x_{M-1}$ and $t_0, t_1, \ldots, t_{N-1}$ are equally spaced, i.e., $x_0 = 1/M, x_1 = 2/M, \ldots, x_{M-1} = 1$ and $t_0 = 0.1/N, t_1 = 0.2/N, \ldots, t_{N-1} = 0.1$. Under this scenario, there exist an absolutely continuous distribution $F(x) = x$ for $x \in [1/M, 1]$ and $G(t) = 0.1t$ for $t \in [0.1/N, 0.1]$, where the empirical c.d.f. of the design points $x_0, x_1, \ldots, x_{M-1}$ and $t_0, t_1, \ldots, t_{N-1}$ will converge to $F(x), G(t)$, respectively, as $M, N \to +\infty$. For the $F(x), G(t)$, we know their first derivatives is bounded for $x \in [1/M, 1]$ and $t \in [0.1/N, 0.1]$, respectively. In the simulation of this paper, we take the equally spaced design points as an illustration example, and its applicability is by no means restricted to this case.

S12.3 Verification of Condition 3.5.

The Condition 3.5 ensures that the smoothing parameter does not tend to zero too rapidly. Silverman (1984) shows that for the equally spaced design points, this condition meets. For other types of design points, for instance, randomly and independently distributed design points, it can also be verified that Condition 3.5 is satisfied (see Silverman, 1984, Section 2).

S13 Details of the Case Study

The header of the CALIPSP dataset and its visualization can be found in Table 3 and Fig. 9(a), respectively. Fig. 9(a) shows presents the curves of the dynamic in the CALIPSP dataset, where the x-axis is the longitude and the y-axis is the value of the observed temperature. Here the black curve plots the observed temperature in January 2017, and the lighter color presents the later month. As seen from Fig. 9, we find overall there is an increasing trend of the temperature in the first half-year and then the temperature decreases.

S14 More details of Fig. 10

The three-dimension surface plot of the observed temperature in 2017 can be found in Fig. 10(a.1), whose fitted value can be found in Fig. 10(a.2). The three-dimension surface plot of the residual between the observed temperature and the fitted temperature can be found in Fig. 10(a.3). Seeing from Fig. 10(a.1)-(a.3), we find the fitted temperature captures the dynamic trend of the raw data well. Although the magnitude of the residual is not small, it is still satisfying given the following reasons. The fitted value by using the explicit Euler method only

Table 3: The header of the CALIPSP dataset

	longitude						
	-177.5	-172.5	-167.5	...	167.5	172.5	177.5
Jan 2017	-46.5103	-48.4720	-44.6581	...	-44.2778	-47.3354	-44.0146
Feb 2017	-46.2618	-43.1994	-47.6370	...	-46.8409	-46.2727	-40.5556
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Dec 2017	-47.3145	-47.6505	-50.8332	...	-43.9705	-43.0475	-46.0618

[1] The data is downloaded from `https://asdc.larc.nasa.gov/data/CALIPSO/LID_L3_Tropospheric_APro_CloudFree-Standard-V4-20/` (registration is required).
[2] The negative and positive longitude refer to the west and east longitude, respectively.

serves as the baseline method, which is not accurate enough for fitting. In this paper, we focus on PDE identification, so we use the most simple method – the explicit Euler method – to check if the identified PDE model in (5.16) can capture the features of the underlying PDE model. The more advanced method – Runge-Kutta fourth order method (RK4) – is not implementable in our content, and the reasons are explained in online supplementary material.

The value to plot Fig. 10(a.2) is calculated as follows. First, we use the identified PDE model in (5.16) to predict the value of $\frac{\partial}{\partial t}u(x,t)$ in January 2017. Then, we use the explicit Euler method (Butcher and Goodwin, 2008) to predict the future value from February 2017 to December 2017, i.e., we have $u(x,t+\Delta t) = u(x,t) + \frac{\partial}{\partial t}u(x,t)\Delta t$.

References

Aydin, D., M. Memmedli, and R. E. Omay (2013). Smoothing parameter selection for nonparametric regression using smoothing spline. *European Journal of Pure and applied mathematics 6*(2), 222–238.

Butcher, J. C. and N. Goodwin (2008). *Numerical methods for ordinary differential equations*, Volume 2. Wiley Online Library.

Craven, P. and G. Wahba (1978). Smoothing noisy data with spline functions. *Numerische mathematik 31*(4), 377–403.

Fan, J., T. Gasser, I. Gijbels, M. Brockmann, and J. Engel (1997). Local polynomial regression: optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics 49*(1), 79–99.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software 33*(1), 1.

Hudson, H. (1974). Empirical bayes estimation technical report no 58.

Lambert, J. D. et al. (1991). *Numerical methods for ordinary differential systems*, Volume 146. Wiley New York.

Mack, Y.-p. and B. W. Silverman (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 61*(3), 405–415.

Mallows, C. L. (2000). Some comments on cp. *Technometrics 42*(1), 87–94.

Messer, K. (1991). A comparison of a spline estimate to its equivalent kernel estimate. *The Annals of Statistics 19*(2), 817–829.

Reinsch, C. H. (1967). Smoothing by spline functions. *Numerische mathematik 10*(3), 177–183.

Rice, J. and M. Rosenblatt (1983). Smoothing splines: regression, derivatives and deconvolution. *The annals of Statistics*, 141–156.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The annals of mathematical statistics 23*(3), 470–472.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 177–184.

Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *The Annals of Statistics*, 898–916.

Tusnády, G. (1977). A remark on the approximation of the sample DF in the multidimensional case. *Periodica Mathematica Hungarica 8*(1), 53–55.

Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische mathematik 24*(5), 383–393.

Winkelbauer, A. (2012). Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*.