Single-Particle Dynamics in Nanoscopic Systems: Statistical Modeling and Inference

by

Feiyu Zhu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Alexandros Beskos Reader Department of Statistical Science University College London
Supervisor(s):	Martin Lysy Associate Professor Department of Statistics and Actuarial Science University of Waterloo
Internal Member(s):	Paul Marriott Professor Department of Statistics and Actuarial Science University of Waterloo
	Gregory Rice Associate Professor Department of Statistics and Actuarial Science University of Waterloo
Internal-External Member:	Roger G. Melko Professor Department of Physics and Astronomy University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

I am the sole author of Chapter 1, Chapter 4 and Chapter 5.

Chapter 2 has been prepared for submission to a peer-reviewed journal. It has been edited by my supervisor Dr. Martin Lysy. I contributed to all proofs in this Chapter and drafted the initial version of the manuscript.

Chapter 3 is adapted from the following paper:

Lysy, M., Zhu, F., Yates, B., and Labuda, A. (2022). Robust and efficient parametric spectral density estimation for high-throughput data. *Technometrics*, 64(1):30–51.

My contribution to this paper was development of the theoretical content, simulated and real data analysis, and lead authorship of the accompanying R/C++ package:

Zhu, F. and Lysy, M. (2021). *realPSD: Robust and Efficient Calibration of Parametric Power Spectral Density Models*. https://github.com/mlysy/realPSD.

Chapter 3 presents only the elements of the aforementioned paper in which I was directly involved in developing and writing.

Abstract

Our work aims to solve some of the most significant and fundamental theoretical problems involved in the current statistical modeling of stochastic processes in single-molecule experiments, for which a well recognized yet mathematically very difficult model is the generalized Langevin equation (GLE). We mainly focus on the following three directions.

In Chapter 2, we prove a remarkable representation theorem that fundamentally connects a continuous stationary process, a widely adopted statistical model in various applications, with the GLE, a ubiquitous tool in physics to model stochastic dynamics in a thermodynamic system. However, there are two important statistical challenges. First, the dynamics of the observed particle must be deconvoluted from a postulated covariance structure of an unobservable thermal force, rendering statistical modeling typically intractable. Second, for a given covariance structure of the latent force, the likelihood function for parameter estimation can rarely be written in closed form in the time domain. Parameter estimation that involves numerical approximation of the given memory kernel via repeated application of the Fast Fourier Transform (FFT) often incurs significant information loss and computational burden. We aim to fill the gaps by establishing a representation theorem that any continuous stationary process can be represented by a physically valid GLE. The upshot is that statistical modeling and inference can be performed entirely in the time domain with the guarantee of satisfying the fundamental laws of physics. The result can also be extended to continuous process with only stationary increments.

In Chapter 3, we carefully study the asymptotic properties of some important spectral density estimators for high-throughput (HTP) data commonly obtained in modern nanoscopic scientific experiments where the sampling frequency of an underlying process set to be extremely large and the recordings of such a continuous process can also be extended for a very long time, giving us more and more observations. Traditional asymptotic results with fixed sampling frequency would break down in such a situation. In the current literature, the asymptotic results for the spectral density estimator given HTP data are rarely seen to the best of our knowledge at the time of writing our work (Lysy et al., 2022). We fill this gap by laying the theoretical foundations for high-frequency sampled stationary processes, based on which a novel and effective two-stage approach was proposed by Lysy et al. (2022) to get a robust and efficient parametric estimation for the noisy HTP data.

In Chapter 4, we design an original non-degenerate sampling scheme using the particle filter method with bridge proposal to better estimate parameters in the quasi-Markovian approximation of the GLE which exhibits hypoellipticity. The proposed method if evaluated numerically would be extremely helpful for efficient parameter estimation in statistical modeling and inference problems tackling the nonlinear GLE.

Acknowledgements

First and foremost, I would like to express my eternal gratitude to my Ph.D. supervisor, Martin. I am privileged to collaborate with him working on some of the most difficult yet fascinating problems in the interdisciplinary field of statistics and physics. He generously gave me the extraordinary idea that any continuous stationary process can be represented as the generalized Langevin equation that he had been pondering for more than a decade. I am incredibly honored that I could be the one who collaborated with him to rigorously realize such a remarkable conjecture bit by bit in the past three years. It would never be exaggerated how much I learned from him in both statistical reasoning and programming paradigm. This thesis would never be possible without the patient guidance and constant help from him. His passion and rigorous altitude towards academic research as well as his sense of responsibility have always been an outstanding example for me to follow.

