Nonlinear Numerical Methods for Stochastic Differential Equations

by

Yuchen Dong

A Dissertation

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Mathematical Sciences

by

August 2020

APPROVED:

Dr. Zhongqiang Zhang, Advisor
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Marcus Sarkis
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Gu Wang
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Qingshuo Song
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Xiu Yang
Department of Industrial and Systems Engineering
Lehigh University

Acknowledgements

First and foremost, I would like to thank my advisor and mentor, Professor Zhongqiang Zhang, for his guidance and support in my Ph.D. program. My work would not have been possible without his help. His advice on my research as well as on my career is valuable. I would also like to express my sincere thanks to my dissertation committee members, Dr. Marcus Sarkis, Dr. Gu Wang, Dr. Qingshuo Song, and Dr. Xiu Yang. I greatly appreciate the time and effort you put forth in helping to improve my work.

I would also like to express my appreciation to all of the faculty members, staff from Mathematical Sciences Department of Worcester Polytechnic Institute for all their supports, especially Dr. Luca Capogna, Dr. Roger Lui, Dr. Darko Volkov, Dr. Homer Walker, Dr. Min Wu, Dr. Stephan Sturm, Dr. Marcel Blais, and Rhonda Podell.

I would like to thank my friends for their kindness and encouragement. In particular, I express my sincere gratitude to all those who have listened and contributed ideas, including Xiangyu Zhu, Zhaopeng Hao, Jiaxuan Ye, Jiamin Jian, Hedy Cheng and Xiao Shen. I am also grateful to all friends of mine who I met at WPI.

Finally, I would like to thank my parents, my father Weimin Dong and my mother Xiaolei Chen, for their love, support, and effort to understand my study. Without them, I would not be where I am today.

Abstract

Turbulence modeling is one of the most important and challenging scientific problems. One significant development for theoretical turbulence is the fractional modeling, where the fractional calculus is applied to describe the turbulence mixing. However, there are several issues in the application of fractional models. First, the order of fractional operators and its coefficients are unknown. Second, there are no efficient solvers for the Navier-Stokes equation with fractional operators. We propose methods of parameter estimation for the fractional operators for the stochastic Navier-Stokes (SNS) equation from simulation data and some nonlinear methods for SNS, addressing mainly the aforementioned first issue.

Our first approach of parameter estimation is based on spectral methods for SNS. The key of this method is splitting the SNS into two subproblems: one is a linear parabolic equation and the other one corresponds to the nonlinear term in SNS. For the linear subproblem, we develop consistent maximum likelihood estimators for the coefficients of fractional operators and prove the convergence order when the frequency goes to infinity. We also develop estimators based on the quadratic variation of the underlying stochastic process. For the nonlinear subproblem, we apply the maximum likelihood estimations.

Though spectral methods work well in very regular domains, they are not flexible enough to fit many applications, such as for problems in complex domains. For wider applications, we plan to use the neural network approximation to solve the SNS and perform parameter estimation with the network approximation. The key component is to develop a mesh-free network solver for the SNS.

Positivity-preserving is critical to achieving efficient numerical methods with correct solutions, such as in two-phase flow modeled together with stochastic phase-field models, e.g., stochastic Allen-Cahn equations. The main issue we focus on here is to preserve the positivity, a property that both analytical and numerical solutions should have. We investigate nonlinear schemes for stochastic ordinary differential equations which have the similar nonlinearity as in the Allen-Cahn equation. The nonlinearity has polynomial growth and thus can deteriorate the convergence of numerical schemes. Several numerical schemes are developed to tame the polynomial growth and guarantee the convergence.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Stochastic Partial differential equations are widely used to describe the evolution of dynamical systems with spatial and temporal uncertainties in the field of pure and applied mathematics. There are many important applications in physics and finance. Meanwhile, fractional calculus arises in many fields of engineering and science. For example, the spatial fractional stochastic Naiver-Stokes equation describes the fluid dynamics with turbulence mixing. The external force acting on the fluid is randomly fluctuating [35, 61].

Given the observations, parameter estimation becomes a valuable problem to be considered. The parameters in the models represent various physical phenomena. For example, the coefficient of the Laplacian term in the Navier-Stokes equation represents the viscosity of fluid. This procedure introduces a method to compare the estimated values with the physical quantities, and it can also validate and test the correctness of these models. The second reason is to give more precise information on the model, if we don't have full knowledge of the model. This step provides a way to measure these unknowns based on the observations mathematically.

Fractional calculus has several definitions introduced in the following section. In chapters 2 and 3, solutions are computed in various numerical methods, such as the spectral methods and implicit Euler schemes. We derive the maximum likelihood estimators for the fractional stochastic heat equations in 1D and fractional stochastic Navier Stokes equation in 2D. To solve the nonlinear issues in the Navier-Stokes equation, we use the divergence-free operator. We present several numerical examples to show the accuracy of estimators.

Furthermore, coefficients of stochastic differential equations (SDEs) grow nonlinearly in many applications of SDEs. When these SDEs with coefficients of superlinear growth are solved using numerical methods, explicit numerical schemes usually fail to converge in the sense of mean-square and moments, e.g., [24, 41]. We introduce several modified Euler schemes with a half order convergence. The numerical examples are presented to show the computational performance and convergence order.

1.2 Some preliminaries

1.2.1 Sobolev spaces

In this subsection, let's introduce the Sobolev spaces and characterization of Sobolev spaces using Fourier transform. [17]

Definition 1.2.1 (Definition of Sobolev spaces)**.** *Let k be a nonnegative integer and fix $1 \leq p \leq \infty$. The Sobolev space $W^{k,p}(U)$ consists of all locally summable functions $u : U \to \mathbb{R}$ such that for each multiindex α with $|\alpha| \leq k$, $D^\alpha u$ exists in the weak sense and belongs to $L^p(U)$.*

Definition 1.2.2. *If $u \in W^{k,p}(U)$, we define its norm to be*

$$||u||_{W^{k,p}(U)} := \begin{cases} \left(\sum_{|\alpha| \leq k} \int_U |D^\alpha u|^p dx \right)^{\frac{1}{p}}, & 1 \leq p < \infty \\ \sum_{|\alpha| \leq k} ess \sup_U |D^\alpha u|, & p = \infty. \end{cases} \tag{1.1}$$

Remark 1.2.3. *We usually write $H^k(U) = W^{k,2}(U)$. The letter H stands for a Hilbert space. Note that $H^0(U) = L^2(U)$. The norm of $H^k(U)$ is $||u||_{H^k(U)} = \left(\sum_{|\alpha| \leq k} \int_U |D^\alpha u|^2 dx \right)^{\frac{1}{2}}$.*

Remark 1.2.4. *Introduce the multiindex notation: Given a multiindex α, define*

$$D^\alpha u(x) := \frac{\partial^{|\alpha|} u(x)}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} ... \partial x_n^{\alpha_n}} = \partial_{x_1}^{\alpha_1} ... \partial_{x_n}^{\alpha_n} u, \text{ where } |\alpha| = \alpha_1 + ... + \alpha_n. \tag{1.2}$$

The following functions in Theorem 1.2.5 are complex valued.

Theorem 1.2.5 (Characterization of H^k by Fourier Transform)**.** *Let k be a nonnegative integer. (1) A function $u \in L^2(\mathbb{R}^n)$ belongs to $H^k(\mathbb{R}^n)$ if and only if*

$$(1 + |y|^k)\hat{u} \in L^2(\mathbb{R}^n). \tag{1.3}$$

(2) In addition, there exists a positive constant C such that

$$\frac{1}{C}||u||_{H^k(\mathbb{R}^n)} \leq ||(1 + |y|^k)\hat{u}||_{L^2(\mathbb{R}^n)} \leq C||u||_{H^k(\mathbb{R}^n)}. \tag{1.4}$$

Proof. (1) Assume that $u \in L^2(\mathbb{R}^n) \subset H^k(\mathbb{R}^n)$. We want to show that $(1 + |y|^k)\hat{u} \in L^2(\mathbb{R}^n)$. That is to show

$$\int_{\mathbb{R}^n} (1 + |y|^k)^2 |\hat{u}|^2 dy < \infty. \tag{1.5}$$

Since $u \in H^k(\mathbb{R}^n)$, for each multiindex $|\alpha| \leq k$, we have $D^\alpha u \in L^2(\mathbb{R}^n)$. If $u \in C^k$ has compact support and properties of Fourier transform, we have

$$\widehat{D^\alpha u} = (iy)^\alpha \hat{u}. \tag{1.6}$$

According to the approximation theory by smooth functions, the Fourier transform (1.6) is true if $u \in H^k(\mathbb{R}^n)$. Thus, $|(iy)^\alpha \hat{u}| \in L^2(\mathbb{R}^n)$ for each $|\alpha| \leq k$. Choose $\alpha = (k, 0, ..., 0), (0, k, ..., 0), ..., (0, ..., k)$. We obtain that

$$\int_{\mathbb{R}^n} |iy|^{2k} |\hat{u}|^2 dy = \int_{\mathbb{R}^n} |y|^{2k} |\hat{u}|^2 dy \leq C \int_{\mathbb{R}^n} |D^k u|^2 dy < \infty. \tag{1.7}$$

Then,

$$\int_{\mathbb{R}^n} (1 + |y|^k)^2 |\hat{u}|^2 dy \leq C \int_{\mathbb{R}^n} (1 + |y| + ... + |y|^{2k}) |\hat{u}|^2 dy \leq C ||u||^2_{H^k(\mathbb{R}^n)} < \infty, \tag{1.8}$$

so $(1 + |y|^k)|\hat{u}| \in L^2(\mathbb{R}^n)$.
Conversely, we suppose that $(1 + |y|^k)|\hat{u}| \in L^2(\mathbb{R}^n)$. For $|\alpha| \leq k$,

$$||(iy)^\alpha \hat{u}||^2_{L^2(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |iy|^{2\alpha} |\hat{u}|^2 dy \leq \int_{\mathbb{R}^n} |y|^{2|\alpha|} |\hat{u}|^2 dy \leq C \int_{\mathbb{R}^n} (1+|y|^k)^2 |\hat{u}|^2 dy = C||(1+|y|^k)\hat{u}||^2_{L^2(\mathbb{R}^n)}. \tag{1.9}$$

For each test function $\phi \in C_c^\infty(\mathbb{R}^n)$, denote $u_\alpha := (iy\check{)}^\alpha \hat{u}$

$$\int_{\mathbb{R}^n} (D^\alpha \phi) \bar{u} dx = \int_{\mathbb{R}^n} \widehat{D^\alpha \phi} \bar{\hat{u}} = \int_{\mathbb{R}^n} (iy)^\alpha \hat{\phi} \bar{\hat{u}} dy = (-1)^{|\alpha|} \int_{\mathbb{R}^n} \phi \overline{u_\alpha} dx. \tag{1.10}$$

Thus, $u_\alpha = D^\alpha u$ in the weak sense. By (1.9), $|D^\alpha u| \in L^2(\mathbb{R}^n)$. Hence $u \in H^k(\mathbb{R}^n)$. □

Remark 1.2.6. *From Theorem 1.2.5 (2), these two norms are equivalent. Thus, the Sobolev space $H^k(\mathbb{R}^n)$ can be equivalently defined by*

$$H^k(\mathbb{R}^n) := \{u \in L^2(\mathbb{R}^n) : (1 + |y|^k)|\hat{u}| \in L^2(\mathbb{R}^n)\} \tag{1.11}$$

Remark 1.2.7. *Assume that $0 < s < \infty$ and $u \in L^2(\mathbb{R}^n)$. Then $u \in H^s(\mathbb{R}^n)$, if $(1 + |y|^s)\hat{u} \in L^2(\mathbb{R}^n)$. For noninteger s, we set*

$$||u||_{H^s(\mathbb{R}^n)} := ||(1 + |y|^s)|\hat{u}|||_{L^2(\mathbb{R}^n)}. \tag{1.12}$$

This is also called fractional Sobolev spaces.

When the domain $\mathcal{D}$ is regular, the Sobolev-Hilbert spaces can also be characterized by the eigen-pairs of the operator $-\Delta$ with vanishing Dirichlet boundary conditions. For example, on the interval $[0, 1]$, the eigenfunctions are

$$e_k(x) = \sqrt{2} \sin(k\pi x), \quad k \geq 1, \tag{1.13}$$

and the corresponding eigenvalues $\lambda_k = \pi^2 k^2$. The eigenfunctions $\{e_m\}_{m=1}^\infty$ form an orthonormal basis in $L^2(\mathcal{D})$.

For any $s \leq 2$, we define

$$\dot{H}^s = \dot{H}^s(\mathcal{D}) = \mathcal{D}((-\Delta)^{s/2}) = \left\{ v \,\middle|\, \|v\|_s = \left\|(-\Delta)^{s/2} v\right\| = \Big(\sum_{k=1}^\infty \lambda_k^s [(v, e_k)]^2\Big)^{1/2} < \infty \right\}.$$

It is known (see e.g. [59]) that $\dot{H}^s = H^s(\mathcal{D})$, where $H^s(\mathcal{D})$ is the classical Sobolev-Hilbert space over $\mathcal{D}$.

1.2.2 Fractional Derivatives

The fractional differential operator is a nonlocal operator used in various real-world models, such as fluid dynamics [40], finance [14], material science [4] etc. Let's recall the definitions and some useful properties. The first definition is related to spectral/Fourier representation.

Definition 1.2.8. *Let $\alpha \in (0,1)$, the fractional Laplacian of order α can be defined on functions $u : \mathbb{R}^d \to \mathbb{R}$ as a Fourier multiplier given the formula*

$$\mathcal{F}((-\Delta)^{\alpha} u)(\xi) = |\xi|^{2\alpha} \mathcal{F}(u)(\xi), \tag{1.14}$$

where $\mathcal{F}$ is a Fourier transform.

The spectral representation is used in Chapter 2 and 3. One of another definition is related to directional representation. The integral characterization of the fractional Laplacian can be found in [56]. The operator is written as

$$(-\Delta)^{\alpha/2} u(x) = C_{\alpha,d} \int_{|\theta|=1} D_\theta^\alpha u(x) d\theta, \quad x, \theta \in \mathbb{R}^d, \alpha \in (0,2]\backslash\{1\}, \tag{1.15}$$

where $C_{\alpha,d}$ is a scaling constant related to Gamma function before the integral [56, 47].

$$C_{\alpha,d} = \frac{\Gamma(\frac{1-\alpha}{2})\Gamma(\frac{d+\alpha}{2})}{2\pi^{\frac{1+d}{2}}}. \tag{1.16}$$

The fractional Laplacian operator is consistent with the integer ones. In particular, we have the following pointwise limits

$$\lim_{\alpha \to 0} (-\Delta)^{\alpha/2} u(x) = u(x),$$
$$\lim_{\alpha \to 2} (-\Delta)^{\alpha/2} u(x) = -\Delta u(x).$$

1.2.3 Parameter estimations for SDEs

In this section, we present the parameter estimation for SDEs. The main idea here is to apply Girsanov's theorem to connect two probability measures. Then the likelihood function is defined by the Radon Nykodym derivatives. Consider the following two scalar diffusion processes X_t and Y_t driven by the same Brownian motion.

$$dX = A(t, X(t))\, dt + \sigma(t, X(t))\, dw(t), \quad X(0) = X_0, \tag{1.17}$$
$$dY = a(t, Y(t))\, dt + \sigma(t, Y(t))\, dw(t), \quad Y(0) = X_0. \tag{1.18}$$

Here the functions A, a and σ satisfy the conditions to ensure existence of a unique strong solution. Assume that the initial conditions are independent of $w(t)$ and $\sigma \geq \sigma_0 > 0$. Let

$$B(t, x) = \frac{A(t, x) - a(t, x)}{\sigma(t, x)}, \tag{1.19}$$

then by Girsanov's theorem (e.g., Theorem 8.6.8 of [45]),

$$\widetilde{w}(t) = -\int_0^t B(s, Y(s))\, ds + w(t), \quad 0 \leq t \leq T \tag{1.20}$$

is a standard Brownian motion under P_T^Y, where the measure P_T^Y is defined by

$$P_T^Y(A) = \mathbb{E}^{\mathbb{P}}\left[\mathbf{1}_A \exp\left(\int_0^T B(t, Y(t))\, dw(t) - \frac{1}{2}\int_0^T B^2(t, Y(t))\, dt\right)\right]. \tag{1.21}$$

Here $\mathbb{P}$ or P_T^X is the measure generated by the process X on $C([0,T];\mathbb{R})$. We can write $Y(t)$ [1] as

$$dY = A(t, Y(t))\, dt + \sigma(t, Y(t))\, d\widetilde{w}(t). \tag{1.22}$$

Then we may define the likelihood (Radon-Nykodym derivative, see e.g. [32])

$$\begin{aligned}
\frac{dP_T^X}{dP_T^Y} &= \exp\left(\int_0^T B(t, Y(t))\, dw - \frac{1}{2}\int_0^T B^2(t, Y(t))\, dt\right) \\
&= \exp\left(\int_0^T \frac{A(t, Y(t)) - a(t, Y(t))}{\sigma^2(t, Y(t))}\, dY(t) - \frac{1}{2}\int_0^T \frac{A^2(t, Y(t)) - a^2(t, Y(t))}{\sigma^2(t, Y(t))}\, dt\right) \\
&= \exp\left(\int_0^T \frac{A(t, X(t)) - a(t, X(t))}{\sigma^2(t, X(t))}\, dX(t) - \frac{1}{2}\int_0^T \frac{A^2(t, X(t)) - a^2(t, X(t))}{\sigma^2(t, X(t))}\, dt\right)
\end{aligned} \tag{1.23}$$

Consider the estimation of the parameter θ in the following scalar equation

$$\mathrm{d}X(t) = \theta f(X(t))\, \mathrm{d}t + \sigma b(X(t))\, \mathrm{d}w(t), \tag{1.24}$$

where $w(t)$ is a standard Brownian motion and $f(\cdot)$, $b(\cdot)$ are suitable real-valued functions such that a strong solution X is well-posed. Let $A(t,x) = \theta f(x)$ and $a(t,x) = \theta_0 f(x)$. Then by taking the logarithm of the likelihood (1.23) and letting the derivative of the log-likelihood be zero, we have

$$\widehat{\theta} = \frac{\int_0^T \frac{f(X(t))}{\sigma^2(X(t))}\, dX(t)}{\int_0^T \frac{f^2(X(t))}{\sigma^2(X(t))}\, dt}. \tag{1.25}$$

Multiplying $f(X)/\sigma^2(X)$ over both sides of the equation of X, we obtain that

$$\theta = \frac{\int_0^T \frac{f(X(t))}{\sigma^2(X(t))}\, dX(t) + \int_0^T \frac{f(X(t))}{\sigma(X(t))}\, dw(t)}{\int_0^T \frac{f^2(X(t))}{\sigma^2(X(t))}\, dt}. \tag{1.26}$$

Thus, we have

$$\widehat{\theta} - \theta = \frac{\int_0^T \frac{f(X(t))}{\sigma(X(t))}\, dw(t)}{\int_0^T \frac{f^2(X(t))}{\sigma^2(X(t))}\, dt}. \tag{1.27}$$

Under certain conditions on f and σ, we may obtain that $\widehat{\theta} - \theta$ converges in distribution to a normal random variable with zero mean and a certain variance. For convergence and its rate, we will defer discussions to the next section, where we will use the Ornstein–Uhlenbeck process as an example.

[1] Here we then have a weak solution $(Y(t), \widetilde{w}(t))$ to equation.

1.2.4 Parameter estimations for SPDEs

Compared with stochastic differential equations, the SPDEs are driven by the Q-cylindrical Gaussian process.

Definition 1.2.9. *[35] Let $X = X(t)$ be zero-mean Gaussian process with the covariance function $R(t,s) = \mathbb{E}[X(t)X(s)]$. A Q-cylindrical X-process on (or over) a Hilbert space H is a collection $\mathbb{X}^Q = \{X_f^Q(t), f \in H, t \in [0,T]\}$ of zero mean Gaussian random variables with the following property: there exists a bounded linear non-negative self-adjoint operator $Q : H \to H$ such that*

$$\mathbb{E}[X_f^Q(t)X_g^Q(s)] = (Qf, g)_H R(t,s)$$

for all $f, g \in H$ and all $t, s \in [0,T]$. If Q is the identity operator, then X^Q is called a cylindrical x-process and denoted simply by X.

In particular, a Q-cylindrical Brownian motion $W^Q = W^Q(t), t \in [0,T]$ on a Hilbert space H, is a collection $\{W_f^Q(t), f \in H, t \in [0,T]\}$ of zero-mean Gaussian random variables such that

$$\mathbb{E}[W_f^Q(t)W_g^Q(s)] = (Qf, g)_H \min(t,s).$$

It can also be represented in the series form.

$$W^Q(t) = \sum_{k=1}^{\infty} \sqrt{q_k} m_k W_k(t),$$

where $\{m_k\}$ is a complete orthonormal basis. For $\theta \in \mathbb{R}$, consider the equation

$$\dot{u}(t) + (A_0 + \theta A_1)u(t) = \sigma \dot{W}^Q(t), \tag{1.28}$$

with initial condition $u(0)$. Then, the dynamcis of the Fourier coefficients u_k is

$$du_k(t) = -\mu_k(\theta)u_k(t)dt + \sigma\sqrt{q_k}dw_k(t). \tag{1.29}$$

Denote $A(t, u_k(t)) = -\mu_k(\theta)u_k(t)$ and $\sigma(t, u_k(t)) = \sigma\sqrt{q_k}$. Regard U as the measure related to the observations and V as the measure related to the exact value. By Girsanov's theorem, we have the log-likelihood function defined by Radon-Nykodym derivatives

$$\frac{d\mathbb{P}^U}{d\mathbb{P}^V} = \exp\left(\int_0^T \frac{A(t, u_k(t)) - a(t, u_k(t))}{\sigma^2(t, u_k(t))} du_k(t) - \int_0^T \frac{A^2(t, u_k) - a^2(t, u_k(t))}{\sigma^2(t, u_k(t))} dt\right). \tag{1.30}$$

With the same procedures as the estimation of SDEs, we have the maximum likelihood estimation on θ. The consistency and asymptotic normality theorem are well-established in [35]. However, the operator of A_1 only includes the operators with integer orders. In my work, we consider the parameter estimation in the stochastic heat equation with the fractional term. There are no suggested estimators on the fractional order α in the existing literature.

There are extensive works on parameter estimations on SDEs. Nevertheless, the field of parameter estimations for stochastic partial differential equations is still developing. see e.g. [10] – many important problems haven't been resolved, see e.g. [11] for stochastic Navier-Stokes, where a simple and seemingly working maximum likelihood estimation is proposed but the convergence and its convergence rate hasn't been established. A new class of estimators, called trajectory fitting estimators was introduced in [12]. The infinite dimensional parameter estimation for stochastic heat diffusion equations is considered using the method of sieves and the consistency property is also studied for the long run data in [1]. The traditional maximum likelihood estimations on the diagonalizable stochastic parabolic equation with fractional noise with any Hurst parameter $H \in (0, 1)$ was investigated in [9]. A completely new closed-form exact estimator was constructed in [13] for a stochastic parabolic equation with multiplicative noise under the assumption that the SPDEs can be reduced to an infinite system of uncoupled diffusion processes.

Besides the maximum likelihood estimators, there are also different types of estimators introduced in this field. Bayes estimation for some stochastic partial differential equations was introduced in [49]. In this paper, the Bayes estimation is limited to the SPDEs with linear drift. Nonparametric estimation was introduced in [50] based on the SPDEs with the same properties.

1.2.5 Neural networks approximation for fractional PDEs with inference on parameters

Deep learning methods are efficient in solving the forward and inverse problems on PDEs. This is still a fairly new research field. It connects the deep learning techniques and classic PDE. There are still many problems to be solved.

Several works are done in the past few years. The physics-informed neural networks is introduced in [52]. In this paper, they introduced the efficient mechanism for regularizing the training of deep neural networks in small data regimes and put forth a deep learning framework on the combination of data and mathematical models. In the paper, several classic problems are demonstrated including the Navier-Stokes equation, reaction-diffusion systems and so on. The works in [63] consider the parametric uncertainty as a stochastic process. It combines the arbitrary polynomial chaos with PINNs for both forward and inverse problems. [48] worked on the fractional physics-informed neural networks. In this paper, it applied the deep learning techniques to find the numerical solutions to fractional PDEs in both forward and inverse problems. The fractional operator is defined by the directional derivatives. To discretize the fractional operators, they employ the Grünwald-Letnikov (GL) formula in one-dimensional fractional ADEs and the vector GL formula in conjunction with the directional fractional Laplacian in two- and three-dimensional fractional ADEs. However, it required more mesh points to get the accuracy of $10^{-3} \sim 10^{-4}$. Our work is to evaluate the integral by generalized Gauss-Laguerre quadrature rule. To get the same accuracy, it will require fewer data points and save more computational cost.

1.3 Outline

The outline of this document is as follows:

In chapter 2, we discuss the parameter estimation for the stochastic heat equation. Thanks to the Girsanov theorems and Radon Nykodym derivatives, we have the likelihood functions between two probability measures and derive the analytic form of maximum likelihood estimations on these parameters. We verify these estimators achieve its maximum value of the log-likelihood function by the second-order derivative test. We state and prove the consistency and asymptotic normality for estimators. We develop the estimators on the case of several parameters. The numerical examples are provided.

In chapter 3, we focus on the fractional Navier-Stokes equation with additive noise. We apply the spectral method to solve this PDE numerically. We derive the maximum likelihood estimation for the generalized viscosity. The related numerical examples are also demonstrated.

In chapter 4, we propose a neural network method to solve the fractional PDE. The fractional Laplacian operator is defined by the directional derivatives. We apply the generalized Gauss-Laguerre quadrature rule and derive the numerical schemes.

In chapter 5, we develop explicit schemes preserving the positivity of solutions to SDEs with non-globally Lipschitz drift and Hölder continuous diffusion coefficients. We present five explicit positivity-preserving schemes. These schemes are modified symmetrized Euler schemes. We discuss several choices for non-globally Lipschitz drift among the state-of-the-art tamed schemes and truncation schemes. We present several numerical examples using these schemes and make comparisons in computational performance and convergence.

Chapter 2

Parameter Estimation for Stochastic Heat Equations

In this chapter, we will discuss the following problem

$$u_t - \beta\Delta u + \theta(-\Delta)^\alpha u + \lambda u + f = \sigma \dot{W}^Q, \quad (t,x) \in (0,T] \times \mathcal{D},\, u(0,x) = u_0(x), \tag{2.1}$$

with periodic boundary conditions when $\mathcal{D}$ is a periodic domain or vanishing Dirichlet boundary conditions when $\mathcal{D}$ is a bounded.

In numerous applications, we can measure u (observations data), while we have no knowledge of the fractional order and coefficients in the PDE. Our goal is to estimate the parameters α, β, θ and λ, given the observed data of the solution u. These coefficients have their own physical meanings. In fluid applications, the coefficient of the laplacian term β represents the viscosity of the fluid and α here represents the turbulence mixing [16, 57]; θ is the generalized kinematic viscosity and λ is the friction coefficient.

Here we assume that

$$\dot{W}^Q = \sum_{|\mathbf{k}|=0}^{\infty} \sqrt{q_{\mathbf{k}}} m_{\mathbf{k}}(x) \dot{W}_{\mathbf{k}}(t), \quad \mathbf{k} = (k_1, k_2, ..., k_d) \tag{2.2}$$

where $\{m_{\mathbf{k}}(x)\}_{\mathbf{k}}$ is a complete orthonormal basis in $L^2(\mathcal{D})$ and $W_{\mathbf{k}}$'s are independent standard Brownian motions.

In this chapter, we first prove the regularity of the stochastic heat equations. With the problem settings in (2.1), we derive the maximum likelihood estimation on each parameter. We also prove the consistency and asymptotic normality of parameter θ in (2.1). Moreover, we also introduced different types of estimators for α. The numerical examples are shown in Chapter 2.5. The detailed calculations steps are listed in Chapter 2.6.

2.1 Regularity of the solution

Let's consider the regularity of the solution when $(-\Delta)^{\alpha/2}$ is defined via eigen-pairs of $-\Delta$ on the same domain.

Lemma 2.1.1. *For $\lambda > 0$ and $t > 0$,*

$$\mathbb{E}\left[\left|\int_0^t e^{-\lambda(t-\theta)}\,dW(\theta) - \int_0^s e^{-\lambda(s-\theta)}\,dW(\theta)\right|^2\right] \le \frac{1-e^{-2\lambda(t-s)}}{\lambda}. \tag{2.3}$$

Proof. The inequality (2.3) holds, because by the convexity $e^{-\lambda(t+s)} \le \frac{1}{2}(e^{-2\lambda t} + e^{-2\lambda s})$ and

$$\mathbb{E}\left[\left(\int_0^t e^{-\lambda(t-\theta)}dW(\theta) - \int_0^s e^{-\lambda(s-\theta)}dW(\theta)\right)^2\right] = Var\left[\left(\int_0^t e^{-\lambda(t-\theta)}dW(\theta) - \int_0^s e^{-\lambda(s-\theta)}dW(\theta)\right)^2\right]$$
$$= \mathbb{E}\left(\int_0^t e^{-2\lambda(t-\theta)}d\theta\right) + \mathbb{E}\left(\int_0^s e^{-2\lambda(s-\theta)}d\theta\right) - 2\mathbb{E}\left(\int_0^t e^{-2\lambda(t-\theta)}dW(\theta)\int_0^s e^{-2\lambda(s-\theta)}dW(\theta)\right)$$
$$= \frac{1}{2\lambda}\left(1-e^{-2\lambda t}\right) + \frac{1}{2\lambda}\left(1-e^{-2\lambda s}\right) - \frac{2}{2\lambda}\left(e^{-\lambda|t-s|} - e^{-\lambda(t+s)}\right)$$
$$= \frac{1}{\lambda}(1-e^{-2\lambda(t-s)}) + \frac{1}{2\lambda}(2e^{-\lambda(t+s)} - e^{-2\lambda s} - e^{-2\lambda t}) \le \frac{1}{\lambda}(1-e^{-2\lambda(t-s)}).$$

□

Let $\{e_{\mathbf{k}}(x)\}_{\mathbf{k}}$ be eigenfunctions of the laplacian operator $-\Delta$ with given boundary conditions and the corresponding eigenvalues are $\lambda_{\mathbf{k}}$:

$$-\Delta e_{\mathbf{k}} = \lambda_{\mathbf{k}}^{\Delta} e_{\mathbf{k}} \tag{2.4}$$

and $e_{\mathbf{k}}$'s satisfy the given boundary conditions. For vanishing Dirichlet boundary conditions,

$$\lambda_{\mathbf{k}}^{\Delta} = \left(\frac{\pi}{l}\right)^2 \sum_{i=1}^{d} k_i^2 := |\mathbf{k}|^2\left(\frac{\pi}{l}\right)^2, \quad e_{\mathbf{k}}(x) = \prod_{i=1}^{d}\sqrt{\frac{2}{l}}\sin\left(k_i\frac{\pi}{l}x_i\right), \quad \mathbf{k} = (k_1, k_2, ..., k_d). \tag{2.5}$$

For periodic boundary conditions, $e_{\mathbf{k}}$ is the normalized $\exp(\sqrt{-1}\frac{2\pi}{l}\mathbf{k}^\top x)$. The eigenvalues $\lambda_{\mathbf{k}}^{\Delta} \sim |\mathbf{k}|^2$. The eigenvalues of the operator $(-\Delta) + (-\Delta)^{\frac{\alpha}{2}} + 1$ are proportional to $|\mathbf{k}|^2 + |\mathbf{k}|^\alpha + 1$. As $e_{\mathbf{k}}$'s form an complete orthonormal basis on $L^2(\mathcal{D})$, we take $m_{\mathbf{k}} = e_{\mathbf{k}}$ in (2.2).

Theorem 2.1.2. *Let $f = 0$. Consider the problem (2.1) with the noise defined by (2.2). Let the domain $\mathcal{D}$ be the periodic domain with periodicity l in each direction or the cube $[0, l]^d$. When $\mathcal{D}$ is the cube, we consider vanishing Dirichlet boundary conditions.*

When $\sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}} < \infty$ and $u_0 \in L^2(\mathcal{D})$, for $p \ge 1$.

$$\mathbb{E}\left[\sup_{0\le t\le T}|u(t,x)|^p\right] < \infty. \tag{2.6}$$

When $u_0 \in H^\gamma(\mathcal{D})$ and $\sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}}\lambda_{\mathbf{k}}^{\gamma-1} < \infty$ with $0 < \gamma \le 1$, the solution is Hölder continuous in time with exponent less than $\gamma/2$ and Hölder continuous in space with exponent less than γ.

Proof. We apply the *method of eigenfunction expansion*, i.e., write the solution in the following form

$$u(t,\mathbf{x}) = \sum_{|\mathbf{k}|=0}^{\infty} u_{\mathbf{k}}(t) e_{\mathbf{k}}(x).$$

Plugging this eigen-expansion into (2.1) and multiplying by e_i before integrating over both sides of the equation, we have

$$du_{\mathbf{k}}(t) = -\lambda_{\mathbf{k}} u_{\mathbf{k}}(t)\,dt + \sqrt{q_{\mathbf{k}}}\,dW_{\mathbf{k}}(t), \quad u_{\mathbf{k}}(0) = \iint_{\mathcal{D}} u_0(x) e_{\mathbf{k}}(x)\,dx.$$

Here $\lambda_{\mathbf{k}} = \beta\lambda_{\mathbf{k}}^{\Delta} + \theta(\lambda_{\mathbf{k}}^{\Delta})^{\alpha/2} + \lambda$. This is the Ornstein-Uhlenbeck process and its solution is

$$u_{\mathbf{k}}(t) = u_{0,\mathbf{k}} e^{-\lambda_{\mathbf{k}} t} + \sqrt{q_{\mathbf{k}}} \int_0^t e^{-\lambda_{\mathbf{k}}(t-s)}\,dW_{\mathbf{k}}(s), \quad u_{0,\mathbf{k}} = \iint_{\mathcal{D}} u_0(x) e_{\mathbf{k}}(x)\,dx.$$

Thus the solution is

$$u(t,x) = \sum_{|\mathbf{k}|=0}^{\infty} \left[\iint_{\mathcal{D}} u_0(x) e_{\mathbf{k}}(x)\,dx e^{-\lambda_{\mathbf{k}} t} + \sqrt{q_{\mathbf{k}}} \int_0^t e^{-\lambda_{\mathbf{k}}(t-s)}\,dW_{\mathbf{k}}(s) \right] e_{\mathbf{k}}(x). \tag{2.7}$$

By the Burkholder-Davis-Gundy inequality, we have for $p \geq 1$,

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \left| u(t,x) - \sum_{|\mathbf{k}|=0}^{\infty} u_{0,\mathbf{k}} e_{\mathbf{k}}(x) e^{-\lambda_{\mathbf{k}} t} \right|^p \right] \leq C_p \left| \sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}} e_{\mathbf{k}}^2(x) \frac{1-e^{-2\lambda_{\mathbf{k}} T}}{2\lambda_{\mathbf{k}}} \right|^{p/2} \leq C_p \left| T \sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}} e_{\mathbf{k}}^2(x) \right|^{p/2}.$$

As long as $\sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}}$ converges, $\mathbb{E}\left[\sup_{0\leq t\leq T} |u(t,x)|^p \right] < \infty$. Now, we show the regularity of the solution in t. As $|e^{-\lambda_{\mathbf{k}} t} - e^{-\lambda_{\mathbf{k}} s}| \leq 2^{\beta} |\lambda_{\mathbf{k}}(t-s)|^{\beta}$, for any $\beta \in [0,1]$. Then, by the orthonormality of $e_{\mathbf{k}}$ and Lemma 2.1.1,

$$\begin{aligned}
&\mathbb{E}\left[\|u(t,x) - u(s,x)\|^2 \right] = \sum_{|\mathbf{k}|=0}^{\infty} \mathbb{E}[|u_{\mathbf{k}}(t) - u_{\mathbf{k}}(s)|^2] \\
= \;& \sum_{|\mathbf{k}|=0}^{\infty} \left| u_{0,\mathbf{k}} \big(e^{-\lambda_{\mathbf{k}} t} - e^{-\lambda_{\mathbf{k}} s}\big) \right|^2 + q_{\mathbf{k}} \mathbb{E}[\left| \int_0^t e^{-\lambda_{\mathbf{k}}(t-\theta)}\,dW_{\mathbf{k}}(\theta) - \int_0^s e^{-\lambda_{\mathbf{k}}(s-\theta)}\,dW_{\mathbf{k}}(\theta) \right|^2] \\
\leq \;& \sum_{|\mathbf{k}|=0}^{\infty} [2^{2\beta} \lambda_{\mathbf{k}}^{2\beta} |u_{0,\mathbf{k}}|^2 (t-s)^{2\beta} + q_{\mathbf{k}} \frac{1-e^{-2\lambda_{\mathbf{k}}(t-s)}}{\lambda_{\mathbf{k}}}] \\
\leq \;& 2^{2\beta}(t-s)^{2\beta} \sum_{|\mathbf{k}|=0}^{\infty} |u_{0,\mathbf{k}}|^2 \lambda_{\mathbf{k}}^{2\beta} + 2^{\gamma}(t-s)^{\gamma} \sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}} \lambda_{\mathbf{k}}^{\gamma-1}, \quad \gamma \in [0,1].
\end{aligned}$$

Taking $2\beta = \gamma$, then

$$\mathbb{E}\left[\|u(t,x) - u(s,x)\|^2 \right] \leq 2^{\gamma}(t-s)^{\gamma} \Big(\sum_{|\mathbf{k}|=0}^{\infty} |u_{0,\mathbf{k}}|^2 \lambda_{\mathbf{k}}^{\gamma} + \sum_{|\mathbf{k}|=0}^{\infty} q_{\mathbf{k}} \lambda_{\mathbf{k}}^{\gamma-1} \Big).$$

Thus when the summation is valid (finite), we have

$$\mathbb{E}\left[\|u(t,x)-u(s,x)\|^2\right] \le C(t-s)^\gamma, \quad 0<\gamma\le 1.$$

Then, by Kolmogorov's continuity theorem for Gaussian processes, the solution is Hölder continuous in time with exponent less than $\gamma/2$.

Now, we show that the solution is Hölder continuous in x with exponent less than 1. By the fact $|e_{\mathbf{k}}(x)-e_{\mathbf{k}}(y)| \le C\lambda_{\mathbf{k}}^{\frac{\gamma}{2}}|x-y|^\gamma$ for any $\gamma\in[0,1]$, it can be readily checked that

$$\begin{aligned}
\mathbb{E}\left[|u(t,x)-u(t,y)|^2\right] &= \sum_{|\mathbf{k}|=0}^{\infty} \mathbb{E}[|u_{\mathbf{k}}(t)|^2]\,|e_{\mathbf{k}}(x)-e_{\mathbf{k}}(y)|^2 \\
&\le C|x-y|^{2\gamma}\sum_{|\mathbf{k}|=0}^{\infty} \mathbb{E}[|u_{\mathbf{k}}(t)|^2]\lambda_{\mathbf{k}}^{\gamma} \\
&\le C|x-y|^{2\gamma}\sum_{|\mathbf{k}|=0}^{\infty}\left(|u_{0,\mathbf{k}}|^2 e^{-2\lambda_{\mathbf{k}}t} + q_{\mathbf{k}}\frac{1-e^{-2\lambda_{\mathbf{k}}t}}{2\lambda_{\mathbf{k}}}\right)\lambda_{\mathbf{k}}^{\gamma} \\
&\le C|x-y|^{2\gamma}\sum_{|\mathbf{k}|=0}^{\infty}\left(|u_{0,\mathbf{k}}|^2\lambda_{\mathbf{k}}^{\gamma} + \frac{1}{2}q_{\mathbf{k}}\lambda_{\mathbf{k}}^{\gamma-1}\right)
\end{aligned}$$

Recall the condition $u_0\in H^\gamma(\mathcal{D})$ and $\sum_{|\mathbf{k}|} q_k\lambda_{\mathbf{k}}^{\gamma-1}<\infty$, we then have

$$\mathbb{E}\left[|u(t,x)-u(t,y)|^2\right] \le C|x-y|^{2\gamma}. \tag{2.8}$$

Then, by Kolmogorov's continuity theorem for Gaussian processes, the solution is Hölder continuous in space with exponent less than γ.

□

Remark 2.1.3. *The derivatives of the solution such as $-\Delta u$ should be understood as a distribution instead of a function. Let's suppose that $u_0(x)=0$ and $\theta=\lambda=0$. The solution becomes*

$$u(t,x)=\sum_{|\mathbf{k}|=0}^{\infty}\sqrt{q_{\mathbf{k}}}\int_0^t e^{-\lambda_{\mathbf{k}}(t-s)}\,dW_{\mathbf{k}}(s)e_{\mathbf{k}}(x). \tag{2.9}$$

Then,

$$-\Delta u(t,x)=\sum_{|\mathbf{k}|=0}^{\infty}\lambda_{\mathbf{k}}\sqrt{q_{\mathbf{k}}}\int_0^t e^{-\lambda_{\mathbf{k}}(t-s)}\,dW_{\mathbf{k}}(s)e_{\mathbf{k}}(x). \tag{2.10}$$

For this Gaussian process to have a a bounded second-order moment, we need

$$\mathbb{E}[\|-\Delta u(t)\|^2]=\sum_{|\mathbf{k}|=0}^{\infty}q_{\mathbf{k}}\lambda_{\mathbf{k}}\frac{1-e^{-2\lambda_{\mathbf{k}}t}}{2} \ge \frac{1-e^{-2(\min_{\mathbf{k},|\mathbf{k}|\neq 0}\lambda_{\mathbf{k}})t}}{2}\sum_{|\mathbf{k}|=1}^{\infty}q_{\mathbf{k}}\lambda_{\mathbf{k}}.$$

Thus if $q_{\mathbf{k}}$ is proportional to $\frac{1}{|\mathbf{k}|^p}$, for $p\le d$, $\sum_{|\mathbf{k}|=0}^{\infty}q_{\mathbf{k}}\lambda_{\mathbf{k}}$ diverges. The condition on $\sum_{|\mathbf{k}|=0}^{\infty}q_{\mathbf{k}}<\infty$ will not give us second-order derivatives in a classical sense.