My sincere thanks go to Dr. Paul Marriott for offering me admission to the doctoral program and kindly being one of my research supervisors for the first year. I am also grateful for his constructive comments on my thesis.

I truly appreciate the help from Dr. Gregory Rice, and thank him for all the insightful discussions of the project covered by Chapter 3, as well as his inspiring teaching in STAT 929: Time Series. I can still remember the meeting we had in his office where we discussed some arguably the most delicate mathematical details involved in the discrete Fourier transform, periodogram and high-frequency asymptotics, etc. I am also grateful that he lent me one of his most treasured books — *Stochastic Curve Estimation* written by Murray Rosenblatt, from which I benefited a lot.

I would also like to thank the other committee members Dr. Alexandros Beskos and Dr. Roger Melko with my sincere gratitude for serving on my thesis committee and also for their time and effort to review my thesis, providing thoughtful and valuable comments.

The faculty and staff in the Department of Statistics and Actuarial Science are amazing. I acknowledge all their collaborative efforts to build our department such a lovely place. In particular, I offer my heartfelt thanks to Ms. Mary Lou Dufton and Ms. Lisa Baxter who helped me with all the necessary paperwork for graduation.

My collaborators Mr. Bryan Yates and Dr. Aleksander Labuda definitely deserve my special acknowledgement. Without their previous work, it would be impossible for me to get a paper published before my graduation in *Technometrics*, a leading journal in the subfield of statistics for physical, chemical, and engineering sciences.

I will not forget how much I was inspired by my brilliant peers. Thanks Mohan Wu for your help in the derivation of the bridge proposal designed in Chapter 4. Thanks Michelle

Ko for our discussions about programming in Julia language. Thank you to all the fellow members in the Differential Equation (DE) research group led by Martin.

Moreover, no words can express my sense of gratitude towards all my Ph.D. colleagues and lifelong friends, especially to those in my office M3 4233, Zhiyi Shen, Yumin Wang, and Qiuqi Wang, as well as my best roommate Han Weng. It is the serendipity that led me to meet all of you. Thank you all for making my Ph.D. life much more colorful and meaningful.

Finally, I am indebted to my parents in Wuhan, China, for their enduring love and selfless support for almost 30 years, especially for their courage and optimism at the very beginning of the COVID-19 outbreak when thousands of people died around them within merely a month back in February 2020. Even in their hardest time, they were always showing me their spiritual strength and caring love, encouraging me to continue my research and overcome all the difficulties in my life. My deepest gratitude and warmest affection are always dedicated to them.

Dedication

To my parents, for everything

Table of Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

There's Plenty of Room at the Bottom.
— Richard Feynman, 1959

Back in 1959, the famous theoretical physicist Richard Feynman gave a talk titled "There's Plenty of Room at the Bottom" at an annual meeting of the American Physical Society at Caltech. In this lecture, Feynman imagined a day when a new field of scientific research full of potential opportunities may be all about nanoparticles or even atoms, and a process by which the ability to manipulate individual atoms and molecules could be developed.

It was nearly 40 years later since then when an inspiring field of nanotechnology began to flourish.

1.1 Background and Challenges

Over the last two decades, advances in nanoscopic instrumentation have allowed researchers to constantly push down the bottom level of single particles that can be observed and traced in various scientific experiments (e.g. Nie and Zare, 1997; Radmacher, 1997; Xie and Trautman, 1998; Weiss, 1999; Loudet and Burgess, 2007; Sugimoto et al., 2007; Marchetto et al., 2008; Li et al., 2010), creating a proliferation of opportunities to study the dynamics of single particles interacting with their environment, leading to remarkable discoveries in a wide range of scientific disciplines (e.g. Xie and Lu, 1999; Tamarat et al., 2000; Weiss, 2000; Seisenberger et al., 2001; Valentine et al., 2001; Moerner, 2002; Yang et al., 2003;

Sikula and Levinshtein, 2004; Ebbinghaus et al., 2007; Zhang et al., 2007; Gerardi et al., 2009; Go et al., 2012). As the resolution of such experiments approaches the limits of deterministic modeling, the intrinsic stochastic dynamics of complex interacting particle systems emerges as a dominating factor, which on the one hand motivates the necessity of applying probability distribution and statistical models to extract the hidden information given the random or even noisy observations, but on the other hand poses new challenges for statistical modeling and inference in many new scientific fields as well.