Remark 2.1.4. *With the deterministic function $f \neq 0$, assume further that $f \in L^2(\mathcal{D})$, to guarantee the existence of the solutions.*

2.2 Parameter Estimation

Our goal is to estimate the parameter $\beta, \lambda, \theta, \alpha$ for the fractional advection-diffusion-reaction equation (2.1). Let's consider the periodic boundary condition in 1D with periodicity 2π: $u(t, 0) = u(t, 2\pi)$. Here f is a deterministic function for simplicity.

We apply the eigenfunction expansion $u(x,t) = \sum_{|k|=0}^{\infty} u_k(t)e_k(x)$, and then we have

$$du_k(t) = -[\mu_k(\Theta)u_k(t) + f_k(t)]dt + \sigma\sqrt{q_k}dw_k(t), \tag{2.11}$$

where $v_k \geq 0$ is the eigenvalue of $-\Delta$ and its corresponding eigenfunction is $e_k(x)$. Here we denote

$$\mu_k(\Theta) = \beta v_k + \theta|v_k|^\alpha + \lambda, \text{ where } \Theta = (\beta, \lambda, \theta, \alpha). \tag{2.12}$$

Then we can apply the maximum log-likelihood method as in Section 1.2.3. The log-likelihood function is

$$\ln(L_{N,T}(\Theta)) = -\sum_{k=1}^{N} \left((\mu_k(\Theta) - \mu_k(\Theta_0))(a_{k,T} + c_{k,T}) + \frac{1}{2}(\mu_k^2(\Theta) - \mu_k^2(\Theta_0))b_{k,T} \right), \tag{2.13}$$

where $\Theta_0 = (\beta_0, \lambda_0, \theta_0, \alpha_0)$ represents the true parameters and

$$a_{k,T} = \frac{1}{\sigma^2}\int_0^T q_k^\rho u_k(t)du_k(t), \qquad b_{k,T} = \frac{1}{\sigma^2}\int_0^T q_k^\rho u_k^2(t)dt, \qquad c_{k,T} = \frac{1}{\sigma^2}\int_0^T q_k^\rho u_k(t)f_k(t)dt. \tag{2.14}$$

Here ρ is a free parameter to be tuned for better convergence orders.

As this log-likelihood function is smooth, optimal parameters need to have vanishing first-order partial derivative with respect to each parameter. Here we list these partial derivatives.

$$\frac{\partial \ln(L_{N,T})}{\partial \beta} = -\sum_{k=1}^{N} v_k(a_{k,T} + c_{k,T} + \mu_k(\Theta)b_{k,T}), \tag{2.15}$$

$$\frac{\partial \ln(L_{N,T})}{\partial \lambda} = -\sum_{k=1}^{N}(a_{k,T} + c_{k,T} + \mu_k(\Theta)b_{k,T}), \tag{2.16}$$

$$\frac{\partial \ln(L_{N,T})}{\partial \theta} = -\sum_{k=1}^{N} |v_k|^\alpha(a_{k,T} + c_{k,T} + \mu_k(\Theta)b_{k,T}), \tag{2.17}$$

$$\frac{\partial \ln(L_{N,T})}{\partial \alpha} = -\theta\sum_{k=1}^{N} |v_k|^\alpha \ln(|v_k|)(a_{k,T} + c_{k,T} + \mu_k(\Theta)b_{k,T}). \tag{2.18}$$

Due to the fact that $b_{k,T} > 0$, for all k and T,

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \beta^2} = -\sum_{k=1}^{N} v_k^2 b_{k,T} < 0, \tag{2.19}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \lambda^2} = -\sum_{k=1}^{N} b_{k,T} < 0, \tag{2.20}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \theta^2} = -\sum_{k=1}^{N} v_k^{2\alpha} b_{k,T} < 0, \tag{2.21}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \alpha^2} = -\sum_{k=1}^{N} \left\{\theta |v_k|^{\alpha} (\ln(|v_k|))^2 (a_{k,T} + c_{k,T} + \mu_k(\Theta) b_{k,T}) + \theta^2 |v_k|^{2\alpha} (\ln(|v_k|)^2 b_{k,T}\right\}. \tag{2.22}$$

The mixed derivatives are all negative:

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \beta \partial \lambda} = -\sum_{k=1}^{N} v_k b_{k,T} < 0, \tag{2.23}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \beta \partial \theta} = -\sum_{k=1}^{N} v_k^{\alpha+1} b_{k,T} < 0, \tag{2.24}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \theta \partial \lambda} = -\sum_{k=1}^{N} v_k^{\alpha} b_{k,T} < 0, \tag{2.25}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \alpha \partial \lambda} = -\sum_{k=1}^{N} \theta |v_k|^{\alpha} \ln(|v_k|) b_{k,T} < 0, \tag{2.26}$$

$$\frac{\partial^2 \ln(L_{N,T})}{\partial \alpha \partial \beta} = -\sum_{k=1}^{N} \theta |v_k|^{\alpha+1} \ln(|v_k|) b_{k,T} < 0. \tag{2.27}$$

The above formulas show that each second order derivative on the log-likelihood function is negative. The maximum can be achieved if we estimate these parameters individually with other two parameters known. It remains to show the negative definiteness of the Hessian of the log-likelihood function, assuming that β, θ and λ are all unknown.

2.3 Estimating β, θ, λ with given α

We will need to show that the following Hessian matrix associated with the log-likelihood function is negative definite.

$$\begin{vmatrix} \frac{\partial^2 \ln(L_{N,T})}{\partial \beta^2} & \frac{\partial^2 \ln(L_{N,T})}{\partial \beta \partial \lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial \beta \partial \theta} \\ \frac{\partial^2 \ln(L_{N,T})}{\partial \beta \partial \lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial \lambda^2} & \frac{\partial^2 \ln(L_{N,T})}{\partial \theta \partial \lambda} \\ \frac{\partial^2 \ln(L_{N,T})}{\partial \beta \partial \theta} & \frac{\partial^2 \ln(L_{N,T})}{\partial \theta \partial \lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial \theta^2} \end{vmatrix}. \tag{2.28}$$

We need the following theorem.

Theorem 2.3.1 (Theorem for the totally positive matrix)**.** *If* $0 < a_1 < a_2 < ... < a_n$ *and* $0 \leq \lambda_1 < \lambda_2 < ... < \lambda_n$. *The determinant of the following matrix is positive.*

$$\begin{vmatrix} a_1^{\lambda_1} & a_2^{\lambda_1} & \cdots & \cdots & a_n^{\lambda_1} \\ a_1^{\lambda_2} & a_2^{\lambda_2} & \cdots & \cdots & a_n^{\lambda_2} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ a_1^{\lambda_n} & a_2^{\lambda_n} & \cdots & \cdots & a_n^{\lambda_n} \end{vmatrix} > 0. \tag{2.29}$$

The detailed proof of theorem 2.3.1 is in Chapter 2.6.

Theorem 2.3.2. *Let* $L_{N,T}$ *be the log-likelihood function defined in* (2.13). *Then its Hessian matrix* (2.28) *is negative definite.*

Proof. We want to show that the determinant of Hessian matrix (2.28) is negative definite. Clearly, the Hessian matrix is symmetric due to the continuity of the log-likelihood function. A symmetric matrix is negative definite if and only if all of its principal minors of even order are positive and all of its principal minors of odd order are negative. Then, need to check the sign of leading principal minors. It is straightforward that the first leading principal minor is negative: $\frac{\partial^2 \ln(L_{N,T})}{\partial\beta^2} = -\sum_{k=1}^{N} v_k^2 b_{k,T} < 0$. The second leading principal minor is

$$\begin{aligned}
&\begin{vmatrix} \frac{\partial^2 \ln(L_{N,T})}{\partial\beta^2} & \frac{\partial^2 \ln(L_{N,T})}{\partial\beta\partial\lambda} \\ \frac{\partial^2 \ln(L_{N,T})}{\partial\beta\partial\lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial\lambda^2} \end{vmatrix} \\
= &\left(\sum_{k=1}^{N} v_k^2 b_{k,T}\right)\left(\sum_{k=1}^{N} b_{k,T}\right) - \left(\sum_{k=1}^{N} v_k b_{k,T}\right)^2 \\
= &\sum_{k,l}^{N} \left(v_k^2 b_{k,T} b_{l,T} - v_k v_l b_{k,T} b_{l,T}\right) = \sum_{k\neq l}^{N} \left(v_k^2 b_{k,T} b_{l,T} - v_k v_l b_{k,T} b_{l,T}\right) \\
= &\sum_{k>l}^{N} \left(v_k^2 b_{k,T} b_{l,T} - v_k v_l b_{k,T} b_{l,T}\right) + \sum_{k<l}^{N} \left(v_k^2 b_{k,T} b_{l,T} - v_k v_l b_{k,T} b_{l,T}\right) \\
= &\sum_{k>l}^{N} \left(v_k^2 b_{k,T} b_{l,T} - v_k v_l b_{k,T} b_{l,T}\right) + \sum_{k>l}^{N} \left(v_l^2 b_{k,T} b_{l,T} - v_k v_l b_{k,T} b_{l,T}\right) \\
= &\sum_{k>l}^{N} \left(v_k - v_l\right)^2 b_{k,T} b_{l,T} > 0.
\end{aligned} \tag{2.30}$$

The third leading principal minor is the determinant of the Hessian matrix.

$$\begin{vmatrix} \frac{\partial^2 \ln(L_{N,T})}{\partial\beta^2} & \frac{\partial^2 \ln(L_{N,T})}{\partial\beta\partial\lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial\beta\partial\theta} \\ \frac{\partial^2 \ln(L_{N,T})}{\partial\beta\partial\lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial\lambda^2} & \frac{\partial^2 \ln(L_{N,T})}{\partial\theta\partial\lambda} \\ \frac{\partial^2 \ln(L_{N,T})}{\partial\beta\partial\theta} & \frac{\partial^2 \ln(L_{N,T})}{\partial\theta\partial\lambda} & \frac{\partial^2 \ln(L_{N,T})}{\partial\theta^2} \end{vmatrix}$$

$$
\begin{aligned}
&= -\begin{vmatrix} \sum_{k=1}^{N} v_k^2 b_{k,T} & \sum_{k=1}^{N} v_k b_{k,T} & \sum_{k=1}^{N} v_k^{\alpha+1} b_{k,T} \\ \sum_{k=1}^{N} v_k b_{k,T} & \sum_{k=1}^{N} b_{k,T} & \sum_{k=1}^{N} v_k^{\alpha} b_{k,T} \\ \sum_{k=1}^{N} v_k^{\alpha+1} b_{k,T} & \sum_{k=1}^{N} v_k^{\alpha} b_{k,T} & \sum_{k=1}^{N} v_k^{2\alpha} b_{k,T} \end{vmatrix} \\
&= -\sum_{k,l,m}^{N} (v_k^2 v_m^{2\alpha} + v_k v_l^{\alpha} v_m^{\alpha+1} + v_l v_m^{\alpha} v_k^{\alpha+1} - v_k^{\alpha+1} v_m^{\alpha+1} - v_k v_l v_m^{2\alpha} - v_k^2 v_l^{\alpha} v_m^{\alpha}) b_{k,T} b_{l,T} b_{m,T} \\
&= -\sum_{k,l,m}^{N} \begin{vmatrix} v_k^2 & v_k & v_k^{\alpha+1} \\ v_l & 1 & v_l^{\alpha} \\ v_m^{\alpha+1} & v_m^{\alpha} & v_m^{2\alpha} \end{vmatrix} b_{k,T} b_{l,T} b_{m,T} = -\sum_{k,l,m}^{N} v_k v_m^{\alpha} \begin{vmatrix} v_k & 1 & v_k^{\alpha} \\ v_l & 1 & v_l^{\alpha} \\ v_m & 1 & v_m^{\alpha} \end{vmatrix} b_{k,T} b_{l,T} b_{m,T} \\
&= -\sum_{k,l,m}^{N} v_k v_m^{\alpha} \begin{vmatrix} 1 & v_k^{\alpha} & v_k \\ 1 & v_l^{\alpha} & v_l \\ 1 & v_m^{\alpha} & v_m \end{vmatrix} b_{k,T} b_{l,T} b_{m,T} < 0. \qquad (2.31)
\end{aligned}
$$

Since the matrix in the last step of (2.31) is a generalized Vandermonde matrix for $0 < \alpha < 1$, its determinant is positive. By Theorem 2.3.1 on total positive matrices, the third leading principal minor is negative. Thus, we have proved that the Hessian matrix is negative definite. □

As the Hessian matrix is negative definite, the log-likelihood function can achieve its global maximum.

2.3.1 Algorithm for estimation of β, θ, λ

Based on the negative definiteness of the Hessian matrix verified in the previous subsection, there exists a global maximum for the log-likelihood function. Given the information on α, we now work on the estimation of β, θ and λ. Construct the linear system on these three variables. We set the first order partial derivatives (2.15), (2.18) and (2.16) to be zero. The solution to this linear system is the critical point of the log-likelihood function.

$$
\begin{cases}
\left(\sum_{k=1}^{N} v_k^2 b_{k,T}\right)\beta + \left(\sum_{k=1}^{N} v_k b_{k,T}\right)\lambda + \left(\sum_{k=1}^{N} v_k^{\alpha+1} b_{k,T}\right)\theta = \sum_{k=1}^{N} -v_k(a_{k,T} + c_{k,T}), \\
\left(\sum_{k=1}^{N} v_k b_{k,T}\right)\beta + \left(\sum_{k=1}^{N} b_{k,T}\right)\lambda + \left(\sum_{k=1}^{N} v_k^{\alpha} b_{k,T}\right)\theta = \sum_{k=1}^{N} -(a_{k,T} + c_{k,T}), \\
\left(\sum_{k=1}^{N} v_k^{\alpha+1} b_{k,T}\right)\beta + \left(\sum_{k=1}^{N} v_k^{\alpha} b_{k,T}\right)\lambda + \left(\sum_{k=1}^{N} v_k^{2\alpha} b_{k,T}\right)\theta = \sum_{k=1}^{N} -v_k^{\alpha}(a_{k,T} + c_{k,T}).
\end{cases} \qquad (2.32)
$$

Let's denote the determinant of the coefficients matrix in (2.32) as $D_{N,T}$, which is computed in (2.31).

$$
D_{N,T} = \sum_{k,l,m}^{N} v_k v_m^{\alpha} \begin{vmatrix} 1 & v_k^{\alpha} & v_k \\ 1 & v_l^{\alpha} & v_l \\ 1 & v_m^{\alpha} & v_m \end{vmatrix} b_{k,T} b_{l,T} b_{m,T}.
$$

Solving this system (2.32) by the Cramer rule, we obtain that

$$\hat{\beta}_{N,T} = \frac{D^{\beta}_{N,T}}{D_{N,T}}, \quad \hat{\lambda}_{N,T} = \frac{D^{\lambda}_{N,T}}{D_{N,T}}, \quad \hat{\theta}_{N,T} = \frac{D^{\theta}_{N,T}}{D_{N,T}}, \tag{2.33}$$

where

$$\begin{aligned}
D^{\beta}_{N,T} &= -\begin{vmatrix} \sum\limits_{k=1}^{N} v_k(a_{k,T}+c_{k,T}) & \sum\limits_{k=1}^{N} v_k b_{k,T} & \sum\limits_{k=1}^{N} v_k^{\alpha+1} b_{k,T} \\ \sum\limits_{k=1}^{N} (a_{k,T}+c_{k,T}) & \sum\limits_{k=1}^{N} b_{k,T} & \sum\limits_{k=1}^{N} v_k^{\alpha} b_{k,T} \\ \sum\limits_{k=1}^{N} v_k^{\alpha}(a_{k,T}+c_{k,T}) & \sum\limits_{k=1}^{N} v_k^{\alpha} b_{k,T} & \sum\limits_{k=1}^{N} v_k^{2\alpha} b_{k,T} \end{vmatrix}, \\
D^{\lambda}_{N,T} &= -\begin{vmatrix} \sum\limits_{k=1}^{N} v_k^2 b_{k,T} & \sum\limits_{k=1}^{N} v_k(a_{k,T}+c_{k,T}) & \sum\limits_{k=1}^{N} v_k^{\alpha+1} b_{k,T} \\ \sum\limits_{k=1}^{N} v_k b_{k,T} & \sum\limits_{k=1}^{N} (a_{k,T}+c_{k,T}) & \sum\limits_{k=1}^{N} v_k^{\alpha} b_{k,T} \\ \sum\limits_{k=1}^{N} v_k^{\alpha+1} b_{k,T} & \sum\limits_{k=1}^{N} v_k^{\alpha}(a_{k,T}+c_{k,T}) & \sum\limits_{k=1}^{N} v_k^{2\alpha} b_{k,T} \end{vmatrix}, \\
D^{\theta}_{N,T} &= -\begin{vmatrix} \sum\limits_{k=1}^{N} v_k^2 b_{k,T} & \sum\limits_{k=1}^{N} v_k b_{k,T} & \sum\limits_{k=1}^{N} v_k(a_{k,T}+c_{k,T}) \\ \sum\limits_{k=1}^{N} v_k b_{k,T} & \sum\limits_{k=1}^{N} b_{k,T} & \sum\limits_{k=1}^{N} (a_{k,T}+c_{k,T}) \\ \sum\limits_{k=1}^{N} v_k^{\alpha+1} b_{k,T} & \sum\limits_{k=1}^{N} v_k^{\alpha} b_{k,T} & \sum\limits_{k=1}^{N} v_k^{\alpha}(a_{k,T}+c_{k,T}) \end{vmatrix}.
\end{aligned}$$

2.3.2 Consistency and Asymptotic Normality of the estimator θ

In this subsection, we present the consistency of the estimator $\hat{\theta}_N(T)$, and asymptotic normality of the error. The consistency and asymptotic normality of other parameter λ can be achieved with the same procedures as θ.

To derive the closed form of the error term, we first solve this SDE analytically by applying Itô's formula on this Ornstein–Uhlenbeck process (2.11) and obtain

$$u_k(t) = u_k(0)e^{-\mu_k(\Theta)t} - \int_0^t e^{-\mu_k(\Theta)(t-s)} f_k(s)ds + \sigma\sqrt{q_k}\int_0^t e^{-\mu_k(\Theta)(t-s)} dW_k(s). \tag{2.34}$$

We now find a closed form of the optimal value $\hat{\theta}_{N,T}$, given the observed data. Let (2.17) be zero. Then it gives

$$\hat{\theta}_N(T) = -\frac{\sum\limits_{k=1}^{N} \left[|v_k|^{\alpha} a_{k,T} + (\beta|v_k|^{\alpha+1} + \lambda|v_k|^{\alpha}) b_{k,T} + |v_k|^{\alpha} c_{k,T}\right]}{\sum\limits_{k=1}^{N} q_k^{\rho} |v_k|^{2\alpha} b_{k,T}} \tag{2.35}$$

According to (2.33), derive the form of θ.

$$\theta = -\frac{\sum_{k=1}^{N}\left[|v_k|^\alpha a_{k,T} + (\beta|v_k|^{\alpha+1} + \lambda|v_k|^\alpha)b_{k,T} + |v_k|^\alpha c_{k,T}\right]}{\sum_{k=1}^{N} q_k^\rho |v_k|^{2\alpha} b_{k,T}} + \frac{\sum_{k=1}^{N}\sigma\int_0^T q_k^{\rho+\frac{1}{2}}|v_k|^\alpha u_k(t)dW_k(t)}{\sum_{k=1}^{N}\int_0^T q_k^\rho |v_k|^{2\alpha}u_k^2(t)dt}, \quad (2.36)$$

We then obtain that

$$\hat{\theta}_N(T) - \theta = -\frac{\sigma\sum_{k=1}^{N}\int_0^T q_k^{\rho+\frac{1}{2}} v_k^\alpha u_k(t)dW_k(t)}{\sum_{k=1}^{N}\int_0^T q_k^\rho v_k^{2\alpha} u_k^2(t)dt}. \tag{2.37}$$

The error term above will be used to show the consistency and asymptotic normality. The following part is to introduce the notations and basic settings for the main theorem on the consistency and asymptotic normality.

We follow the idea of Chapter 6.3 of [35] to show the desired conclusion.

Consider the diagonal parabolic equation of order $2m$

$$\dot{u} + (A_0 + \Theta A_1)u = \dot{W}^Q. \tag{2.38}$$

Denote $\omega_\alpha = \dfrac{2(\alpha - m)}{d}$. Here α is the order of the operator A_1, $2m$ is the order of the operator $A_0 + \Theta A_1$.

Assumption 2.3.3 (Eigenvalues of operators)**.** *Assume that the following limits exist:*

$$\bar{v} = \lim_{k\to\infty}\frac{v_k}{k^{\alpha/d}} \neq 0, \tag{2.39}$$

$$\bar{\mu}(\Theta) = \lim_{k\to\infty}\frac{\mu_k(\Theta)}{k^{2m/d}} > 0, \tag{2.40}$$

Moreover, the eigenvalues of the operators have a particular growth rate in k:

$$k^{th} \text{ eigenvalue } \sim k^{\text{ order of the operator}/d}. \tag{2.41}$$

Assumption 2.3.4 (Initial condition)**.** *Assume that the modes of the initial condition $u_0(x)$ satisfy that $u_k(0) \sim \mathcal{N}(m_k, \sigma_k^2)$ for $k \geq 1$ and for some $\rho \geq -1$,*

$$\lim_{k\to\infty}(m_k^2 + \sigma_k^2)q_k^\rho = 0, \tag{2.42}$$

$$\lim_{k\to\infty}(m_k^2 + \sigma_k^2)q_k^{2\rho+1} = 0, \tag{2.43}$$

$$\lim_{k\to\infty} q_k^{2\rho}\mu_k(\Theta)(\sigma_k^2 + 2m_k^2) = 0. \tag{2.44}$$

Here are some examples when the assumption is satisfied.

- When $q_k \equiv 1$, $u_k(0)$ is a constant when k is large enough.

- When $q_k = k^{-1}$, $\rho = -1$, $u_k(0)$ is a normal distributed random variable with mean 0 and variance $\frac{1}{k^4}$, for $k \geq 1$.

Define the number $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\Theta) > 0$ by

$$\boldsymbol{\sigma}^2(\Theta) = \begin{cases} \dfrac{2(\omega_\alpha + 1)\bar{\mu}(\Theta)}{\sigma^2 \bar{v}^{2\alpha}}, & \text{if } \omega_\alpha > -1, \\ \dfrac{2\bar{\mu}(\Theta)}{\sigma^2 \bar{v}^{2\alpha}}, & \text{if } \omega_\alpha = -1. \end{cases} \tag{2.45}$$

Theorem 2.3.5 (c.f.[10]). *For equation 2.38, assume that Assumption* (2.3.3) *and* (2.3.4) *hold. Define* $I(n) := \sum_{k=1}^{n} q_k^{\rho+1} k^{\omega_\alpha}$ *and*

$$\sum_{n=1}^{\infty} \frac{q_n^{2\rho+2} n^{\omega_\alpha}}{I^2(n)} < \infty. \tag{2.46}$$

Assumption 3:

$$\sum_{n=1}^{\infty} \frac{q_n^{2\rho+2} n^{2\omega_\alpha}}{I^2(n)} < \infty. \tag{2.47}$$

If $\omega_\alpha \geq -1$, *then the estimator* $\hat{\theta}_N$ *of* θ *is strongly consistent and asymptotically normal as* $N \to \infty$. *More precisely,*

$$\lim_{N\to\infty} \hat{\theta}_N(T) = \theta, \qquad \text{in probability, for every } T > 0, \tag{2.48}$$

and

$$\lim_{N\to\infty} \sqrt{I(N)}(\hat{\theta}_N - \theta) = \mathcal{N}(0, \boldsymbol{\sigma}^2(\theta)/T). \tag{2.49}$$

2.3.3 Proof of the main theorem 2.3.5

In this subsection, we state and present the proof of consistency and asymptotic normality of the parameters. In this proof, we will apply the strong law of large numbers and Martingale Representation theorem. The main idea of the proof follows the step in [35].

Theorem 2.3.6 (The Strong Law of Large Numbers). *Let* ζ_n, $n \geq 1$, *be a sequence of independent random variables and* b_n, $n \geq 1$, *a sequence of positive numbers such that* $b_{n+1} \geq b_n$ *and* $\lim_{n\to\infty} b_n = +\infty$, *and*

$$\sum_{n\geq 1} \frac{\mathrm{Var}(\zeta_n)}{b_n^2} < \infty. \tag{2.50}$$

Then,

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} (\zeta_n - \mathbb{E}[\zeta_n])}{b_n} = 0 \tag{2.51}$$

with probability one.

Theorem 2.3.7 (Martingale Central Limit theorem). *For $t \geq 0$ and $\epsilon > 0$, if $X_\epsilon = X_\epsilon(t)$ and $X = X(t)$ are real-valued, continuous square-integrable martingales such that X is a Gaussian process, $X_\epsilon(0) = X(0) = 0$, and, for some $t_0 > 0$,*

$$\lim_{\epsilon \to 0} \langle X_\epsilon \rangle (t_0) = \langle X \rangle (t_0)$$

in probability, then $\lim_{\epsilon \to 0} X_\epsilon(t_0) = X(t_0)$ in distribution.

Below is the proof of the main theorem. The detailed calculation steps are listed in Chapter 2.6.

Proof. From (2.37), let's denote

$$\zeta_k = \int_0^T q_k^{\rho+\frac{1}{2}} v_k^{\alpha} u_k(t) dW_k(t), \tag{2.52}$$

$$\eta_k = \int_0^T q_k^{\rho} v_k^{2\alpha} u_k^2(t) dt, \tag{2.53}$$

and

$$b_n = \sum_{k=1}^{n} \mathbb{E}[\eta_k] = \sum_{k=1}^{n} q_k^{\rho} v_k^{2\alpha} \int_0^T \mathbb{E}[u_k^2(t)] dt. \tag{2.54}$$

With the notation (2.52) to (2.54), the error term (2.37) becomes

$$\hat{\theta}_N - \theta = -\sigma \frac{\sum_{k=1}^{N} \zeta_k}{b_N} \times \frac{\sum_{k=1}^{N} \mathbb{E}[\eta_k]}{\sum_{k=1}^{N} \eta_k} =: -\sigma I_1 I_2. \tag{2.55}$$

Consistency follows from Theorem 2.3.6. The only non-trivial conditions to check are the convergence of the following series

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(\zeta_n)}{b_n^2} < \infty, \qquad \sum_{n=1}^{\infty} \frac{\mathrm{Var}(\eta_n)}{b_n^2} < \infty, \tag{2.56}$$

which will imply, respectively, that $I_1 \to 0$ a.s. and $I_2 \to 1$ a.s., as $N \to \infty$, and by (2.55) the consistency of $\hat{\theta}_N$ follows. To verify (2.56), it can be done by establishing precise asymptotic behavior of $Var(\zeta_n)$, $Var(\eta_n)$ and b_n in Step 1 and 2. The asymptotic normality is shown in Step 3.

Step 1: Start with evaluating b_n, the denominator in (2.56). For simplicity, we assume that $u_k(0) = 0$ is constant, the expectation m_k and the variance σ_k^2 are 0. For fixed $T > 0$, using the explicit form of $\int_0^T \mathbb{E}[u_k^2(t)]dt$ computed in Chapter 2.6,

$$\begin{aligned}
b_n &= \sum_{k=1}^n v_k^{2\alpha} q_k^\rho \int_0^T \mathbb{E}[u_k^2(t)]dt \\
&\approx \sum_{k=1}^n v_k^{2\alpha} q_k^\rho \left[\frac{m_k^2+\sigma_k^2}{2\mu_k(\Theta)} + \frac{\sigma^2 q_k T}{2\mu_k(\Theta)}\right] \\
&= \sum_{k=1}^n \frac{v_k^{2\alpha} T}{2\mu_k(\Theta)} \left[\frac{(m_k^2+\sigma_k^2)q_k^\rho}{T} + \sigma^2 q_k^{\rho+1}\right] \\
&= \sum_{k=1}^n \frac{\sigma^2 v_k^{2\alpha} q_k^{\rho+1} T}{2\mu_k(\Theta)} \\
&\approx \frac{\sigma^2 T \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} \sum_{k=1}^n \frac{q_k^{\rho+1} k^{2\alpha/d}}{k^{2m/d}} \\
&= \frac{\sigma^2 T \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} \sum_{k=1}^n q_k^{\rho+1} k^{\frac{2(\alpha-m)}{d}} := \frac{\sigma^2 T \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} I(n).
\end{aligned} \tag{2.57}$$

Recall the definition of $\zeta_k = \int_0^T q_k^{\rho+\frac{1}{2}} v_k^\alpha u_k(t) dW_k(t)$ and $b_n = \sum_{k=1}^n q_k^\rho v_k^{2\alpha} \int_0^T \mathbb{E}[u_k^2(t)]dt \sim I(n)$ in (2.52) and (2.54).

$$\begin{aligned}
\mathrm{Var}(\zeta_n) &= q_n^{2\rho+1} v_n^{2\alpha} \int_0^T \mathbb{E}[u_n^2(t)]dt \approx q_n^{2\rho+1} v_n^{2\alpha}\left[\frac{m_n^2+\sigma_n^2}{2\mu_n(\Theta)} + \frac{\sigma^2 q_n T}{2\mu_n(\Theta)}\right] \\
&\approx \frac{\sigma^2 q_n^{2\rho+2} v_n^{2\alpha} T}{2\mu_n(\Theta)} \approx \frac{\sigma^2 T q_n^{2\rho+2} \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} n^{\frac{2(\alpha-m)}{d}} \approx q_n^{2\rho+2} n^{\omega_\alpha}.
\end{aligned} \tag{2.58}$$

It follows that

$$\frac{\mathrm{Var}\zeta_n}{b_n^2} \sim \frac{\mathrm{Var}\zeta_n}{I^2(n)} \sim \frac{q_n^{2\rho+2} n^{\omega_\alpha}}{I^2(n)}. \tag{2.59}$$

With the assumption that the series $\sum_{n=1}^\infty \frac{q_n^{2\rho+2} n^{\omega_\alpha}}{I^2(n)}$ converges for all $\omega_\alpha \geq -1$. We have $I_1 \to 0$ a.s. as $n \to \infty$, that implies

$$\lim_{N\to\infty} \frac{\sum_{k=1}^N \sigma \int_0^T q_k^{\rho+\frac{1}{2}} v_k^\alpha u_k(t) dW_k(t)}{\sum_{k=1}^N \int_0^T q_k^\rho v_k^{2\alpha} \mathbb{E}[u_k^2(t)]dt} = 0 \text{ with probability one.} \tag{2.60}$$

Step 2: In this step, we want to prove the series $\sum_{n=1}^\infty \frac{\mathrm{Var}(\eta_n)}{b_n^2}$ converges. Recall the numerator $\eta_k = q_k^\rho v_k^{2\alpha} \int_0^T u_k^2(t)dt$ and $b_n = \sum_{k=1}^n \mathbb{E}[\eta_k]$. We know that $b_n \sim I(n)$, from Step 1. It remains to show that

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}(\eta_n)}{I^2(n)} < \infty. \tag{2.61}$$

To get bound on $\mathrm{Var}(\eta_n)$, note that

$$\begin{aligned}\mathrm{Var}\left(\int_0^T u_k^2(t)dt\right) &= \mathbb{E}\left[\left(\int_0^T u_k^2(t)dt\right)^2\right] - \left(\mathbb{E}\left[\int_0^T u_k^2(t)dt\right]\right)^2 \\ &= \mathbb{E}\left[\left(\int_0^T (u_k^2(t) - \mathbb{E}[u_k^2(t)])dt\right)^2\right].\end{aligned} \tag{2.62}$$

By Cauchy Schwartz inequality, we have

$$\mathrm{Var}\left(\int_0^T u_k^2(t)dt\right) \le T\int_0^T \mathbb{E}[(u_k^2(t) - \mathbb{E}[u_k^2(t)])^2]dt = T\int_0^T \mathrm{Var}(u_k^2(t))dt. \tag{2.63}$$

Here, $\int_0^T \mathrm{Var}(u_k^2(t))dt$ is computed in Chapter 2.6. Rearrange the terms to find

$$\begin{aligned}\mathrm{Var}(\eta_k) &= v_k^{4\alpha} q_k^{2\rho}\int_0^T \mathrm{Var}(u_k^2(t))dt \\ &\approx v_k^{4\alpha} q_k^{2\rho}\left(\frac{2\sigma^2(\sigma_k^2 + 2m_k^2)}{4\mu_k(\Theta)} + \frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}\left[T - \frac{3}{4\mu_k(\Theta)}\right] + \frac{(m_k^2+\sigma_k^2)\sigma^2 q_k}{2\mu_k^2(\Theta)}\right) \\ &= \frac{\sigma^2 v_k^{4\alpha}}{2\mu_k^2(\Theta)}\left(q_k^{2\rho}\mu_k(\Theta)(\sigma_k^2 + 2m_k^2) + \sigma^2 q_k^{2\rho+2}\left[T - \frac{3}{4\mu_k(\Theta)}\right] + (m_k^2+\sigma_k^2)q_k^{2\rho+1}\right).\end{aligned} \tag{2.64}$$

As $k \to \infty$, with Assumptions 2.3.3 and 2.3.4,

$$v_k^{4l} q_k^{2\rho}\int_0^T Var(u_k^2(t))dt \sim \frac{\sigma^4 q_k^{2\rho+2} v_k^{4\alpha}}{2\mu_k^2(\Theta)}\left[T - \frac{3}{4\mu_k(\Theta)}\right] \sim \frac{\sigma^4 q_k^{2\rho+2}\bar{v}^{4\alpha}T}{2\bar{\mu}^2}k^{2\omega_\alpha}, \tag{2.65}$$

As a result,

$$\sum_{n=1}^{\infty}\frac{\mathrm{Var}(\eta_n)}{I^2(n)} \approx \sum_{n=1}^{\infty}\frac{q_n^{2\rho+2}n^{2\omega_\alpha}}{I^2(n)}, \tag{2.66}$$

With the assumption that $\sum_{n=1}^{\infty}\frac{q_n^{2\rho+2}n^{2\omega_\alpha}}{I^2(n)} < \infty$ and by the strong law of large numbers, we obtain that

$$\lim_{N\to\infty}\frac{\sum_{k=1}^{N}\int_0^T q_k^{2\rho+1}v_k^{2\alpha}u_k^2(t)dt}{\sum_{k=1}^{N}\int_0^T q_k^{\rho}v_k^{2\alpha}\mathbb{E}[u_k^2(t)]dt} = 1 \text{ with probability one.} \tag{2.67}$$

This completes the proof of consistency.

Step 3: Finally, we focus on the proof of the asymptotic normality using Martingale Central Limit

Theorem (see Theorem 2.3.7). Thinking of $\frac{1}{N}$ as ϵ to be consistent with the notations of theorem 2.3.7, define

$$X_N(t) = \frac{1}{\sqrt{I(N)}} \sum_{k=1}^{N} q_k^{\rho+\frac{1}{2}} v_k^{\alpha} \int_0^T u_k(t) dW_k(t). \tag{2.68}$$

Then X_N is a continuous square-integrable martingale, and the quadratic variation of X_N

$$\langle X_N \rangle (t) = \frac{1}{I(N)} \sum_{k=1}^{N} q_k^{2\rho+1} v_k^{2\alpha} \int_0^T u_k^2(t) dt. \tag{2.69}$$

From the error term between θ_N and θ (2.37), we have

$$\sqrt{I(N)}(\hat{\theta}_N - \theta) = -\frac{\sigma I(N)}{\sum_{k=1}^{N} q_k^{\rho} v_k^{2\alpha} \int_0^T u_k^2(t) dt} X_N(t). \tag{2.70}$$

By (2.57),

$$\lim_{N\to\infty} \langle X_N \rangle (t) = \frac{t}{\boldsymbol{\sigma}^2(\Theta)}, \tag{2.71}$$

where $\boldsymbol{\sigma}^2(\Theta)$ is defined in (2.45). Since $\frac{t}{\sigma^2(\Theta)} = \langle X \rangle (t)$, where $X(t) = \frac{W(t)}{\boldsymbol{\sigma}(\Theta)}$ and $W(t)$ is a standard Brownian motion, the result follows:

$$\lim_{N\to\infty} X_N(T) = \lim_{N\to\infty} \frac{\sum_{k=1}^{N} q_k^{\rho+\frac{1}{2}} v_k^{\alpha} \int_0^T u_k(t) dW_k(t)}{\sqrt{I(N)}} = \mathcal{N}\left(0, \frac{T}{\boldsymbol{\sigma}^2(\Theta)}\right) \text{ in distribution} \tag{2.72}$$

and

$$\lim_{N\to\infty} \frac{I(N)}{\sum_{k=1}^{N} q_k^{\rho} v_k^{2\alpha} \int_0^T u_k^2(t) dt} = \frac{\boldsymbol{\sigma}^2(\Theta)}{T} \text{ with probability one.} \tag{2.73}$$

This completes the proof of theorem 2.3.5. □

2.3.4 Special case when $q_k = 1$ for all k

When $q_k = 1$ for all k, a $Q-$cylindrical Brownian motion becomes a cylindrical Brownian motion. Here Q is an identity operator. b_n in (2.57) becomes

$$\frac{\sigma^2 T \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} \sum_{k=1}^{n} q_k^{\rho+1} k^{\frac{2(\alpha-m)}{d}} = \frac{\sigma^2 T \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} \sum_{k=1}^{n} k^{\frac{2(\alpha-m)}{d}} = \frac{\sigma^2 T \bar{v}^{2\alpha}}{2\bar{\mu}(\Theta)} \sum_{k=1}^{n} k^{\omega_\alpha} \approx \frac{\tilde{I}(n) T}{\boldsymbol{\sigma}^2(\Theta)}, n \to \infty. \tag{2.74}$$

where $\tilde{I}(n)$ is defined as

$$\tilde{I}(n) = \begin{cases} n^{\omega_\alpha+1}, \text{ if } \omega_\alpha > -1, \\ \ln(n), \text{ if } \omega_\alpha = -1. \end{cases} \tag{2.75}$$

Then, in Assumption 2.3.3,

$$\frac{q_n^{2\rho+2}n^{\omega_\alpha}}{I^2(n)} = \frac{n^{\omega_\alpha}}{I^2(n)} \sim \begin{cases} n^{-2-\omega_\alpha}, \text{ if } \omega_\alpha > -1, \\ \dfrac{1}{n(\ln(n))^2}, \text{ if } \omega_\alpha = -1. \end{cases}$$

As a result, by the integral test, the series $\sum_{n=1}^{\infty} \frac{n^{\omega_\alpha}}{I^2(n)} < \infty$, for all $\omega_\alpha \geq -1$. Similarly, for Assumption 3, $\sum_{n=1}^{\infty} \frac{q_n^{2\rho+2}n^{\omega_\alpha}}{I^2(n)} = \sum_{n=1}^{\infty} \frac{n^{2\omega_\alpha}}{I^2(n)} < \infty$. Thus, in the case of $q_k = 1$ for all k, Assumption 2.3.3 and 3 in Theorem 2.3.5 are automatically satisfied.

2.3.5 Several Parameters

In this subsection, we work on the multi-parameter cases given the value of α. Applying the eigenfunction expansion, we have (2.11). Denote the vector

$$\Theta = [\beta, \theta, \lambda]^T, \text{ and } \mathbf{v} = [v_k, |v_k|^\alpha, 1]^T. \tag{2.76}$$

The OU process (2.12) can be written into a vector form.

$$\mu_k(\Theta) = \beta v_k + \theta |v_k|^\alpha + \lambda = \Theta^T \mathbf{v}. \tag{2.77}$$

Thanks to the Girsanov Theorem and Radon Nykodym derivatives, we derive the closed form of parameters (2.33) from solving the linear system (2.32).

Theorem 2.3.8 (Martingale Central Limit Theorem)**.** *[35] Let $M = (M_1(t), M_2(t), ..., M_d(t))$, $0 \leq t \leq T$, be a d-dimensional continuous Gaussian martingale with $M(0) = 0$, and let $M_\epsilon = (M_{\epsilon,1}(t), ..., M_{\epsilon,d}(t))$, $\epsilon \geq 0$, $0 \leq t \leq T$, be a family of continuous square-integrable d-dimensional martingales such that $M_\epsilon(0) = 0$ for all ϵ and, for some $t_0 \in [0, 1]$ and all $i, j = 1, ..., d$,*

$$\lim_{\epsilon\to\infty} \langle M_{\epsilon,i}, M_{\epsilon,j} \rangle (t) = \langle M_i, M_j \rangle (t) \tag{2.78}$$

in probability. Then $\lim_{\epsilon\to\infty} M_\epsilon(t_0) \stackrel{\mathcal{D}}{=} M(t_0)$.

Theorem 2.3.9. *For every $\beta, \theta, \lambda \in \mathbb{R}$, in $d = 2$,*

$$\lim_{N\to\infty} \begin{bmatrix} \hat{\beta_N} \\ \hat{\theta_N} \\ \hat{\lambda_N} \end{bmatrix} = \begin{bmatrix} \beta \\ \theta \\ \lambda \end{bmatrix}. \tag{2.79}$$

with probability one and

$$\lim_{N\to\infty} \begin{bmatrix} N(\hat{\beta_N} - \beta) \\ N^\alpha(\hat{\theta_N} - \theta) \\ \sqrt{\ln(N)}(\hat{\lambda_N} - \lambda) \end{bmatrix} \stackrel{\mathcal{D}}{=} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix}, \tag{2.80}$$

with the zero-mean Gaussian random variables ζ_1, ζ_2 and ζ_3 independent.