Du and Kou (2019, 2020) recently gave an up-to-date survey of major statistical methodologies used in single-particle experiments and the current challenges faced by scientists in various research studies, e.g. subdiffusion dynamics (and anomalous diffusion phenomenon), heterogeneity in single-molecule trajectories, Fluorescence correlation spectroscopy (FCS), single-molecule signal analysis, polymer dynamics, etc. Moreover, Didier et al. (2012) also presented a great overview of the statistical challenges in microrheology which is the study of the properties of a complex fluid through the diffusion dynamics of small particles by using small tracers to probe material structure at micron length scale or smaller. All the statistical challenges can be summarized into three categories:

(i) Analytically complicated mathematical models are used extensively in single-molecule study. These models are often developed from the well-established fundamental principles of physics, in order to provide a clear interpretation or convincing intuition of the mechanisms of the underlying system. Many of the too intricate models cannot fit into the existing statistical inference framework due to either the intractable likelihood functions, or failure of efficient approximations in high-dimensional settings, or violation of classic assumptions (e.g. Markov property, linearity, independence etc.) behind many standard statistical models. Moreover, statistical modeling is typically restricted by the laws of physics. A valid statistical model should not violate first principles of physics, which to some extent limit the choices available to statisticians.

(ii) There is no well-established model comparison framework in single-particle studies. The existing model selection tools in statistics are designed for relatively generic linear models and therefore may not be suitable for comparing complicated nonlinear stochastic models. By using Bayesian approach, indeed, one may address the gap. However, it may be hard to design a suitable sampling algorithm to incorporate competing models under the same framework for comparison. A recent study by Lysy et al. (2016) made some contributions in this realm by developing rigorous and computationally efficient Bayesian methodology to compare two prevailing models in the study of single particle tracking.

(iii) Complex data with additional layers of complexity are usually produced in modern single-molecule experiments that involve the use of sophisticated measurement techniques. For example, in some experiments, the strong interactions between the measurement methods and the dynamics of molecules may alter the data generating process. Statisticians should attempt to incorporate the relevant information into the inference framework. Moreover, data at different coordinates can evolve on a wide range of timescales from fast and small vibrations between neighboring atoms linked by chemical bonds with a characteristic timescale of $t \approx 10^{-13}$s to large-scale conformational changes caused by folding of a protein molecule on timescales of at least $t \approx 10^{-6}$s, which gives rise to problems known as multiscale diffusions (Pavliotis and Stuart, 2007). For an overview of this application area, please see Schlick (2010).

Regarding the first type of challenges, in many applications under dynamic thermal environment, the fundamental laws of physics to govern a single particle trajectory are written in the form of a generalized Langevin equation (GLE) (e.g. Zwanzig, 2001; Tuckerman, 2010). As a ubiquitous model of single particle dynamics, the GLE is a stochastic integro-differential equation

$$m\ddot{x}(t) = -U'(x(t)) - \int_0^t K(t-u)\dot{x}(u)\mathrm{d}u + F(t) \tag{1.1}$$

for which parameter inference is an open challenge, even in the linear setting. Here $x(t)$ denotes the trajectory of a particle with mass m, and $\dot{x}(t)$ and $\ddot{x}(t)$, respectively, represent the first and second order derivatives of $x(t)$ with respect to time t (i.e. velocity and acceleration), $U(x(t))$ is the external potential energy (e.g. a magnetic field), $K(t)$ is a function called memory kernel which characterizes the non-Markovian frictional force and $F(t)$ is the latent random force.

To understand the GLE (1.1) better, let us conceptually decompose the total force exerted on the particle as follows:

$$\begin{aligned} \text{total force} &= \text{potential force} + \text{kinetic force} \\ F_{\text{total}} &= -U'(x(t)) + (F_{\text{friction}} + F_{\text{thermal}}) \end{aligned}$$

By the Newton's second law of motion,

$$F_{\text{total}} = \text{mass} \times \text{acceleration} = m\ddot{x}(t).$$

The potential force is set outside of the system depending on the experiment. The kinetic force comes from the particle's collisions with its environment, which can be further de-

composed into deterministic and stochastic components. For systems in thermodynamic equilibrium, the stochastic "thermal" force $F_{\text{thermal}} = F(t)$ is modeled as a mean-zero stationary process with autocorrelation function $\gamma_F(t)$. The deterministic "friction" force

$$F_{\text{friction}} = \int_0^t K(t-u)\dot{x}(u)du$$

depends on the history of the particle velocity, weighted by the memory kernel $K(t)$.