Proof. The idea of proof is similar with Theorem 2.3.5. Denote the following terms used in the estimators (2.33).

$$A_{p,N} = \frac{1}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^\rho v_k^p u_k(t)du_k(t), \quad B_{p,N} = \frac{1}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^\rho v_k^p u_k^2(t)dt, \quad C_{p,N} = \frac{1}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^\rho v_k^p u_k(t)f_k(t)dt. \tag{2.81}$$

Denote

$$\overline{A_{p,N}} = \frac{1}{\sigma}\sum_{k=1}^{N}\int_0^T q_k^{\rho+\frac{1}{2}} v_k^p u_k(t)dW_k(t). \tag{2.82}$$

For simplicity, suppose that $f_k(t) = 0$. The determinant of the coefficient matrix is

$$D_{N,T} = \sum_{k,l,m}^{N} v_k v_m^\alpha \begin{vmatrix} 1 & v_k^\alpha & v_k \\ 1 & v_l^\alpha & v_l \\ 1 & v_m^\alpha & v_m \end{vmatrix} b_{k,T}b_{l,T}b_{m,T}.$$

We obtain that

$$\hat{\beta}_{N,T} = \frac{D_{N,T}^\beta}{D_{N,T}}, \quad \hat{\lambda}_{N,T} = \frac{D_{N,T}^\lambda}{D_{N,T}}, \quad \hat{\theta}_{N,T} = \frac{D_{N,T}^\theta}{D_{N,T}}, \tag{2.83}$$

where

$$\begin{aligned} D_{N,T}^\beta &= -\begin{vmatrix} A_{1,N} & B_{1,N} & B_{\alpha+1,N} \\ A_{0,N} & B_{0,N} & B_{\alpha,N} \\ A_{\alpha,N} & B_{\alpha,N} & B_{2\alpha,N} \end{vmatrix}, \\ D_{N,T}^\lambda &= -\begin{vmatrix} B_{2,N} & A_{1,N} & B_{\alpha+1,N} \\ B_{1,N} & A_{0,N} & B_{\alpha,N} \\ B_{\alpha+1,N} & A_{\alpha,N} & B_{2\alpha,N} \end{vmatrix}, \\ D_{N,T}^\theta &= -\begin{vmatrix} B_{2,N} & B_{1,N} & A_{1,N} \\ B_{1,N} & B_{0,N} & A_{0,N} \\ B_{\alpha+1,N} & B_{\alpha,N} & A_{\alpha,N} \end{vmatrix}. \end{aligned}$$

With the explicit form of $\int_0^T \mathbb{E}[u_k^2(t)]dt$ computed in Chapter 2.6, we have

$$\mathbb{E}\left[\int_0^T u_k^2(t)dt\right] \approx \frac{\sigma^2 q_k T}{2\mu_k(\Theta)}. \tag{2.84}$$

By (2.84) and the strong law of large numbers, when $\rho = -1$,

$$\begin{aligned}
B_{p,N} &= \frac{1}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^{-1} v_k^p u_k^2(t)dt \\
&\approx \frac{1}{\sigma^2}\sum_{k=1}^{N} q_k^{-1} v_k^p \mathbb{E}\left[\int_0^T u_k^2(t)dt\right] \\
&\approx \begin{cases} \frac{T}{2}\ln(N), \quad p = 0 \\ \frac{T}{2}N^p, \quad p \in (0,1) \\ \frac{T}{2}N - \ln N, \quad p = 1 \\ \frac{T}{2}N^p - N^{p-1}, \quad p \in (1,2] \end{cases}.
\end{aligned} \tag{2.85}$$

The determinant in the denominator of each estimator (2.33) is

$$\begin{aligned}
D_{N,T} &= \begin{vmatrix} B_{2,N} & B_{1,N} & B_{\alpha+1,N} \\ B_{1,N} & B_{0,N} & B_{\alpha,N} \\ B_{\alpha+1,N} & B_{\alpha,N} & B_{2\alpha,N} \end{vmatrix} \\
&= B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}.
\end{aligned} \tag{2.86}$$

Take $\hat{\beta}_N$ as example.

$$\hat{\beta}_N = -\frac{A_{1,N}(B_{0,N}B_{2\alpha,N} - B_{\alpha,N}^2) + A_{\alpha,N}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + A_{0,N}(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}. \tag{2.87}$$

Then, the error term is

$$\hat{\beta_N} - \beta = -\frac{\overline{A_{1,N}}(B_{0,N}B_{2\alpha,N} - B_{\alpha,N}^2) + \overline{A_{\alpha,N}}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + \overline{A_{0,N}}(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}. \tag{2.88}$$

From the calculations in Chap 2.6, we have

$$\lim_{N\to\infty} \frac{2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}{B_{2,N}B_{0,N}B_{2\alpha,N} - B_{\alpha+1,N}^2 B_{0,N}} = 0. \tag{2.89}$$

Similarly, we can obtain that $\lim_{N\to\infty}(\hat{\theta}_N - \theta) = 0$ and $\lim_{N\to\infty}(\hat{\lambda}_N - \lambda) = 0$. The consistency follows that

$$\lim_{N\to\infty} \begin{bmatrix} \hat{\beta_N} \\ \hat{\theta_N} \\ \hat{\lambda_N} \end{bmatrix} = \begin{bmatrix} \beta \\ \theta \\ \lambda \end{bmatrix}. \tag{2.90}$$

Denote $M_{N,1} := N(\hat{\beta}_N - \beta)$, $M_{N,2} = N^\alpha(\hat{\theta}_N - \theta)$, and $M_{N,3} := \sqrt{\ln(N)}(\hat{\lambda}_N - \lambda)$. Due to the distribution of $\overline{A_{p,N}}$ for $p = 0, \alpha, 1$, $M_{N,1}, M_{N,2}, M_{N,3}$ construct a three-dimensional continuous Gaussian martingale $M_N = (M_{N,1}, M_{N,2}, M_{N,3})^T$. The cross variation is computed in Chap 2.6.

$$\lim_{N\to 0} \langle M_{N,1}, M_{N,2}\rangle (t) = 0. \tag{2.91}$$

The derivation for other cross variations is similar. This shows the independence between two components. Thanks to Theorem 2.3.8, the asymptotic normality can be proved. □

Remark 2.3.10. *More generally, in a d dimensional space, the consistency still holds. The asymptotic normality is still true, but the rates depend on d.*

2.4 Estimation of α with given β, θ and λ

With the estimation of β, θ and λ or the given information on these parameters, we estimate the order of the fractional Laplacian α. Unlike the β, θ, λ, α is the power of v_k in the log-likelihood function, which is a nonlinear function with respect to α. Our goal is to solve the following equation:

$$\frac{\partial \ln(L_{N,T})}{\partial \alpha} = \sum_{k=1}^{N} -\theta |v_k|^\alpha \ln(|v_k|)(a_{k,T} + c_{k,T} + \mu_k(\Theta) b_{k,T}) = 0.$$

In order to determine that the maximum value is achievable, we now investigate the sign of the second order derivative $\dfrac{\partial^2 \ln(L_{N,T})}{\partial \alpha^2}$. From (2.11), we have, for some $f_k(t) = 0$,

$$\begin{aligned} \int_0^T q_k^\rho u_k(t) du_k(t) &= -\int_0^T (\beta v_k + \theta v_k^\alpha + \lambda) u_k^2(t) dt + \sigma \int_0^T q_k^{\rho+\frac{1}{2}} u_k(t) dW_k(t) \\ \Longrightarrow a_{k,T} + \mu_k(\Theta) b_{k,T} &= \frac{1}{\sigma} \int_0^T q_k^{\rho+\frac{1}{2}} u_k(t) dW_k(t). \end{aligned} \tag{2.92}$$

With $c_{k,T} = 0$ (i.e. $f_k(t) = 0$), we have, using the definition of $a_{k,T}$ and $b_{k,T}$ from (2.14),

$$\begin{aligned} \frac{\partial^2 \ln(L_{N,T})}{\partial \alpha^2} &= \sum_{k=1}^{N} \{ -\theta |v_k|^\alpha (\ln(|v_k|))^2 (a_{k,T} + c_{k,T}) - \theta |v_k|^\alpha (\ln(|v_k|))^2 \mu_k(\Theta) b_{k,T} \\ &\quad -\theta^2 |v_k|^{2\alpha} (\ln(|v_k|)^2 b_{k,T} \} \\ &= -\theta \sum_{k=1}^{N} |v_k|^\alpha (\ln(v_k))^2 \{ (a_{k,T} + \mu_k(\Theta) b_{k,T}) + \theta |v_k|^\alpha b_{k,T} \} \\ &= -\theta \sum_{k=1}^{N} |v_k|^\alpha (\ln(v_k))^2 \left\{ \frac{1}{\sigma} \int_0^T q_k^{\rho+\frac{1}{2}} u_k(t) dW_k(t) + \theta |v_k|^\alpha b_{k,T} \right\}. \end{aligned} \tag{2.93}$$

Theorem 2.4.1 (Strong Law of Large Numbers for Martingales)**.** *[15, 31] Let M_t be a continuous squared-integrable martingale with quadratic variation $\langle M \rangle_t$. Then $\dfrac{M_t}{\langle M \rangle_t} \to 0, a.s.$ on the event $\{\langle M \rangle_\infty = \infty\}$.*

From the second order derivative with respect to α (2.93) and the notation $b_{k,t}$ introduced in (2.14), we denote that $M_t := \dfrac{1}{\sigma} \sum_{k=1}^{N} |v_k|^\alpha q_k^{\rho+\frac{1}{2}} (\ln(v_k))^2 \int_0^t u_k(s) dW_k(s)$. The quadratic variation of M_t is

$$\langle M \rangle_t = \frac{1}{\sigma^2} \sum_{k=1}^{N} v_k^{2\alpha} q_k^{2\rho+1} (\ln(v_k))^4 \int_0^t u_k^2(s) ds = \sum_{k=1}^{N} v_k^{2\alpha} q_k^{\rho+1} (\ln(v_k))^4 b_{k,t}.$$

Since M_t is a squared-integrable martingale, by Theorem 2.4.1, $\dfrac{M_t}{\langle M\rangle_t} \to 0 \quad a.s.$, as $t \to \infty$. It follows that, for sufficient large t,

$$M_t \lesssim C \sum_{k=1}^{N} v_k^{2\alpha} q_k^{\rho+1} (\ln(v_k))^4 b_{k,t}, \tag{2.94}$$

Now, the second order derivative (2.93) becomes,

$$\sum_{k=1}^{N} v_k^{2\alpha} (\ln(v_k))^2 b_{k,T} \left[\theta - C q_k^{\rho+1} (\ln(v_k))^2\right] > 0 \quad a.s. \text{ if } \rho+1>0. \tag{2.95}$$

Then, $\dfrac{\partial^2 \ln(L_{N,T})}{\partial \alpha^2} < 0 \quad a.s.$. This implies that we can find the global maximum of log-likelihood function for α, given the knowledge of β, θ and λ.

In the following subsections, we estimate the value of the fractional Laplacian order α, given the value of other parameters β, θ, λ. The following suggested estimation are derived from the properties of u_k.

2.4.1 Find α by each k

Construct the nonlinear system using (2.15), (2.16), (2.17) and (2.18). Set them to be zero to figure out the value of $(\alpha, \beta, \theta, \lambda)$. For simplicity, let's introduce some notations.

$$\Omega^1_{k,T} = a_{k,T} + c_{k,T}, \qquad \Omega^2_{k,T} = (\beta v_k + \theta |v_k|^\alpha + \lambda) b_{k,T},$$

$$\begin{cases} \sum_{k=1}^{N} (\Omega^1_{k,T} + \Omega^2_{k,T}) & = 0 \\ \sum_{k=1}^{N} v_k (\Omega^1_{k,T} + \Omega^2_{k,T}) & = 0 \\ \sum_{k=1}^{N} |v_k|^\alpha (\Omega^1_{k,T} + \Omega^2_{k,T}) & = 0 \\ \sum_{k=1}^{N} |v_k|^\alpha \ln(|v_k|)(\Omega^1_{k,T} + \Omega^2_{k,T}) & = 0 \end{cases} \tag{2.96}$$

When $N = 1$, $\Omega^1_{1,T} + \Omega^2_{1,T} = 0$. It is easily to deduce by induction that

$$\Omega^1_{k,T} + \Omega^2_{k,T} = 0, \text{ for all } k \in \mathbb{N}. \tag{2.97}$$

That means for $k = 1, 2, ...$

$$a_{k,T} + c_{k,T} = (\beta v_k + \theta |v_k|^\alpha + \lambda) b_{k,T}. \tag{2.98}$$

Now, we could find the least square solutions numerically for

$$\min_{(\alpha,\beta,\theta,\lambda)} \sum_{k=1}^{N} \left\| \beta v_k + \theta |v_k|^\alpha + \lambda - \frac{a_{k,T} + c_{k,T}}{b_{k,T}} \right\|_2^2 \tag{2.99}$$

As N goes to infinity, the objective function will go to infinity as well, due to the value of v_k. Thus, this objective function (2.99) cannot be used.

2.4.2 Find α by using the log-likelihood function

Another possible way is to find the estimator of α, that is to minimize the first order derivatives of log-likelihood function with respect to α defined in (2.18). Given the exact value of other parameters θ, β, λ, we solve this equation below for α.

$$\frac{\partial \ln(L_{N,T})}{\partial \alpha} = \sum_{k=1}^{N} -|v_k|^{\alpha} \ln(|v_k|)(a_{k,T} + c_{k,T} + \mu_k(\Theta) b_{k,T}) = 0.$$

The solution to the equation above is the estimation of α, used in the numerical example.

2.4.3 Find α by using the quadratic variation

Assume that we have the observations $u_k(t_i)$. We set the difference between quadratic variations as our new loss functions. By the definition of quadratic variation, based on the observation of u_k, we have

$$\langle u_k, u_k \rangle_t \approx \sum_{i=1}^{N_T} (u_k(t_i) - u_k(t_{i-1}))^2 . \tag{2.100}$$

From the dynamics of (2.11), the quadratic variation of u_k is

$$\langle u_k, u_k \rangle_t = \sigma^2 q_k t. \tag{2.101}$$

We then apply the optimization method to find the estimation of α. We define the loss function as follows.

$$\text{argmin}_{\alpha} \frac{1}{N} \sum_{k=1}^{N} \left[\sum_{i=1}^{N_T} (u_k(t_i) - u_k(t_{i-1}))^2 - \sigma^2 q_k T \right]^2 . \tag{2.102}$$

or

$$\text{argmin}_{\alpha} \frac{1}{N} \sum_{k=1}^{N} \left[\sum_{i=1}^{N_T} (u_k(t_i) - u_k(t_{i-1}))^2 - \sum_{i=1}^{N_T} (\tilde{u}_k(t_i) - \tilde{u}_k(t_{i-1}))^2 \right]^2 , \tag{2.103}$$

where $\tilde{u}_k(t_i)$ depends on different α and probability spaces.

2.4.4 Find α by using the first moment

In this subsection, we estimate α by comparing the moments in the long term. With the ergodicity of the process, the time average converges in squared mean to the ensemble average $\mathbb{E}[u_k^2(t)]$. Based on the ergodicity, we can construct the loss function as follows.

$$\text{argmin}_{\alpha} \frac{1}{N} \sum_{k=1}^{N} \left[\frac{1}{T} \int_0^T u_k^2(t) dt - \mathbb{E}[u_k^2(\infty)] \right]^2 , \tag{2.104}$$

and due to $\mu_k(\Theta) > 0$, with the assumption $u_k^2(0)$ is finite,

$$\mathbb{E}[u_k^2(\infty)] = \lim_{t \to \infty} \left\{ \mathbb{E}[u_k^2(0)] e^{-\mu_k(\Theta) t} + \frac{\sigma^2 q_k}{2\mu_k(\Theta)} \left[1 - e^{-2\mu_k(\Theta) t} \right] \right\} = \frac{\sigma^2 q_k}{2\mu_k(\Theta)}. \tag{2.105}$$

Based on (2.105) and the discretization of the integrals in (2.104), the loss function becomes

$$\operatorname{argmin}_\alpha \frac{1}{N}\sum_{k=1}^{N}\left[\frac{1}{T}\sum_{i=1}^{N_T}\left(u_k^2(t_i)-u_k^2(t_{i-1})\right)\Delta t-\frac{\sigma^2 q_k}{2(\beta v_k+\theta|v_k|^\alpha+\lambda)}\right]^2. \tag{2.106}$$

The optimal value to this problem is the estimation of α.

2.4.5 Find α by using the trajectory fitting estimators

In this subsection, we consider to apply the trajectory fitting estimators for SPDEs [12], in order to find the estimator α. Apply the Itô formula on the dynamic of $u_k(t)$ and find the $u_k^2(t)$.

$$u_k^2(t)=u_k^2(0)+\int_0^t\left[\sigma^2 q_k-2(\beta v_k+\theta v_k^\alpha+\lambda)u_k^2(t)-2u_k(s)f_k(s)\right]ds+2\sigma\sqrt{q_k}\int_0^t u_k(s)dW_k(s), \tag{2.107}$$

and we let

$$V_k(t,\alpha):=u_k^2(0)+\int_0^t\left[\sigma^2 q_k-2(\beta v_k+\theta v_k^\alpha+\lambda)u_k^2(s)-2u_k(s)f_k(s)\right]ds. \tag{2.108}$$

The trajectory fitting estimator (TFE) for the unknown parameter α is defined as

$$\tilde{\alpha}_N^{TFE}:=\operatorname{argmin}_\alpha\sum_{k=1}^{N}\int_0^T (V_k(t,\alpha)-u_k^2(t))^2 dt. \tag{2.109}$$

Take the first order derivative with respect to α to find the critical points.

$$\frac{\partial V_k}{\partial\alpha}=-2\theta v_k^\alpha\ln(v_k)\int_0^t u_k^2(s)ds.$$

The loss function in (2.109) can be simplified into the following form.

$$\begin{aligned}
&2\sum_{k=1}^{N}\int_0^T (V_k(t,\alpha)-u_k^2(t))\frac{\partial V_k}{\partial\alpha}dt=0,\\
&2\sum_{k=1}^{N}\int_0^T (V_k(t,\alpha)-u_k^2(t))\left(-2\theta v_k^\alpha\ln(v_k)\int_0^t u_k^2(s)ds\right)dt=0,\\
&\sum_{k=1}^{N}\left\{v_k^\alpha\ln(v_k)\int_0^T\xi_k(t)\left(u_k^2(0)+\int_0^t\left[\sigma^2 q_k-2(\beta v_k+\theta v_k^\alpha+\lambda)u_k^2(s)-2u_k(s)f_k(s)\right]ds-u_k^2(t)\right)dt\right\}=0,\\
&\sum_{k=1}^{N}v_k^\alpha\ln(v_k)\left\{u_k^2(0)Y_k(T)+\sigma^2 q_k X_k(T)-2(\beta v_k+\theta v_k^\alpha+\lambda)Z_k(T)\right.\\
&\qquad\qquad\left.-2\int_0^T\xi_k(t)F_k(t)dt-\int_0^T\xi_k(t)u_k^2(t)dt\right\}=0,
\end{aligned} \tag{2.110}$$

where $\xi_k(t)=\int_0^t u_k^2(s)ds$, $F_k(t):=\int_0^t u_k(s)f_k(s)ds$, $X_k(t):=\int_0^t s\xi_k(s)ds$, $Y_k(t):=\int_0^t\xi_k(s)ds$ and $Z_k(t):=\int_0^t\xi_k^2(s)ds$. The goal is to find α satisfying (2.110), a nonlinear equation of α.

2.5 Numerical Examples

We work on the numerical simulations on MATLAB. We fix the random seed generator by Matlab command `rng(100,'twister')` and work on the time period $T = 1$. Given the simulated data with parameter settings, generate the OU process from (2.11). Due to the stiffness, we generate the simulated data by using the implicit Euler schemes. We set the time increment $\Delta t = 10^{-3}$, and we truncate $N = 1024$. For PDE settings, we consider the domain $x \in [0, 2\pi]$. To eliminate the statistical error, we generate $M = 1500$ Monte Carlo paths. With the periodic boundary condition, the eigenfunctions e^{ikx} are chosen for $k = 1, 2, ...,$ where i is the imaginary unit $\sqrt{-1}$. Suppose that $u(t,x) = \sum_{k=-\infty}^{\infty} u_k(t)e^{ikx}$. Then, $u_{xx}(t,x) = -\sum_{k=-\infty}^{\infty} k^2 u_k(t)e^{ikx}$. Therefore, the eigenvalues v_k are k^2 for each k.

Consider the following fractional advection-diffusion equation with the periodic boundary condition $u(t,0) = u(t,2\pi)$. In the numerical example, we set $f = 0$ and $\sigma = 0.01$.

$$u_t - \beta\Delta u + \theta(-\Delta)^\alpha u + \lambda u = \sigma \dot{W}^Q, \tag{2.111}$$

It follows the dynamics of Fourier frequency u_k. Using the backward Euler schemes, (2.111) can be discretized as follows. For $j = 0, 1, ..., N_T$,

$$u_k(t_{j+1}) = u_k(t_j) - [(\beta k^2 + \theta|k|^{2\alpha} + \lambda)u_k(t_{j+1}) + f_k(t_{j+1})]\Delta t + \sigma\sqrt{q_k}(W_k(t_{j+1}) - W_k(t_j)). \tag{2.112}$$

Then $a_{k,T}, b_{k,T}$ and $c_{k,T}$ can be discretized by the Ito-Riemann sum in the (2.14).

$$a_{k,T} = \frac{1}{\sigma^2}\int_0^T q_k^\rho u_k(t)du_k(t) \approx \frac{1}{\sigma^2}\sum_{p=1}^{N_T} q_k^\rho u_k(t_p)(u_k(t_p) - u_k(t_{p-1})). \tag{2.113}$$

Similarly, we have

$$b_{k,T} = \frac{1}{\sigma^2}\int_0^T q_k^\rho u_k^2(t)dt \approx \frac{1}{\sigma^2}\sum_{p=1}^{N_T} q_k^\rho u_k^2(t_p)\Delta t. \tag{2.114}$$

$$c_{k,T} = \frac{1}{\sigma^2}\int_0^T q_k^\rho u_k(t)f_k(t)dt \approx \frac{1}{\sigma^2}\sum_{p=1}^{N_T} q_k^\rho u_k(t_p)f_k(t_p)\Delta t. \tag{2.115}$$

For simplicity, we set f in the PDE to be zero. Here, q_k is pre-selected depending on the frequency k. The frequency of Fourier expansion k is $[0, 1, ..., N/2, 1 - N/2, .., -1]$. The length of frequency is still N. In our numerical example, we set $q_k = \frac{1}{k^2}$, and $\rho = -1$. To avoid the zero denominator, we use the Fourier frequency starting from 1. Suppose that $u_k(0) = 0$, which is deterministic and constant. Assumption 2.3.4 in Theorem 2.3.5 is satisfied. With the choice of q_k and ρ, from the (2.49) in the theorem 2.3.5, we have the convergence rate

$$\lim_{N\to\infty} \sqrt{I(N)}(\hat{\theta}_N - \theta) = \mathcal{N}(0, \boldsymbol{\sigma}^2(\theta)/T), \tag{2.116}$$

where $I(N) = \sum_{k=1}^{N} q_k^{\rho+1}k^{\omega_\alpha}$ and $\omega_\alpha = \frac{2(\alpha - m)}{d}$. Recall that $2m$ is the order of the operator $A_0 + \Theta A_1$. As $\omega_\alpha \geq -1$ in Theorem 2.3.5 and $\alpha \in (0,1)$, it follows that $m - \frac{d}{2} \leq \alpha < 1$. In our case, suppose that β is known. $A_0 = \beta(-\Delta)$ and $A_1 = \theta(-\Delta)^\alpha + \lambda\mathbf{I}$. Thus, we have dimension $d = 1$, so that $\omega_\alpha = 2\alpha - 2$. It follows that

$$I(N) = \sum_{k=1-N/2, k\neq 0}^{N/2} k^{-2(\rho+1)+\omega_\alpha} = \sum_{k=1-N/2, k\neq 0}^{N/2} k^{2\alpha-2}. \tag{2.117}$$

We assume that $\omega_\alpha \geq -1$ in Theorem 2.3.5. It implies $\alpha \geq \frac{1}{2}$ in our case. Thus, for a fixed α, we obtain that

$$I(N) = \sum_{k=1-N/2, k\neq 0}^{N/2} k^{2\alpha-2} \approx \mathcal{O}(N^{2\alpha-1}). \tag{2.118}$$

2.5.1 Case 0: Given α, θ, λ, estimate β

We start with the simple case. Consider the following stochastic heat equation driven by additive noise problem. For $x \in [0, \pi]$ and $t \in [0, 1]$,

$$du(t,x) - \beta\Delta u(t,x)dt = \sigma dW(t,x), \tag{2.119}$$

with zero initial condition $u(0, x) = 0$, zero boundary conditions and driven by space-time white noise. We assume the true but unknown parameter $\theta = 1$. Based on the spectral method, we have the dynamics of Fourier coefficients:

$$du_k(t) = -\beta v_k u_k dt + \sigma dW_k(t). \tag{2.120}$$

Compared with the dynamics (2.11), we set $q_k = \frac{1}{k^2}, v_k = k^2$, $\theta = \lambda = 0, \alpha = 0, \sigma = 0.01$ and $f = 0$. With the different choice of q_k, we will have different results. Discretize it by implicit Euler schemes with time step Δt, which is equal to 10^{-3}. The MLE estimator derived from the first order derivatives of log-likelihood functions (2.15) will be simplified as follows. With the notation of $a_{k,T}$ and $b_{k,T}$ in (2.14),

$$\frac{\partial \ln(L_{N,T})}{\partial \beta} = 0 \Longrightarrow \sum_{k=1}^{N} -v_k a_{k,T} - \beta v_k^2 b_{k,T} = 0 \Longrightarrow \hat{\beta_N} = -\frac{\sum_{k=1}^{N} v_k a_{k,T}}{\sum_{k=1}^{N} v_k^2 b_{k,T}} = -\frac{\sum_{k=1}^{N} v_k \int_0^T u_k(t)du_k(t)}{\sum_{k=1}^{N} v_k^2 \int_0^T u_k^2(t)dt}. \tag{2.121}$$

Then,

$$\hat{\beta_N} = -\frac{\sum_{k=1}^{N} k^2 \int_0^T u_k(t)du_k(t)}{\sum_{k=1}^{N} k^4 \int_0^T u_k^2(t)dt} \approx -\frac{\sum_{k=1}^{N} k^2 \sum_{j=0}^{N_T-1} u_k(t_j)(u_k(t_{j+1}) - u_k(t_j))}{\sum_{k=1}^{N} k^4 \sum_{j=0}^{N_T-1} u_k^2(t_j)\Delta t}. \tag{2.122}$$

Based on the simulations in MATLAB, we get the following relationship between the estimations and the first 200 Fourier modes. As Figure 2.1 and Figure 2.2 shown below, when the number of Fourier modes N increases, the error between the maximum likelihood estimator (MLE) and the exact value are getting smaller. When $N = 200$, the error $\hat{\beta_N} - \beta$ is equal to -1.0282×10^{-5}.

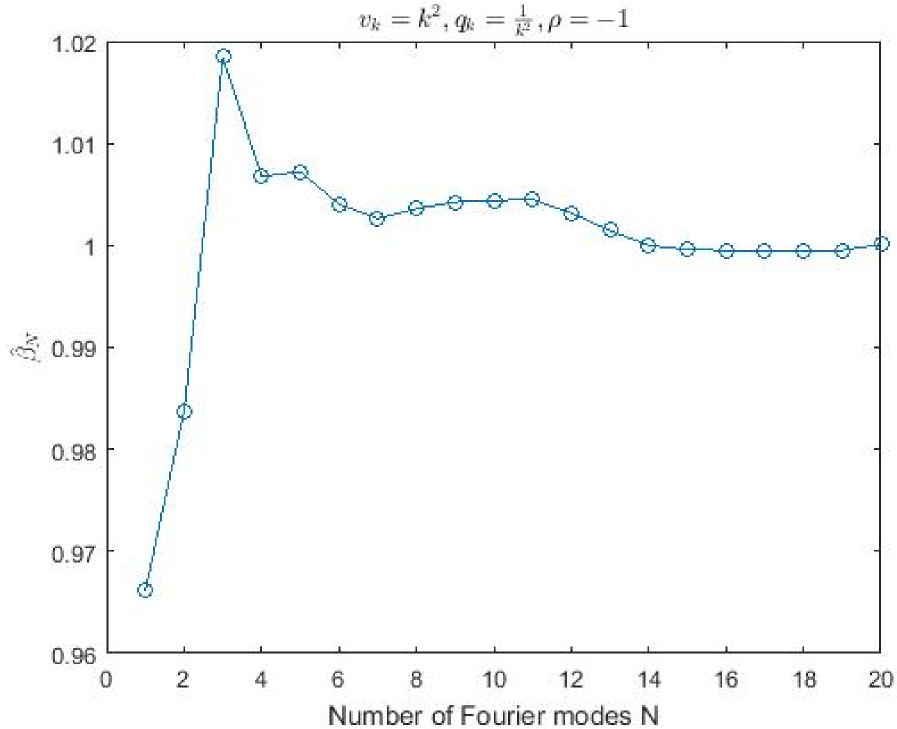

Figure 2.1: The estimation $\hat{\beta}_N$ on the coefficients of the Laplacian and different number of Fourier modes N from 1 to 20 for the stochastic heat equation. The estimation $\hat{\beta}_N$ is an ensemble average of 100 simulations with different random paths.

2.5.2 Case 1: Given α, estimate θ and λ

In this case, we estimate the value of β, θ and λ, given the information of α. As the numerical setting mentioned above, to eliminate the errors coming from the stochastic issues, we ran $M = 100$ paths, and for each path, we computed its estimators respectively. We consider two different situations when α is less than 0.5 and greater than 0.5. The following numerical results are derived from the linear system (3.14). For solving this linear system, I used the backslash in Matlab. The initial condition of $u(t, x)$ is 0.

Suppose that $\alpha = 0.3$, $\beta = \theta = \lambda = 1$, and set the time step $\Delta t = 10^{-3}$. When $N = 200$, we get the estimation $\hat{\theta}_n = 0.9587947$. The error on this estimation is 4×10^{-2} . The estimations of $\hat{\theta}_n$ versus the different number of Fourier modes N are in Figure 2.3. Based on these data, I fit one linear regression line. We can observe from the orange line that the estimation is getting closer to the exact value as the Fourier mode increases.

Run this test again with a smaller time step $\Delta t = 10^{-4}$, with the same settings. When $N = 200$, we get the estimation $\hat{\theta}_n = 1.025071$. The error is 2×10^{-2}. When $\Delta t = 10^{-4}$, the estimation is more accurate. From Figure 2.4, we observe that the estimation goes closer to the exact values as the Fourier mode N increases. Compared with Figure 2.3, using the smaller time step, the slope of the regression line is smaller, which means the smaller time step gives the more accurate estimation with limited Fourier modes N.

With the different value of $\alpha = 0.8$, the estimation $\hat{\theta}_n$ is 0.9985065. The error is 1.5×10^{-3}.

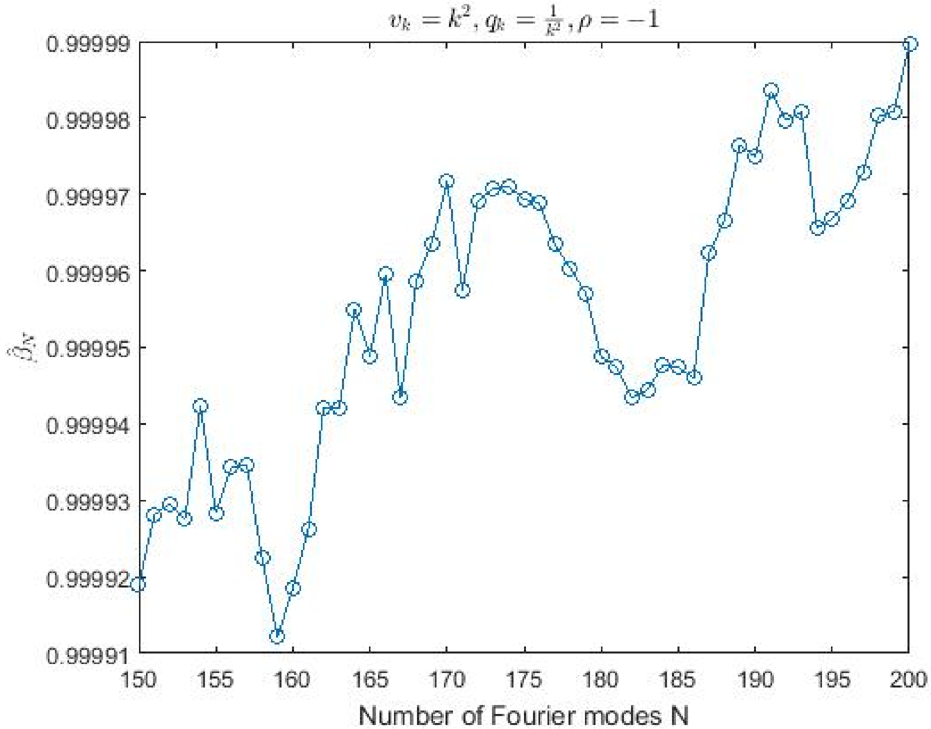

Figure 2.2: The estimation $\hat{\beta}_N$ on the coefficients of the Laplacian and larger number of Fourier modes N from 150 to 200 for the stochastic heat equation. The estimation $\hat{\beta}_N$ is an ensemble average of 100 simulations with different random paths.

When N is large, the estimation is less volatile than the case when $\alpha = 0.3$. Since θ is the coefficient of the fractional Laplacian term, so the estimation is dependent on the value of α. Again, when N becomes larger, the estimation is still closer than the exact value. We can conclude that when α is close to 1, i.e. the fractional Laplacian term becomes the ordinary Laplacian term, less N can be used than the case when α is close to 0, to get the same accuracy.

2.5.3 Case 2: Given β, θ and λ, estimate α

In this case, we estimate the alpha, given the information β, θ and λ. The estimator value is evaluated by solving a nonlinear equation $\dfrac{\partial \ln(L)}{\partial \alpha} = 0$. In Matlab, we solve this one variable nonlinear equation by the built-in function `fsolve`.

Suppose that $\theta = 1, \beta = \lambda = 0$ and the exact value of α is 0.3. The estimation $\hat{\alpha}_N = 0.2986590$. The error is 1.3×10^{-3}. We use the initial guess 0.4 when applying `fsolve`. In Figure 2.7, we observe from the regression line that the estimation becomes closer to the exact value when N becomes larger. With the same setting except the time step $\Delta t = 10^{-4}$, the estimation is 0.3002675, and the error is -2.6750×10^{-4}. See Figure 2.9. We observe that the estimation with smaller time steps is more accurate.

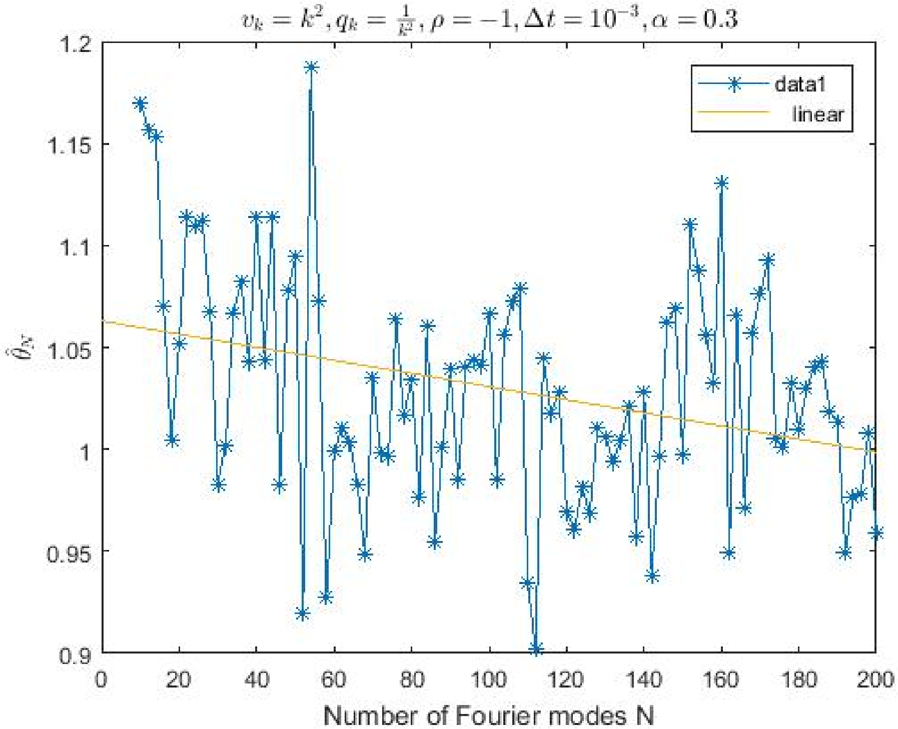

Figure 2.3: The estimation $\hat{\theta}_N$ on the coefficients of the fractional Laplacian term and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-3}$. The estimation $\hat{\theta}_N$ is an ensemble average of 100 simulations with different random paths.

2.6 Appendix

2.6.1 Proof of the totally positive matrix

Proof of Theorem 2.3.1:

Proof. The proof follows the ideas in [8]. Recall the fact that the determinant of the ordinary Vandermonde matrix is positive. There exists a continuous path that goes from the determinant (2.29) stated in the theorem to the determinant of the ordinary Vandermonde matrix. A continuous function can not go from a negative value to a positive value without vanishing some points. Hence, the determinant of a generalized Vandermonde matrix is positive. It remains to show that the determinant of it could not be zero.

Proof by induction. When $n = 1$, it is trivial, since $a_1^{\lambda_1} > 0$.

Induction Hypothesis: Assume that $\det(A_n) \neq 0$, for all $n < N - 1$.

Now, want to prove that $\det(A_N) \neq 0$.. Arguing by contradiction, we assume that $\det(A_N) = 0$, which means that if we had a vanishing generalized Vandemonde determinant of size n, there would be a linear combination of different powers of x vanishing at least n positive values of $x \in \{a_1, a_2, ..., a_n\}$. That is to say that we would have coefficients c_i (not all vanishing) for which

$$0 = c_1 x^{\lambda_1} + c_2 x^{\lambda_2} + ... + c_n x^{\lambda_n}, \text{ for at least } n \text{ positive values of } x.$$

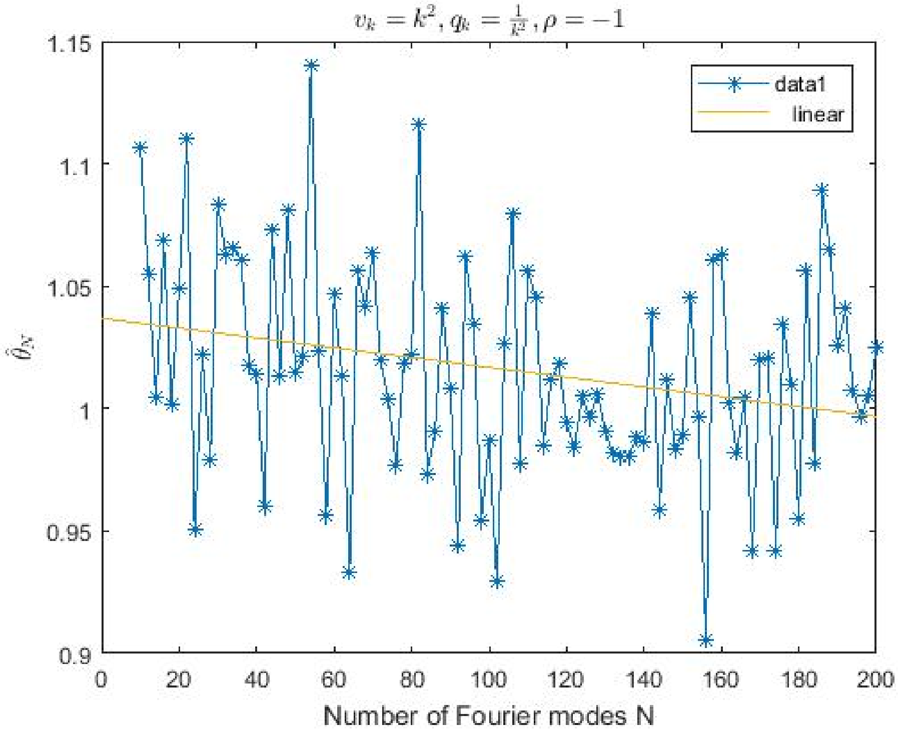

Figure 2.4: The estimation $\hat{\theta}_N$ on the coefficients of the fractional Laplacian term and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-4}$. The estimation $\hat{\theta}_N$ is an ensemble average of 100 simulations with different random paths.