Remarkably, the precise GLE form (1.1) can be derived directly from the Hamiltonian dynamics — the fundamental laws governing any physical motion — for the so-called Kac–Zwanzig "heat-bath" interacting particle model (Ford and Kac, 1987; Ford et al., 1965; Kupferman, 2004). Moreover, due to the fluctuation-dissipation theorem (Kubo, 1966), the thermal force dynamics and the memory kernel are intrinsically related to each other, such that

$$\gamma_F(t) = k_B T \cdot K(t)$$

Due to the ubiquitous nature of the Kac–Zwanzig model, the GLE (1.1) has found applications in a wide range of scientific disciplines, including chemistry (Kantorovich, 2008; Kantorovich and Rompotis, 2008; Xing and Kim, 2011), climatology (Gritsun and Branstator, 2007), molecular biology (Kou and Xie, 2004; McKinley et al., 2009), molecular dynamics (Xiang et al., 1991; Lange and Grubmüller, 2006) and quantum physics (Celli et al., 2002; Ceriotti et al., 2011).

Given the prevalence of the GLE in single-particle dynamics modeling, however, there are two main challenges faced by statisticians. First, given some observations of particle trajectory, if we want to fit a GLE model, the dynamics of the observed particle must be deconvoluted from a postulated covariance structure of an unobservable thermal force [1], rendering statistical modeling typically intractable. Second, for a given covariance structure [2] of the latent force, the likelihood function for parameter estimation can rarely be written in closed form in the time domain. As a consequence, likelihood-based parameter estimation involves numerical approximation of the given memory kernel via repeated application of the Fast Fourier Transform (FFT), which incurs significant information loss.

Moreover, in systems far from equilibrium (Zwanzig, 1980), the analytic solution of the

[1] When the particle is under the influence of a harmonic potential, i.e., in the linear GLE case, the GLE can be solved analytically via Fourier analysis with a fractional Brownian motion noise embedded, as given by Kou (2008). But for a relatively general settings, without a good postulate of the random force, statistical modeling would be hard.

[2] For some experiments, researchers may have a good candidate of memory kernel or force structure in mind based on their domain knowledge.

nonlinear GLE is usually intractable. With the exhibit of non-Markovian property due to the convolution term involved, it is of huge difficulty to perform parameter inference for the nonlinear GLE. Some methods to approximate the GLE have been proposed, among which a very promising one is the quasi-Markovian approximation (Ottobre and Pavliotis, 2010) which convert the original GLE to a Markovian system of stochastic differential equations (SDEs) by introducing some auxiliary variables. However, the resulting SDE system — which we call qmGLE — is a hypoelliptic diffusion (Ottobre and Pavliotis, 2010), for which numerical discretization must be done with great care to perform parameter inference correctly (Pokern et al., 2009; Ditlevsen and Samson, 2019). To the best of our knowledge, the only existing method of inference for the qmGLE is the Bayesian Gibbs sampling algorithm proposed by Pokern et al. (2009), which can become arbitrarily inefficient as the desired accuracy of the numerical discretization increases (Roberts and Stramer, 2001).

1.2 Our Contribution and Outline

In this thesis, our focus is mainly on the theoretical challenges posed by the GLE, while we also discuss how to deal with a special type of complex data introduced by increasingly large sample size and ultra-high sampling frequency.

In Chapter 2, we rigorously prove a striking result that every twice differentiable weakly stationary process is the solution of the linear GLE arising from many possible interacting particle systems. Any such process can thus be used for statistical inference within a modeling framework which does not violate the laws of physics. We give explicit forms for the GLE parameters in terms of the autocorrelation of the observed particle process, and examine the accuracy of a computational approximation.

For a large sequence of complex data generated with ultra-high sampling frequency from complex (but stationary) underlying dynamics, e.g. the GLE process or CARFIMA models (as explained in Chapter 3), we can easily perform statistical inference without worrying about violating the underlying physical dynamics. Moreover, we derive the asymptotic properties of some most widely used parametric/semi-parametric estimators which extends the classical time series theory of asymptotics, which is given in Chapter 3.

As for fitting the nonlinear GLE, as long as the memory kernel can be written as a sum of exponential functions with some proper regularity (e.g. Fricks et al., 2009; McKinley et al., 2009; McKinley and Nguyen, 2018), the GLE can be well approximated by the qmGLE, for which we design an efficient parameter inference scheme using particle filters in Chapter 4.