Since x^{λ_1} is positive,

$$0 = c_1 + c_2 x^{\lambda_2 - \lambda_1} + ... + c_n x^{\lambda_n - \lambda_1}, \text{ for those same values.}$$

By Rolle's theorem, we have

$$0 = c_2(\lambda_2 - \lambda_1) x^{\lambda_2 - \lambda_1} + ... + c_n(\lambda_n - \lambda_1) x^{\lambda_n - \lambda_1}, \text{ for at least } n-1 \text{ values of } x > 0,$$

which contradicts with the induction hypothesis that $\det(A_{N-1}) \neq 0$. Thus, $\det(A_n) \neq 0$, for all $n \in \mathbb{N}$. $\square$

2.6.2 Properties of OU process u_k

In this subsection, we calculate the moments of the Fourier modes u_k. These moments are used in the proof of Theorem 2.3.5 to evaluate η_n, b_n. Let the OU process be

$$u_k(t) = u_k(0) e^{-\mu_k(\Theta)t} - \int_0^t e^{-\mu_k(\Theta)(t-s)} f_k(s) ds + \sigma \sqrt{q_k} \int_0^t e^{-\mu_k(\Theta)(t-s)} dW_k(s). \tag{2.123}$$

With Assumption 2.3.4 that $u_k(0) \sim \mathcal{N}(m_k, \sigma_k^2)$ and independence of $W_k(t)$, denote $I_k(t) :=$

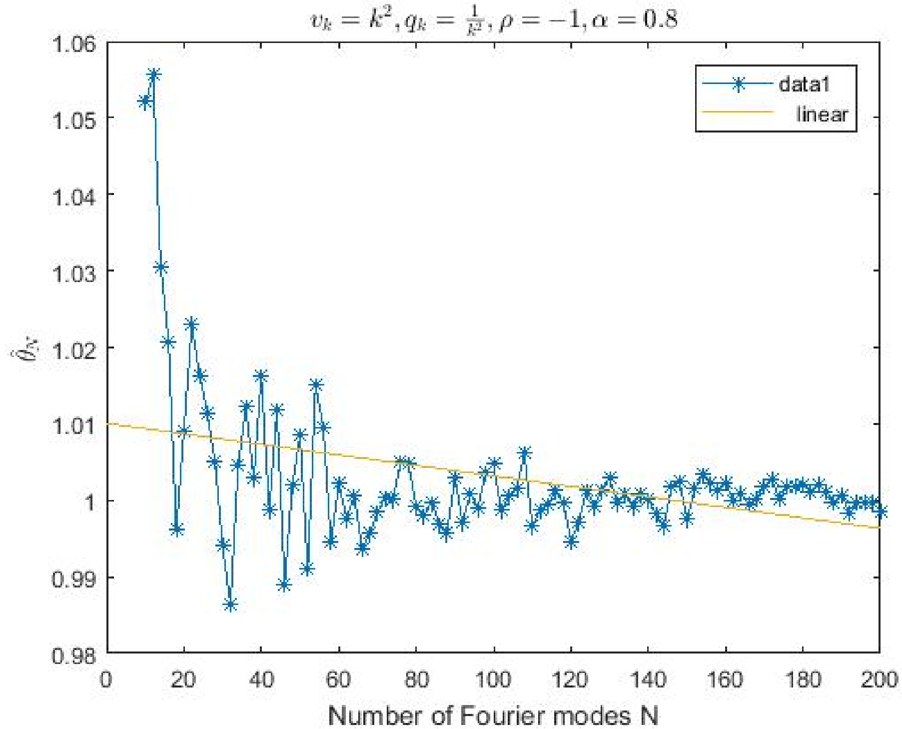

Figure 2.5: The estimation $\hat{\theta}_N$ on the coefficients of the fractional Laplacian term and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-3}$, $\alpha = 0.8$. The estimation $\hat{\theta}_N$ is an ensemble average of 100 simulations with different random paths.

$\int_0^t e^{-\mu_k(\Theta)(t-s)} f_k(s)ds$. Then, we obtain that

$$\begin{aligned}\mathbb{E}[u_k^2(t)] &= (m_k^2 + \sigma_k^2)e^{-2\mu_k(\Theta)t} + I_k^2(t) + \sigma^2 q_k \int_0^t e^{-2\mu_k(\Theta)(t-s)}ds - 2m_k I_k(t) e^{-\mu_k(\Theta)t} \\ &= (m_k^2 + \sigma_k^2)e^{-2\mu_k(\Theta)t} + I_k^2(t) + \frac{\sigma^2 q_k}{2\mu_k(\Theta)}\left[1 - e^{-2\mu_k(\Theta)t}\right] - 2m_k I_k(t) e^{-\mu_k(\Theta)t},\end{aligned} \tag{2.124}$$

and

$$\begin{aligned}\int_0^T \mathbb{E}[u_k^2(t)]dt &= \frac{m_k^2 + \sigma_k^2}{-2\mu_k(\Theta)} e^{-2\mu_k(\Theta)t}\Big|_0^T + \int_0^T I_k^2(t)dt + \frac{\sigma^2 q_k T}{2\mu_k(\Theta)} + \frac{\sigma^2 q_k}{4\mu_k^2(\Theta)} e^{-2\mu_k(\Theta)t}\Big|_0^T \\ &\quad - 2m_k \int_0^T I_k(t) e^{-\mu_k(\Theta)t} dt \\ &= \frac{m_k^2 + \sigma_k^2}{2\mu_k(\Theta)}\left[1 - e^{-2\mu_k(\Theta)T}\right] + \int_0^T I_k^2(t)dt + \frac{\sigma^2 q_k T}{2\mu_k(\Theta)} - \frac{\sigma^2 q_k}{4\mu_k^2(\Theta)}\left[1 - e^{-2\mu_k(\Theta)T}\right] \\ &\quad - 2m_k \int_0^T I_k(t) e^{-\mu_k(\Theta)t} dt.\end{aligned} \tag{2.125}$$

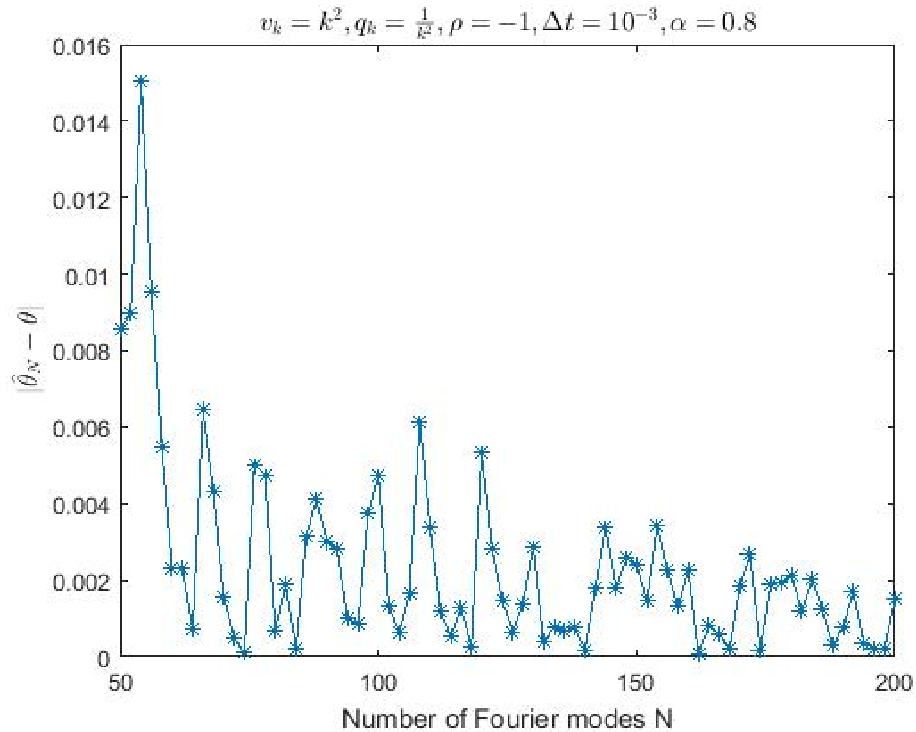

Figure 2.6: The error $|\hat{\theta}_N - \theta|$ on the coefficients of the fractional Laplacian term and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-3}$, $\alpha = 0.8$, The estimation $\hat{\theta}_N$ is an ensemble average of 100 simulations with different random paths.

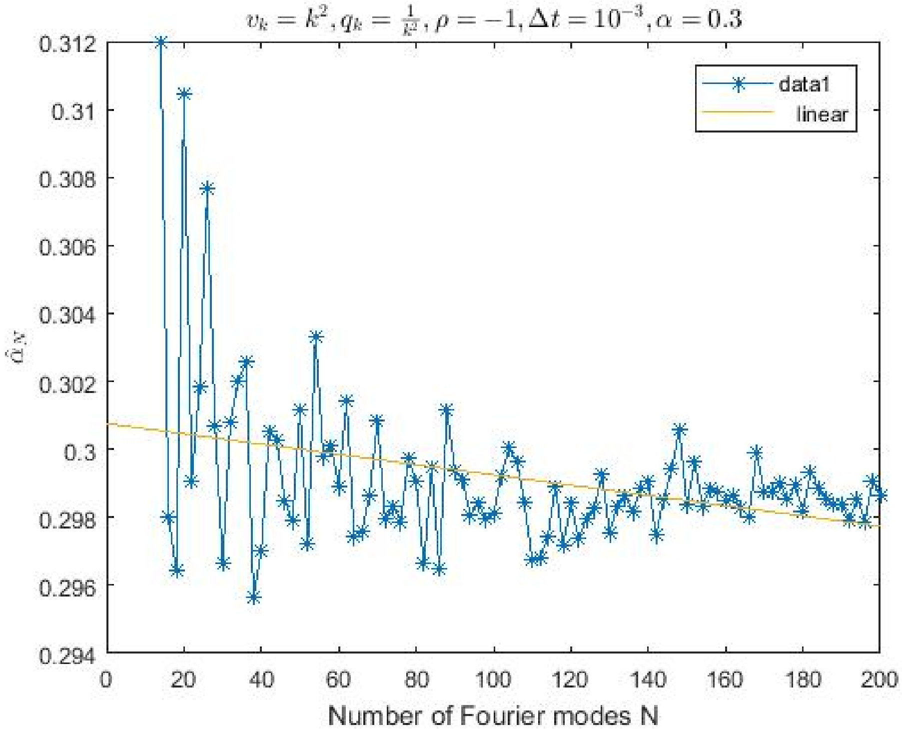

Figure 2.7: The estimation $\hat{\alpha}_N$ on the coefficients of the fractional Laplacian order and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-3}$, $\alpha = 0.3$. The estimation $\hat{\alpha}_N$ is an ensemble average of 100 simulations with different random paths.

Here, we need that $f \in L^2((0,\infty))$. For simplicity, we set f to be zero. It follows that

$$\int_0^T \mathbb{E}[u_k^2(t)]dt = \frac{m_k^2 + \sigma_k^2}{2\mu_k(\Theta)} \left[1 - e^{-2\mu_k(\Theta)T}\right] + \frac{\sigma^2 q_k T}{2\mu_k(\Theta)} - \frac{\sigma^2 q_k}{4\mu_k^2(\Theta)} \left[1 - e^{-2\mu_k(\Theta)T}\right]. \tag{2.126}$$

Using (2.34), let $u_k(t) := A + B$, where $A := u_k(0)e^{-\mu_k(\Theta)t}$, and with the assumption that $f = 0$,

$$B := -\int_0^t e^{-\mu_k(\Theta)(t-s)} f_k(s)ds + \sigma\sqrt{q_k}\int_0^t e^{-\mu_k(\Theta)(t-s)}dW_k(s) = \sigma q_k^{\frac{1}{2}} \int_0^t e^{-\mu_k(\Theta)(t-s)}dW_k(s).$$

With the distribution on the initial data $u(0)$, $\mathbb{E}[A] = m_k e^{-\mu_k(\Theta)t}, Var(A) = \sigma_k^2 e^{-2\mu_k(\Theta)t}$. Since $u_k(0) - m_k \sim \mathcal{N}(0, \sigma_k^2)$, we similarly find that, using $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$,

$$\begin{aligned} \mathbb{E}[A^4] &= \mathbb{E}[(u_k(0)e^{-\mu_k(\Theta)t})^4] = e^{-4\mu_k(\Theta)t}\mathbb{E}[(u_k(0) - m_k + m_k)^4] \\ &= e^{-4\mu_k(\Theta)t}\left(\mathbb{E}[(u_k(0) - m_k)^4] + 6\mathbb{E}[(u_k(0) - m_k)^2]m_k^2 + m_k^4\right) \\ &= e^{-4\mu_k(\Theta)t}\left(3\sigma_k^4 + 6m_k^2\sigma_k^2 + m_k^4\right). \end{aligned} \tag{2.127}$$

Then,

$$\begin{aligned} Var(A^2) &= \mathbb{E}[A^4] - (\mathbb{E}[A^2])^2 = e^{-4\mu_k(\Theta)t}\left(3\sigma_k^4 + 6m_k^2\sigma_k^2 + m_k^4\right) - \left[(m_k^2 + \sigma_k^2)e^{-2\mu_k(\Theta)t}\right]^2 \\ &= 2\sigma_k^2(\sigma_k^2 + 2m_k^2)e^{-4\mu_k(\Theta)t}. \end{aligned} \tag{2.128}$$

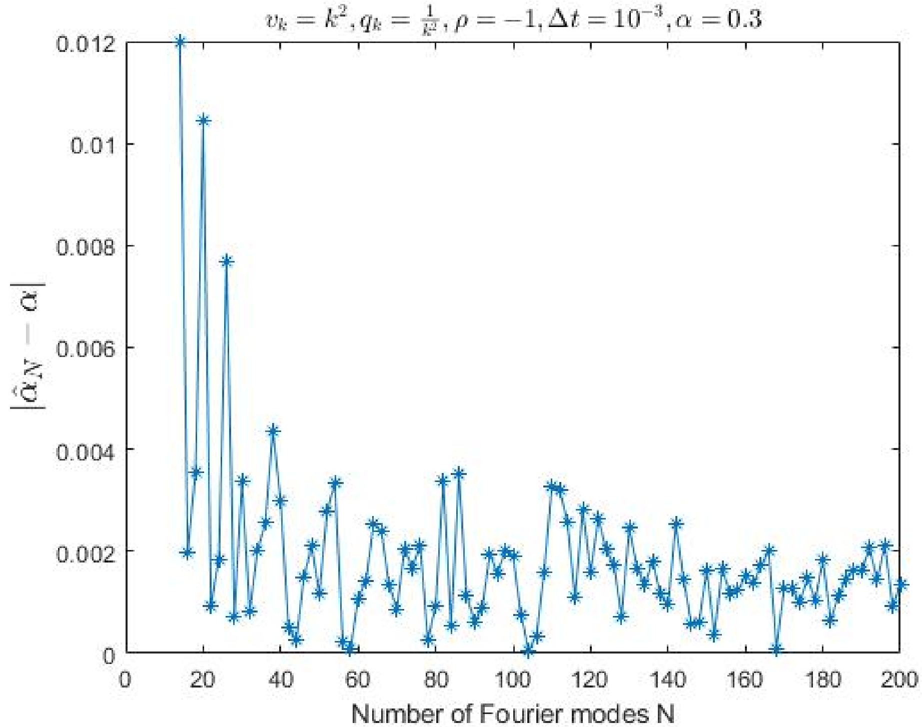

Figure 2.8: The error $|\hat{\alpha}_N - \alpha|$ on the coefficients of the fractional Laplacian order and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-3}$, $\alpha = 0.3$. The estimation $\hat{\alpha}_N$ is an ensemble average of 100 simulations with different random paths.

For the term B, we have $\mathbb{E}[B] = \mathbb{E}[B^3] = 0$, and $Var(B) = \mathbb{E}[B^2] = \dfrac{\sigma^2 q_k}{2\mu_k(\Theta)}(1 - e^{-2\mu_k(\Theta)t})$. Recall that if $X \sim \mathcal{N}(0, \sigma^2)$, then $\mathbb{E}[X^4] = 3\sigma^2$. Then,

$$\mathbb{E}[B^4] = 3\left(\frac{\sigma^2 q_k}{2\mu_k(\Theta)}(1 - e^{-2\mu_k(\Theta)t})\right)^2 = \frac{3\sigma^4 q_k^2}{4\mu_k^2(\Theta)}(1 - e^{-2\mu_k(\Theta)t})^2.$$

Then,

$$\begin{aligned} Var(B^2) = \mathbb{E}[B^4] - (\mathbb{E}[B^2])^2 &= \frac{3\sigma^4 q_k^2}{4\mu_k^2(\Theta)}(1 - e^{-2\mu_k(\Theta)t})^2 - \left[\frac{\sigma^2 q_k}{2\mu_k(\Theta)}(1 - e^{-2\mu_k(\Theta)t})\right]^2 \\ &= \frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}(1 - e^{-2\mu_k(\Theta)t})^2. \end{aligned} \tag{2.129}$$

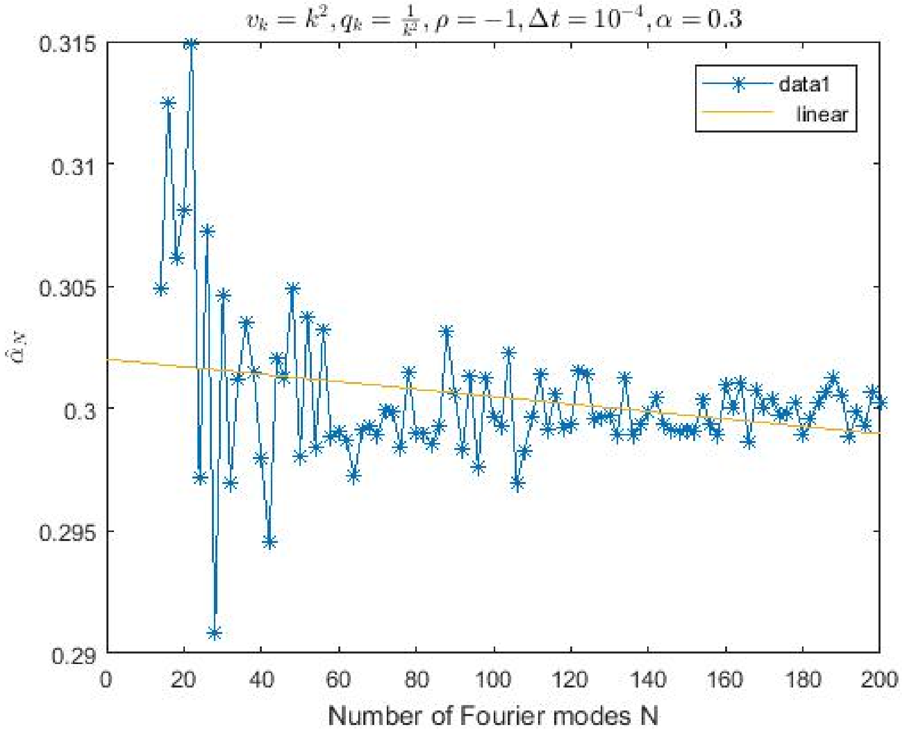

Figure 2.9: The estimation $\hat{\alpha}_N$ on the coefficients of the fractional Laplacian order and different number of Fourier modes N for the stochastic heat equation with fractional derivatives $\Delta t = 10^{-4}$, $\alpha = 0.3$. The estimation $\hat{\alpha}_N$ is an ensemble average of 100 simulations with different random paths.

From (2.128), (2.129), since A and B are independent Gaussian distributed, we have

$$
\begin{aligned}
&Var(u_k^2(t)) = Var((A+B)^2) = Var(A^2) + Var(B^2) + 4Var(AB) \\
&= Var(A^2) + Var(B^2) + 4(\mathbb{E}[(AB)^2] - (\mathbb{E}[AB])^2) = Var(A^2) + Var(B^2) + 4\mathbb{E}(A^2)\mathbb{E}[B^2] \\
&= \underbrace{2\sigma_k^2(\sigma_k^2 + 2m_k^2)e^{-4\mu_k(\Theta)t}}_{I_1} + \underbrace{\frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}(1 - e^{-2\mu_k(\Theta)t})^2}_{I_2} + \underbrace{4\left[(m_k^2 + \sigma_k^2)e^{-2\mu_k(\Theta)t}\right]\left[\frac{\sigma^2 q_k}{2\mu_k(\Theta)}(1 - e^{-2\mu_k(\Theta)t})\right]}_{I_3}.
\end{aligned}
\tag{2.130}
$$

We split (2.130) into three parts. Integrating each term with respect to time on $[0, T]$, we have

$$
I_1 : 2\sigma_k^2(\sigma_k^2 + 2m_k^2)\int_0^T e^{-4\mu_k(\Theta)t}dt = \frac{2\sigma_k^2(\sigma_k^2 + 2m_k^2)}{4\mu_k(\Theta)}\left(1 - e^{-4\mu_k(\Theta)T}\right) \approx \frac{2\sigma_k^2(\sigma_k^2 + 2m_k^2)}{4\mu_k(\Theta)}, \tag{2.131}
$$

$$
\begin{aligned}
I_2 :&\frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}\int_0^T (1-e^{-2\mu_k(\Theta)t})^2 dt = \frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}\int_0^T (1-2e^{-2\mu_k(\Theta)t}+e^{-4\mu_k(\Theta)t})dt \\
&= \frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}\left[T - \frac{2e^{-2\mu_k(\Theta)t}}{-2\mu_k(\Theta)}\Big|_0^T + \frac{e^{-4\mu_k(\Theta)t}}{-4\mu_k(\Theta)}\Big|_0^T\right] \\
&= \frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}\left[T - \frac{1-e^{-2\mu_k(\Theta)T}}{\mu_k(\Theta)} + \frac{1-e^{-4\mu_k(\Theta)T}}{4\mu_k(\Theta)}\right] \\
&\approx \frac{\sigma^4 q_k^2}{2\mu_k^2(\Theta)}\left[T - \frac{3}{4\mu_k(\Theta)}\right],
\end{aligned}
\tag{2.132}
$$

$$
\begin{aligned}
I_3 :&\int_0^T 4\left[(m_k^2+\sigma_k^2)e^{-2\mu_k(\Theta)t}\right]\left[\frac{\sigma^2 q_k}{2\mu_k(\Theta)}(1-e^{-2\mu_k(\Theta)t})\right]dt = \frac{4(m_k^2+\sigma_k^2)\sigma^2 q_k}{2\mu_k(\Theta)}\int_0^T \left(e^{-2\mu_k(\Theta)t}-e^{-4\mu_k(\Theta)t}\right)dt \\
&= \frac{4(m_k^2+\sigma_k^2)\sigma^2 q_k}{2\mu_k(\Theta)}\left[\frac{e^{-2\mu_k(\Theta)t}}{-2\mu_k(\Theta)} - \frac{e^{-4\mu_k(\Theta)t}}{-4\mu_k(\Theta)}\right]\Bigg|_0^T \\
&= \frac{4(m_k^2+\sigma_k^2)\sigma^2 q_k}{2\mu_k(\Theta)}\left[\frac{1-e^{-2\mu_k(\Theta)T}}{2\mu_k(\Theta)} - \frac{1-e^{-4\mu_k(\Theta)T}}{4\mu_k(\Theta)}\right] \approx \frac{4(m_k^2+\sigma_k^2)\sigma^2 q_k}{2\mu_k(\Theta)}\left[\frac{1}{2\mu_k(\Theta)} - \frac{1}{4\mu_k(\Theta)}\right] \\
&= \frac{(m_k^2+\sigma_k^2)\sigma^2 q_k}{2\mu_k^2(\Theta)}, \text{ when } T \text{ is large enough.}
\end{aligned}
\tag{2.133}
$$

The equations (2.131), (2.132) and (2.133) givens the variance of η_k in (2.64) in the proof of Theorem 2.3.5.

2.6.3 Error term $\hat{\beta_N} - \beta$

In this subsection, we compute the error term $\hat{\beta}_N - \beta$ in (2.88). This error term is used in the proof of convergence.

$$
\begin{aligned}
\hat{\beta}_N - \beta &= -\frac{A_{1,N}(B_{0,N}B_{2\alpha,N} - B_{\alpha,N}^2) + A_{\alpha,N}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + A_{0,N}(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}} - \beta \\
&= -\frac{(A_{1,N}+\beta B_{2,N})(B_{0,N}B_{2\alpha,N} - B_{\alpha,N}^2) + (A_{\alpha,N}+\beta B_{\alpha+1,N})(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}} \\
&\quad - \frac{(A_{0,N}+\beta B_{1,N})(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}.
\end{aligned}
\tag{2.134}
$$

We compute the numerator of (2.134) into 3 parts. Recall the SDE (2.11). It follows that

$$\begin{aligned}
&\underbrace{\frac{1}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^{\rho} v_k u_k(t) du_k(t)}_{A_{1,N}} = \underbrace{-\frac{\beta}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^{\rho} v_k^2 u_k^2(t) dt}_{-\beta B_{2,N}} \underbrace{-\frac{\theta}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^{\rho} v_k^{\alpha+1} u_k^2(t) dt}_{-\theta B_{\alpha+1,N}} \\
&\quad \underbrace{-\frac{\lambda}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^{\rho} v_k u_k^2(t) dt}_{-\lambda B_{1,N}} \underbrace{+\frac{1}{\sigma}\sum_{k=1}^{N}\int_0^T q_k^{\rho+\frac{1}{2}} v_k u_k^2(t) dw_k(t)}_{\overline{A_{1,N}}}, \qquad (2.135) \\
&A_{1,N} + \beta B_{2,N} = \overline{A_{1,N}} - \theta B_{\alpha+1,N} - \lambda B_{1,N}.
\end{aligned}$$

Similarly, we have other two terms.

$$\begin{aligned}
A_{\alpha,N} + \beta B_{\alpha+1,N} &= \overline{A_{\alpha,N}} - \theta B_{2\alpha,N} - \lambda B_{\alpha,N}. \\
A_{0,N} + \beta B_{1,N} &= \overline{A_{0,N}} - \theta B_{\alpha,N} - \lambda B_{0,N}.
\end{aligned} \qquad (2.136)$$

Then, (2.135) becomes

$$\begin{aligned}
\hat{\beta}_N - \beta = &-\frac{(\overline{A_{1,N}} - \theta B_{\alpha+1,N} - \lambda B_{1,N})(B_{0,N}B_{2\alpha,N} - B_{\alpha,N}^2)}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}} \\
&-\frac{(\overline{A_{\alpha,N}} - \theta B_{2\alpha,N} - \lambda B_{\alpha,N})(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}} \\
&-\frac{(\overline{A_{0,N}} - \theta B_{\alpha,N} - \lambda B_{0,N})(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}, \\
= &-\frac{\overline{A_{1,N}}(B_{0,N}B_{2\alpha,N} - B_{\alpha,N}^2) + \overline{A_{\alpha,N}}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + \overline{A_{0,N}}(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{\alpha+1,N}^2 B_{0,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}.
\end{aligned} \qquad (2.137)$$

It is similar to have the following error terms.

$$\hat{\lambda}_{N,T} - \lambda = \frac{\overline{D_{N,T}^{\lambda}}}{D_{N,T}}, \quad \hat{\theta}_{N,T} - \theta = \frac{\overline{D_{N,T}^{\theta}}}{D_{N,T}}, \qquad (2.138)$$

where

$$\begin{aligned}
\overline{D_{N,T}^{\lambda}} &= -\begin{vmatrix} B_{2,N} & \overline{A_{1,N}} & B_{\alpha+1,N} \\ B_{1,N} & \overline{A_{0,N}} & B_{\alpha,N} \\ B_{\alpha+1,N} & \overline{A_{\alpha,N}} & B_{2\alpha,N} \end{vmatrix}, \\
\overline{D_{N,T}^{\theta}} &= -\begin{vmatrix} B_{2,N} & B_{1,N} & \overline{A_{1,N}} \\ B_{1,N} & B_{0,N} & \overline{A_{0,N}} \\ B_{\alpha+1,N} & B_{\alpha,N} & \overline{A_{\alpha,N}} \end{vmatrix}.
\end{aligned}$$

For sufficiently large T and $\mu_k(\Theta) \sim v_k$. When $d = 2$, eigenvalues $v_k \sim k$. More precisely, we obtain that

$$\begin{aligned}
B_{p,N} &= \frac{1}{\sigma^2}\sum_{k=1}^{N}\int_0^T q_k^{-1}v_k^p u_k^2(t)dt \approx \frac{1}{\sigma^2}\mathbb{E}\left[\sum_{k=1}^{N}\int_0^T q_k^{-1}v_k^p u_k^2(t)dt\right] \\
&\approx \frac{1}{\sigma^2}\sum_{k=1}^{N} q_k^{-1}v_k^p \mathbb{E}\left[\int_0^T u_k^2(t)dt\right] \\
&\approx \frac{1}{\sigma^2}\sum_{k=1}^{N} q_k^{-1}v_k^p \left\{\frac{\sigma^2 q_k T}{2\mu_k(\Theta)} - \frac{\sigma^2 q_k}{4\mu_k^2(\Theta)}\left[1 - e^{-2\mu_k(\Theta)T}\right]\right\} \\
&= \sum_{k=1}^{N} v_k^{p-1}\left\{\frac{T}{2} - \frac{1 - e^{-2\mu_k(\Theta)T}}{4v_k}\right\} \\
&\approx \begin{cases} \frac{T}{2}\ln(N), & p = 0 \\ \frac{T}{2}N^p, & p \in (0,1) \\ \frac{T}{2}N - \ln N, & p = 1 \\ \frac{T}{2}N^p - N^{p-1}, & p \in (1,2] \end{cases}.
\end{aligned} \tag{2.139}$$

Compute the following limit used to prove the convergence.

$$\lim_{N\to\infty} \frac{2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}^2 B_{2\alpha,N} - B_{\alpha,N}^2 B_{2,N}}{B_{2,N}B_{0,N}B_{2\alpha,N} - B_{\alpha+1,N}^2 B_{0,N}}. \tag{2.140}$$

As $N \to \infty$, for $\alpha < 0.5$, the denominator

$$\begin{aligned}
&B_{0,N}\left(B_{2,N}B_{2\alpha,N} - B_{\alpha+1,N}^2\right) \approx \frac{T}{2}\ln(N)\left[\left(\frac{T}{2}N^2 - N\right)\frac{T}{2}N^{2\alpha} - \left(\frac{T}{2}N^{\alpha+1} - N^{\alpha}\right)^2\right] \\
&\approx \frac{2T^2 - T^3}{8}\ln(N)N^{2\alpha+2}.
\end{aligned} \tag{2.141}$$

Then, we split the limit into three parts.

$$\lim_{N\to\infty} \frac{2B_{1,N}B_{\alpha,N}B_{\alpha+1,N}}{B_{0,N}\left(B_{2,N}B_{2\alpha,N} - B_{\alpha+1,N}^2\right)} = \lim_{N\to\infty} \frac{2\left(\frac{T}{2}N - \ln N\right)\frac{T}{2}N^{\alpha}\left(\frac{T}{2}N^{\alpha+1} - N^{\alpha}\right)}{\frac{2T^2 - T^3}{8}\ln(N)N^{2\alpha+2}} = 0. \tag{2.142}$$

$$\lim_{N\to\infty} \frac{B_{1,N}^2 B_{2\alpha,N}}{B_{0,N}\left(B_{2,N}B_{2\alpha,N} - B_{\alpha+1,N}^2\right)} = \lim_{N\to\infty} \frac{\left(\frac{T}{2}N - \ln N\right)^2 \frac{T}{2}N^{2\alpha}}{\frac{2T^2 - T^3}{8}\ln(N)N^{2\alpha+2}} = 0, \tag{2.143}$$

$$\lim_{N\to\infty} \frac{B_{\alpha,N}^2 B_{2,N}}{B_{0,N}\left(B_{2,N}B_{2\alpha,N} - B_{\alpha+1,N}^2\right)} = \lim_{N\to\infty} \frac{\frac{T}{2}N^{2\alpha}\left(\frac{T}{2}N^2 - N\right)}{\frac{2T^2 - T^3}{8}\ln(N)N^{2\alpha+2}} = 0. \tag{2.144}$$

Thus, the limit (2.140) is equal to 0. We can get the results when $\alpha \geq 0.5$. Here are the computation of cross variations used to show the asymptotic normality for several parameters (2.80).

$$
\begin{aligned}
&\lim_{N\to 0} \langle M_{N,1}, M_{N,2}\rangle (t)\\
&= \lim_{N\to 0} N \frac{\overline{A_{1,N}}(B_{0,N}B_{2\alpha,N} - B^2_{\alpha,N}) + \overline{A_{\alpha,N}}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + \overline{A_{0,N}}(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B^2_{\alpha+1,N}B_{0,N} - B^2_{1,N}B_{2\alpha,N} - B^2_{\alpha,N}B_{2,N}}\\
&\cdot N^\alpha \frac{\overline{A_{\alpha,N}}(B_{0,N}B_{2,N} - B^2_{1,N}) + \overline{A_{1,N}}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + \overline{A_{0,N}}(B_{\alpha,N}B_{2,N} - B_{1,N}B_{\alpha+1,N})}{B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B^2_{\alpha+1,N}B_{0,N} - B^2_{1,N}B_{2\alpha,N} - B^2_{\alpha,N}B_{2,N}}\\
&= \lim_{N\to\infty} N^{\alpha+1} \frac{T_1 \cdot T_2}{(B_{2,N}B_{0,N}B_{2\alpha,N} + 2B_{1,N}B_{\alpha,N}B_{\alpha+1,N} - B^2_{\alpha+1,N}B_{0,N} - B^2_{1,N}B_{2\alpha,N} - B^2_{\alpha,N}B_{2,N})^2}.
\end{aligned}
\tag{2.145}
$$

Denote the numerator of $\hat{\beta}_N - \beta$ and the numerator of $\hat{\theta}_N - \theta$ as T_1 and T_2 respectively.

$$
\begin{aligned}
T_1 &= \overline{A_{1,N}}(B_{0,N}B_{2\alpha,N} - B^2_{\alpha,N}) + \overline{A_{\alpha,N}}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + \overline{A_{0,N}}(B_{\alpha,N}B_{\alpha+1,N} - B_{1,N}B_{2\alpha,N})\\
&\approx \overline{A_{1,N}}(\ln(N)N^{2\alpha} - N^{2\alpha}) + \overline{A_{\alpha,N}}[(N - \ln(N))N^\alpha - (N^{\alpha+1} - N^\alpha)\ln(N)]\\
&\quad + \overline{A_{0,N}}[N^\alpha(N^{\alpha+1} - N^\alpha) - (N - \ln(N))N^{2\alpha}]\\
&\approx \overline{A_{1,N}}N^{2\alpha}(\ln(N) - 1) + \overline{A_{\alpha,N}}N^\alpha(1 - \ln(N)) + \overline{A_{0,N}}N^{2\alpha}(\ln(N) - 1),
\end{aligned}
\tag{2.146}
$$

$$
\begin{aligned}
T_2 &= \overline{A_{\alpha,N}}(B_{0,N}B_{2,N} - B^2_{1,N}) + \overline{A_{1,N}}(B_{1,N}B_{\alpha,N} - B_{\alpha+1,N}B_{0,N}) + \overline{A_{0,N}}(B_{\alpha,N}B_{2,N} - B_{1,N}B_{\alpha+1,N})\\
&\approx \overline{A_{1,N}}[(N - \ln(N))N^\alpha - (N^{\alpha+1} - N^\alpha)\ln(N)] + \overline{A_{\alpha,N}}[(\ln(N))N^{2\alpha} - N^{2\alpha}]\\
&\quad + \overline{A_{0,N}}[N^\alpha(N^2 - N) - (N - \ln(N))(N^{\alpha+1} - N^\alpha)]\\
&\approx \overline{A_{1,N}}N^{\alpha+1}(1 - \ln(N)) + \overline{A_{\alpha,N}}N^{2\alpha}(\ln(N) - 1) + \overline{A_{0,N}}\ln(N)N^\alpha(N - 1)
\end{aligned}
\tag{2.147}
$$

The highest order of the denominator is $\ln(N)N^{2\alpha+2}$ derived in (2.141). Thus,

$$
\lim_{N\to\infty} \langle M_{N,1}, M_{N,2}\rangle (t) \approx \lim_{N\to\infty} \frac{-N^{4\alpha+2}(\ln(N) - 1)^2}{[\ln(N)N^{2\alpha+2}]^2} = 0. \tag{2.148}
$$

This shows the cross variation is 0, which implies the independence between $M_{N,1}$ and $M_{N,2}$. Similarly, the components of the continuous Gaussian martingale are pairwise independent.

Chapter 3

Parameter Estimation for Fractional Navier-Stokes Equations

In this chapter, we consider the parameter estimation problem for the following fractional Navier-Stokes equation with the additive noise in 2D. This equation describes the turbulence flow of a viscous, incompressible fluid. We solve this PDE numerically by the spectral method. Based on the log-likelihood function, we derive the analytic form of the coefficient of the Laplacian term and fractional Laplacian term. The fundamental idea is to find the maximizer of the log-likelihood function. Numerical simulations are presented in the last section in this chapter.

3.1 Parameter estimations on Fractional Navier-Stokes equation

3.1.1 Problem settings

In this subsection, we consider the fractional Navier-Stokes equations forced with the additive noise in 2D.

$$dU_t + (U,\nabla)U + \beta(-\Delta)u + \theta(-\Delta)^{\alpha}U + \nabla P = \sigma W_t^Q, \tag{3.1}$$

$$\nabla \cdot U = 0, \tag{3.2}$$

$$U(0) = U_0, \tag{3.3}$$

which describe the turbulence flow of a viscous, incompressible fluid. Here $U = (u_1, u_2)$ represents the velocity field and P represents the pressure.

Suppose that the flow occur over all of $\mathbb{R}^2$. We take $\mathcal{D} = [-L/2, L/2]^2$ for some $L > 0$ and the periodic boundary condition is prescribed as

$$U(\mathbf{x} + L\mathbf{e}_j, t) = U(\mathbf{x}, t), \text{ for } \mathbf{x} \in \mathbb{R}^2, t \geq 0; \int_{\mathcal{D}} U(\mathbf{x})d\mathbf{x} = 0. \tag{3.4}$$

Let's consider the spaces associated with the periodic boundary conditions. Define the space $L^2_{per}(\mathcal{D})^2$ and $H^1_{per}(\mathcal{D})^2$ to be the families of vector fields $U = U(\mathbf{x})$ which are L periodic in each direction and which is the respective subset of $L^2(\mathcal{O})^2$ and $H^1(\mathcal{O})^2$, for every open bounded set $\mathcal{O} \subset \mathbb{R}^2$. We further define the spaces H and V based on the $L^2_{per}(\mathcal{D})^2$ and $H^1_{per}(\mathcal{D})^2$ with the divergence free condition and zero mean condition as follows.

$$H := \left\{ U \in L^2_{per}(\mathcal{D})^2 : \nabla \cdot U = 0, \int_{\mathcal{D}} U(\mathbf{x})d\mathbf{x} = 0 \right\}, \tag{3.5}$$

and

$$V := \left\{ U \in H^1_{per}(\mathcal{D})^2 : \nabla \cdot U = 0, \int_{\mathcal{D}} U(\mathbf{x}) d\mathbf{x} = 0 \right\}. \tag{3.6}$$

The associated norm with H is $|U| = (U,U)^{\frac{1}{2}} = \left(\int_{\mathcal{D}} U^2(\mathbf{x}) d\mathbf{x} \right)^{\frac{1}{2}}$, and the norm of V is $||U|| = (\nabla U, \nabla U)^{\frac{1}{2}} = \left(\int_{\mathcal{D}} \nabla U(\mathbf{x}) \cdot \nabla U(\mathbf{x}) d\mathbf{x} \right)^{\frac{1}{2}}$. The linear portion of (3.1) in the Stokes operator $A_1 = -P_H \Delta$, The Leray-Hopf operator P_H is defined as the orthogonal projection of $L^2(\mathcal{D})^2$ onto H. Now let's define $A_2 = -P_H \Delta^\alpha = A_1^\alpha$, which is also unbounded. Given $\alpha > 0$, take $D(A_1^\alpha) := \left\{ U \in H : \sum_k \lambda_k^{2\alpha} |u_k|^2 < \infty. \right\}$, where $u_k = (U, \mathbf{e}_k)$. The fraction power of A_1 is defined as follows, by using the same set of orthonormal basis.

$$A_1^\alpha U := \sum_k \lambda_k^\alpha u_k \mathbf{e}_k. \tag{3.7}$$

The stochastic setting is the same as ones in the previous chapter. Given the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. A Q-cylindrical Brownian motion can be represented as

$$\sigma W^Q(t) = \sigma \sum_{|\mathbf{k}|=1}^{\infty} \sqrt{q_{\mathbf{k}}} \mathbf{e}_{\mathbf{k}}(x) W_{\mathbf{k}}(t) = \sum_{|\mathbf{k}|=1}^{\infty} \lambda_k^{-\gamma} q_k^\rho \mathbf{e}_{\mathbf{k}}(x) dW_{\mathbf{k}}(t). \tag{3.8}$$

The range of γ is determined by the value of σ. The nonlinear part in the fractional NS equation is defined as the operator

$$B(U,Y) := P_H((U \cdot \nabla)Y) = P_H(u_1 \partial_1 Y + u_2 \partial_2 Y), \text{ for } U \in V, Y \in D(A_1) \cap D(A_1^\alpha). \tag{3.9}$$

Based on the notation of operators, (3.1) can be rewritten as

$$dU + [(\beta A_1 + \theta A_2(\alpha))U + B(U,U)]dt = \sigma dW^Q. \tag{3.10}$$

3.1.2 Modified Estimations

In this section, we apply the estimators based on the specific version of the Girsanov theorem between two different probability measures. Similar to the deviation in chapter 2, we truncate U into the first N terms. Denote U^N be the projection of the solution U to the original equation (3.1) onto $H_N = P_N H$, which is isomorphic to $\mathbb{R}^N$. From (3.10), we apply the operator P_N, then we get

$$dU^k = -[(\beta A_1 + \theta A_2(\alpha))U^k + \psi_k(U)]dt + \sigma P_k W^Q, \quad U^k(0) = U_0^k, \text{ for } k = 1, 2, ..., N. \tag{3.11}$$

where $\psi_k(U) = P_k(B(U,U))$. We compute the Radon-Nykodym derivative as the likelihood ratio $L^{k,T}$. Denote $G := (P_k \sigma)^{-1}$, and $A_{\beta,\theta,\alpha} = (\beta A_1 + \theta A_2(\alpha))U^k + \psi_k(U)$, which is a vector. $A'_{\beta,\theta,\alpha}$ is the transpose of $A_{\beta,\theta,\alpha}$.

$$\begin{aligned}
L^{k,T} = \frac{d\mathbb{P}^{k,T}_{\beta,\theta}(U^k)}{d\mathbb{P}^{k,T}_{\beta_0,\theta_0}} &= \exp\Bigg(\int_0^T (A_{\beta,\theta,\alpha} - A_{\beta_0,\theta_0,\alpha_0})' G^2 dU^k(t) \\
&\qquad - \int_0^T (A'_{\beta,\theta,\alpha} G^2 A_{\beta,\theta,\alpha} - A'_{\beta_0,\theta_0,\alpha_0} G^2 A_{\beta_0,\theta_0,\alpha_0}) dt\Bigg) \\
&= \exp\Bigg(-\int_0^T [((\beta-\beta_0)A_1 + (\theta A_2(\alpha) - \theta_0 A_2(\alpha_0)))U^k]' G^2 dU^k(t) \\
&-\frac{1}{2}\int_0^T \left[(\beta A_1 + \theta A_2(\alpha))U^k\right]' G^2 \left[(\beta A_1 + \theta A_2(\alpha))U^k\right] dt \\
&+\frac{1}{2}\int_0^T \left[(\beta_0 A_1 + \theta_0 A_2(\alpha_0))U^k\right]' G^2 \left[(\beta_0 A_1 + \theta_0 A_2(\alpha_0))U^k\right] dt \\
&- \int_0^T [((\beta-\beta_0)A_1 + (\theta A_2(\alpha) - \theta_0 A_2(\alpha_0))U^k]' G^2 \psi_k(U) dt\Bigg).
\end{aligned}$$

Take logarithm and compute its maximum likelihood estimators $\beta_k, \theta_k, \alpha_k$ of parameters β, θ, α. Define the log-likelihood function $\ln(L_T(\beta,\theta,\alpha))$, which is expressed by the operator form.

$$\frac{\partial \ln(L^T)}{\partial \beta} = -\sum_{k=1}^N \int_0^T (A_1 U^k)' G^2 dU^k(t) - \sum_{k=1}^N \int_0^T (A_1 U^k)' G^2 (\beta A_1 + \theta A_2(\alpha)) U^k dt - \sum_{k=1}^N \int_0^T (A_1 U^k)' G^2 \psi_k(U) dt = 0. \tag{3.12}$$

$$\begin{aligned}
\frac{\partial \ln(L^T)}{\partial \theta} = &-\sum_{k=1}^N \int_0^T (A_2(\alpha) U^k)' G^2 dU^k(t) - \sum_{k=1}^N \int_0^T (A_2(\alpha) U^k)' G^2 (\beta A_1 + \theta A_2(\alpha)) U^k dt \\
&- \sum_{k=1}^N \int_0^T (A_2(\alpha) U^k)' G^2 \psi_k(U) dt = 0.
\end{aligned} \tag{3.13}$$

Then, we have a linear system for β and θ. We denote

$$\begin{aligned}
a_{11} &= \sum_{k=1}^N \int_0^T (A_1 U^k)' G^2 (A_1 U^k) dt, \text{ and } a_{22} = \int_0^T (A_2(\alpha) U^k)' G^2 A_2(\alpha) U^k dt, \\
a_{12} &= \sum_{k=1}^N \int_0^T (A_1 U^k)' G^2 A_2(\alpha) U^k dt = a_{21}, \\
b_1 &= \sum_{k=1}^N \int_0^T (A_1 U^k)' G^2 [dU^k(t) + \psi_k(U) dt] \text{ and } b_2 = \sum_{k=1}^N \int_0^T (A_2(\alpha) U^k)' G^2 [dU^k(t) + \psi_k(U) dt].
\end{aligned}$$

Construct (3.12) and (3.13) into a linear system and write it into a matrix form,

$$-\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} \beta \\ \theta \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \tag{3.14}$$

Solving this linear system, we have the closed form of the estimations for β and θ.

$$\begin{pmatrix} \widehat{\beta_{k,T}} \\ \widehat{\theta_{k,T}} \end{pmatrix} = -\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}^{-1} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} -\frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}^2} \\ -\frac{a_{11}b_2 - a_{12}b_1}{a_{11}a_{22} - a_{12}^2} \end{pmatrix}. \tag{3.15}$$

3.2 Numerical Implementations

In this section, we work on the numerical schemes on solving the fractional Navier Stokes equation. The main steps to derive the numerical algorithms are to apply for the fast Fourier transform to get the frequencies. Then use the frequency $u_{\mathbf{k}}$ to compute the estimations. The section split into two parts. The first part is to generate $u_{\mathbf{k}}$, and the second part is to use $u_{\mathbf{k}}$ to get $\hat{\theta}_N$ and $\hat{\alpha}_N$.

3.2.1 Numerical Implementations for $\hat{\mathbf{u}}_{\mathbf{k}}(t)$

The solutions to the fractional Navier Stokes equations have the Fourier series representation

$$U(\mathbf{x}, t) = \sum_{|\mathbf{k}|=1}^{\infty} \hat{\mathbf{u}}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{3.16}$$

$$P(\mathbf{x}, t) = \sum_{|\mathbf{k}|=1}^{\infty} \hat{p}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}}, \tag{3.17}$$

$$W^Q(t) = \sigma \sum_{|\mathbf{k}|=1}^{\infty} \sqrt{q_{\mathbf{k}}} e^{i\mathbf{k}\cdot\mathbf{x}} W_{\mathbf{k}}(t). \tag{3.18}$$

The power of $q_{\mathbf{k}}$ in (3.18) can use either $\frac{1}{2}$ or another free parameter ρ. Now, use the representation (3.16), (3.17) and (3.18) into the equation (3.10). Then we have

$$\left(\frac{d}{dt} + \beta|\mathbf{k}|^2 + \theta|\mathbf{k}|^{\alpha}\right) \hat{u}_{\mathbf{k}} + i\mathbf{k} \cdot \hat{p}_{\mathbf{k}} + \widehat{(u, \nabla) u_{\mathbf{k}}} = \sigma \sqrt{q_{\mathbf{k}}} dW_{\mathbf{k}}(t). \tag{3.19}$$

With the divergence free condition, we have

$$i\mathbf{k} \cdot \hat{u}_{\mathbf{k}} = 0. \tag{3.20}$$

Dot product with $i\mathbf{k}$ on (3.19) and we get

$$\left(\frac{d}{dt} + \beta|\mathbf{k}|^2 + \theta|\mathbf{k}|^{\alpha}\right) i\mathbf{k} \cdot \hat{u}_{\mathbf{k}} - |\mathbf{k}|^2 \cdot \hat{p}_{\mathbf{k}} + i\mathbf{k} \cdot \widehat{(u, \nabla) u_{\mathbf{k}}} = \sigma i\mathbf{k} \cdot \sqrt{q_{\mathbf{k}}} dW_{\mathbf{k}}(t). \tag{3.21}$$

Due to the divergence free condition (3.20), we obtain that

$$-|\mathbf{k}|^2 \hat{p}_{\mathbf{k}} = i\mathbf{k} \cdot [-\widehat{(u, \nabla) u_{\mathbf{k}}} + \sigma \sqrt{q_{\mathbf{k}}} dW_{\mathbf{k}}(t)] \tag{3.22}$$

Denote

$$\hat{f}_{\mathbf{k}} = -\widehat{(u, \nabla) u_{\mathbf{k}}} + \sigma \sqrt{q_{\mathbf{k}}} dW_{\mathbf{k}}(t). \tag{3.23}$$

Then, the Fourier coefficient $\hat{p}_{\mathbf{k}}$ can be expressed as

$$\hat{p}_{\mathbf{k}} = -\frac{1}{|\mathbf{k}|^2} i\mathbf{k} \cdot \hat{f}_{\mathbf{k}}. \tag{3.24}$$

Hence, the Fourier representation of the equation becomes

$$\left(\frac{d}{dt} + \beta|\mathbf{k}|^2 + \theta|\mathbf{k}|^{\alpha}\right) \hat{u}_{\mathbf{k}} = \hat{f}_{\mathbf{k}} - \mathbf{k} \frac{(\mathbf{k}, \hat{f}_{\mathbf{k}})}{|\mathbf{k}|^2}. \tag{3.25}$$

For a fixed frequency $\mathbf{k} = (k_x, k_y) = \left(\frac{2\pi i}{L}N_x, \frac{2\pi i}{L}N_y\right)$ and $i = \sqrt{-1}$ the imaginary unit, we could also set $q_{\mathbf{k}} = |\mathbf{k}|^2$. Then, we have

$$\hat{f}_{\mathbf{k}} = \begin{pmatrix} \hat{f}_{1\mathbf{k}} \\ \hat{f}_{2\mathbf{k}} \end{pmatrix} = \begin{pmatrix} -[\widehat{u_1(u_1)_x + u_2(u_1)_y}] + \sigma\sqrt{q_{\mathbf{k}}}dW_1(t) \\ -[\widehat{u_1(u_2)_x + u_2(u_2)_y}] + \sigma\sqrt{q_{\mathbf{k}}}dW_2(t) \end{pmatrix} \tag{3.26}$$

We can rewrite (3.19) into the linear system form,

$$\begin{cases} \frac{d}{dt}(\hat{u_1})_{\mathbf{k}} + \beta(k_x^2 + k_y^2)(\hat{u_1})_{\mathbf{k}} + \theta(k_x^2 + k_y^2)^{\frac{\alpha}{2}}(\hat{u_1})_{\mathbf{k}} = \hat{f}_{1\mathbf{k}} - k_x \frac{k_x \hat{f}_{1\mathbf{k}} + k_y \hat{f}_{2\mathbf{k}}}{k_x^2 + k_y^2}, \\ \frac{d}{dt}(\hat{u_2})_{\mathbf{k}} + \beta(k_x^2 + k_y^2)(\hat{u_2})_{\mathbf{k}} + \theta(k_x^2 + k_y^2)^{\frac{\alpha}{2}}(\hat{u_2})_{\mathbf{k}} = \hat{f}_{2\mathbf{k}} - k_y \frac{k_x \hat{f}_{2\mathbf{k}} + k_y \hat{f}_{2\mathbf{k}}}{k_x^2 + k_y^2} \end{cases}. \tag{3.27}$$

As mentioned in the previous section, discretize (3.27) by implicit Euler schemes for the derivatives on time. For instance, the first equation in the (3.27) can be discretized as follows. We apply Fast Fourier Transform (FFT) on $u(t, x, y)$ to get the value of $\hat{u}_n(t)$. Conversely, we can apply inverse Fourier transform to convert the frequency back to the function values, that is to say, given a vector of $\hat{u}_n(t)$ at different n with a fixed time, we can derive $u(t, x, y)$ by using the inverse Fourier transform.

For the two dimensional case, we can use the built-in function `fft2` for fast Fourier transform and `ifft2` for inverse fast Fourier transform. Suppose you have function values and N. The frequency we get from `fft2` is arranged in the order $[0, 1, 2, ..., N/2, (1 - N/2), ..., -1]$. In order to get the normalized value, when applying `fft2`, we also need to divide N^2. Similarly, applying `ifft2`, we adjust it by multiplying N^2. Introduce some notations: the Fourier operator $\mathcal{F}$ and the inverse Fourier operator $\mathcal{F}^{-1}$. Set ξ as a standard normal distributed random variables. Denote the time increment $\Delta t = t_{j+1} - t_j$.

$$\begin{aligned} \hat{f}_{1\mathbf{k}}(t_{j+1}) &= -\left[\widehat{u_1(t_{j+1})(u_1(t_{j+1}))_x + u_2(t_{j+1})(u_1(t_{j+1}))_y}\right] + \sigma\sqrt{k_x}(W_1(t_{j+1}) - W_1(t_j)) \\ &= -\mathcal{F}\left[\mathcal{F}^{-1}(\hat{u}_n(t_{j+1})) \cdot \mathcal{F}^{-1}\left(k_x \hat{u}_n(t_{j+1})\right) + \mathcal{F}^{-1}(\hat{v}_n(t_{k+1})) \cdot \mathcal{F}^{-1}\left(k_y \hat{u}_n(t_{j+1})\right)\right] + \sigma\sqrt{q_{\mathbf{k}_x}\Delta t}\xi_{1j}. \end{aligned} \tag{3.28}$$

Then,

$$\begin{aligned} \frac{\hat{u}_{1\mathbf{k}}(t_{j+1}) - \hat{u}_{1\mathbf{k}}(t_j)}{\Delta t} &= -\beta(k_x^2 + k_y^2)\hat{u}_{1\mathbf{k}}(t_{j+1}) - \theta(k_x^2 + k_y^2)^{\frac{\alpha}{2}}\hat{u}_{1\mathbf{k}}(t_{j+1}) \\ &\quad + \hat{f}_{1\mathbf{k}}(t_{j+1}) - k_x \frac{k_x \hat{f}_{1\mathbf{k}}(t_{j+1}) + k_y \hat{f}_{2\mathbf{k}}(t_{j+1})}{k_x^2 + k_y^2}. \end{aligned} \tag{3.29}$$

This is a nonlinear equation for $\hat{u}_{1\mathbf{k}}(t_{j+1})$, which can be numerically solved by iteration methods.

From (3.30), derive the form of $\hat{u}_{\mathbf{k}}^{(m+1)}(t_{j+1}) = G(\hat{u}_{\mathbf{k}}^{(m)}(t_{j+1}))$, where m is the number of iterations.

$$
\begin{aligned}
\frac{\hat{u_1}_{\mathbf{k}}^{(m+1)}(t_{j+1}) - \hat{u_1}_{\mathbf{k}}(t_j)}{\Delta t} &= -\beta(k_x^2+k_y^2)\hat{u_1}_{\mathbf{k}}^{(m+1)}(t_{j+1}) - \theta(k_x^2+k_y^2)^{\frac{\alpha}{2}}\hat{u_1}_{\mathbf{k}}^{(m+1)}(t_{j+1}) \\
&\quad + \hat{f_1}_{\mathbf{k}}^{(m)}(t_{j+1}) - k_x\frac{k_x\hat{f_1}_{\mathbf{k}}^{(m)}(t_{j+1}) + k_y\hat{f_2}_{\mathbf{k}}^{(m)}(t_{j+1})}{k_x^2+k_y^2}, \\
\hat{u_1}_{\mathbf{k}}^{(m+1)}(t_{j+1}) &= \left(1+\Delta t\beta(k_x^2+k_y^2)+\Delta t\theta(k_x^2+k_y^2)^{\frac{\alpha}{2}}\right)^{-1}\Bigg[\hat{u_1}_{\mathbf{k}}(t_j) + \Delta t\hat{f_1}_{\mathbf{k}}^{(m)}(t_{j+1}) \\
&\quad -\Delta t k_x\frac{k_x\hat{f_1}_{\mathbf{k}}^{(m)}(t_{j+1}) + k_y\hat{f_2}_{\mathbf{k}}^{(m)}(t_{j+1})}{k_x^2+k_y^2}\Bigg].
\end{aligned}
\tag{3.30}
$$

Based on (3.28) and (3.30), we can get the similar schemes for $\hat{u_2}_{\mathbf{k}}^{(m+1)}(t_{j+1})$ and then apply the fixed point iterations to get $\hat{u}_{\mathbf{k}}^{(m+1)}(t_{j+1})$, given $\hat{u}_{\mathbf{k}}^{(m)}(t_{j+1})$.

3.2.2 Numerical Implementations for $\hat{\theta}_N$ and $\hat{\alpha}_N$

Let's rewrite the PDE systems (3.27) into the dynamics form of $\hat{u_{\mathbf{k}}}$ For simplicity, let's split $\hat{f_{\mathbf{k}}}$ into two parts. Denote $\hat{f_{\mathbf{k}}} = \hat{g_{\mathbf{k}}} + \sigma\sqrt{q_{\mathbf{k}}}dW_{\mathbf{k}}(t)$, where

$$
\hat{g_{\mathbf{k}}} = \begin{pmatrix} -[u_1(u_1)_x \widehat{+\, u_2}(u_1)_y] \\ -[u_1(u_2)_x \widehat{+\, u_2}(u_2)_y] \end{pmatrix},
\tag{3.31}
$$

Based on (3.25) and (3.26) and the notation above, we have

$$
\begin{aligned}
& d\hat{u}_{\mathbf{k}} = -\left(\beta|\mathbf{k}|^2+\theta|\mathbf{k}|^\alpha\right)\hat{u}_{\mathbf{k}}dt + \hat{f}_{\mathbf{k}} - \mathbf{k}\frac{(\mathbf{k},\hat{f}_{\mathbf{k}})}{|\mathbf{k}|^2} \\
& = -\left(\beta|\mathbf{k}|^2+\theta|\mathbf{k}|^\alpha\right)\hat{u}_{\mathbf{k}}dt + \hat{g_{\mathbf{k}}}dt + \sigma\sqrt{q_{\mathbf{k}}}dW_{\mathbf{k}}(t) - \mathbf{k}\frac{(\mathbf{k},\hat{g_{\mathbf{k}}}dt+\sigma\sqrt{q_{\mathbf{k}}}dW_{\mathbf{k}}(t))}{|\mathbf{k}|^2} \\
& = -\left(\beta|\mathbf{k}|^2+\theta|\mathbf{k}|^\alpha\right)\hat{u}_{\mathbf{k}}dt + \left(\hat{g_{\mathbf{k}}} - \mathbf{k}\frac{(\mathbf{k},\hat{g_{\mathbf{k}}})}{|\mathbf{k}|^2}\right)dt + \sigma\left(\sqrt{q_{\mathbf{k}}} - \mathbf{k}\frac{(\mathbf{k},\sqrt{q_{\mathbf{k}}})}{|\mathbf{k}|^2}\right)dW_{\mathbf{k}}(t).
\end{aligned}
\tag{3.32}
$$

Though $\hat{u}_{\mathbf{k}}(t)$ is equivalent to (3.30), it is easier to write its log-likelihood function using the form of dynamics of OU process. We separate this into dt terms and $dW_{\mathbf{k}}(t)$ terms. Similarly, we can define the log-likelihood function. Denote $\mu_{\mathbf{k}}(\Theta) := \beta|\mathbf{k}|^2+\theta|\mathbf{k}|^\alpha$ and $\sigma(t,\hat{u}_{\mathbf{k}}(t)) := \sigma\left(\sqrt{q_{\mathbf{k}}} - \mathbf{k}\frac{(\mathbf{k},\sqrt{q_{\mathbf{k}}})}{|\mathbf{k}|^2}\right)$, where $\Theta = (\beta,\theta,\alpha)$. We now introduce the following notations, and the dot product in these notations is element-wise,

$$
a_{\mathbf{k},T} = \frac{1}{\sigma^2(t,\hat{u}_{\mathbf{k}})(t)}\int_0^T \hat{u}_{\mathbf{k}}(t)\cdot d\hat{u}_{\mathbf{k}}(t), \quad b_{\mathbf{k},T} = \frac{1}{\sigma^2(t,\hat{u}_{\mathbf{k}})(t)}\int_0^T (\hat{u}_{\mathbf{k}}(t))^2dt,
\tag{3.33}
$$

and

$$
c_{\mathbf{k},T} = \frac{1}{\sigma^2(t,\hat{u_{\mathbf{k}}})(t)}\int_0^T \hat{u_{\mathbf{k}}}(t)\cdot\left(\hat{g_{\mathbf{k}}} - \mathbf{k}\frac{(\mathbf{k},\hat{g_{\mathbf{k}}})}{|\mathbf{k}|^2}\right)dt,
\tag{3.34}
$$

$$\ln(L_{N,\mathbf{k}}(\beta,\theta,\alpha)) = -\sum_{|\mathbf{k}|=1}^{N}(\mu_{\mathbf{k}}(\Theta)-\mu_{\mathbf{k}}(\Theta_0))(a_{\mathbf{k},T}+c_{\mathbf{k},T})+\frac{1}{2}(\mu_{\mathbf{k}}^2(\Theta)-\mu_{\mathbf{k}}^2(\Theta_0))b_{\mathbf{k},T}. \tag{3.35}$$

Similarly, we take partial derivatives with respect to each variable β, θ and α.

$$\begin{aligned}
\frac{\partial \ln(L_{N,\mathbf{k}})}{\partial\beta} &= -\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^2(a_{\mathbf{k},T}+c_{\mathbf{k},T})+\mu_{\mathbf{k}}(\Theta)|\mathbf{k}|^2 b_{\mathbf{k},T},\\
\frac{\partial \ln(L_{N,\mathbf{k}})}{\partial\theta} &= -\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^\alpha(a_{\mathbf{k},T}+c_{\mathbf{k},T})+\mu_{\mathbf{k}}(\Theta)|\mathbf{k}|^\alpha b_{\mathbf{k},T},\\
\frac{\partial \ln(L_{N,\mathbf{k}})}{\partial\alpha} &= -\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^\alpha(\ln(|\mathbf{k}|))(a_{\mathbf{k},T}+c_{\mathbf{k},T})+\mu_{\mathbf{k}}(\Theta)|\mathbf{k}|^\alpha(\ln(|\mathbf{k}|))b_{\mathbf{k},T}.
\end{aligned}$$

We need to solve this system (3.36), (3.36) and (3.36) to find the estimation of these parameters. The linear system have the similar form as in the previous chapter, except the notations of $a_{\mathbf{k},T}, b_{\mathbf{k},T}$ and $c_{\mathbf{k},T}$. For given α, we get

$$\begin{cases}
\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^4 b_{\mathbf{k},T}\right)\beta+\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{\alpha+2} b_{\mathbf{k},T}\right)\theta = -\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^2(a_{\mathbf{k},T}+c_{\mathbf{k},T})\\
\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{\alpha+2} b_{\mathbf{k},T}\right)\beta+\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{2\alpha} b_{\mathbf{k},T}\right)\theta = -\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^\alpha(a_{\mathbf{k},T}+c_{\mathbf{k},T})
\end{cases} \tag{3.36}$$

Applying the Cramer rule, we have the approximated values of β and θ as follows, given the value of α.

$$\hat{\beta}_{N,T} = \frac{\left(-\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^2(a_{\mathbf{k},T}+c_{\mathbf{k},T})\right)\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{2\alpha} b_{\mathbf{k},T}\right)+\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{\alpha+2} b_{\mathbf{k},T}\right)\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^\alpha(a_{\mathbf{k},T}+c_{\mathbf{k},T})\right)}{\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^4 b_{\mathbf{k},T}\right)\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{2\alpha} b_{\mathbf{k},T}\right)-\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{\alpha+2} b_{\mathbf{k},T}\right)^2} \tag{3.37}$$

$$\hat{\theta}_{N,T} = \frac{\left(-\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^\alpha(a_{\mathbf{k},T}+c_{\mathbf{k},T})\right)\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{4} b_{\mathbf{k},T}\right)+\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{\alpha+2} b_{\mathbf{k},T}\right)\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^2(a_{\mathbf{k},T}+c_{\mathbf{k},T})\right)}{\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^4 b_{\mathbf{k},T}\right)\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{2\alpha} b_{\mathbf{k},T}\right)-\left(\sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^{\alpha+2} b_{\mathbf{k},T}\right)^2} \tag{3.38}$$

Next, we estimate α, given the value of β and θ. Letting the first-order derivative in α be zero gives

$$f(\alpha) := \sum_{|\mathbf{k}|=1}^{N}|\mathbf{k}|^\alpha(\ln(|\mathbf{k}|))(a_{\mathbf{k},T}+c_{\mathbf{k},T})+\left(\beta|\mathbf{k}|^2+\theta|\mathbf{k}|^\alpha\right)|\mathbf{k}|^\alpha(\ln(|\mathbf{k}|))b_{\mathbf{k},T}=0. \tag{3.39}$$

From the first order derivatives (3.36), we get a nonlinear equation (3.39) with respect to α. The solutions to this equation is the estimation of α.

3.3 Numerical Simulations

Consider the following fractional Navier-Stokes equation in two dimensional spaces.

$$\begin{cases} du_t + uu_x + vu_y - \beta(u_{xx} + u_{yy}) + \theta(-\Delta)^\alpha u + p_x = \sigma dW_t^{Q,(1)}, \\ dv_t + uv_x + vv_y - \beta(v_{xx} + v_{yy}) + \theta(-\Delta)^\alpha v + p_y = \sigma dW_t^{Q,(2)}, \\ u_x + u_y = 0. \end{cases} \tag{3.40}$$

The initial condition is

$$u(0, x, y) = v(0, x, y) = A \sin\left(\frac{2\pi\kappa x}{L}\right) \cos\left(\frac{2\pi\kappa y}{L}\right), \tag{3.41}$$

where $A = 1, \kappa = 1$ and the period $L = 2\pi$. In simulation we set the time increment $\Delta t = 0.005$. Apply the numerical schemes mentioned above. We applied the fixed point iteration in each time step. To compute $\hat{u}_n(t_{k+1})$, the initial guess in the iteration method is set to be the frequency at the time t_k, i.e. $\hat{u}_n^{(0)}(t_{k+1}) = \hat{u}_n(t_k)$. We set the maximum iteration number as 1000 and the error tolerance to be 10^{-12} for the fixed point iteration. We then have the sequence of $\hat{u}_k$ and $\hat{v}_k$. Finally, by solving the linear system (3.36), we then get the estimation of $\hat{\theta}_N$ and $\hat{\alpha}_N$ for a fixed number of Fourier modes N.

3.3.1 Estimate the coefficient of the fractional Laplacian term: θ

Set the parameters in this equation (3.40): $\beta = 1$ and $\alpha = 0.2$. The coefficient of the stochastic term σ is set to be 0.01.

We obtain the data by letting $\theta = 1$. The estimation $\hat{\theta}_N$ is 1.009509, and the error is around 9×10^{-3}. When N goes close to 100 from a small number, the estimation is getting close to 1. When N is greater than 100, the estimation of θ becomes relatively stable. Though the estimation is still volatile when N increases, it can be seen the estimated value is volatile around 1. As N becomes larger and time step Δt becomes smaller, the estimation is much closer to the exact value.

When changing another set of parameters, we will get the similar results for θ. Suppose that $\theta = 5$, $\beta = 1$ and $\alpha = 0.2$. The coefficient of the stochastic term σ is set to be 0.05. In this settings, the stochastic term has a less impact on the solutions than the other terms. See Figures 3.4 and 3.5. The estimation $\hat{\theta}_N$ is 5.080691. The error is around 8×10^{-2}, when $N = 256$.

3.3.2 Estimate the fractional order α

In this subsection, we estimate of the fractional order α, by solving a nonlinear equation (3.36) numerically. Consider the parameters $\beta = 0.01, \theta = 1, \sigma = 0.01$ in the fractional Navier-Stokes Equation. The exact value of the fractional order α is 0.8. In Matlab, we solve this one variable nonlinear equation by the built-in function `fsolve`. The initial guess is 0.9. The performance of the estimation depends on the choice of the initial guess. When the number of Fourier modes $N = 200$ and the time step $\Delta t = 10^{-3}$, the estimation $\hat{\alpha}_N = 0.8401927$. The error is around 4×10^{-2}. In Figure 3.6, although the estimation $\hat{\alpha}_N$ is oscillating, they are getting closer to the exact value.

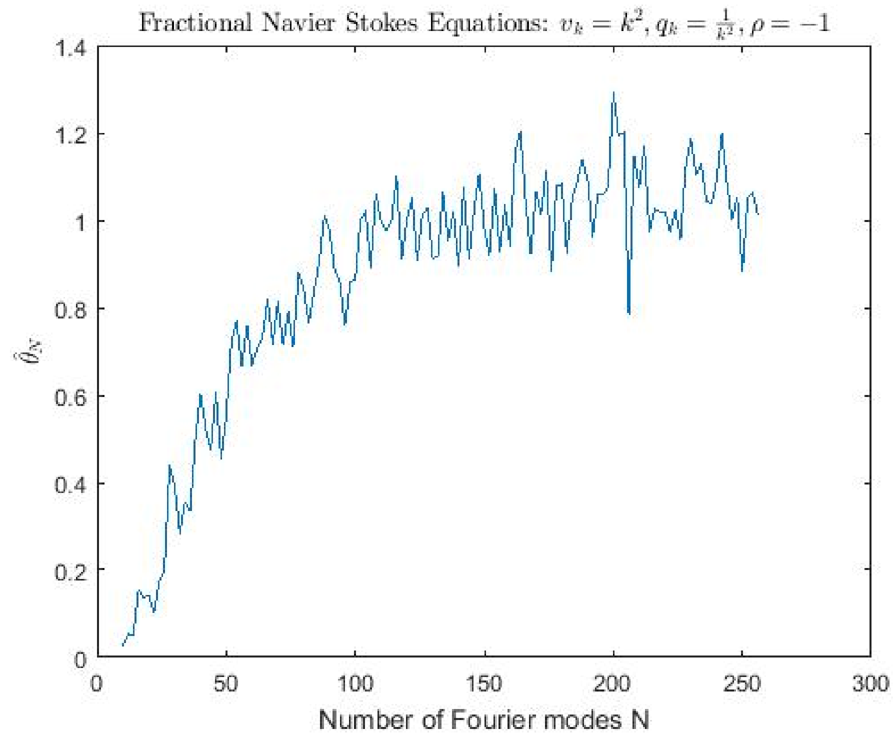

Figure 3.1: Fractional NS Equations: MLE estimation for θ for different number of Fourier modes N, $\theta = 1$

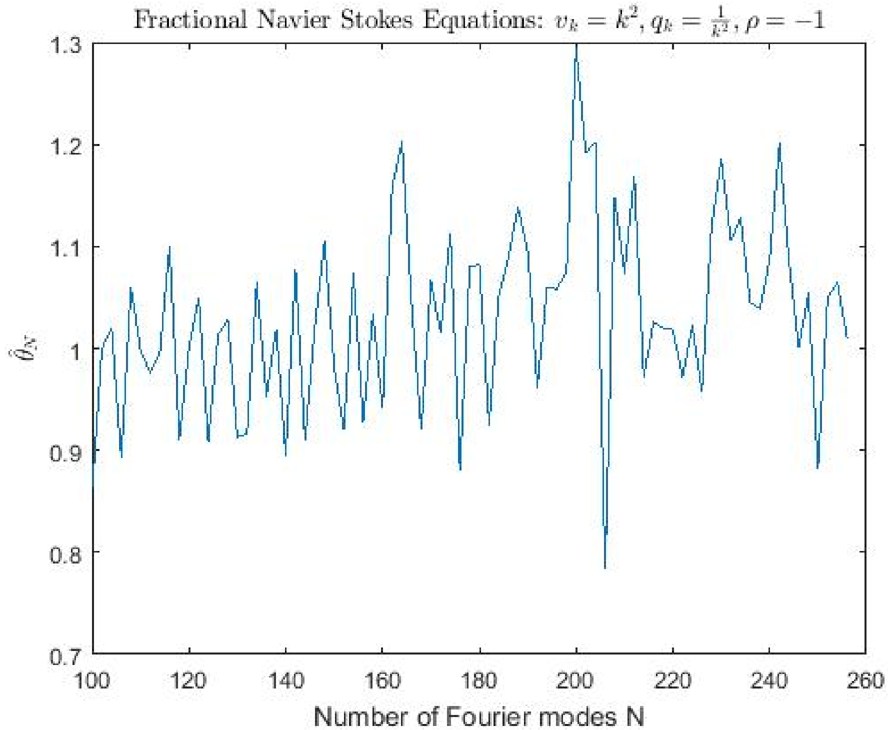

Figure 3.2: Fractional NS Equations: MLE estimation for θ for different number of Fourier modes N between 100 and 256, $\theta = 1$

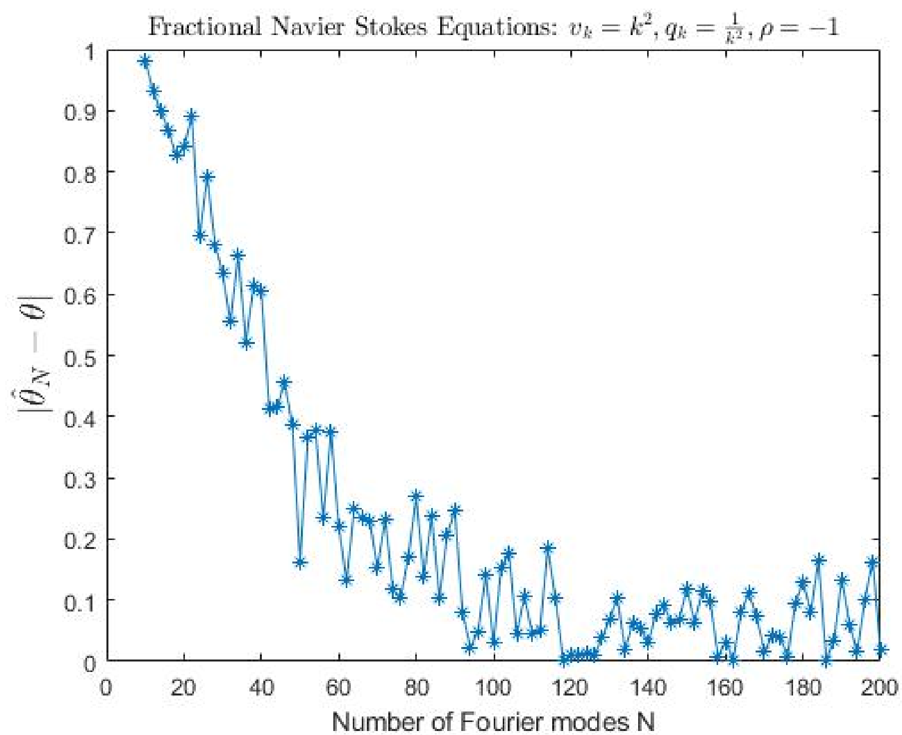

Figure 3.3: Fractional NS Equations: Error $|\hat{\theta}_N - \theta|$ for different number of Fourier modes.

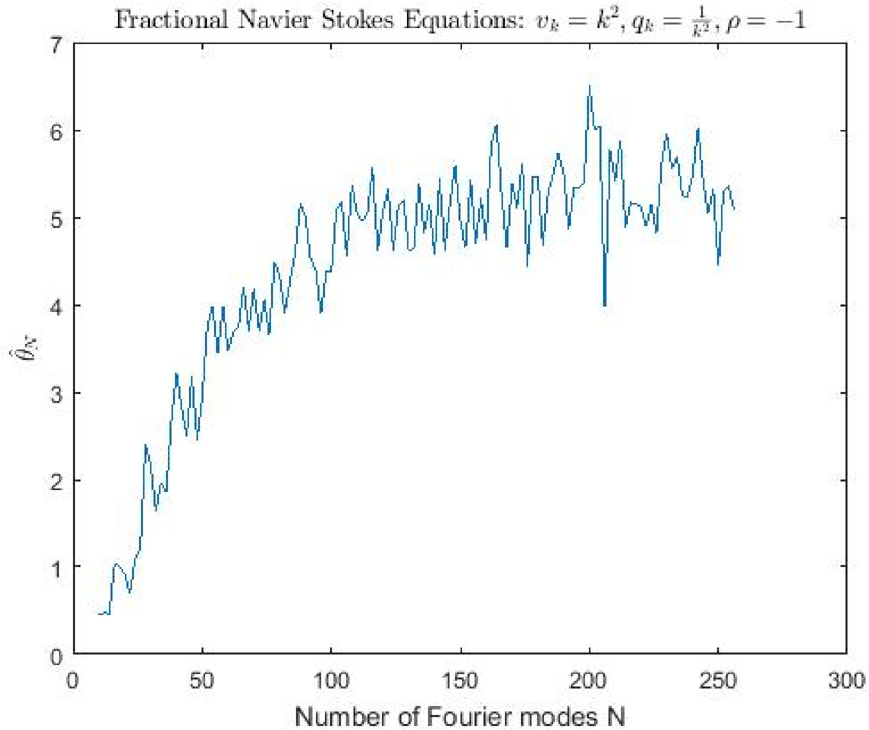

Figure 3.4: Fractional NS Equations: MLE estimation θ for different number of Fourier modes, $\theta = 5$

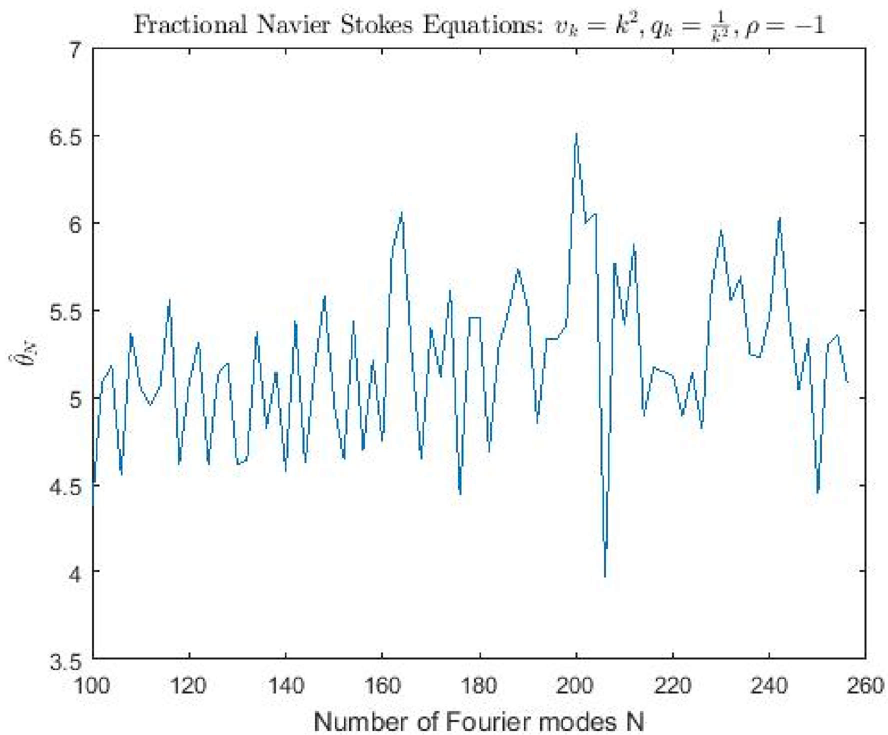

Figure 3.5: Fractional NS Equations: MLE estimation for θ MLE estimation θ for number of Fourier modes N between 100 and 256, $\theta = 5$

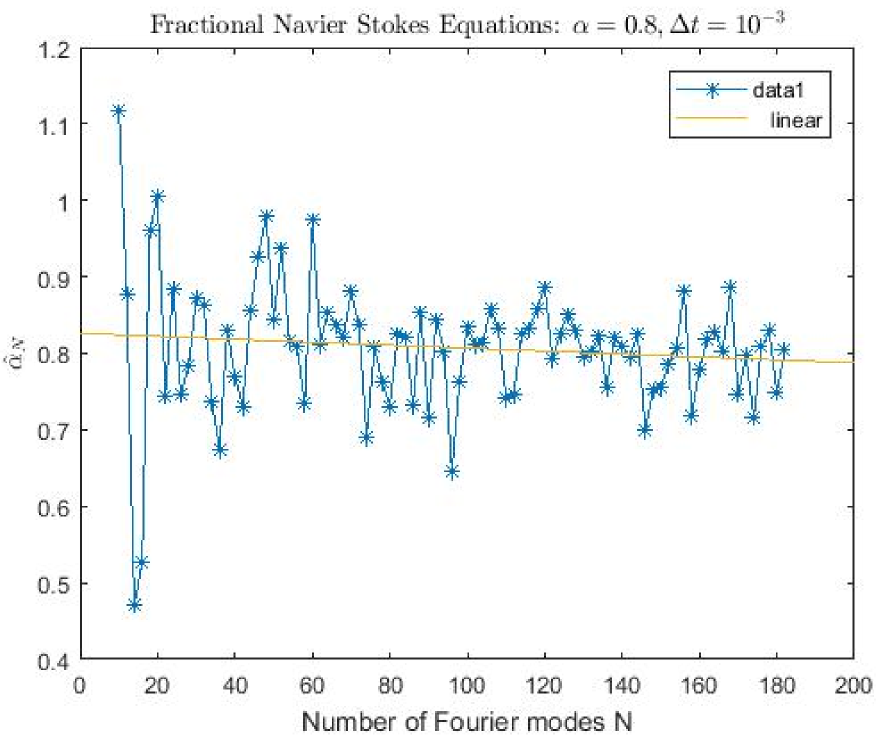

Figure 3.6: Fractional NS Equations: MLE estimation for α for different number of Fourier modes N

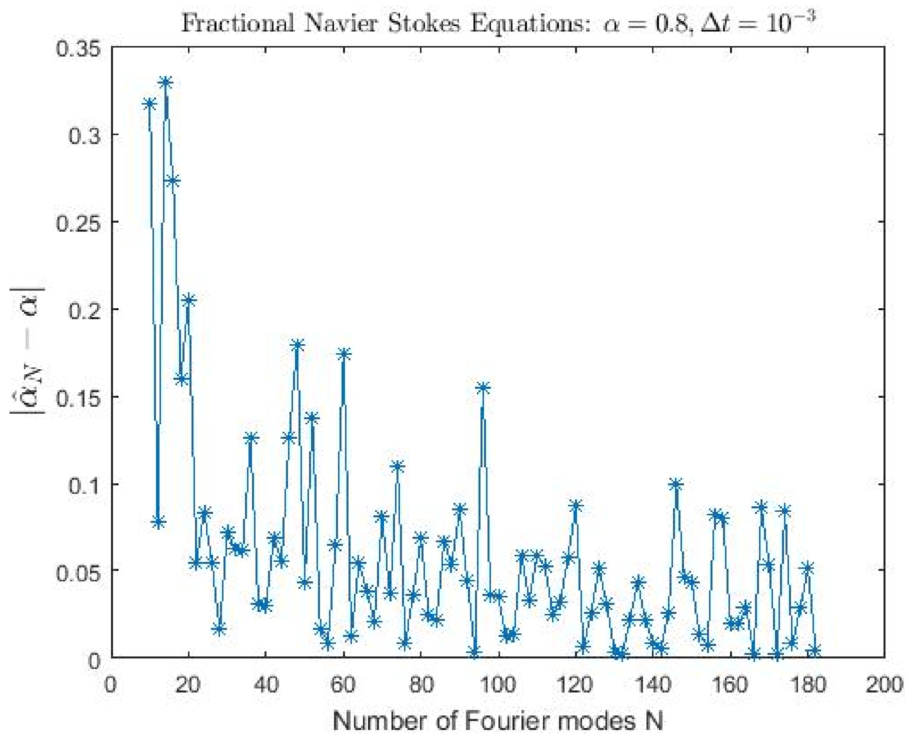

Figure 3.7: Fractional NS Equations: Error $|\hat{\alpha}_N - \alpha|$ for different number of Fourier modes N.