Chapter 2

Generalized Langevin Equation Representation of Continuous Stationary Processes

2.1 Introduction

Over the last two decades, advances in nanoscopic instrumentation have allowed researchers to observe the dynamics of single particles interacting with their environment at unprecedented resolution and accuracy (e.g. Nie and Zare, 1997; Radmacher, 1997; Xie and Trautman, 1998; Weiss, 1999; Loudet and Burgess, 2007; Sugimoto et al., 2007; Marchetto et al., 2008; Li et al., 2010). This has led to remarkable discoveries in a wide range of scientific disciplines like physics, biophysics, chemistry, electrical engineering etc. (Xie and Lu, 1999; Tamarat et al., 2000; Weiss, 2000; Seisenberger et al., 2001; Valentine et al., 2001; Moerner, 2002; Yang et al., 2003; Sikula and Levinshtein, 2004; Ebbinghaus et al., 2007; Zhang et al., 2007; Gerardi et al., 2009; Go et al., 2012). As the resolution of such experiments approaches the limits of deterministic modeling, stochastic modeling of nanoscopic phenomena satisfying the fundamental laws of physics becomes of critical importance (Yang et al., 2003; Kou, 2008; Didier et al., 2012). In many applications, these laws are expressed in the form of a so-called generalized Langevin equation (GLE) (Kubo, 1966; Zwanzig, 2001).

Let $x(t)$ denote the displacement of a single particle in a thermodynamic environment.

The GLE for $x(t)$ is a stochastic integro-differential equation given by

$$m\ddot{x}(t) = -U'(x(t)) - \int_0^t K(t-u)\dot{x}(u)\mathrm{d}u + F(t) \tag{2.1}$$

In (2.1), m is the mass of the particle, and $\dot{x}(t) = \frac{\mathrm{d}}{\mathrm{d}t}x(t)$ and $\ddot{x}(t) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}x(t)$ are its velocity and acceleration. Heuristically, the GLE decomposes the total force acting on the particle — expressed using Newton's second law of motion on the left-hand side of (2.1) — into three terms on the right: the force $U'(x(t))$ due to the spatially-dependent potential energy $U(x)$ in the system (e.g., due to magnetic, electric or gravitational field etc.), the frictional force $\int_0^t K(t-u)\dot{x}(u)\mathrm{d}u$ relying on the history of the particle's velocity weighted by the so-called memory kernel $K(t)$, and the thermal force $F(t)$, modeled as a stationary stochastic process with mean zero and autocorrelation function (ACF)

$$\mathcal{C}_F(t) = \mathrm{cov}(F(t), F(0)).$$

In his description of Brownian motion, a key observation made by Einstein (Einstein, 1905, 1956; Nelson, 2001) is that the frictional and thermal forces acting on the particle originate from the same source, i.e., collisions with the enormous number of particles in the surrounding environment. Thus, if the system is in some sense stationary, the frictional and thermal forces ought to balance out. Indeed, according to the fluctuation-dissipation theorem (FDT) (Kubo, 1966), the time correlation $\mathcal{C}_F(t)$ of $F(t)$ and the memory kernel $K(t)$ of the friction are intrinsically related to each other, such that

$$\mathcal{C}_F(t) = k_B T \cdot K(t), \tag{2.2}$$

where k_B is Boltzmann's constant and T is the temperature (in Kelvin) of the system.

Remarkably, the heuristic decomposition above has a solid underpinning in theoretical physics. That is, the GLE can be rigorously derived from Hamiltonian dynamics — the fundamental laws governing any physical motion — for the so-called Kac–Zwanzig or "heat bath" interacting particle model (Ford et al., 1965; Mazur and Oppenheim, 1970; Albers et al., 1971; Zwanzig, 1973, 1980; Ciccotti and Ryckaert, 1981; Xiang et al., 1991; Zwanzig, 2001; Tuckerman, 2010). In a system of $N+1$ particles, the Kac–Zwanzig model described by the Hamiltonian function

$$H(x, v, \mathsf{q}, \mathsf{p}) = U(x) + \frac{v^2}{2m} + \frac{1}{2}\sum_{i=1}^{N} p_i^2 + (\alpha_i q_i - \gamma_i x)^2,$$

where x and v are the position and momentum of a "distinguished particle" in the system, and $\mathsf{q} = (q_1, \ldots, q_N)$, $\mathsf{p} = (p_1, \ldots, p_N)$ are the positions and momenta of the N heat bath particles in the surrounding environment. Letting $\Gamma_t = (x_t, v_t, \mathsf{q}_t, \mathsf{p}_t)$ denote the state of the system at time t, the evolution of the system is given by the Hamiltonian equations of motion

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} x_t &= \frac{\partial}{\partial v} H(\Gamma_t), & \frac{\mathrm{d}}{\mathrm{d}t} q_{it} &= \frac{\partial}{\partial p_i} H(\Gamma_t) \\ \frac{\mathrm{d}}{\mathrm{d}t} v_t &= -\frac{\partial}{\partial x} H(\Gamma_t), & \frac{\mathrm{d}}{\mathrm{d}t} p_{it} &= -\frac{\partial}{\partial q_i} H(\Gamma_t). \end{aligned} \tag{2.3}$$

When the initial state Γ_0 is drawn from the Boltzmann distribution of the system,

$$p_B(\Gamma) \propto \exp\left\{\frac{-H(\Gamma)}{k_B T}\right\},$$

an elementary calculation (e.g., Zwanzig, 2001) shows that the solution of (2.3) produces the GLE (2.1) with

$$K(t) = \sum_{i=1}^{N} \gamma_i^2 \cos(\alpha_i t),$$

and for which $F(t)$ a Gaussian process. The GLE with arbitrary $K(t)$ and Gaussian $F(t)$ can be obtained either by carefully taking $N \to \infty$ (Ariel and Vanden-Eijnden, 2009) or using infinite-dimensional Hilbert spaces (e.g., Rey-Bellet, 2006; Pavliotis, 2014). It is this latter approach we adopt for the developments in Sections 2.2.

Due to the ubiquitous nature of the heat-bath model, the GLE has found applications in a wide range of scientific topics and interdisciplinary fields, including atom-solid surface scattering (Adelman and Doll, 1974; Doll et al., 1975; Adelman and Doll, 1976; Doll and Dion, 1976), climatology (Gritsun and Branstator, 2007), conformational dynamics in proteins (Kou and Xie, 2004; Kou, 2008; Lange and Grubmüller, 2006), microrheology/polymer physics (Mason and Weitz, 1995; Mason et al., 1997; Fricks et al., 2009; McKinley et al., 2009; Panja, 2010; Squires and Mason, 2010; Démery et al., 2014; Hohenegger and McKinley, 2017; McKinley and Nguyen, 2018), molecular dynamics (Kantorovich, 2008; Kantorovich and Rompotis, 2008; Li et al., 2015), sampling of molecular systems (Celli et al., 2002; Ceriotti et al., 2011), and transport in magnetized plasmas (Krommes, 2018b,a), and even quantum mechanics (Cortés et al., 1985; Ford and Kac, 1987; McDowell, 2000; de Oliveira, 2020).

From a statistical perspective, the GLE is not only a fundamental physical model with countless applications, it also provides an interpretable framework for modeling $x(t)$.

Indeed, $x(t)$ is a stationary process with invariant (Boltzmann) distribution

$$p(x) \propto \exp\left\{\frac{-U(x)}{k_B T}\right\}$$

when the above is integrable. In this sense, physically valid statistical modeling of $x(t)$ reduces to specifying the stationary distribution of $x(t)$ and the autocorrelation of the thermal force $F(t)$. In many applications (e.g., Chandler, 1987; Kou and Xie, 2004), the potential energy can be well approximated by a harmonic potential,

$$U(x(t)) = \frac{1}{2}\kappa x^2(t),$$

in which case we obtain the "linear" GLE

$$m\ddot{x}(t) = -\kappa x(t) - \int_0^t K(t-u)\dot{x}(u)\mathrm{d}u + F(t). \tag{2.4}$$

If $F(t)$ is taken to be Gaussian, then so is $x(t)$. The statistical objective is now to infer the dynamics of $x(t)$ from discrete time observations $\boldsymbol{x} = (x(t_1), \ldots, x(t_n))$, where $t_n = n\Delta t$. However, two significant complications typically stand in the way.

First, what can be observed empirically are the particle positions $\boldsymbol{x}$. However, statistical modeling of the GLE requires one to postulate an ACF $\mathcal{C}_F(t)$ for the unobservable thermal force, of which the precise effect on position dynamics is difficult to deconvolute. As a result, statistical modeling in GLEs is often limited to asymptotic behavior (e.g., Morgado et al., 2002; Kou and Xie, 2004; Kneller, 2011; McKinley and Nguyen, 2018).

Second, even upon assuming that a parametric model $\mathcal{C}_F(t \mid \boldsymbol{\theta})$ for the force ACF is given, the likelihood function $L(\boldsymbol{\theta} \mid \boldsymbol{x})$ can rarely be written in closed form. This is because the ACF $\mathcal{C}_x(t \mid \boldsymbol{\theta})$ of $x(t)$ is typically available only in the Fourier domain. Therefore, likelihood-based parameter estimation involves numerical approximation of $\mathcal{C}_x(t \mid \boldsymbol{\theta})$ via repeated application of the Fast Fourier Transform (FFT), which is both computationally intensive and prone to information loss (Mason and Weitz, 1995; Solomon and Lu, 2001).