Chapter 4

Neural Networks

In this chapter, we work further on the numerical method of fractional Navier Stokes equation. We apply the spectral method on solving the Navier Stokes equation with the periodic boundary condition in the previous chapter. Physical-informed neural networks were first introduced in [52]. With the consideration of the fractional derivative terms, the fractional PINNs was introduced, with shorthand fPINNs. In the first section, we first review the fractional PINNs. The fractional Laplacian is defined by the directional derivative. It is approximated by using the generalized Gauss-Laguerre quadrature rule.

4.1 Fractional Physics-informed Neural Networks

Physics-informed neural networks (PINNs) was a new way to solve a partial differential equation. It convert a classic problem into an optimization problem. The solutions to the partial differential equation can be regarded as a neural network. The weights of the points $x_1, x_2, \ldots, x_n$, and bias in the neural networks are the parameters to be estimated. Let's review the PINNs or FPINNs briefly in the following subsection.

4.1.1 PINNs

To illustrate the mechanisms of PINNs, let's consider the following PDE.

$$\begin{aligned}
&\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(t,x)}{\partial x^2} + f_{BB}(x,t), \qquad (x,t) \in (0,1) \times (0,1], \\
&u(x,0) = g(x), \\
&u(0,t) = u(1,t) = 0,
\end{aligned} \tag{4.1}$$

where f_{BB} is a black-box forcing function. We approximate the solutions by $u_{NN}(x,t)$. The neural network (NN) is parametrized by the weights w and biases b.
To guarantee that it satisfy the boundary condition. We select an auxiliary function $\rho(x)$ such that $\rho(0) = \rho(1) = 0$, and construct the solutions as $\tilde{u}(x,t) = \rho(x) u_{NN}(x,t)$. With the initial condition $u(x,0) = g(x)$, the approximated solutions will be changed as

$$\tilde{u}(x,t) = t\rho(x) u_{NN}(x,t) + g(x). \tag{4.2}$$

The neural network consists of three layers: input layers, hidden layers and output layers. From

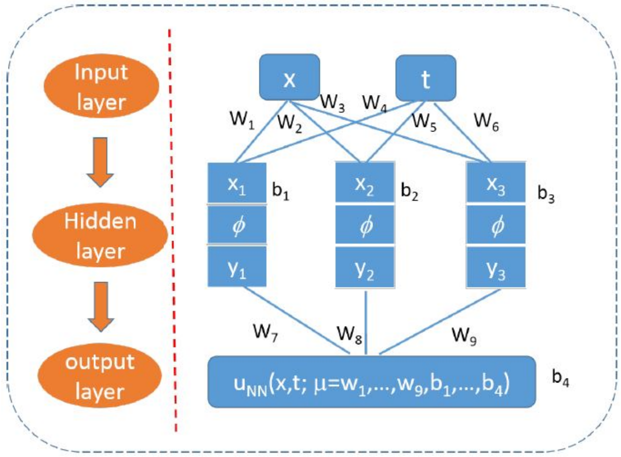

Figure 4.1: Framework of a simple neural network [52]

Figure 4.1, suppose that we have only one hidden layer. This is a fully connected neural networks with 3 neurons. For example, the inputs are one location x and one time t. It follows that

$$\begin{cases} x_1 = w_1 x + w_4 t \\ x_2 = w_2 x + w_5 t \\ x_3 = w_3 x + w_6 t \end{cases} \Longrightarrow \begin{cases} y_1 = \phi(w_1 x + w_4 t) \\ y_2 = \phi(w_2 x + w_5 t) \\ y_3 = \phi(w_3 x + w_6 t) \end{cases} \Longrightarrow \tilde{u}(\mathbf{x}, t) = w_7 y_1 + w_8 y_2 + w_9 y_3 + b_4. \tag{4.3}$$

Here, ϕ is an activation function, which is a nonlinear function. b_1, b_2, b_3 are biases in the hidden layer. b_4 is the bias in the output layer. The parameters $w_1, ..., w_9$ and $b_1, ..., b_4$ are to be optimized by minimizing the loss functions. There are several candidates for the activation function, such as sigmoid functions, Rectified Linear Unit(ReLU) functions and hyperbolic function. Due to the auto differentiation package used in the numerical simulation, hyperbolic function is recommended. We denote the parameter $\mu := (w_1, ..., w_9, b_1, ..., b_4)$. Now we applied the neural network approximation to the forward problem (4.1). The loss function is defined as follows by the mean-squared-error.

$$L(\mu) = \frac{1}{N} \sum_{k=1}^{N} \left(\frac{\partial \tilde{u}(x_k, t_k)}{\partial t_k} - \frac{\partial^2 \tilde{u}(x_k, t_k)}{\partial x^2} - f_{BB}(x_k, t_k) \right)^2, \tag{4.4}$$

where (x_k, t_k) for $k = 1, 2, ..., N$ are N training points. The training points are selected either lattice-like or scatter grid points. The scatter grid points are drawn from the quasi-random sequences. We employ auto differentiation to compute the spatial temporal derivatives in the loss function.

4.1.2 FPINNs

For the partial differential equation with fraction term, we apply FPINNs method to solve it numerically. Let's consider the following cases.

$$\begin{aligned} &\mathcal{L}(u(\mathbf{x}, t)) = f_{BB}(\mathbf{x}, t), \quad \text{for } (\mathbf{x}, t) \in \Omega \times (0, T] \\ &u(\mathbf{x}, 0) = g(\mathbf{x}), \quad \mathbf{x} \in \Omega \\ &u(\mathbf{x}, t) = 0, \quad \mathbf{x} \in \partial\Omega, \end{aligned} \tag{4.5}$$

where g is a given function as the initial condition. As mentioned above, to satisfy the initial and boundary condition, we construct the neural network solutions as above (4.2). The operator $\mathcal{L}$ can either be a linear or nonlinear operator. Here, the fractional Navier Stokes equation is

$$\mathcal{L} := \frac{\partial}{\partial t} + \beta(-\Delta) + \theta(-\Delta)^{\alpha/2} + (\cdot, \nabla)\cdot, \tag{4.6}$$

This operator can be split by two parts. One is $\mathcal{L}_{AD} := \frac{\partial}{\partial t} + \beta(-\Delta)$, which can be applied by auto differentiation and the other is $\mathcal{L}_{nonAD} := \theta(-\Delta)^{\alpha/2} + (\cdot, \nabla)\cdot$. For the non auto-differentiation operator, we applied the classic numerical approximation, such as finite difference method [48] and quadrature rule what we did here. Recall the definition of fractional laplacian operator in nonAD operator [33, 51].

Definition 4.1.1. *The directional representation of fractional laplacian is defined as follows.*

$$(-\Delta)^{\alpha/2} u(\mathbf{x}, t) = \frac{\Gamma(\frac{1-\alpha}{2})\Gamma(\frac{D+\alpha}{2})}{2(\pi)^{\frac{D+1}{2}}} \int_{||\theta||_2=1} D_\theta^\alpha u(\mathbf{x}, t) d\theta, \quad \theta \in \mathbb{R}^D, 1 < \alpha \leq 2, \tag{4.7}$$

where $||\cdot||_2$ is the L^2 norm of a vector. The symbol D_θ^α denotes the directional fractional differential operator, where θ is the differentiation direction vector.

The definition of the Riemann-Liouville directional derivative of a sufficiently properly defined function $u(\mathbf{x})$ is ($\alpha \in (1, 2]$) [33, 51]

$$D_\theta^\alpha u(\mathbf{x}) = \frac{1}{\Gamma(2-\alpha)} (\theta \cdot \nabla)^2 \int_0^{+\infty} \xi^{1-\alpha} u(\mathbf{x} - \xi\theta) d\xi, \quad \mathbf{x}, \theta \in \mathbb{R}^D, \tag{4.8}$$

where the differentiation direction θ is defined as follows.

$$\theta = \begin{cases} \cos\theta = \pm 1, \text{ where } \theta = 0 \text{ or } \pi, \text{ for } D = 1 \\ [\cos\theta, \sin\theta], \text{ where } \theta \in [0, 2\pi), \text{ for } D = 2 \\ [\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi], \text{ where } \theta \in [0, 2\pi), \phi \in [0, \pi], \text{ for } D = 3 \end{cases} \tag{4.9}$$

The symbol ∇ represents the gradient operator, and $\theta \cdot \nabla$ represents the inner product of two vectors. In one dimensional case, $(\theta \cdot \nabla)^2 = \frac{\partial^2}{\partial x^2}$. In the 2D case, $\theta \cdot \nabla = \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y}$.

Definition 4.1.2. *The general Gauss-Laguerre quadrature rule in one dimensional space is defined as follows. For $\gamma > -1$,*

$$\int_0^\infty x^\gamma e^{-x} f(x) dx \approx \sum_{i=1}^n w_i f(x_i), \tag{4.10}$$

where the weights are given in the terms of the generalized Laguerre polynomials $w_i = \dfrac{\Gamma(n+\gamma+1)x_i}{n!(n+1)^2[L_{n+1}^{(\gamma)}(x_i)]^2}$ and x_i are the i^{th} root of Laguerre polynomials $L_n^{(\gamma)} = \dfrac{x^{-\gamma}e^x}{n!}\dfrac{d^n}{dx^n}(e^{-x}x^{n+\gamma})$.

Instead of using the finite difference method and the shifted vector Grunwald-Letnikov (GL) formula in [48], we apply the generalized Gauss-Laguerre quadrature rule to approximate the directional part in (4.8) [21]. The nodes and weights of Gauss Laguerre rule are derived from the

Elhay-Kautsky method [29]. From the definition, the power of the integral variable γ should be strictly greater than -1. In our problem settings, the fractional order $\alpha/2$ of Laplacian operator is in the range of $(1/2, 1]$. It follows that $1-\alpha \in [-1, 1)$, which means that it works only for the fractional order if we want to apply for the generalized Gauss-Laguerre quadrature rule. Let's denote $\tilde{u}(x)$ to be the neural networks solutions satisfying the initial and boundary conditions. Then, the integral of $\tilde{u}$ to be estimated in 2D is

$$\int_0^\infty \xi^{1-\alpha}\tilde{u}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \xi \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}\right) d\xi = \int_0^\infty \xi^{1-\alpha}e^{-\xi}\underbrace{e^{\xi}\tilde{u}\left(\begin{bmatrix} x \\ y \end{bmatrix} - \xi \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}\right)}_{:=f(x,y;\xi,\theta)} d\xi \approx \sum_{i=1}^{N_\xi} w_i f(x,y;\xi_i,\theta), \tag{4.11}$$

where the weights $w_i := \dfrac{\Gamma(n+2-\alpha)\xi_i}{n!(n+1)^2[L_{n+1}^{(1-\alpha)}(\xi_i)]^2}$ and ξ_i are the i^{th} root of Laguerre polynomials $L_{n+1}^{(1-\alpha)} = \dfrac{\xi^{-(1-\alpha)}e^{\xi}}{(n+1)!}\dfrac{d^{n+1}}{d\xi^{n+1}}(e^{-\xi}\xi^{n+2-\alpha})$. We have estimated the integral part in the definition of fractional Laplacian (4.8). Then, in 2D, it becomes

$$D_\theta^\alpha u(\mathbf{x}) = \frac{1}{\Gamma(2-\alpha)}\sum_{i=1}^{N_\xi} w_i\left(\cos^2\theta\frac{\partial^2}{\partial x^2} + 2\cos\theta\sin\theta\frac{\partial^2}{\partial x\partial y} + \sin^2\theta\frac{\partial^2}{\partial y^2}\right)f(x,y;\xi_i,\theta), \quad \mathbf{x},\theta \in \mathbb{R}^D, \tag{4.12}$$

where f is derived in (4.11). Then, the fractional term in 2D cases becomes

$$\begin{aligned}(-\Delta)^{\alpha/2}u(\mathbf{x},t) &= C_{2,\alpha}\sum_{i=1}^{N_\xi} w_i \int_{||\theta||_2=1}\left(\cos^2\theta\frac{\partial^2}{\partial x^2} + 2\cos\theta\sin\theta\frac{\partial^2}{\partial x\partial y} + \sin^2\theta\frac{\partial^2}{\partial y^2}\right)f(x,y;\xi_i,\theta)d\theta, \\ &\approx C_{2,\alpha}\sum_{j=1}^{N_\theta} J_2\nu_j\sum_{i=1}^{N_\xi} w_i\left(\cos^2\theta_j\frac{\partial^2}{\partial x^2} + 2\cos\theta_j\sin\theta_j\frac{\partial^2}{\partial x\partial y} + \sin^2\theta_j\frac{\partial^2}{\partial y^2}\right)f(x,y;\xi_i,\theta_j), \text{ for } \theta_j \in (0, 2\pi].\end{aligned} \tag{4.13}$$

where the constant $C_{2,\alpha} = \dfrac{\Gamma(\frac{1-\alpha}{2})\Gamma(1+\alpha/2)}{2(\pi)^{\frac{3}{2}}\Gamma(2-\alpha)}$. The determinant of the Jacobian matrix is $J_2 = 1$ for the polar-Cartesian coordinate transformation. ν_j are Gauss-Legendre quadrature weights and θ_j are the corresponding quadrature nodes. (4.11) can be simplified in one dimensional case

$$\begin{aligned}&\int_0^\infty \xi^{1-\alpha}[u(x+\xi) + u(x-\xi)]d\xi \\ &= \int_0^\infty \xi^{1-\alpha}e^{-\xi}e^{\xi}[u(x+\xi) + u(x-\xi)]d\xi \approx \sum_{i=1}^{N_\xi(x)} w_i e^{\xi_i}[u(x+\xi_i) + u(x-\xi_i)].\end{aligned} \tag{4.14}$$

(4.13) in one dimensional case becomes

$$(-\Delta)^\alpha u(x,t) \approx \frac{1}{2\cos\left(\frac{\pi\alpha}{2}\right)}\sum_{i=1}^{N_\xi(x)} w_i e^{\xi_i}\left[\frac{\partial^2 u}{\partial x^2}\Big|_{(x+\xi_i)} + \frac{\partial^2 u}{\partial x^2}\Big|_{(x-\xi_i)}\right]. \tag{4.15}$$

Again, ξ_i and w_i are the nodes and weights of the generalized Gauss Laguerre quadrature rule. $N_\xi(x_j)$ is the number of nodes ξ used for x_j. Similarly, x_j and ν_j are the nodes and weights of the Gauss Legendre rule. The differential operator can be applied in auto-differentiation package in python. We want to optimize the weights and bias in the neural networks. These three parts in (4.13) computed separately. The loss function of FPINNs for the forward problem is defined as

$$L_{FW}(\mu) = \frac{1}{|N|} \sum_{(\mathbf{x},t)\in N} [\mathcal{L}_{AD}^{N_\xi,N_\theta}\{\tilde{u}(\mathbf{x},t)\} - f_{BB}(\mathbf{x},t)]^2, \tag{4.16}$$

where N is the set of all training points $N \subset \Omega \times [0,T]$ and $|N|$ denotes the number of training points. On the other hand, FPINNs can also be applied in the inverse problem to find the estimation of the parameters α, β and θ in the Navier Stokes equation. Given the data at $t = T$,

$$u(\mathbf{x},t) = h_{BB}(\mathbf{x},t), \quad (\mathbf{x},t) \in \Omega \times \{t = T\}. \tag{4.17}$$

The loss function for this inverse problem is

$$\begin{aligned} L_{INV}(\mu) = \frac{w_1}{|N_1|} &\sum_{(\mathbf{x},t)\in N_1} [\mathcal{L}_{AD}^{N_\xi,N_\theta}\{\tilde{u}(\mathbf{x},t)\} - f_{BB}(\mathbf{x},t)]^2 \\ + \frac{w_2}{|N_2|} &\sum_{(\mathbf{x},T)\in N_2} [\tilde{u}(\mathbf{x},T) - h_{BB}(\mathbf{x},T)]^2. \end{aligned} \tag{4.18}$$

N_1 and N_2 are two distinct sets of data. w_1 and w_2 are two preselected weights. This loss function is a weighted sum of two parts. The first part is a measurement on training a neural network and the second part is to minimize the mean-squared error between the solutions and observations.

4.2 Numerical examples

In this section, we demonstrate the numerical simulations for the fractional laplacian equation using FPINNs with the generalized Gauss Laguerre rule.

$$(-\Delta)^{\frac{\alpha}{2}} u(x) + \beta \Delta u(x) = f(x) \qquad x \in (0,1), \tag{4.19}$$

with the boundary condition $u(0) = u(1) = 0$. Suppose that the forcing term is [65]

$$\begin{aligned} f(x) = \frac{1}{2\cos(\pi\alpha/2)} \Bigg[&\frac{\Gamma(4)}{\Gamma(4-\alpha)}(x^{3-\alpha} + (1-x)^{3-\alpha}) - \frac{3\Gamma(5)}{\Gamma(5-\alpha)}(x^{4-\alpha} + (1-x)^{4-\alpha}) \\ + &\frac{3\Gamma(6)}{\Gamma(6-\alpha)}(x^{5-\alpha} + (1-x)^{5-\alpha}) - \frac{\Gamma(7)}{\Gamma(7-\alpha)}(x^{6-\alpha} + (1-x)^{6-\alpha}) \Bigg]. \end{aligned} \tag{4.20}$$

Suppose that the coefficient of the Laplacian β is 0, the exact solutions to (4.19) is $u(x) = x^3(1-x)^3$. Our goal is to estimate the parameter β. From (4.15), in order to evaluate the value $u(x_j)$, we need to have the approximated solution $\tilde{u}(x_j) = x(1-x)u_{NN}(x)$ and evaluations at some supported points additionally. We construct it into a matrix form. The points are arranged in order. The first part are the points $\tilde{u}(x_j)$ to be evaluated, and the remaining parts are the corresponding supporting points of $\tilde{u}(x_j)$. The following matrix is to evaluate $\tilde{u}(x_j)$

$$\frac{1}{2\cos\left(\frac{\pi\alpha}{2}\right)}\begin{bmatrix}0 & \dots & \nu_j w_1 e^{\xi_1} & \dots & \nu_j w_{N_{j,\xi}} e^{\xi_{N_{j,\xi}}} & \nu_j w_1 e^{\xi_1} & \dots & \nu_j w_{N_{j,\xi}} e^{\xi_{N_{j,\xi}}} & \dots & 0\end{bmatrix}\begin{bmatrix}\tilde{u}_{xx}(x_j)\\ \tilde{u}_{xx}(x_j+\xi_1)\\ \vdots\\ \tilde{u}_{xx}(x_j+\xi_{N_{1,\xi}})\\ \tilde{u}_{xx}(x_j-\xi_1)\\ \vdots\\ \tilde{u}_{xx}(x_j-\xi_{N_{1,\xi}})\end{bmatrix} \tag{4.21}$$

Extend to the evaluation at different training points x_j. The vector is arranged as follows.

$$\tilde{\mathbf{u}} := \begin{bmatrix}\tilde{u}_{xx}(x_1)\\ \vdots\\ \tilde{u}_{xx}(x_N)\\ \tilde{u}_{xx}(x_1+\xi_1)\\ \vdots\\ \tilde{u}_{xx}(x_1+\xi_{N_{1,\xi}})\\ \tilde{u}_{xx}(x_1-\xi_1)\\ \vdots\\ \tilde{u}_{xx}(x_1-\xi_{N_{1,\xi}})\\ \vdots\\ \tilde{u}_{xx}(x_N+\xi_1)\\ \vdots\\ \tilde{u}_{xx}(x_N-\xi_{N_{N,\xi}})\end{bmatrix}_{\left(N+2\sum\limits_{j=1}^{N}N_{j,\xi}\right)\times 1} \tag{4.22}$$

Then construct a corresponding sparse coefficient matrix $\mathbf{A}$. The loss function of FPINNs evaluated at $\mathbf{x} = [x_1, ..., x_N]^T$ is defined as

$$L(\mu) = MSE\left(\mathbf{A}\tilde{\mathbf{u}} - f(\mathbf{x})\right). \tag{4.23}$$

Our goal is to find the estimation β such that it minimizes the loss function.(4.23). Unfortunately, we have no numerical results on this example.

Chapter 5

Positivity-preserving Methods for SDEs with Non-Lipschitz coefficients

In this chapter, we consider positivity-preserving explicit schemes for one-dimensional nonlinear stochastic differential equations. The drift coefficients satisfy the one-sided Lipschitz condition, and the diffusion coefficients are Hölder continuous. To control the fast growth of moments of solutions, we introduce several explicit schemes including the tamed and truncated Euler schemes. The fundamental idea is to guarantee the non-negativity of solutions. The proofs rely on the boundedness for negative moments and exponential of negative moments. We present several numerical schemes for a modified Cox-Ingersoll-Ross model and a two-factor Heston model and demonstrate their half-order convergence rate.

5.1 Introduction

In many applications of stochastic differential equations (SDEs), coefficients of SDEs grow nonlinearly. When these SDEs with coefficients of superlinear growth are solved using numerical methods, explicit numerical schemes usually fail to converge in the sense of mean-square and moments, e.g., [24, 41]. The failure of explicit schemes lies in the fact that the moments of numerical solutions explode. To control the fast explosion of these moments of solutions to SDEs with coefficients of superlinear growth, several approaches have been proposed when the coefficients are assumed to be *one-sided Lipschitz continuous*: a) Euler schemes with variable step sizes, b) implicit schemes [20, 22, 39, 38, 60] instead of explicit schemes, c) balanced implicit schemes, d) tamed schemes [23, 24, 26, 27, 55, 54, 58, 60, 62, 66, 64], e) truncated Euler schemes [34, 37, 42], f) semi-discrete schemes [19]. A review of recent literature on this topic is presented in [25].

SDEs with coefficients of sublinear growth are also used in applications such as in finance. For example, the Cox-Ingersoll-Ross (CIR) model and the Heston model have a *Hölder continuous* diffusion coefficients. For these models, most explicit schemes converge very slowly. It was proved in [18] that the Euler scheme for the CIR model converges at the rate of $1/\ln(n)$, where n is the number of time steps. See a similar conclusion in [44] for an SDE with coefficients of both one-sided Lipschitz and Hölder continuity. A standard solution to improve the convergence order is to ensure the *positivity* of numerical solutions such that the negative moments of numerical solutions are bounded, see e.g., [5, 7, 30].

We consider the following one-dimensional Itô SDE

$$X_t = X_0 + \int_0^t b(X_s)\,ds + \int_0^t \sigma X_s^{\gamma}\,dW_s, \quad \frac{1}{2} < \gamma < 1. \tag{5.1}$$

where $X_0 \geq 0$ and $\sigma \in \mathbb{R}$ are given constants and $(W_t)_{t\geq 0}$ is a standard one-dimensional Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here, the drift coefficient is one-sided Lipschitz continuous and grows polynomially at infinity. The diffusion coefficient is Hölder continuous when $1/2 < \gamma < 1$, see Section 5.2 for detailed assumptions. When $b(x) = 1 - x^3$, it is shown in [44] that the convergence order is $\gamma - 1/2$ if $1/2 < \gamma < 1$. When γ is close to $1/2$, the convergence order is low. Moreover, if the solution is non-negative, the numerical solution of forward Euler scheme is not positivity-preserving [53]. Inspired by the work in [5], we take the absolute value at each time step so that the numerical solution is nonnegative if the solution to (5.1) is nonnegative.

To the best of our knowledge, explicit schemes of half-order convergence for Equation (5.1) with a non-Lipschitz drift coefficient have not been investigated. In literature, positivity-preserving explicit schemes are discussed either for SDEs with coefficients of superlinear growth such as in [58] or for SDEs with Hölder continuous diffusion coefficients and Lipschitz continuous drift coefficients, see a review in [30]. Also, positivity of numerical solutions by the forward Euler scheme for SDEs with superlinearly growing locally Lipschitz coefficients is discussed in [53].

For Equation (5.1), we combine the tamed/balanced or projection schemes and positivity-preserving scheme proposed in [5], see (5.9). In our setting, the tamed schemes are sufficient to control the fast growth of moments of solutions caused by the drift coefficients of superlinear growth as in [44] while the positivity of numerical solution is essential to recover the half-order convergence as in numerical schemes for Hölder continuous coefficients, see e.g., [2, 3, 5, 19],

To prove the half-order convergence of our schemes, we follow ideas of the proofs in [5], where the drift coefficients are assumed to be Lipschitz continuous. However, this assumption is violated in our setting, and two technical issues arise. *First*, the drift coefficients with polynomial growth can not be bounded linearly. For locally Lipschitz drift coefficients, we only have a Lipschitz-like condition for the drift in the numerical scheme and the drift in absolute value can be bounded by $(\Delta t)^{-1}|x|$ or $(\Delta t)^{-\frac{1}{2}}|x|$ plus a constant proportional to some positive powers of Δt, where Δt is the time step size. *Second*, due to the nonlinear growth of drift coefficients, the boundedness of exponential moment of solutions cannot be proved via a known comparison theorem and by comparing with solutions to a CIR model as in [5]. We prove the bounded exponential moment using a surprisingly simple approach, which can significantly simplify the proof of bounded exponential moments in [5]. See details in Lemma 5.3.11.

The main contributions of this chapter are listed as follows.

• We develop explicit schemes preserving positivity of solutions to SDEs with non-globally Lipschitz drift and Hölder continuous diffusion coefficients.

• We present five explicit positivity-preserving schemes. These schemes are modified symmetrized Euler schemes. We discuss several choices for non-globally Lipschitz drift among the state-of-the-art tamed schemes and truncation schemes.

• We present several numerical examples using these schemes and make comparisons in computational performance and convergence.

In this chapter, we only consider explicit schemes because the additional computational effort is required for implicit schemes to solve nonlinear equations at each time step. Moreover, it is unclear whether implicit schemes preserve positivity, although positivity-preserving is possible in some special cases, e.g., backward (implicit) Euler scheme in [28, 39, 43].

The rest of the chapter is organized as follows. In Section 5.2, we introduce the assumptions on the coefficients of (5.1) and present positivity-preserving explicit schemes. We also state our main result of this chapter on the half-order strong convergence. In Section 5.3, we present all proofs of bounded moments and the strong convergence order. We provide some numerical experiments and verify our prediction in Section 4.

5.2 Preliminaries and numerical scheme

Let $X_t, 0 \le t \le T$, be a strong solution of (5.1). We assume the following conditions.

Assumption 5.2.1. *(i) The initial condition is such that*

$$\mathbb{E}[|X_0|^{2p}] \le K < \infty, \quad \textit{for all} \;\; p \ge 1. \tag{5.2}$$

(ii) There is a positive constant β such that

$$(x-y)(b(x)-b(y)) \le \beta|x-y|^2. \tag{5.3}$$

(iii) There exist $\mathcal{K}_1 > 0$ and $\alpha \ge 1$ such that for $t \in [0,T]$

$$|b(x)-b(y)|^2 \le \mathcal{K}_1(1+|x|^{2\alpha-2}+|y|^{2\alpha-2})|x-y|^2, \quad x,y \in \mathbb{R}. \tag{5.4}$$

(iv) The function $b(x)$ is positive when $x=0$, i.e., $b(0)>0$.

We note that (5.3) implies that there exists a constant β_1 such that

$$xb(x) \le xb(0) + \beta x^2 \le \beta_1(1+x^2), \; t \in [0,T], \quad x \in \mathbb{R}. \tag{5.5}$$

Also, this inequality implies that if $x \ge 0$,

$$b(x) \le b(0) + \beta x, \tag{5.6}$$

which is also observed in [58]. By Itô's formula, there exists a constant $\mathcal{C} > 0$ such that

$$\mathbb{E}[|X_t|^p] \le \mathcal{C}(1+\mathbb{E}[|X_0|^p]) \quad t \in [0,T], \quad p \ge 2. \tag{5.7}$$

From (5.4) and letting $y=0$, we have $|b(x)-b(0)| \le \sqrt{\mathcal{K}_1}(1+|x|^{2\alpha-2})^{1/2}\,|x|$ and $|b(x)| \le \sqrt{\mathcal{K}_1}(|x|+|x|^\alpha)+b(0)$. Also, there exists a constant $\mathcal{K}_2 > 0$ such that

$$|b(x)|^2 \le \mathcal{K}_2(1+|x|^{2\alpha}) \quad x \in \mathbb{R}. \tag{5.8}$$

5.2.1 Numerical schemes

Let $N\Delta t = T$ and $t_k = k\Delta t$ where $k = 0,1,...,N$. Define Y_{t_k} by the following equation

$$Y_{t_{k+1}} = |Y_{t_k} + \bar{b}(Y_{t_k})\Delta t + \sigma Y_{t_k}^{\gamma}(W_{t_{k+1}} - W_{t_k})|, \, k = 0,1,...,N-1, \tag{5.9}$$

where $Y_0 = X_0 > 0$ and we consider five cases of $\bar{b}(x)$:

$$
\begin{aligned}
a) \qquad & \bar{b}(y) &=& \frac{b(y)}{1+|y|^{\alpha}(\Delta t)^{1/2}}, & (5.10)\\
b) \qquad & \bar{b}(y) &=& b(\hat{y}), \quad \text{where } \hat{y} = \min(1,(\Delta t)^{-\eta}|y|^{-1})y, \quad 2\eta(\alpha-1)=1, & (5.11)\\
c) \qquad & \bar{b}(y) &=& \frac{b(y)}{1+|y|^{\alpha}\Delta t}, & (5.12)\\
d) \qquad & \bar{b}(y) &=& \frac{b(y)}{1+|b(y)|\Delta t}, & (5.13)\\
e) \qquad & \bar{b}(y) &=& \frac{\tanh(b(y)\Delta t)}{\Delta t}. & (5.14)
\end{aligned}
$$

We remark that these treatments of $b(x)$ to control the growth of numerical solution are not new. In literature, Cases a),c),d),e) are from tamed Euler schemes for SDEs with coefficients of superlinear growth. Cases a),d),e) are from [55], [26], [64], respectively. Case b) is from the projection Euler scheme in [6] (c.f. [34]).

Define $\theta(t) = \sup_{k\in\{1,2,\ldots,N\}}\{t_k : t_k \le t\}$. Let

$$Z_t = Y_{\theta(t)} + \bar{b}(Y_{\theta(t)})(t-\theta(t)) + \sigma Y^{\gamma}_{\theta(t)}(W_t - W_{\theta(t)}), \tag{5.15}$$

the numerical solution can be obtained by $Y_t = |Z_t|$ for $t \in [0,T]$.

Theorem 5.2.2. *Suppose that Assumption 5.2.1 holds and $X_t > 0$ when $t \in [0,T]$. Suppose also that $X_0 = x > 0$. Then there exists a positive constant C depending on σ, p and T but not on Δt such that*

$$\sup_{t\in[0,T]} \mathbb{E}[|X_t - Y_t|^{2p}] \le C\Delta t^{p}, \quad p \ge 1.$$

5.3 Proof of Theorem 5.2.2

In this section, we present the proof of our main result Theorem 5.2.2. Before the proof, we show the Lipschitz-like condition for the modified drift coefficient $\bar{b}$ from (5.10) -(5.14) in Section 5.3.1. We then prove the moment bounds of solutions to the numerical scheme (5.9) with (5.10)-(5.14) in Section 5.3.2. In Section 5.3.3, we prove that expectations of negative powers of the solution to (5.1) are bounded when the solution is nonnegative. At last, we present the proof of Theorem 5.2.2 in Section 5.3.4.

5.3.1 Properties of the modified drift coefficient $\bar{b}(x)$

Let's first present some properties of the modified drift coefficients.

Lemma 5.3.1. *Under Assumption 5.2.1, we have for* (5.10) *that*

$$\begin{aligned}
\bar{b}(x) &\le b(0)+\beta x, \quad x>0, && (5.16)\\
\left|\bar{b}(x)-\bar{b}(0)\right| &\le \frac{|b(x)-b(0)|}{1+|x|^{\alpha}\sqrt{\Delta t}}+\frac{b(0)|x|^{\alpha}\sqrt{\Delta t}}{1+|x|^{\alpha}\sqrt{\Delta t}}, && (5.17)\\
\left|\bar{b}(x)\right| &= \frac{|b(x)|}{1+|x|^{\alpha}\sqrt{\Delta t}} \le C\min(\frac{1}{\sqrt{\Delta t}},1+|x|^{\alpha}), && (5.18)\\
\left|\bar{b}(x)-b(x)\right| &\le |b(x)|\,|x|^{\alpha}\sqrt{\Delta t} \le C(1+|x|^{2\alpha})\sqrt{\Delta t}, && (5.19)\\
\bar{b}(x) &\ge -\sqrt{\mathcal{K}_1}(|x|+|x|^{\alpha})+\tilde{b}, \quad \tilde{b}=\frac{b(0)}{1+|x|^{\alpha}\sqrt{\Delta t}}. && (5.20)
\end{aligned}$$

Proof. By the assumption (5.3), we have $xb(x)\le b(0)x+\beta\,|x|^2$. When $x>0$, $\bar{b}(x)=\frac{b(x)}{1+|x|^{\alpha}\sqrt{\Delta t}}\le b(0)+\beta x$ which is (5.16).

The inequality (5.17) follows from (5.4) and $b(0)>0$ as we have

$$\left|\bar{b}(x)-\bar{b}(0)\right| = \left|\frac{b(x)}{1+|x|^{\alpha}\sqrt{\Delta t}}-b(0)\right| = \left|\frac{b(x)-b(0)}{1+|x|^{\alpha}\sqrt{\Delta t}}-\frac{b(0)\,|x|^{\alpha}\sqrt{\Delta t}}{1+|x|^{\alpha}\sqrt{\Delta t}}\right|.$$

By (5.17), we have

$$\left|\bar{b}(x)\right| \le \left|\bar{b}(x)-\bar{b}(0)\right|+\bar{b}(0)\le \frac{\sqrt{\mathcal{K}_1}(1+|x|^{\alpha-1})|x|}{1+|x|^{\alpha}\sqrt{\Delta t}}+\frac{b(0)|x|^{\alpha}\sqrt{\Delta t}}{1+|x|^{\alpha}\sqrt{\Delta t}}+b(0).$$

Then we obtain the upper bound (5.18). The inequality (5.19) follows from the definition of $\bar{b}$ (5.10) and (5.8). The inequality (5.20) follows from (5.17) and (5.4). □

Lemma 5.3.2. *Consider $\alpha>1$. Under Assumption 5.2.1, we have for* (5.10) *that when $\sqrt{\Delta t}\le((\frac{\alpha-1}{\alpha+1})^{(\alpha-1)}+1)^{-1}$,*

$$\begin{aligned}
\sqrt{\Delta t}\left|\bar{b}(x)-\bar{b}(0)\right| &\le \sqrt{\mathcal{K}_1}\,|x|+\frac{b(0)|x|^{\alpha}\sqrt{\Delta t}}{1+|x|^{\alpha}\sqrt{\Delta t}}\sqrt{\Delta t},\\
\sqrt{\Delta t}\bar{b}(x) &\ge -\sqrt{\mathcal{K}_1}\,|x|+\frac{b(0)\sqrt{\Delta t}}{1+|x|^{\alpha}\sqrt{\Delta t}}.
\end{aligned}$$

Proof. We have from (5.17) that

$$\sqrt{\Delta t}\left|\bar{b}(x)-\bar{b}(0)\right| \le \frac{\sqrt{\mathcal{K}_1}(1+|x|^{\alpha-1})|x|}{1+|x|^{\alpha}\sqrt{\Delta t}}\sqrt{\Delta t}+\frac{b(0)|x|^{\alpha}\sqrt{\Delta t}}{1+|x|^{\alpha}\sqrt{\Delta t}}\sqrt{\Delta t},$$

It only requires to show that

$$\frac{(1+|x|^{\alpha-1})|x|}{1+|x|^{\alpha}\sqrt{\Delta t}}\sqrt{\Delta t}\le|x|\,, \text{ when } \sqrt{\Delta t}\le((\frac{\alpha-1}{\alpha+1})^{(\alpha-1)}+1)^{-1}.$$

Denote $g(z)=z(1-\sqrt{\Delta t})+z^{\alpha}(z-1)\sqrt{\Delta t}$. Then it suffices to show that $g(|x|)\ge 0$. Observe that $g(z)>0$ whenever $\sqrt{\Delta t}\le 1$ and $z\ge 1$. We then focus on the case $0<z<1$. Here

$g'(z) = (1-\sqrt{\Delta t}) + z^{\alpha-1}\sqrt{\Delta t}[(\alpha+1)z-\alpha]$ and $g''(z) = \alpha\sqrt{\Delta t}z^{\alpha-2}(z(\alpha+1)-(\alpha-1))$. The only nontrivial root of $g''(z^*)$ is $z^* = \frac{\alpha-1}{\alpha+1}$. Then $g''(z) < 0$ when $0 < z < z^*$ and $g''(z) > 0$ when $z > z^*$. Thus by $g'(0) \geq 1-\sqrt{\Delta t}$ ($\alpha > 1$) and $g'(1) = 1 > 0$, we have, for any $0 < z < 1$ and $z \neq z^*$,

$$g'(z) > g'(z^*) = (1-\sqrt{\Delta t}) + (z^*)^{\alpha-1}[(\alpha+1)z^*-\alpha]\sqrt{\Delta t} = 1-(1+(z^*)^{\alpha-1})\sqrt{\Delta t}.$$

When $\sqrt{\Delta t} \leq [(\frac{\alpha-1}{\alpha+1})^{(\alpha-1)}+1]^{-1}$, i.e., $(1+(z^*)^{\alpha-1})\sqrt{\Delta t} \leq 1$, we have $g'(z) \geq 0$ and thus $g(z) \geq g(0) = 0$. When $\alpha = 1$, $g(z) = z^2\sqrt{\Delta t} - 2z\sqrt{\Delta t} + z \geq g(1) = 1-\sqrt{\Delta t} \geq 0$. This ends the proof. $\square$

Lemma 5.3.3. *Suppose Assumption 5.2.1 holds and $\Delta t \in (0,1]$, and $\alpha \geq 1$. For (5.11), we have that*

$$\begin{aligned}
|\hat{x}-\hat{y}| &\leq |x-y|, \quad \text{for any } x,y \in \mathbb{R}, && (5.21)\\
|\bar{b}(x)-\bar{b}(y)| &\leq \sqrt{\mathcal{K}_1}(1+2(\Delta t)^{-\eta(\alpha-1)})\,|x-y|, \quad \text{for any } x,y \in \mathbb{R}, && (5.22)\\
|\bar{b}(x)-\bar{b}(y)| &\leq \sqrt{\mathcal{K}_1}(1+|\hat{x}|^{\alpha-1}+|\hat{y}|^{\alpha-1})\,|x-y|, \quad \text{for any } x,y \in \mathbb{R}, && (5.23)\\
\bar{b}(x) &= b(\hat{x}) \leq b(0)+\beta\hat{x} \leq b(0)+\beta x, \quad x > 0, && (5.24)\\
|\bar{b}(x)-b(x)| &\leq \sqrt{\mathcal{K}_1}(1+|\hat{x}|^{\alpha-1}+|x|^{\alpha-1})\,|\hat{x}-x| \leq \sqrt{\mathcal{K}_1}(1+2\,|x|^{\alpha-1})\,|\hat{x}-x|, && (5.25)\\
\bar{b}(x) &\geq -\sqrt{\mathcal{K}_1}(|x|+|x|^{\alpha})+\bar{b}(0), && (5.26)\\
\bar{b}(x) &\geq -\sqrt{\mathcal{K}_1}(1+2(\Delta t)^{-\eta(\alpha-1)})|x|+\bar{b}(0). && (5.27)
\end{aligned}$$

The proof of first two properties can be found in [6]. The proof of the third, the fourth and the fifth properties follows from Assumption 5.2.1 and (5.11). The inequality (5.26) follows from (5.23) with $y = 0$ and the triangle inequality. The last inequality follows from (5.22) by letting $y = 0$ when $2\eta(\alpha-1) = 1$.

Lemma 5.3.4. *Under Assumption 5.2.1, we have for (5.12) that*

$$\begin{aligned}
\bar{b}(x) &\leq b(0)+\beta x, \quad x > 0, && (5.28)\\
|\bar{b}(x)-\bar{b}(0)| &\leq \frac{|b(x)-b(0)|}{1+|x|^{\alpha}\Delta t}+\frac{b(0)|x|^{\alpha}\Delta t}{1+|x|^{\alpha}\Delta t}, && (5.29)\\
|\bar{b}(x)| &= \frac{|b(x)|}{1+|x|^{\alpha}\Delta t} \leq C\min(\frac{1}{\Delta t}, 1+|x|^{\alpha}), && (5.30)\\
|\bar{b}(x)-b(x)| &\leq |b(x)|\,|x|^{\alpha}\Delta t \leq C(1+|x|^{2\alpha})\Delta t, && (5.31)\\
\Delta t\bar{b}(x) &\geq -\sqrt{\mathcal{K}_1}\,|x|+\frac{b(0)\Delta t}{1+|x|^{\alpha}\Delta t}, \quad \text{when } \Delta t \leq \left[(\frac{\alpha-1}{\alpha+1})^{(\alpha-1)}+1\right]^{-1}, && (5.32)\\
\bar{b}(x) &\geq -\sqrt{\mathcal{K}_1}(|x|+|x|^{\alpha})+\tilde{b}, \quad \tilde{b} = \frac{b(0)}{1+|x|^{\alpha}\Delta t}. && (5.33)
\end{aligned}$$

Here we use the convention that $0^0 \equiv 0$.

The proof is similar to those of Lemmas 5.3.1 and 5.3.2 and thus is omitted.