In the face of such challenges, one might wonder to what extent the theoretical appeal of the GLE model limits its use in statistical practice. The main result of this paper suggest the answer is "very little". Indeed, let

$$\mathfrak{L}\{f(t)\} = \mathfrak{L}\{f(t)\}(z) = \int_0^\infty e^{tz} f(t)\mathrm{d}t$$

denote the Laplace transform of a locally integrable exponentially bounded $f(t)$. Then we prove the following:

Theorem 2.1.1. *Let $x(t)$ be a twice differentiable weakly stationary process. Then there exists a Hamiltonian interacting particle system for which $x(t)$ is one of the observables, and $x(t)$ satisfies the linear GLE (2.4), with parameters explicitly given by*

$$\begin{aligned}
\kappa &= \frac{k_B T}{\mathrm{var}(x(t))}, & \mathrm{var}(x(t)) &= \langle x(t), x(t) \rangle \\
m &= \frac{k_B T}{\mathrm{var}(\dot{x}(t))}, & \mathrm{var}(\dot{x}(t)) &= \langle \dot{x}(t), \dot{x}(t) \rangle \\
\check{K}(z) &= m \left(\frac{-(z^2 + \kappa/m)\check{\mathcal{C}}_x(z) + z\mathcal{C}_x(0)}{z\check{\mathcal{C}}_x(z) - \mathcal{C}_x(0)} \right), & K(t) &= \mathfrak{L}^{-1}\{\check{K}(z)\} \\
\check{\mathcal{C}}_F(z) &= k_B T \cdot \check{K}(z), & \mathcal{C}_F(t) &= \mathfrak{L}^{-1}\{\check{\mathcal{C}}_F(z)\}.
\end{aligned}$$

The upshot of Theorem 2.1.1 is that *any* twice differentiable (weakly) stationary process $x(t)$ — Gaussian or otherwise — can be used to model nanoscopic phenomena without violating the laws of physics. This is a considerably stronger result (see Theorem 2.4.1) than related ones for only the autocorrelation functions (Okabe, 1986) or restricted to Gaussian processes (Pavliotis, 2014). Inference about $x(t)$ can be performed using any statistical method of choice, with the GLE parameters of scientific interest being recoverable post-inference using the formulas in Theorem (2.1.1).

The remainder of this Chapter is organized as follows. Section 2.2 defines Hamiltonian dynamics on infinite dimensional Hilbert spaces using semigroup theory. Sections 2.3 and 2.4 show how to construct arbitrary continuous stationary processes in such systems. Sections 2.5 and 2.6 derive the GLE for twice continuously differentiable stationary processes. Section 2.7 provides a means of computing certain GLE memory kernels using the fast Fourier transform (FFT). Section 2.8 closes with concluding remarks.

2.2 Hamiltonian Dynamics on Hilbert Spaces

We begin this section with a motivating example. Suppose that each possible state of a physical system is represented by a function of the form

$$\Gamma : \mathbb{R}^+ \to \mathbb{R}^2, \qquad \Gamma(x) = (\mathsf{q}(x), \mathsf{p}(x)).$$

in the Hilbert space $\mathscr{H}$ defined by the inner product

$$\langle \Gamma_1, \Gamma_2 \rangle = \int_0^\infty x \cdot \big[\mathsf{q}_1(x)\mathsf{q}_2(x) + \mathsf{p}_1(x)\mathsf{p}_2(x)\big]\mathrm{d}x, \qquad \Gamma_1, \Gamma_2 \in \mathscr{H}.$$

In the context of Hamiltonian systems, we may think of $\mathsf{q} = \mathsf{q}(x)$ as the positions of particles indexed by $x \in \mathbb{R}^+$, for which $\mathsf{p} = \mathsf{p}(x)$ denotes the (conjugate) momenta. In this context, $\Gamma = \Gamma(x)$ is called an element of the *phase space* $\mathscr{H}$.