Lemma 5.3.5. *Under Assumption 5.2.1, we have for* (5.13) *that*

$$
\begin{aligned}
\bar{b}(x) &\leq b(0) + \beta x, \quad x > 0, && (5.34)\\
\left|\bar{b}(x) - \bar{b}(0)\right| &\leq |b(x) - b(0)|, && (5.35)\\
\left|\bar{b}(x)\right| &= \frac{|b(x)|}{1 + |b(x)|\,\Delta t} \leq C \min(\frac{1}{\Delta t}, 1 + |x|^{\alpha}), && (5.36)\\
\left|\bar{b}(x) - b(x)\right| &\leq |b(x)|^2 \Delta t \leq C(1 + |x|^{2\alpha})\Delta t, && (5.37)\\
\Delta t \bar{b}(x) &\geq -2(\sqrt{\mathcal{K}_1}\Delta t)^{\frac{1}{\alpha}}(|x| + |x|^{\frac{1}{\alpha}}) + \Delta t \bar{b}(0), && (5.38)\\
\bar{b}(x) &\geq -\sqrt{\mathcal{K}_1}(|x| + |x|^{\alpha}) + \bar{b}(0). && (5.39)
\end{aligned}
$$

Proof. The proof of first four inequalities are similar to that in Lemmas 5.3.1 and 5.3.4. Since the derivative of the function $y/(1 + |y|\,\Delta t)$ is bounded by 1, we have from the mean value theorem that $\left|\bar{b}(x) - \bar{b}(0)\right| \leq |b(x) - b(0)|$. Also,

$$
\Delta t \left|\bar{b}(x) - \bar{b}(0)\right| \leq \Delta t(\left|\bar{b}(x)\right| + \left|\bar{b}(0)\right|) = \frac{|b(x)\Delta t|}{1 + |b(x)|\,\Delta t} + \frac{|b(0)|\,\Delta t}{1 + |b(0)|\,\Delta t} \leq 2,
$$

By convexity, we have $(a + b)^x \leq a^x + b^x$, for $a, b > 0$ and $0 < x \leq 1$. It follows that for $\alpha \geq 1$,

$$
\begin{aligned}
\Delta t \left|\bar{b}(x) - \bar{b}(0)\right| &= (\Delta t \left|\bar{b}(x) - \bar{b}(0)\right|)^{\frac{1}{\alpha}} \cdot (\Delta t \left|\bar{b}(x) - \bar{b}(0)\right|)^{1-\frac{1}{\alpha}} \leq (\Delta t \left|\bar{b}(x) - \bar{b}(0)\right|)^{\frac{1}{\alpha}} \cdot 2^{1-\frac{1}{\alpha}}\\
&\leq 2(\Delta t)^{\frac{1}{\alpha}}(\left|\bar{b}(x) - \bar{b}(0)\right|)^{\frac{1}{\alpha}} \leq 2(\Delta t)^{\frac{1}{\alpha}}(\sqrt{\mathcal{K}_1}(1 + |x|^{2\alpha-2})^{\frac{1}{2}}|x|)^{\frac{1}{\alpha}}\\
&= 2(\Delta t\sqrt{\mathcal{K}_1})^{\frac{1}{\alpha}}(|x|^2 + |x|^{2\alpha})^{\frac{1}{2\alpha}} \leq 2(\Delta t\sqrt{\mathcal{K}_1})^{\frac{1}{\alpha}}(|x|^{\frac{1}{\alpha}} + |x|).
\end{aligned}
$$

Then we have (5.38). The estimate (5.39) follows from (5.35) and (5.4). $\square$

Lemma 5.3.6. *Under Assumption 5.2.1, we have for* (5.14) *that*

$$
\begin{aligned}
\bar{b}(x) &\leq b(0) + \beta x, \quad x > 0, && (5.40)\\
\left|\bar{b}(x) - \bar{b}(0)\right| &\leq |b(x) - b(0)|, && (5.41)\\
\left|\bar{b}(x)\right| &= \frac{|\tanh(b(x)\Delta t)|}{\Delta t} \leq C \min(\frac{1}{\Delta t}, 1 + |x|^{\alpha}), && (5.42)\\
\left|\bar{b}(x) - b(x)\right| &\leq |b(x)|^2 \Delta t \leq C(1 + |x|^{2\alpha})\Delta t, && (5.43)\\
\Delta t \bar{b}(x) &\geq -2(\sqrt{\mathcal{K}_1}\Delta t)^{\frac{1}{\alpha}}(|x| + |x|^{\frac{1}{\alpha}}) + \bar{b}(0)\Delta t, && (5.44)\\
\bar{b}(x) &\geq -\sqrt{\mathcal{K}_1}(|x| + |x|^{\alpha}) + \bar{b}(0). && (5.45)
\end{aligned}
$$

Proof. To show the first inequality, we obtain from the monotonicity of $\tanh(\cdot)$ and (5.6) that

$$
\bar{b}(x) = \frac{\tanh(b(x)\Delta t)}{\Delta t} \leq \frac{\tanh((b(0) + \beta x)\Delta t)}{\Delta t} \leq \frac{(b(0) + \beta x)\Delta t}{\Delta t} = b(0) + \beta x,\ x > 0. \qquad (5.46)
$$

The inequality (5.41) follows from applying the mean value theorem:

$$
|\bar{b}(x) - \bar{b}(0)| = \frac{1}{\Delta t}\,|\tanh(b(x)\Delta t) - \tanh(b(0)\Delta t)| \leq |1 - \tanh^2(\xi)||b(x) - b(0)| \leq |b(x) - b(0)|,
$$

where ξ lies in between $b(x)\Delta t$ and $b(0)\Delta t$. The inequality (5.42) follows from the fact that $|\tanh(y)| \le |y|$ and (5.8):

$$|\bar{b}(x)| = \frac{1}{\Delta t}\left|\tanh(b(x)\Delta t)\right| \le |b(x)| \le C\min(\frac{1}{\Delta t}, 1+|x|^{\alpha}). \tag{5.47}$$

By Taylor's formula and $|\tanh(y)| \le |y|$, we have $|\tanh(y)-y| \le \tanh^2(\theta y)\,|y| \le |y|$ for some $0 \le \theta \le 1$. We also have

$$|\tanh(y)-y| \le |y|^2. \tag{5.48}$$

By (5.48), we have $\Delta t\left|\bar{b}(x)-b(x)\right| = \left|\tanh\big(b(x)\Delta t\big) - b(x)\Delta t\right| \le b^2(x)(\Delta t)^2$. Then (5.43) follows from (5.8). Moreover, we can have (5.44) as in the proof of Lemma 5.3.5 since

$$\Delta t\left|\bar{b}(x)-b(x)\right| \le \Delta t(\left|\bar{b}(x)\right| + \left|\bar{b}(0)\right|) = |\tanh(b(x)\Delta t)| + |\tanh(b(0)\Delta t)| \le 2.$$

The inequality (5.45) follows from (5.41) and the local Lipschitz condition (5.4). $\square$

5.3.2 Properties of the solution to (5.9)

By $Y_t = |Z_t|$ and (5.15), we have from Tanaka's formula (See page 58 in [46]) from that

$$Y_t = X_0 + \int_0^t \operatorname{sgn}(Z_s)\bar{b}(Y_{\theta(s)})ds + \int_0^t \operatorname{sgn}(Z_s)\sigma Y_{\theta(s)}^{\gamma}dW_s + L_t(Y), \tag{5.49}$$

where Y_t is a continuous semimartingale with a continuous local time $L_t(Y) = \int_0^t \delta(Y_s)\sigma^2 Y_{\theta(s)}^{2\gamma}\,ds$, $\operatorname{sgn}(x)$ is the sign function, and $\delta(x)$ is the Dirac delta function. Moreover, we have

$$Y_t = Y_{\theta(t)} + \int_{\theta(t)}^t \operatorname{sgn}(Z_s)\bar{b}(Y_{\theta(s)})ds + \int_{\theta(t)}^t \operatorname{sgn}(Z_s)\sigma(Y_{\theta(s)})dW_s + \int_{\theta(t)}^t \delta(Y_s)\sigma^2 Y_{\theta(s)}^{2\gamma}\,ds. \tag{5.50}$$

Lemma 5.3.7. *Assume that $\{Z_t\}_{0\le t\le T}$ is given by (5.15) and assumptions 5.2.1 holds. If $\frac{1}{2} < \gamma < 1$ and $2\sqrt{\mathcal{K}_1}\Delta t \le \min(1, 2\sqrt{\mathcal{K}_1})$, then for Z_t from* (5.15) *with* (5.10)-(5.14),

$$\sup_{t\in[0,T]} \mathbb{P}(Z_t \le 0) \le C\exp(-\Delta t^{1-2\gamma}).$$

Proof. We first prove the desired conclusion for the scheme (5.9) with (5.11) or (5.13) or (5.14). From Lemmas 5.3.3, 5.3.5 and 5.3.6, we have $\bar{b}(y) \ge \bar{b}(0) - \sqrt{\mathcal{K}_1}(|y|+|y|^{\alpha})$. Then by (5.9), we have

$$\begin{aligned}
\mathbb{P}(Z_t \le 0) &\le \mathbb{P}(Y_{\theta(t)} + \bar{b}(Y_{\theta(t)})(t-\theta(t)) + \sigma Y_{\theta(t)}^{\gamma}(W_t - W_{\theta(t)}) \le 0)\\
&= \mathbb{P}\left(W_t - W_{\theta(t)} \le \frac{-Y_{\theta(t)} - \bar{b}(Y_{\theta(t)})(t-\theta(t))}{\sigma Y_{\theta(t)}^{\gamma}}, Y_{\theta(t)} > 0\right)\\
&\le \mathbb{P}\left(W_t - W_{\theta(t)} \le \frac{-Y_{\theta(t)} + \sqrt{\mathcal{K}_1}(Y_{\theta(t)} + \left|Y_{\theta(t)}\right|^{\alpha})(t-\theta(t)) - \bar{b}(0)(t-\theta(t))}{\sigma Y_{\theta(t)}^{\gamma}}, Y_{\theta(t)} > 0\right)\\
&\le \mathbb{P}\left(W_t - W_{\theta(t)} \le \frac{(\sqrt{\mathcal{K}_1}\Delta t - 1)Y_{\theta(t)} + \sqrt{\mathcal{K}_1}Y_{\theta(t)}^{\alpha}\Delta t - \bar{b}(0)(t-\theta(t))}{\sigma Y_{\theta(t)}^{\gamma}}, Y_{\theta(t)} > 0\right).
\end{aligned}$$

By conditioning on the natural filtration at $\theta(t)$, i.e., $\mathcal{F}^W_{\theta(t)}$ that

$$\begin{aligned}\mathbb{P}(Z_t \le 0) &\le \mathbb{P}\left(W_t - W_{\theta(t)} \le \frac{(\sqrt{\mathcal{K}_1}\Delta t - 1)Y_{\theta(t)} + \sqrt{\mathcal{K}_1}Y^{\alpha}_{\theta(t)}\Delta t - \bar{b}(0)(t-\theta(t))}{\sigma Y^{\gamma}_{\theta(t)}}, Y_{\theta(t)} > 0\right) \\ &= \mathbb{E}[\mathbb{P}\left(W_t - W_{\theta(t)} \le \frac{(\sqrt{\mathcal{K}_1}\Delta t - 1)Y_{\theta(t)} + \sqrt{\mathcal{K}_1}Y^{\alpha}_{\theta(t)}\Delta t - \bar{b}(0)(t-\theta(t))}{\sigma Y^{\gamma}_{\theta(t)}}, Y_{\theta(t)} > 0|\mathcal{F}^W_{\theta(t)}\right)]\end{aligned}$$

Then by the facts that $\mathbb{P}(\xi \le k) \le \frac{1}{2}\exp(-\frac{k^2}{2\mathbb{V}\mathrm{ar}[\xi]})$ for a centered Gaussian random variable ξ and $0 < t - \theta(t) \le \Delta t$, we have

$$\mathbb{P}(Z_t \le 0) \le \frac{1}{2}\mathbb{E}[\exp(-f(Y_{\theta(t)})\mathbb{1}_{Y_{\theta(t)}>0}] \le \frac{1}{2}\mathbb{E}[\exp(-g(Y_{\theta(t)})\mathbb{1}_{Y_{\theta(t)}>0}], \tag{5.51}$$

where

$$g(y) = \frac{C_1}{\Delta t}y^{2-2\gamma} + C_2\Delta t y^{2\alpha-2\gamma} - C_3\Delta t y^{\alpha-2\gamma} + C_4 y^{1-2\gamma} - C_5 y^{\alpha+1-2\gamma}, \quad y > 0,$$

and

$$\begin{aligned} f(y) &= \frac{(\sqrt{\mathcal{K}_1}\Delta t - 1)^2 y^2 + \mathcal{K}_1\Delta t^2 y^{2\alpha} + \bar{b}^2(0)(t-\theta(t))^2}{2\sigma^2 y^{2\gamma}(t-\theta(t))} \\ &\quad + \frac{2\sqrt{\mathcal{K}_1}\Delta t(\sqrt{\mathcal{K}_1}\Delta t - 1)y^{\alpha+1} - 2\bar{b}(0)\sqrt{\mathcal{K}_1}\Delta t y^{\alpha}(t-\theta(t)) - 2\bar{b}(0)(\sqrt{\mathcal{K}_1}\Delta t - 1)y(t-\theta(t))}{2\sigma^2 y^{2\gamma}(t-\theta(t))} \\ &\ge g(y) + \frac{\bar{b}^2(0)(t-\theta(t))}{2\sigma^2}y^{-2\gamma}, \quad y > 0. \end{aligned}$$

Here we denote

$$C_1 = \frac{(\sqrt{\mathcal{K}_1}\Delta t - 1)^2}{2\sigma^2}, C_2 = \frac{\mathcal{K}_1}{2\sigma^2}, C_3 = \frac{\bar{b}(0)\sqrt{\mathcal{K}_1}}{\sigma^2}, C_4 = \frac{\bar{b}(0)(1-\sqrt{\mathcal{K}_1}\Delta t)}{\sigma^2}, C_5 = \frac{\sqrt{\mathcal{K}_1}(1-\sqrt{\mathcal{K}_1}\Delta t)}{\sigma^2}. \tag{5.52}$$

By calculations, in what follows, we prove that $g(y) \ge C(\Delta t)^{1-2\gamma}$ for all $y > 0$.

If $0 < y \le 1$, we have $-y^{\alpha-1} \ge -1$ and thus

$$g(y) \ge \tfrac{C_1}{\Delta t}y^{2-2\gamma} - (C_3\Delta t - C_4)y^{1-2\gamma} - C_5 = y^{1-2\gamma}\big(\tfrac{C_1}{\Delta t}y - (C_3\Delta t - C_4)\big) - C_5 =: h(y) - C_5.$$

The only root of $h'(y)$ is $y^* = \frac{(C_3\Delta t - C_4)(1-2\gamma)}{C_1(2-2\gamma)}\Delta t$ as $C_3\Delta t - C_4 = \frac{\bar{b}(0)(2\sqrt{\mathcal{K}_1}\Delta t - 1)}{\sigma^2} < 0$ by the assumption. Then the minimum of $h(y)$ is the minimum of $h(1) = \frac{C_1}{\Delta t} - (C_3\Delta t - C_4)$ and $h(y^*) = C\Delta t^{1-2\gamma}$ for some $C > 0$. Thus $g(y) \ge C\Delta t^{1-2\gamma}$.

If $y \ge 1$, then $-y^{\alpha-2\gamma} \ge -y^{\alpha+1-2\gamma}$ and thus

$$\begin{aligned} g(y) &= \Big(\sqrt{\frac{C_1}{\Delta t}}y^{1-\gamma} - \sqrt{C_2\Delta t}y^{\alpha-\gamma}\Big)^2 - C_3\Delta t y^{\alpha-2\gamma} + C_4 y^{1-2\gamma} \\ &\ge \Big(\sqrt{\frac{C_1}{\Delta t}}y^{1-\gamma} - \sqrt{C_2\Delta t}y^{\alpha-\gamma}\Big)^2 - C_3\Delta t y^{\alpha+1-2\gamma} =: h(y) \end{aligned}$$

Observe that $h'(y) = y^{1-2\gamma}(\frac{C_1}{\Delta t}(2-2\gamma) + C_2\Delta t(2\alpha-2\gamma)y^{2\alpha-2} - 2(\sqrt{C_1C_2} + C_3\Delta t)(\alpha+1-2\gamma)y^{\alpha-1})$ and then the only zero of $h'(y)$ on $[1,\infty)$ satisfies that $(y^*)^{\alpha-1}$ is at the order of $\frac{\sqrt{C_1C_2}}{C_2\Delta t}$. Then,

$h(y^*)$ is at the order of $\Delta t^{\frac{2\gamma-\alpha-1}{\alpha-1}}$. Since $\frac{2\gamma-\alpha-1}{\alpha-1} < -1$, $g(y) \geq h(y) \geq \min(h(y^*), h(1)) \geq C(\Delta t)^{-1}$. In conclusion, $g(y) \geq C\Delta t^{1-2\gamma}$ when $0 < y \leq 1$, while $g(y) \leq C\Delta t^{-1}$ when $y \geq 1$. By the fact that $\Delta t^{1-2\gamma} < \Delta t^{-1}$ $(-1 < 1-2\gamma < 0)$, we have that $g(y) \geq C(\Delta t)^{1-2\gamma}$.

Then by (5.51) and that that $g(y) \geq C(\Delta t)^{1-2\gamma}$ for $y > 0$, we have

$$\mathbb{P}(Z_t \leq 0) \quad \leq \quad \frac{1}{2}\mathbb{E}[\exp(-g(Y_{\theta(t)})\mathbb{1}_{Y_{\theta(t)}>0}] \leq \frac{1}{2}\mathbb{E}[\exp(-C(\Delta t)^{1-2\gamma})].$$

We now prove the desired conclusion for the scheme (5.9) with (5.10) or (5.12). We claim that an inequality similar to (5.51) still holds. In fact,

$$\mathbb{P}(Z_t \leq 0) \leq \frac{1}{2}\mathbb{E}[\exp(-g(Y_{\theta(t)})\mathbb{1}_{Y_{\theta(t)}>0}], \tag{5.53}$$

where

$$g(y) = \frac{C_1}{\Delta t}y^{2-2\gamma} + C_2\Delta t y^{2\alpha-2\gamma} - \tilde{C}_3\Delta t y^{\alpha-2\gamma} + \tilde{C}_4 y^{1-2\gamma} - C_5 y^{\alpha+1-2\gamma}, \quad y > 0,$$

Here C_1, C_2, C_5 are from (5.52) and $\tilde{C}_3 = \dfrac{\tilde{b}\sqrt{\mathcal{K}_1}}{\sigma^2}$, $\tilde{C}_4 = \dfrac{\tilde{b}(1-\sqrt{\mathcal{K}_1}\Delta t)}{\sigma^2}$. Recall that $\tilde{b} = \frac{b(0)}{1+|y|^\alpha\sqrt{\Delta t}}$ or $\frac{b(0)}{1+|y|^\alpha\Delta t}$ from Lemmas 5.3.1 and 5.3.4, and $\bar{b}(0) = b(0)$. Similar to the above discussion, we can readily show that when $\Delta t \leq 1$,

$$g(y) \geq \left\{ \begin{array}{ll} y^{1-2\gamma}\big(\frac{C_1}{\Delta t}y - \frac{1}{1+\Delta t}(C_3\Delta t - C_4)\big) - C_5, & 0 < y \leq 1 \\ \big(\sqrt{\frac{C_1}{\Delta t}}y^{1-\gamma} - \sqrt{C_2\Delta t}y^{\alpha-\gamma}\big)^2 - C_3\Delta t y^{\alpha+1-2\gamma}, & y \geq 1 \end{array}\right.,$$

where C_3, C_4 are from (5.52) ($\bar{b}(0)$ may vary with the considered schemes though). We then can conclude as above that $g(y) \geq C\Delta t^{1-2\gamma}$ when $0 < y \leq 1$, and $g(y) \geq C\Delta t^{-1}$ when $y \geq 1$, which implies that $g(y) \geq C(\Delta t)^{1-2\gamma}$. Then (5.53) leads to the desired conclusion. □

Throughout the paper, we denote $a \wedge b = \min\{a, b\}$.

Lemma 5.3.8. *Let $p \geq 1$. Then the numerical scheme $Y_t = |Z_t|$ has bounded positive moments. i.e. $\mathbb{E}[|Y_t|^p] < \infty$, for $0 \leq t \leq T$. Here Z_t is defined in* (5.15).

Proof. Here we prove the conclusion when $p > 2$. For any $1 \leq p \leq 2$, the conclusion follows from Hölder's inequality. Define $\zeta_n = T \wedge \inf\{t \in [0,T] : Y_t \geq n\}$ $(n \geq 1)$, where we set $\inf\emptyset = +\infty$. Then $0 \leq Y_t \leq n$ on $[0, \zeta_n]$. By $Y_t = |Z_t|$, (5.15), and Itô's formula, we have

$$\mathbb{E}\left[Y^p_{t\wedge\zeta_n}\right] \quad = \quad \mathbb{E}\left[X_0^p + \int_0^{t\wedge\zeta_n} pY_s^{p-1}\operatorname{sgn}(Z_s)\bar{b}(Y_{\theta(s)})ds + \frac{p(p-1)}{2}Y_s^{p-2}\sigma^2 Y_{\theta(s)}^{2\gamma}\,ds\right].$$

Here we have used the fact that $\mathbb{E}[\int_0^{t\wedge\zeta_n}\operatorname{sgn}(Z_s)\sigma Y^\gamma_{\theta(s)}dW_s] = 0$. Then by the Young's inequality and the fact that $\mathbb{E}[Y^q_{\theta(s)}] \leq \sup_{s\in[0,T]}\mathbb{E}[Y^q_s]$ $(q > 0)$, we have

$$\begin{aligned}\mathbb{E}\left[Y^p_{t\wedge\zeta_n}\right] \quad &\leq \quad \mathbb{E}\left[X_0^p + \int_0^{t\wedge\zeta_n} pY_s^{p-1}\operatorname{sgn}(Z_s)\bar{b}(Y_{\theta(s)})ds\right] + \sigma^2\frac{p(p-1)}{2}\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-2}(1+Y^2_{\theta(s)})\,ds\right] \\ &\leq \quad \mathbb{E}\left[X_0^p + \int_0^{t\wedge\zeta_n} pY_s^{p-1}\operatorname{sgn}(Z_s)\bar{b}(Y_{\theta(s)})ds\right] + \sigma^2\frac{p(p-1)}{2}\int_0^t \sup_{u\in[0,s]}\mathbb{E}[Y^p_{u\wedge\zeta_n}]ds + Ct.\end{aligned}$$

By Lemmas 5.3.1–5.3.6, we have $\bar{b}(Y_{\theta(s)}) \le b(0) + \beta Y_{\theta(s)}$ when $Y_{\theta(s)} > 0$. The same inequality holds for $Y_{\theta(s)} = 0$ as the inequality becomes $\bar{b}(0) \le b(0)$ (recall that $\bar{b}(0), b(0) > 0$) which can be readily verified for all our schemes. Then we have

$$\begin{aligned}
\mathbb{E}\left[Y^p_{t\wedge\zeta_n}\right] &\le \mathbb{E}\left[X_0^p + \int_0^{t\wedge\zeta_n} pY_s^{p-1}\operatorname{sgn}(Z_s)\bar{b}(Y_{\theta(s)})ds\right] + C\int_0^t \sup_{u\in[0,s]} \mathbb{E}[Y^p_{u\wedge\zeta_n}]ds + Ct \\
&\le \mathbb{E}\left[X_0^p + \int_0^{t\wedge\zeta_n} pY_s^{p-1}\mathbb{1}_{\{Z_s\ge 0\}}(b(0)+\beta Y_{\theta(s)})\,ds\right] + C\int_0^t \sup_{u\in[0,s]} \mathbb{E}[Y^p_{u\wedge\zeta_n}]ds \\
&\quad -p\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\bar{b}(Y_{\theta(s)})\,ds\right] + Ct. \qquad (5.54)
\end{aligned}$$

By Young's inequality and the fact that $\mathbb{E}[Y^q_{\theta(s)}] \le \sup_{s\in[0,t]}\mathbb{E}[Y_s^q]$ ($q>0$, $s\le t$), we have

$$\left|\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s\ge 0\}}(b(0)+\beta Y_{\theta(s)})\,ds]\right| \le C\int_0^t \sup_{u\in[0,s]}\mathbb{E}[Y^p_{u\wedge\zeta_n}]\,ds + Ct. \qquad (5.55)$$

To estimate $\left|\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\bar{b}(Y_{\theta(s)})\,ds\right]\right|$, we need to discuss cases of $\bar{b}$.

Case I, $\left|\bar{b}(x)\right|\sqrt{\Delta t} \le C$, which holds for (5.10) by Lemma 5.3.1. By the Young's inequality

$$\begin{aligned}
\left|\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\bar{b}(Y_{\theta(s)})\,ds\right]\right| &\le C\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\Delta t^{-\frac{1}{2}}\,ds\right] \qquad (5.56)\\
&\le C\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^p\,ds] + \int_0^t C\mathbb{P}(Z_{s\wedge\zeta_n}\le 0)\,ds\Delta t^{-\frac{p}{2}}.
\end{aligned}$$

Case II, for the scheme (5.11), we have (5.8) that $\left|\bar{b}(x)\right| = |b(\hat{x})| \le C(1+|\hat{x}|^\alpha) \le C(1+(\Delta t)^{-\eta\alpha})$. Thus by the Young's inequality, we have

$$\begin{aligned}
\left|\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\bar{b}(Y_{\theta(s)})\,ds]\right| &\le C\left|\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}(1+(\Delta t)^{-\eta\alpha})\,ds]\right| \\
&\le C\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^p\,ds] + C\int_0^t \mathbb{P}(Z_{s\wedge\zeta_n}\le 0)\,ds(1+(\Delta t)^{-p\eta\alpha}).
\end{aligned}$$

As we take $2\eta(\alpha-1)=1$, we have

$$\left|\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\bar{b}(Y_{\theta(s)})\,ds]\right| \le C\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^p\,ds] + C\int_0^t \mathbb{P}(Z_{s\wedge\zeta_n}\le 0)\,ds(1+(\Delta t)^{-\frac{p\alpha}{2(\alpha-1)}}). \qquad (5.57)$$

Case III, $\left|\bar{b}(x)\right|\Delta t \le C$, which is valid for (5.12)-(5.14). For (5.12), $\bar{b}(x) = \dfrac{b(x)}{1+|x|^\alpha\Delta t}$. From Lemma 5.3.4, we have that $\left|\bar{b}(x)\right|\Delta t \le C$. For (5.13), we have $\bar{b}(x) = \dfrac{b(x)}{1+|b(x)|\Delta t}$ and by Lemma 5.3.5, $\left|\bar{b}(x)\right|\Delta t \le C$. From Lemma 5.3.6, we have that $\left|\bar{b}(x)\right|\Delta t \le C$ for (5.14), i.e., $\bar{b}(x) = \dfrac{\tanh(b(x)\Delta t)}{\Delta t}$. By Young's inequality, we have

$$\begin{aligned}
\left|\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\bar{b}(Y_{\theta(s)})\,ds\right]\right| &\le C\mathbb{E}\left[\int_0^{t\wedge\zeta_n} Y_s^{p-1}\mathbb{1}_{\{Z_s<0\}}\Delta t^{-1}\,ds\right] \qquad (5.58)\\
&\le C\mathbb{E}[\int_0^{t\wedge\zeta_n} Y_s^p\,ds] + \int_0^t C\mathbb{P}(Z_{s\wedge\zeta_n}\le 0)\,ds\Delta t^{-p}.
\end{aligned}$$

Then by (5.54), (5.55) and (5.57)-(5.58), we obtain

$$\mathbb{E}\left[Y^p_{t\wedge\zeta_n}\right] \leq \mathbb{E}[X^p_0] + Ct + C\int_0^t \sup_{u\in[0,s]} \mathbb{E}[Y^p_{u\wedge\zeta_n}]\,ds + C\int_0^t \mathbb{P}(Z_{s\wedge\zeta_n}\leq 0)\,ds\Delta t^{-p^*},$$

where $p^* = p/2$ for (5.10), $p^* = \frac{p\alpha}{2(\alpha-1)}$ for (5.11) and $p^* = p$ for (5.12)-(5.14). Thus by Lemma 5.3.7, we have $\int_0^t \mathbb{P}(Z_{s\wedge\zeta_n}\leq 0)\,ds\Delta t^{-p^*} \leq \int_0^t \sup_{u\in[0,s]}\mathbb{P}(Z_u\leq 0)\,ds\Delta t^{-p^*}\leq C$ and thus

$$\mathbb{E}\left[Y^p_{t\wedge\zeta_n}\right] \leq \mathbb{E}[X^p_0] + C\int_0^t \sup_{u\in[0,s]} \mathbb{E}[Y^p_{u\wedge\zeta_n}]\,ds + Ct.$$

Then the Gronwall inequality leads to $\mathbb{E}\left[Y^p_{t\wedge\zeta_n}\right]\leq Ce^{Ct}$. The conclusion then follows from Fatou's lemma. □

Lemma 5.3.9. *For all $p\geq 1$, there exists a positive constant C depending on σ, p and T but not on Δt such that*

$$\mathbb{E}[|Y_{\theta(t)}-Y_t|^{2p}]\leq C\Delta t^p.$$

Proof. By the scheme (5.9) and the fact that $Y_{\theta(s)}$ is independent of $W_s - W_{\theta(s)}$, we have

$$\begin{aligned}
\mathbb{E}[|Y_{\theta(s)}-Y_s|^{2p}] &= \mathbb{E}\left[\left|\left|Y_{\theta(s)} + \bar{b}(Y_{\theta(s)})(s-\theta(s)) + \sigma Y^{\gamma}_{\theta(s)}(W_s - W_{\theta(s)})\right| - Y_{\theta(s)}\right|^{2p}\right]\\
&\leq C\mathbb{E}[(\left|\bar{b}(Y_{\theta(s)})\right|^{2p}\Delta t^{2p} + \left|Y_{\theta(s)}\right|^{2\gamma p}|W_s - W_{\theta(s)}|^{2p})]\\
&\leq C\mathbb{E}[\left|\bar{b}(Y_{\theta(s)})\right|^{2p}]\Delta t^{2p} + C\mathbb{E}[\left|Y_{\theta(s)}\right|^{2\gamma p}]\Delta t^p\\
&\leq C\mathbb{E}[1+\left|Y_{\theta(s)}\right|^{2\alpha p}]\Delta t^{2p} + C\mathbb{E}[\left|Y_{\theta(s)}\right|^{2\gamma p}]\Delta t^p.
\end{aligned}$$

In the last step, we applied $\left|\bar{b}(x)\right|\leq C(1+|x|^{\alpha})$ and $(1+|a|)^{2p}\leq C(1+|a|^{2p})$. Here the fact that $\left|\bar{b}(x)\right|\leq C(1+|x|^{\alpha})$ holds, according to Lemmas 5.3.1 and 5.3.3-5.3.6. □

5.3.3 Bounded moments for X_t

We first show bounded moments of the solution to (5.1).

Lemma 5.3.10 (Bounded moments for the solution to (5.1))**.** *Assume X_t is the solution to* (5.1) *and Assumption 5.2.1 hold. For $p\geq 1$, there exists a positive constant $\mathcal{C}_1$, depending on p, σ and $T>0$, such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|X_t|^p\right]\leq \mathcal{C}_1(1+\mathbb{E}[X^p_0]). \tag{5.59}$$

If $X_t>0$ holds for $t\in[0,T]$, then there is a positive constant $\mathcal{C}_2$ depending on σ, p and T such that

$$\sup_{0\leq t\leq T}\mathbb{E}[|X_t|^{-p}]\leq \mathcal{C}_2(1+\mathbb{E}[X^{-p}_0]),\quad p>0. \tag{5.60}$$

Proof. The conclusion (5.59) can be proved by Itô's formula, see e.g. [36]. Here we only prove (5.60).

Let $\tau_n = \inf\{0 < s \le T; X_s \le \frac{1}{n}\}$. It follows from (5.4) that $b(X_s) - b(0) \ge -\sqrt{\mathcal{K}_1}(X_s + X_s^\alpha)$. For any $p > 0$, apply the Itô's formula,

$$\begin{aligned}
\mathbb{E}[X_{t\wedge\tau_n}^{-p}] &= \mathbb{E}[X_0^{-p}] - p\mathbb{E}\left[\int_0^{t\wedge\tau_n} b(X_s)X_s^{-p-1}ds\right] + \frac{p(p+1)\sigma^2}{2}\mathbb{E}\left[\int_0^{t\wedge\tau_n} X_s^{-p-1+(2\gamma-1)}ds\right] \\
&\le \mathbb{E}[X_0^{-p}] - p\mathbb{E}\left[\int_0^{t\wedge\tau_n}[-\sqrt{\mathcal{K}_1}(X_s + X_s^\alpha) + b(0)]X_s^{-p-1}ds\right] \\
&\quad + \frac{p(p+1)\sigma^2}{2}\mathbb{E}\left[\int_0^{t\wedge\tau_n} X_s^{-p-1+(2\gamma-1)}ds\right] \\
&= \mathbb{E}[X_0^{-p}] + p\sqrt{\mathcal{K}_1}\mathbb{E}\left[\int_0^{t\wedge\tau_n} X_s^{-p}ds\right] + \mathbb{E}\left[\int_0^{t\wedge\tau_n} f(X_s)ds\right], \qquad (5.61)
\end{aligned}$$

where we define $f(x) = px^{-p-1}(\sqrt{\mathcal{K}_1}x^\alpha + \frac{(p+1)\sigma^2}{2}x^{2\gamma-1} - b(0))$, where $x > 0, \alpha \ge 1, b(0) > 0$ and $\frac{1}{2} < \gamma < 1$. When $p+1 \ge \alpha$, we claim that there exists a constant C such that $f(x) \le C$, for all $x > 0$. It follows that when $p + 1 \ge \alpha$,

$$\mathbb{E}[X_{t\wedge\tau_n}^{-p}] \le \mathbb{E}[X_0^{-p}] + CT + p\sqrt{\mathcal{K}_1}\mathbb{E}\left[\int_0^{t\wedge\tau_n} X_s^{-p}\,ds\right] \le \mathbb{E}[X_0^{-p}] + CT + p\sqrt{\mathcal{K}_1}\mathbb{E}\left[\int_0^{t} X_{s\wedge\tau_n}^{-p}\,ds\right].$$

Apply the Gronwall inequality, then we have

$$\mathbb{E}[|X_{t\wedge\tau_n}|^{-p}] \le (\mathbb{E}[|X_0|^{-p}] + CT)\exp(p\sqrt{\mathcal{K}_1}T).$$

By Fatou's Lemma, $\mathbb{E}[|X_t|^{-p}] = \mathbb{E}[\lim_{n\to\infty}|X_{t\wedge\tau_n}|^{-p}] \le \mathcal{C}_2(\mathbb{E}[|X_0|^{-p}] + 1)$, where $\mathcal{C}_2$ is a positive constant. For $0 < p \le \alpha - 1$, $\mathbb{E}[X_t^{-p}] < \infty$, since $\mathbb{E}[X_t^{-\alpha}] < \infty$.

It remains to prove that $f(x) \le C$, for all $x > 0$. When $x \ge 1$, it is ready to see that $f(x) < p\sqrt{\mathcal{K}_1} + \frac{p(p+1)}{2}\sigma^2$. Let $\epsilon > 0$ be an arbitrary small number. For $x \in (0, \epsilon)$, f is bounded from above in this case since $f(0+) = -\infty$. When $x \in (\epsilon, 1]$, $x^\alpha \le 1$ and $f(x) \le px^{-p}\left(\frac{(p+1)\sigma^2}{2}x^{2\gamma-2} + \left[\sqrt{\mathcal{K}_1} - b(0)\right]x^{-1}\right) =: g(x)$. Observe that $g'(x) = -(p+2-2\gamma)\frac{p(p+1)}{2}\sigma^2 x^{2\gamma-p-3} + p(p+1)[b(0) - \sqrt{\mathcal{K}_1}]x^{-p-2}$ and the only possible root x^* satisfies that $(x^*)^{2\gamma-1} = \frac{2[b(0)-\sqrt{\mathcal{K}_1}]}{(p+2-2\gamma)\sigma^2}$. The maximum value of $g(x)$ can be achieved only when $x = \epsilon$, $x = 1$ or $x = x^*$ and thus we obtain that $g(x) \le C$. In fact, $g(x^*) \le p(x^*)^{-p-1}\left(\frac{p+1}{p+2-2\gamma}[b(0) - C\sqrt{\mathcal{K}_1}] - b(0)\right)$ if $x^* \in (\epsilon, 1)$, $g(1) = p\left(\frac{(p+1)\sigma^2}{2} + \sqrt{\mathcal{K}_1} - b(0)\right)$, and $g(\epsilon)$ is a finite number. $\square$

When $X_t \ge 0$, we define the following process $(\eta(t))_{t\ge0}$ by

$$\chi(t) := \int_0^t \frac{ds}{(Y_{\theta(s)}^{1-\gamma} + X_s^{1-\gamma})^2}. \qquad (5.62)$$

Lemma 5.3.11. *Suppose that Assumption 5.2.1 holds and $X_t > 0$ almost everywhere in t and that $X_0 > 0$ is a constant. Then there exist positive constants μ and C such that*

$$\mathbb{E}[\exp(\mu\chi(T))] \le \exp(CT). \qquad (5.63)$$

Proof. We prove the conclusion when $\mu = 1$. By (5.62) and Jensen's inequality, we have

$$\mathbb{E}[\exp(\mu\chi(T))] \le \mathbb{E}[\exp(\int_0^T X_s^{2\gamma-2})\,ds] \le \mathbb{E}[\frac{1}{T}\int_0^T \exp(TX_s^{2\gamma-2})\,ds].$$

Thus we only need to check that $\mathbb{E}[\exp(TX_s^{2\gamma-2})] \le \exp(CT)$ for all $s \in [0,T]$. We will show this desired result in two steps.

First, we show the boundedness of $\mathbb{E}[X_t^{-p}]$ for any $p > 0$ with the dependence on p. By (5.61), (5.60) and Fatou's lemma, we have for any fixed $p > 0$,

$$\mathbb{E}[X_t^{-p}] \le \mathbb{E}[X_0^{-p}] + \mathbb{E}\left[\int_0^t F(X_s)ds\right], \tag{5.64}$$

where $F(x) = px^{-p-1}(\sqrt{\mathcal{K}_1}x^\alpha + \sqrt{\mathcal{K}_1}x + \frac{(p+1)\sigma^2}{2}x^{2\gamma-1} - b(0)) = f(x) + p\sqrt{\mathcal{K}_1}x^{-p}$, where $x > 0$, $\alpha \ge 1$ and $b(0) > 0$. Here f is the same as in (5.61). We claim that when $p+1 > \alpha$,

$$F(x) \le C(p\epsilon^{-1-p+\alpha} + p\epsilon^{-p} + p^2\epsilon^{-2-p+2\gamma}). \tag{5.65}$$

In fact, $F(0+) = -\infty$ if $p+1 > \alpha$ and thus there exists $0 < \epsilon < 1$ such that $F(x) \le 0$, when $0 < x \le \epsilon < 1$. When $x \ge \epsilon$, $F(x) \le G(x,p) \le G(\epsilon,p)$, where $G(x,p) = px^{-p-1}(\sqrt{\mathcal{K}_1}x^\alpha + \sqrt{\mathcal{K}_1}x + \frac{(p+1)\sigma^2}{2}x^{2\gamma-1})$. For large enough p ($p+1 \ge \alpha$), $G(\epsilon,p) \le C(p\epsilon^{-1-p+\alpha} + p\epsilon^{-p} + p^2\epsilon^{-2-p+2\gamma})$. Then by (5.64) and (5.65), we have that for any large enough p,

$$\mathbb{E}[X_t^{-p}] \le \mathbb{E}[X_0^{-p}] + tG(\epsilon,p). \tag{5.66}$$

For a small p, $\mathbb{E}[X_t^{-p}] \le 1 + \mathbb{E}[X_t^{-p_0}]$ where p_0 is large enough such that (5.66) holds.

Second, we show the boundedness of the desired exponential moment. For a large integer k_0 ($2k_0(1-\gamma) \ge p_0$), we have from $\mathbb{E}[X_t^{-2k(1-\gamma)}] \le \mathbb{E}[1 + X_t^{-2(k_0+1)(1-\gamma)}]$ ($k \le k_0$) and (5.66) that

$$\begin{aligned}
\sum_{k=0}^\infty \frac{1}{k!}\mathbb{E}[X_t^{-2k(1-\gamma)}] &\le \sum_{k=0}^{k_0}\frac{1}{k!}\mathbb{E}[1 + X_t^{-2(k_0+1)(1-\gamma)}] + \sum_{k=k_0+1}^\infty \frac{1}{k!}\mathbb{E}[X_t^{-2k(1-\gamma)}] \\
&\le C(1 + \mathbb{E}[X_t^{-2(k_0+1)(1-\gamma)}]) + \sum_{k=k_0+1}^\infty \frac{1}{k!}\mathbb{E}[X_0^{-2k(1-\gamma)}] \\
&\quad + Ct\sum_{k=k_0+1}^\infty \frac{1}{k!}(k\epsilon^{-1-2k(1-\gamma)+\alpha} + k\epsilon^{-2k(1-\gamma)} + k^2\epsilon^{-2(k+1)(1-\gamma)}).
\end{aligned}$$

Then by the facts that $\sum_{n=0}^\infty \frac{x^n}{n!} = e^x$, $\sum_{n=1}^\infty \frac{nx^n}{n!} = xe^x$, and $\sum_{n=1}^\infty \frac{n^2x^n}{n!} = (x^2+x)e^x$, we have

$$\begin{aligned}
\sum_{k=0}^\infty \frac{1}{k!}\mathbb{E}[X_t^{-2k(1-\gamma)}] &\le C(1 + \mathbb{E}[X_t^{-2(k_0+1)(1-\gamma)}]) + Ct\epsilon^{-2(1-\gamma)}(\epsilon^{\alpha-1}+1)\exp(\epsilon^{-2(1-\gamma)}) \\
&\quad + Ct(\epsilon^{6\gamma-6} + \epsilon^{4\gamma-4})\exp(\epsilon^{-2(1-\gamma)}) \\
&\le C(1 + \mathbb{E}[X_t^{-2(k_0+1)(1-\gamma)}]) + Ct\exp(2\epsilon^{-2(1-\gamma)}) \le \exp(Ct).
\end{aligned} \tag{5.67}$$

Then the fact that $\mathbb{E}[\exp(TX_t^{2\gamma-2})] = \sum_{k=0}^\infty \frac{T^k}{k!}\mathbb{E}[X_t^{-2k(1-\gamma)}]$ and Lemma 5.3.10 lead to the desired conclusion. □

5.3.4 Proof of Theorem 5.2.2

We follow the idea of the proof for the positivity-preserving scheme for SDEs with Lipschitz drift and Hölder diffusion coefficients in [5]. Denote the error process $\epsilon_t := Y_t - X_t$ and it satisfies

$$\epsilon_t = \int_0^t [\bar{b}(Y_{\theta(s)}) \operatorname{sgn}(Z_s) - b(X_s)]ds + \sigma \int_0^t (Y^\gamma_{\theta(s)} \operatorname{sgn}(Z_s) - X^\gamma_s)dW_s + L_t(Y).$$

For an arbitrary stopping time τ valued in $[0, T]$, we apply the Itô's formula to ϵ_t^{2p}, between 0 and τ, where $p \geq 2$ is an integer. For any other values of p ($p \geq 1$), the desired conclusion follows from Hölder's inequality. As $\int_0^\tau (\epsilon_s)^{2p-1} dL_s(Y) = \int_0^\tau (-X_s)^{2p-1} \delta(Y_s)\sigma^2 Y^{2\gamma}_{\theta(s)} ds \leq 0$, we obtain

$$\begin{aligned}
\mathbb{E}[|\epsilon_\tau|^{2p}] &\leq 2p\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}(\bar{b}(Y_{\theta(s)}) \operatorname{sgn}(Z_s) - b(X_s))ds\right] \\
&\quad + p(2p-1)\sigma^2 \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-2}(Y^\gamma_{\theta(s)} \operatorname{sgn}(Z_s) - X^\gamma_s)^2 ds\right] \\
&\leq 2p\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}(\bar{b}(Y_{\theta(s)}) - b(X_s))ds\right] + \sigma^2 p(2p-1)\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-2}(Y^\gamma_{\theta(s)} - X^\gamma_s)^2 ds\right] \\
&\quad + 2\mathbb{E}\left[\int_0^\tau \left\{2p(\epsilon_s)^{2p-1}(-\bar{b}(Y_{\theta(s)})) + \sigma^2 p(2p-1)(\epsilon_s)^{2p-2}(Y^{2\gamma}_{\theta(s)} + Y^\gamma_{\theta(s)} X^\gamma_s)\right\} \mathbb{1}_{\{Z_s<0\}} ds\right] \\
&=: I + II + III.
\end{aligned} \tag{5.68}$$

Let's consider the term I first in (5.68). We split $\bar{b}(Y_{\theta(s)}) - b(X_s)$ into three parts:

$$\bar{b}(Y_{\theta(s)}) - b(X_s) = [\bar{b}(Y_{\theta(s)}) - b(Y_{\theta(s)})] + [b(Y_{\theta(s)}) - b(Y_s)] + [b(Y_s) - b(X_s)].$$

We denote the corresponding integral by I_1, I_2 and I_3.

$$\begin{aligned}
I_1 &= \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}[\bar{b}(Y_{\theta(s)}) - b(Y_{\theta(s)})]ds\right], \\
I_2 &= \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}[b(Y_{\theta(s)}) - b(Y_s)]ds\right], \\
I_3 &= \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}[b(Y_s) - b(X_s)]ds\right].
\end{aligned}$$

Now we consider the bound for I_1 for all schemes (5.10)-(5.14).
Case A. *Scheme* (5.49) *with* (5.10). We have from Lemma 5.3.1 that

$$\left|\bar{b}(x) - b(x)\right| \leq C(1 + |x|^{2\alpha})\sqrt{\Delta t}. \tag{5.69}$$

Then by Young's inequality and Lemma 5.3.8, we have

$$\begin{aligned}
I_1 &\leq \mathbb{E}\left[\int_0^\tau \left|(\epsilon_s)^{2p-1}[\bar{b}(Y_{\theta(s)}) - b(Y_{\theta(s)})]\right| ds\right] \\
&\leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\int_0^T \mathbb{E}[1 + (Y_{\theta(s)})^{4\alpha p}]ds\Delta t^p \leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\Delta t^p.
\end{aligned} \tag{5.70}$$

Case B. *Scheme* (5.49) *with* (5.11). We have from Lemma 5.3.3 that

$$\left|\bar{b}(x) - b(x)\right| \leq \mathcal{K}_1(1 + 2\left|x\right|^{\alpha-1})\left|\hat{x} - x\right|. \tag{5.71}$$

Then by Young's inequality, Hölder's inequality and Lemma 5.3.8, we have

$$\begin{aligned}
I_1 &\leq \mathbb{E}\left[\int_0^\tau \left|(\epsilon_s)^{2p-1}(\bar{b}(Y_{\theta(s)}) - b(Y_{\theta(s)}))\right| ds\right] \\
&\leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\int_0^T \left(\mathbb{E}[(Y_{\theta(s)} - \widehat{Y}_{\theta(s)})^{4p}\mathbb{1}_{\{|Y_{\theta(s)}|\geq(\Delta t)^{-\eta}\}}]\right)^{1/2} ds \\
&\leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + CT\big(\mathbb{P}(\left|Y_{\theta(s)}\right| \geq (\Delta t)^{-\eta}\big)^{1/2}.
\end{aligned}$$

By Markov inequality and Lemma 5.3.8, $\mathbb{P}(\left|Y_{\theta(s)}\right| \geq (\Delta t)^{-\eta}) \leq C(\Delta t)^{2p}$. Then we have

$$I_1 \leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + CT\sqrt{\mathbb{P}(\left|Y_{\theta(s)}\right| \geq (\Delta t)^{-\eta})} \leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C(\Delta t)^p. \tag{5.72}$$

Case C. *Scheme* (5.49) *with* (5.12) *or* (5.13) *or* (5.14). We have from Lemmas 5.3.4, 5.3.5, 5.3.6 that

$$\left|\bar{b}(x) - b(x)\right| \leq C(1 + |x|^{2\alpha})\Delta t. \tag{5.73}$$

The proof is similar to that for Scheme (5.49) with (5.10). From the discussions in cases A-C, we then conclude that

$$|I_1| \leq \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p}\, ds\right] + C(\Delta t)^p. \tag{5.74}$$

By Young's inequality, Hölder's inequality and Lemma 5.3.8, we have

$$\begin{aligned}
I_2 &= \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}(b(Y_{\theta(s)}) - b(Y_s)) ds\right] \leq \mathbb{E}\left[\int_0^\tau \left|\epsilon_s\right|^{2p-1}\left|b(Y_{\theta(s)}) - b(Y_s)\right| ds\right] \\
&\leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\mathbb{E}\left[\int_0^\tau (1 + |Y_{\theta(s)}|^{2\alpha-2} + |Y_s|^{2\alpha-2})^p |Y_{\theta(s)} - Y_s|^{2p} ds\right] \\
&\leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\int_0^T \mathbb{E}\left[|Y_{\theta(s)} - Y_s|^{4p}\right]^{\frac{1}{2}} ds \\
&\leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\Delta t^p.
\end{aligned} \tag{5.75}$$

In the last step, we have applied Lemma 5.3.9. With Assumption 5.2.1 (to be precise, (5.3)),

$$I_3 = \mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-2}(\epsilon_s)(b(Y_s) - b(X_s)) ds\right] \leq \beta\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right]. \tag{5.76}$$

Combining (5.74), (5.75) and (5.76), we have

$$I = 2p\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p-1}(\bar{b}(Y_{\theta(s)}) - b(X_s)) ds\right] = I_1 + I_2 + I_3 \leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\Delta t^p. \tag{5.77}$$

For III, we have from Assumption 5.2.1 and from Lemmas 5.3.1-5.3.6 that $\left|\bar{b}(x)\right| + |b(x)| \leq C(1+|x|^{\alpha})$. Then by Young's inequality twice and bounded moments (Lemmas 5.3.8 and 5.3.10), we have

$$
\begin{aligned}
III &= 2\mathbb{E}\left[\int_0^{\tau}\left\{2p|\epsilon_s|^{2p-1}(-\bar{b}(Y_{\theta(s)})) + \sigma^2 p(2p-1)(\epsilon_s)^{2p-2}(Y_{\theta(s)}^{2\gamma} + Y_{\theta(s)}^{\gamma}X_s^{\gamma})\right\}\mathbb{1}_{\{Z_s<0\}}ds\right] \\
&\leq C\mathbb{E}\left[\int_0^{\tau}\left\{|\epsilon_s|^{2p-1}(1+(Y_{\theta(s)})^{\alpha}) + (\epsilon_s)^{2p-2}(Y_{\theta(s)}^{2\gamma} + Y_{\theta(s)}^{\gamma}X_s^{\gamma})\right\}\mathbb{1}_{\{Z_s<0\}}ds\right] \\
&\leq C\mathbb{E}\left[\int_0^{\tau}|\epsilon_s|^{2p}\,ds\right] + C\mathbb{E}\left[\int_0^{\tau}\mathbb{1}_{\{Z_s<0\}}\,ds\right].
\end{aligned}
$$

Then by Lemma 5.3.7, we have when Δt is sufficiently small,

$$
\begin{aligned}
III &\leq C\mathbb{E}\left[\int_0^{\tau}|\epsilon_s|^{2p}\,ds\right] + C\mathbb{E}\left[\int_0^{\tau}\mathbb{1}_{\{Z_s<0\}}\,ds\right] \\
&\leq C\mathbb{E}\left[\int_0^{\tau}|\epsilon_s|^{2p}\,ds\right] + C\exp(-C(\Delta t)^{1-2\gamma}) \\
&\leq C\mathbb{E}\left[\int_0^{\tau}|\epsilon_s|^{2p}\,ds\right] + C\Delta t^{p}. \qquad (5.78)
\end{aligned}
$$

Now let's consider the term II. Recall the definition of $(\eta(t))_{t\geq 0}$ in (5.62). Then we have

$$
\begin{aligned}
\mathbb{E}\left[\int_0^{\tau}(\epsilon_s)^{2p-2}(Y_{\theta(s)}^{\gamma} - X_s^{\gamma})^2 ds\right] &= \mathbb{E}\left[\int_0^{\tau}(\epsilon_s)^{2p-2}(Y_{\theta(s)}^{\gamma} - X_s^{\gamma})^2(Y_{\theta(s)}^{1-\gamma} + X_s^{1-\gamma})^2 d\eta(s)\right] \\
&\leq 4\gamma^2\mathbb{E}\left[\int_0^{\tau}(\epsilon_s)^{2p-2}(Y_{\theta(s)} - X_s)^2 d\eta(s)\right] \\
&\leq 8\gamma^2\mathbb{E}\left[\int_0^{\tau}(\epsilon_s)^{2p-2}[(Y_{\theta(s)} - Y_s)^2 + (\epsilon_s)^2]d\eta(s)\right],
\end{aligned}
$$

where we have applied the following inequality in the second line, for $\frac{1}{2} < \gamma < 1$,

$$
\forall x \geq 0, y \geq 0, |x^{\gamma} - y^{\gamma}|(x^{1-\gamma} + y^{1-\gamma}) \leq 2\gamma|x-y|.
$$

Then by Young's inequality we have

$$
\begin{aligned}
&\mathbb{E}[\int_0^{\tau}(\epsilon_s)^{2p-2}(Y_{\theta(s)}^{\gamma} - X_s^{\gamma})^2 ds] \\
\leq\ & C\mathbb{E}\left[\int_0^{\tau}|\epsilon_s|^{2p}d\eta(s)\right] + C\mathbb{E}\left[\int_0^{\tau}(Y_{\theta(s)} - Y_s)^{2p}d\eta(s)\right] \\
\leq\ & C\mathbb{E}\left[\int_0^{\tau}(\epsilon_s)^{2p}d(s+\eta(s))\right] + C\mathbb{E}\left[\int_0^{\tau}(Y_{\theta(s)} - Y_s)^{2p}d(s+\eta(s))\right]. \qquad (5.79)
\end{aligned}
$$

The second term in last inequality can be estimated as follows. By (5.62), Hölder's inequality, Lemma 5.3.9, and Lemma 5.3.10, we have

$$\begin{aligned}
& \mathbb{E}\left[\int_0^\tau (Y_{\theta(s)} - Y_s)^{2p} d(s+\eta(s))\right] \\
= \ & \mathbb{E}\left[\int_0^\tau (Y_{\theta(s)} - Y_s)^{2p} \left(1 + \frac{1}{(Y_{\theta(s)}^{1-\gamma} + X_s^{1-\gamma})^2}\right) ds\right] \\
\leq \ & \mathbb{E}\left[\int_0^T |Y_{\theta(s)} - Y_s|^{2p} \left(1 + \frac{1}{X_s^{2-2\gamma}}\right) ds\right] \leq C\Delta t^p + C\left(\int_0^T \mathbb{E}[|Y_{\theta(s)} - Y_s|^{3p}]^{\frac{2}{3}} (\mathbb{E}[X_s^{(6\gamma-6)}]^{\frac{1}{3}}) ds\right) \\
\leq \ & C\Delta t^p \left(1 + \sup_{s\in[0,T]} \left(\mathbb{E}[X_s^{(6\gamma-6)}]\right)^{\frac{1}{3}}\right).
\end{aligned} \tag{5.80}$$

Then by (5.68), (5.77), (5.79) (with (5.80)), and (5.78), we have

$$\mathbb{E}[\epsilon_\tau^{2p}] \leq C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} ds\right] + C\mathbb{E}\left[\int_0^\tau (\epsilon_s)^{2p} d(s+\eta(s))\right] + C\Delta t^p. \tag{5.81}$$

The inequality (5.81) does not lead to our conclusion. Then similar to the proof of the Theorem 2.2 in [5], we reach the desired conclusion. Here we present the detailed proof for completeness. Define a stopping time $\tau_v = \inf\{s \in [0,T], \eta(s) + s \geq v\}$, where $v \in \mathbb{R}^+$ and $\inf \emptyset = T$, where $\eta(t)$ is defined in (5.62). Observing that $\tau_v + \eta(\tau_v) = v$, we apply the change of time $u = s + \eta(s)$ to obtain that $\mathbb{E}\left[\int_0^{\tau_v} (\epsilon_s)^{2p} d(s+\eta(s))\right] \leq \int_0^v \sup_{s\leq u} \mathbb{E}[(\epsilon_{\tau_s})^{2p}] du$. Then by (5.81), we have

$$\mathbb{E}[|\epsilon_{\tau_v}|^{2p}] \leq C\int_0^v \sup_{s\leq u} \mathbb{E}[|\epsilon_{\tau_s}|^{2p}] du + C(\Delta t)^p. \tag{5.82}$$

By the Gronwall inequality, we have

$$\mathbb{E}[|\epsilon_{\tau_v}|^{2p}] \leq C(\Delta t)^p \exp(Cv). \tag{5.83}$$

It follows from (5.81) that

$$\mathbb{E}[|\epsilon_t|^{2p}] \leq C\mathbb{E}\left[\int_0^{\eta(T)+T} (\epsilon_{\tau_u})^{2p} du\right] + C(\Delta t)^p. \tag{5.84}$$

By (5.83) and the Cauchy-Schwartz inequality, we obtain that

$$\begin{aligned}
\mathbb{E}\left[\int_0^{\eta(T)+T} (\epsilon_{\tau_u})^{2p} du\right] &= \int_0^{+\infty} \mathbb{E}\left[\mathbb{1}_{\{\eta(T)+T\geq u\}} (\epsilon_{\tau_u})^{2p}\right] du \\
&= \int_0^T \mathbb{E}\left[(\epsilon_{\tau_u})^{2p}\right] du + \int_T^{+\infty} \mathbb{E}\left[\mathbb{1}_{\{\eta(T)+T\geq u\}} (\epsilon_{\tau_u})^{2p}\right] du \\
&\leq C(\Delta t)^p \left[1 + \int_0^\infty \mathbb{P}(\eta(T) \geq u)^{\frac{1}{2}} \exp(Cu) du\right].
\end{aligned} \tag{5.85}$$

The proof is complete if we can show that $\int_0^\infty \mathbb{P}(\chi(T) \geq u)^{\frac{1}{2}} \exp(Cu) du < \infty$, where C depends on T, X_0 and coefficients of the SDE (5.1). In fact, by Markov inequality, we observe that for any $\mu > 0$

$$\left(\mathbb{P}(\chi(T) \geq u)\right)^{\frac{1}{2}} \leq \exp(-\mu u)(\mathbb{E}[\exp(2\mu\chi(T))])^{\frac{1}{2}}. \tag{5.86}$$

By Lemma 5.3.11, we have $\mathbb{E}[\exp(2\mu\chi(T))] \leq C_\mu$, where $C_\mu > 0$ depends on μ, T, X_0 and also on the coefficients of the SDE (5.1). Then $\int_0^\infty \mathbb{P}(\chi(T) \geq u)^{\frac{1}{2}} \exp(Cu)du \leq C_\mu^{\frac{1}{2}} \int_0^\infty \exp((C-\mu)u)du < \infty$ when μ is larger than C.

5.4 Numerical Examples

In this section, we test the schemes (5.9) with (5.10)-(5.14) using two examples. One is a scalar equation of the form (5.1). The other is two-dimensional SDE including (5.1) as its first dimension.

To test accuracy, we ran M independent trajectories $X^{(i)}(t)$ and computed the mean squared error

$$(\mathbb{E}[|X_T - X_{t_N}|^2])^{\frac{1}{2}} = \left(\frac{1}{M}\sum_{i=1}^{M}[X_T^{(i)} - X_{t_N}^{(i)}]^2\right)^{\frac{1}{2}}, \tag{5.87}$$

where $M = 10^4$. The reference solutions are computed by small time steps $h = 10^{-4}$ using the same numerical scheme. The number of trajectories $M = 10^4$ is sufficiently large for the statistical errors not to hinder the mean squared errors significantly. In our numerical test, Monte Carlo errors are computed with 95% confidence level and are at least 10 times smaller than the reported mean-squared errors. All experiments are performed using Matlab R2017a and random numbers are generated by Matlab command `rng(100,'twister')`.

In both examples, we test two cases for the power of diffusion coefficients, $\gamma = 0.5$ or 0.8. While we only have a proof for the case $0.5 < \gamma < 1$, we expect the convergence order for $\gamma = 0.5$ will be half as in [5], where the drift coefficient $b(X_t)$ is Lipschitz continuous.

Example 1. *Let $b(x) = 1 - x^3$ in (5.1) and $\sigma = 1$. The SDE then reads*

$$dX_t = (1 - X_t^3)dt + X_t^\gamma dW_t, \quad X_0 = 0.5. \tag{5.88}$$

Here $\alpha = 3$ in (5.4). To compare with the schemes (5.10)-(5.14), we also consider the backward Euler scheme

$$X_{t_{k+1}} = X_{t_k} + \Delta t(1 - X_{t_{k+1}}^3) + \sigma X_{t_k}^\gamma \sqrt{\Delta t}\xi_k, \tag{5.89}$$

where ξ_k's are i.i.d. standard normal random variables. In the following, we present the schemes (5.10)-(5.14) for the equation in the example.

$$X_{t_{k+1}} = \left| X_{t_k} + \frac{(1 - X_{t_k}^3)\Delta t}{1 + |X_{t_k}|^3\sqrt{\Delta t}} + \sigma X_{t_k}^\gamma \sqrt{\Delta t}\xi_k \right|, \tag{5.90}$$

$$X_{t_{k+1}} = \left| X_{t_k} + \Delta t(1 - \widehat{X_{t_k}}^3) + \sigma X_{t_k}^\gamma \sqrt{\Delta t}\xi_k \right|, \quad \widehat{X_{t_k}} = \frac{X_{t_k}}{|X_{t_k}|} \min\left(|X_{t_k}|, \Delta t^{-\frac{1}{4}}\right), \tag{5.91}$$

$$X_{t_{k+1}} = \left| X_{t_k} + \frac{(1 - X_{t_k}^3)\Delta t}{1 + |X_{t_k}|^3\Delta t} + \sigma X_{t_k}^\gamma \sqrt{\Delta t}\xi_k \right|, \tag{5.92}$$

$$X_{t_{k+1}} = \left| X_{t_k} + \tanh(\Delta t(1 - X_{t_k}^3)) + \sigma X_{t_k}^\gamma \sqrt{\Delta t}\xi_k \right|, \tag{5.93}$$

$$X_{t_{k+1}} = \left| X_{t_k} + \frac{(1 - X_{t_k}^3)\Delta t}{1 + \Delta t|1 - X_{t_k}^3|} + \sigma X_{t_k}^\gamma \sqrt{\Delta t}\xi_k \right|. \tag{5.94}$$

In Table 5.1, we test the case $\gamma = 0.5$. The numerical results suggest that the mean-square convergence order is one half. For a fixed time step h, mean-square errors are at the same order of magnitude, while the scheme (5.90) produces the largest errors. It is not clear why the scheme (5.90) is less accurate in this example. From Table 5.1, if we want to have an accuracy of 10^{-2} , we can take h to be 0.005 for most schemes above except (5.90). For (5.90), we need a smaller time step size, at least $h = 0.001$.

Table 5.1: Mean-square errors of the scheme (5.9) with (5.10)-(5.14) for Equation (5.88) and convergence rates when $\gamma = 0.5$ at $T = 1$.

h	(5.90)	rate	(5.91)	rate	(5.92)	rate
0.05	1.26e-01		6.00e-02		6.28e-02	
0.02	7.77e-02	0.53	3.25e-02	0.67	3.35e-02	0.69
0.01	5.29e-02	0.56	2.17e-02	0.58	2.21e-02	0.60
0.005	2.03e-02	0.60	**9.14e-03**	0.54	**9.16e-03**	0.55
0.001	**1.29e-02**	0.65	6.31e-03	0.53	6.32e-03	0.54
h	(5.93)	rate	(5.94)	rate	(5.89)	rate
0.05	5.98e-02		6.30e-02		5.49e-02	
0.02	3.25e-02	0.67	3.37e-02	0.68	3.13e-02	0.61
0.01	2.17e-02	0.58	2.22e-02	0.60	2.13e-02	0.56
0.005	**9.14e-03**	0.54	**9.18e-03**	0.55	**9.12e-03**	0.53
0.001	6.31e-03	0.53	6.33e-03	0.54	6.31e-03	0.52

We present the mean-square errors for the aforementioned scheme in Table 5.2 when $\gamma = 0.8$. We observe similar effects of mean-square errors as in Table 5.1.

Table 5.2: Mean-square errors of the scheme (5.9) with (5.10)-(5.14) for Equation (5.88) and convergence rates when $\gamma = 0.8$ at $T = 1$.

h	(5.90)	rate	(5.91)	rate	(5.92)	rate
0.05	1.57e-01		6.87e-02		7.26e-02	
0.02	9.98e-02	0.49	3.59e-02	0.71	3.72e-02	0.73
0.01	6.70e-02	0.58	2.39e-02	0.59	2.44e-02	0.61
0.005	2.42e-02	0.63	**9.93e-03**	0.54	**9.96e-03**	0.56
0.001	**1.52e-02**	0.67	6.80e-03	0.54	6.82e-03	0.55
h	(5.90)	rate	(5.91)	rate	(5.92)	rate
0.05	6.83e-02		7.19e-02		5.81e-02	
0.02	3.59e-02	0.70	3.70e-02	0.72	3.37e-02	0.59
0.01	2.39e-02	0.59	2.43e-02	0.60	2.31e-02	0.54
0.005	**9.93e-03**	0.54	**9.97e-03**	0.55	**9.90e-03**	0.53
0.001	6.81e-03	0.54	6.82e-03	0.55	6.81e-03	0.54

Example 2. *Consider the following test model*

$$\begin{aligned} dX_t = &\quad (1 - X_t^3)dt + X_t^\gamma dW_t^1, \\ dS_t = &\quad \mu S_t dt + \sqrt{X_t} S_t(\rho dW_t^1 + \sqrt{1-\rho^2} dW_t^2), \end{aligned} \tag{5.95}$$

where W_t^1 *and* W_t^2 *are two independent standard Brownian motions,* $\mu = 0.5, \rho = -0.7$ *and initial values are* $S_0 = 1, X_0 = 0.5$.

Here we test two cases again: $\gamma = 0.5$ and $\gamma = 0.8$. For X_t, we used the same set of schemes (5.10)-(5.14) as in the previous example. For S_t, we used the forward Euler scheme. For this example, we expect that X_t and S_t both converge in the mean-square sense at the order 1/2 as we proved the half-order mean-square convergence of schemes (5.10)-(5.14) and it is known that the forward Euler scheme has a half-order mean-square convergence for SDEs with Lipschitz continuous coefficients. In this example, we consider a two-factor Heston model. We compute both

Table 5.3: Mean-square errors of S_t, using the scheme the scheme (5.9) with (5.10)-(5.14) for X_t plus forward Euler scheme for S_t in Equation (5.95) and convergence rates when $\gamma = 0.5$ at $T = 1$.

h	(5.90)	rate	(5.91)	rate	(5.92)	rate
0.05	3.26e-01		3.31e-01		3.27e-01	
0.02	2.14e-01	0.46	2.17e-01	0.46	2.14e-01	0.46
0.01	1.54e-01	0.47	1.56e-01	0.48	1.54e-01	0.47
0.005	**6.94e-02**	0.50	**7.03e-02**	0.49	**6.93e-02**	0.50
0.001	4.81e-02	0.53	4.89e-02	0.53	4.81e-02	0.53
h	(5.93)	rate	(5.94)	rate	(5.89)	rate
0.05	3.27e-01		3.27e-01		3.25e-01	
0.02	2.14e-01	0.46	2.13e-01	0.46	2.13e-01	0.46
0.01	1.54e-01	0.48	1.54e-01	0.47	1.54e-01	0.47
0.005	**6.93e-02**	0.50	**6.94e-02**	0.50	**6.93e-02**	0.49
0.001	4.81e-02	0.53	4.81e-02	0.53	4.80e-02	0.53

mean-square errors and statistical errors. For $\gamma = 0.5$, the rate of convergence for S_t is close to one half as indicated in Table 5.3, which matches our expectations on the convergence order. We also observe that the accuracy of all schemes for S_t is at the same magnitude when $\gamma = 0.5$ and $\gamma = 0.8$, see Tables 5.3 and 5.4.

Table 5.4: Mean-square errors of S_t using the scheme (5.9) with (5.10)-(5.14) for X_t plus forward Euler scheme for S_t in Equation (5.95) and convergence rates when $\gamma = 0.8$ at $T = 1$.

h	(5.90)	rate	(5.91)	rate	(5.92)	rate
0.05	3.46e-01		3.50e-01		3.47e-01	
0.02	2.25e-01	0.47	2.27e-01	0.47	2.25e-01	0.47
0.01	1.62e-01	0.47	1.63e-01	0.47	1.62e-01	0.47
0.005	7.25e-02	0.49	7.34e-02	0.50	7.25e-02	0.50
0.001	5.05e-02	0.52	5.11e-02	0.52	5.04e-02	0.53

h	(5.93)	rate	(5.94)	rate	(5.89)	rate
0.05	3.46e-01		3.50e-01		3.47e-01	
0.02	2.25e-01	0.47	2.27e-01	0.47	2.25e-01	0.47
0.01	1.62e-01	0.47	1.63e-01	0.47	1.62e-01	0.47
0.005	7.25e-02	0.49	7.34e-02	0.50	7.25e-02	0.50
0.001	5.05e-02	0.52	5.11e-02	0.52	5.04e-02	0.53

Chapter 6

Conclusion and Future Work

6.1 Conclusion

We work on the parameter estimation on fractional stochastic heat equations in one-dimensional space and fractional Navier Stokes equations driven by additive noise in two-dimensional space. We also investigated several positivity-preserving numerical schemes on solving the SDEs with non-Lipschitz coefficients.

In chapter 2, we derived the closed-form of parameters in the fractional heat equation, including the coefficient of the fractional terms, and order of fractional Laplacian term. We state and prove the consistency and asymptotic normality for the parameters. The proof mainly relies on the strong law of large numbers and martingale representation theorem. We verified that the global maximum of the log-likelihood function exists. It is checked by the negative definiteness of its Hessian matrix, the negative definiteness is proved by verifying the sign of principal minors, by using properties of the total positive matrix. As a preview on the parameter estimation, I showed an example of the stochastic heat equations with the error of 10^{-5}. Moreover, we present several numerical examples on the coefficient of the fractional Laplacian terms θ and fractional order α respectively. In the estimation of θ, the estimation is close to the exact value with an accuracy of 4×10^{-2}. With the decrease of the time step Δt, the error decreases to 2×10^{-2}. Given the same number of Fourier modes, the error is less volatile when the fractional order α is close to 1. The estimation of the fractional order α is solved by finding zeros of the first order derivative of the log-likelihood function with respect to α. The accuracy is around 1×10^{-3} with a good choice of the initial guess when the Fourier mode N is 200. The accuracy of the estimation depends on the size of time steps and initial guess on the iteration methods.

In chapter 3, we worked on the stochastic fractional Navier-Stokes Equation with the periodic boundary condition. The goal is to estimate the coefficient of the fractional Laplacian θ and fractional order α. We applied the divergence-free projection and spectral methods on solving the fractional Navier-Stokes equation numerically and computed the Fourier coefficients using the fast Fourier transform and fixed point iterations. We constructed the analytic form of the fractional Laplacian coefficients. The error of θ is around 1×10^{-3}. The fractional order α is estimated by the same procedures in chapter 2. The error of α is around 4×10^{-2}.

In chapter 4, we applied the neural networks techniques on the inverse problem of PDEs with fractional Laplacian terms. The fractional Laplacian term is defined as the directional derivatives. Compared with the current research work, we applied the generalized Gauss-Laguerre quadrature rule instead of using the finite difference schemes and shifted.

In chapter 5, we developed explicit schemes preserving the positivity of solutions to SDEs with non-globally Lipschitz drift and Hölder continuous diffusion coefficients. We present five explicit positivity-preserving schemes. These schemes are modified symmetrized Euler schemes. We discussed several choices for non-globally Lipschitz drift among the state-of-the-art tamed schemes and truncation schemes. We present several numerical examples using these numerical schemes in both 1D and 2D cases and make comparisons in computational performance to show the half-order convergence.

6.2 Future Work

This work is the very first step to what applications of fractional Navier-Stokes equation. There are a lot of issues to explore.

In chapters 2 and 3, we have derived the maximum likelihood estimators for parameters in the fractional heat and fractional Navier-Stokes equations driven by additive noise. However, parameters α and β, θ, λ are estimated separately. Combining the estimator for α and the estimators for β, θ, λ, use alternating direction method to estimate all parameters. This method would require a good initial guess from either α or β, θ, λ. The further numerical study should be investigated before convergence analysis.

In chapter 3, error estimates on the modified estimations for the coefficient of fractional Laplacian θ should be provided.

In both chapters, estimators based on other metrics, such as KL divergence, may be developed, especially the most important parameter α.

In chapter 4, numerical simulations for FPINNs should be performed to obtain the numerical solutions as well as the parameter estimation.

In chapter 5, we have introduced several positivity preserving numerical schemes and proved the half-order convergence for these schemes when $\frac{1}{2} < \gamma < 1$. The case when $\gamma = \frac{1}{2}$ should be analyzed as well, which is used in the classic interest rate CIR model.

Bibliography

[1] Shin'ichi Aihara. Parameter identification for stochastic parabolic systems. In *Systems and control*, pages 1–12. Mita, Tokyo, 1991.

[2] Aurélien Alfonsi. On the discretization schemes for the CIR (and Bessel squared) processes. *Monte Carlo Methods Appl.*, 11(4):355–384, 2005.

[3] Aurélien Alfonsi. Strong order one convergence of a drift implicit Euler scheme: application to the CIR process. *Statist. Probab. Lett.*, 83(2):602–607, 2013.

[4] Peter W. Bates. On some nonlocal evolution equations arising in materials science. In *Nonlinear dynamics and evolution equations*, volume 48 of *Fields Inst. Commun.*, pages 13–52. Amer. Math. Soc., Providence, RI, 2006.

[5] Abdel Berkaoui, Mireille Bossy, and Awa Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM Probab. Stat.*, 12:1–11 (electronic), 2008.

[6] Wolf-Jürgen Beyn, Elena Isaak, and Raphael Kruse. Stochastic C-stability and B-consistency of explicit and implicit Euler-type schemes. *J. Sci. Comput.*, 67(3):955–987, 2016.

[7] Mireille Bossy and Awa Diop. An efficient discretisation scheme for one dimensional SDEs with a diffusion coefficient function of the form $|x|^a$, $a \in [1/2, 1)$, 2007. Version 2.

[8] Young-Ming Chen, Hsuan-Chu Li, and Eng-Tjioe Tan. An explicit factorization of totally positive generalized Vandermonde matrices avoiding Schur functions. *Appl. Math. E-Notes*, 8:138–147, 2008.

[9] Igor Cialenco. Parameter estimation for SPDEs with multiplicative fractional noise. *Stoch. Dyn.*, 10(4):561–576, 2010.

[10] Igor Cialenco. Statistical inference for SPDEs: an overview. *Stat. Inference Stoch. Process.*, 21(2):309–329, 2018.

[11] Igor Cialenco and Nathan Glatt-Holtz. Parameter estimation for the stochastically perturbed Navier-Stokes equations. *Stochastic Process. Appl.*, 121(4):701–724, 2011.

[12] Igor Cialenco, Ruoting Gong, and Yicong Huang. Trajectory fitting estimators for SPDEs driven by additive noise. *Stat. Inference Stoch. Process.*, 21(1):1–19, 2018.

[13] Igor Cialenco and Sergey V. Lototsky. Parameter estimation in diagonalizable bilinear stochastic parabolic equations. *Stat. Inference Stoch. Process.*, 12(3):203–219, 2009.

[14] Rama Cont and Peter Tankov. *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004.

[15] Kacha Dzhaparidze and Peter Spreij. The strong law of large numbers for martingales with deterministic quadratic variation. *Stochastics Stochastics Rep.*, 42(1):53–65, 1993.

[16] Brenden P. Epps and Benoit Cushman-Roisin. Turbulence modeling via the fractional laplacian. *arXiv:1803.05286.*, 2018.

[17] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, second edition, 2010.

[18] István Gyöngy. A note on Euler's approximations. *Potential Anal.*, 8(3):205–216, 1998.

[19] Nikolaos Halidias. Constructing positivity preserving numerical schemes for the two-factor CIR model. *Monte Carlo Methods Appl.*, 21(4):313–323, 2015.

[20] Desmond J. Higham, Xuerong Mao, and Andrew M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063, 2002.

[21] F. B. Hildebrand. *Introduction to numerical analysis.* McGraw-Hill Book Company, Inc., New York-Toronto-London, 1956.

[22] Yaozhong Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. In *Stochastic analysis and related topics*, pages 183–202. Birkhäuser Boston, Boston, MA, 1996.

[23] M. Hutzenthaler and A. Jentzen. On a perturbation theory and on strong convergence rates for stochastic ordinary and partial differential equations with non-globally monotone coefficients. *ArXiv*, Jan 2014.

[24] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. A*, (2130):1563–1576, 2011.

[25] Martin Hutzenthaler and Arnulf Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Mem. Amer. Math. Soc.*, 236(1112):v+99, 2015.

[26] Martin Hutzenthaler, Arnulf Jentzen, and Peter E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22(4):1611–1641, 2012.

[27] Martin Hutzenthaler, Arnulf Jentzen, and Xiaojie Wang. Exponential integrability properties of numerical approximation processes for nonlinear stochastic differential equations. *Math. Comp.*, 87(311):1353–1413, 2018.

[28] C. Kahl, M. Günther, and T. Rossberg. Structure preserving stochastic integration schemes in interest rate derivative modeling. *Appl. Numer. Math.*, 58(3):284–295, 2008.

[29] J. Kautsky and S. Elhay. Gauss quadratures and Jacobi matrices for weight functions not of one sign. *Math. Comp.*, 43(168):543–550, 1984.

[30] Peter Kloeden and Andreas Neuenkirch. Convergence of numerical methods for stochastic differential equations in mathematical finance. In *Recent developments in computational finance*, pages 49–80. World Sci. Publ., Hackensack, NJ, 2013.

[31] R. Sh. Liptser and A. N. Shiryayev. *Theory of martingales*, volume 49 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1989. Translated from the Russian by K. Dzjaparidze [Kacha Dzhaparidze].

[32] Robert S. Liptser and Albert N. Shiryaev. *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition, 2001. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

[33] Anna Lischke, Guofei Pang, Mamikon Gulian, and et al. What is the fractional Laplacian? A comparative review with new results. *J. Comput. Phys.*, 404:109009, 62, 2020.

[34] Wei Liu and Xuerong Mao. Strong convergence of the stopped Euler-Maruyama method for nonlinear stochastic differential equations. *Appl. Math. Comput.*, 223(0):389 – 400, 2013.

[35] Sergey V. Lototsky and Boris L. Rozovsky. *Stochastic partial differential equations.* Universitext. Springer, Cham, 2017.

[36] Xuerong Mao. *Stochastic differential equations and applications.* Horwood Publishing Limited, Chichester, second edition, 2008.

[37] Xuerong Mao. The truncated Euler–Maruyama method for stochastic differential equations. *J. Comput. Appl. Math.*, 290:370–384, 2015.

[38] Xuerong Mao and L. Szpruch. Strong convergence rates for backward Euler-Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients. *Stochastics*, 85(1):144–171, 2013.

[39] Xuerong Mao and Lukasz Szpruch. Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients. *J. Comput. Appl. Math.*, 238:14–28, 2013.

[40] Ralf Metzler and Joseph Klafter. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.*, 339(1):77, 2000.

[41] G. N. Milstein and M. V. Tretyakov. *Stochastic numerics for mathematical physics.* Springer-Verlag, Berlin, 2004.

[42] G. N. Milstein and M. V. Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Numer. Anal.*, 43(3):1139–1154, 2005.

[43] Andreas Neuenkirch and Lukasz Szpruch. First order strong approximations of scalar SDEs defined in a domain. *Numer. Math.*, 128(1):103–136, 2014.

[44] Hoang-Long Ngo and Duc-Trong Luong. Strong rate of tamed euler-maruyama approximation for stochastic differential equations with hölder continuous diffusion coefficient. *Braz. J. Prob. Stat., available online at http://imstat. org/bjps/papers/BJPS301. pdf*, 2015.

[45] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[46] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2005. An introduction with applications.

[47] Guofei Pang, Wen Chen, and K. Y. Sze. Gauss-Jacobi-type quadrature rules for fractional directional integrals. *Comput. Math. Appl.*, 66(5):597–607, 2013.

[48] Guofei Pang, Lu Lu, and George Em Karniadakis. fPINNs: fractional physics-informed neural networks. *SIAM J. Sci. Comput.*, 41(4):A2603–A2626, 2019.

[49] B. L. S. Prakasa Rao. Bayes estimation for some stochastic partial differential equations. volume 91, pages 511–524. 2000. Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998).

[50] B. L. S. Prakasa Rao. Nonparametric inference for a class of stochastic partial differential equations based on discrete observations. *Sankhyā Ser. A*, 64(1):1–15, 2002.

[51] Humberto Rafeiro and Stefan Samko. Fractional integrals and derivatives: mapping properties. *Fract. Calc. Appl. Anal.*, 19(3):580–607, 2016.

[52] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.

[53] Alexandra Rodkina and Henri Schurz. On positivity and boundedness of solutions of nonlinear stochastic difference equations. *Discrete Contin. Dyn. Syst.*, (Dynamical systems, differential equations and applications. 7th AIMS Conference, suppl.):640–649, 2009.

[54] Sotirios Sabanis. A note on tamed Euler approximations. *Electron. Commun. Probab.*, 18:no. 47, 1–10, 2013.

[55] Sotirios Sabanis. Euler approximations with varying coefficients: the case of superlinearly growing diffusion coefficients. *Ann. Appl. Probab.*, 26(4):2083–2105, 2016.

[56] Stefan G. Samko, Anatoly A. Kilbas, and Oleg I. Marichev. *Fractional integrals and derivatives*. Gordon and Breach Science Publishers, Yverdon, 1993. Theory and applications, Edited and with a foreword by S. M. Nikolʹskiĭ, Translated from the 1987 Russian original, Revised by the authors.

[57] Béla J. Szekeres. Turbulence modeling using fractional derivatives. In *Mathematical problems in meteorological modelling*, volume 24 of *Math. Ind.*, pages 47–55. Springer, [Cham], 2016.

[58] Ł ukasz Szpruch and Xiling Zhang. V-integrability, asymptotic stability and comparison property of explicit numerical schemes for non-linear SDEs. *Math. Comp.*, 87(310):755–783, 2018.

[59] Vidar Thomée. *Galerkin finite element methods for parabolic problems*. Springer-Verlag, Berlin, second edition, 2006.

[60] M. V. Tretyakov and Z. Zhang. A fundamental mean-square convergence theorem for SDEs with locally Lipschitz coefficients and its applications. *SIAM J. Numer. Anal.*, 51(6):3135–3162, 2013.

[61] John B. Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 265–439. Springer, Berlin, 1986.

[62] Xiaojie Wang and Siqing Gan. The tamed Milstein method for commutative stochastic differential equations with non-globally Lipschitz continuous coefficients. *J. Difference Equ. Appl.*, 19(3):466–490, 2013.

[63] Dongkun Zhang, Lu Lu, Ling Guo, and George Em Karniadakis. Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *J. Comput. Phys.*, 397:108850, 19, 2019.

[64] Zhongqiang Zhang and Heping Ma. Order-preserving strong schemes for SDEs with locally Lipschitz coefficients. *Appl. Numer. Math.*, 112:1–16, 2017.

[65] Lijing Zhao and Weihua Deng. A series of high-order quasi-compact schemes for space fractional diffusion equations based on the superconvergent approximations for fractional derivatives. *Numer. Methods Partial Differential Equations*, 31(5):1345–1381, 2015.

[66] Xiaofeng Zong, Fuke Wu, and Chengming Huang. Convergence and stability of the semi-tamed Euler scheme for stochastic differential equations with non-Lipschitz continuous coefficients. *Appl. Math. Comput.*, 228(0):240–250, 2014.