Suppose that $\mathscr{H}$ is used to model an interacting particle system in thermal equilibrium, i.e., for which no energy is exchanged between the system and its surrounding environment. Then the laws of physics governing the system are determined by a function

$$\mathsf{H} : \mathbb{R}^5 \to \mathbb{R}; \qquad (x, \mathsf{q}, \mathsf{p}, \nabla\mathsf{q}, \nabla\mathsf{p}) \mapsto \mathsf{H}(x, \mathsf{q}, \mathsf{p}, \nabla\mathsf{q}, \nabla\mathsf{p})$$

and the corresponding functional

$$\mathcal{H} : \mathscr{H} \to \mathbb{R}, \qquad \mathcal{H}(\Gamma) = \int_0^\infty \mathsf{H}(x, \mathsf{q}(x), \mathsf{p}(x), \tfrac{\mathrm{d}}{\mathrm{d}x}\mathsf{q}(x), \tfrac{\mathrm{d}}{\mathrm{d}x}\mathsf{p}(x))\mathrm{d}x.$$

These are called, respectively, the *Hamiltonian density* (H) and the *Hamiltonian energy function* ($\mathcal{H}$) of the system. Let $\Gamma_t = \Gamma_t(x) = (\mathsf{q}_t(x), \mathsf{p}_t(x))$ denote the state of the system at time t. Then the time evolution of the system is given by

$$\begin{aligned} \dot{\mathsf{q}}_t(x) &= \frac{\partial}{\partial t}\mathsf{q}_t(x) = +\frac{\partial}{\partial \mathsf{p}}\mathcal{H}(\Gamma_t), \\ \dot{\mathsf{p}}_t(x) &= \frac{\partial}{\partial t}\mathsf{p}_t(x) = -\frac{\partial}{\partial \mathsf{q}}\mathcal{H}(\Gamma_t), \end{aligned} \tag{2.5}$$

where $\frac{\partial}{\partial \mathsf{q}}\mathcal{H}(\Gamma) = \frac{\partial}{\partial \mathsf{q}}\mathcal{H}(\mathsf{q}, \mathsf{p})$ denotes the functional derivative with respect to q,

$$\begin{aligned} \frac{\partial}{\partial \mathsf{q}}\mathcal{H}(\Gamma) : x \mapsto &\frac{\partial}{\partial \mathsf{q}}\mathsf{H}(x, \mathsf{q}(x), \mathsf{p}(x), \tfrac{\mathrm{d}}{\mathrm{d}x}\mathsf{q}(x), \tfrac{\mathrm{d}}{\mathrm{d}x}\mathsf{p}(x)) \\ &- \frac{\partial}{\partial x}\left[\frac{\partial}{\partial \nabla \mathsf{q}}\mathsf{H}(x, \mathsf{q}(x), \mathsf{p}(x), \tfrac{\mathrm{d}}{\mathrm{d}x}\mathsf{q}(x), \tfrac{\mathrm{d}}{\mathrm{d}x}\mathsf{p}(x))\right], \end{aligned}$$

and $\frac{\partial}{\partial \mathsf{p}}\mathcal{H}(\Gamma)$ is similarly defined.

The precise conditions under which $\mathcal{H}$ defines a valid physical system are given by Chernoff and Marsden (1975). As a concrete example, we take $\mathcal{H}$ to be defined by the norm of

$\mathscr{H}$, such that

$$\mathcal{H}(\Gamma) = \tfrac{1}{2}\langle \Gamma, \Gamma \rangle = \tfrac{1}{2}\|\Gamma\|^2.$$

The Hamiltonian equations of motion (2.5) then become

$$\begin{aligned} \dot{\mathsf{q}}_t &= x \cdot \mathsf{p}_t(x), \\ \dot{\mathsf{p}}_t &= -x \cdot \mathsf{q}_t(x), \end{aligned} \tag{2.6}$$

for which the solution in terms of the initial state of the system $\Gamma_0 = (\mathsf{q}_0, \mathsf{p}_0)$ is

$$\begin{bmatrix} \mathsf{q}_t(x) \\ \mathsf{p}_t(x) \end{bmatrix} = \begin{bmatrix} \cos(xt) & \sin(xt) \\ -\sin(xt) & \cos(xt) \end{bmatrix} \begin{bmatrix} \mathsf{q}_0(x) \\ \mathsf{p}_0(x) \end{bmatrix}. \tag{2.7}$$

More compactly, we may express (2.7) as $\Gamma_t = \Phi_t(\Gamma_0)$, and note that

$$\Phi_u \circ \Phi_t = \Phi_{u+t}, \qquad \|\Gamma_t\| = \|\Gamma_0\|.$$

In most systems of interest, the complete state Γ is far too large to observe directly. Instead, attention is focused on specific *observables* of the system, which are mathematically represented as functionals of the form $X : \mathscr{H} \to \mathbb{R}$. The simplest of these functionals are linear and continuous, the complete set of which is $\mathscr{H}'$, the continuous dual space of $\mathscr{H}$. By the Riesz representation theorem, $\mathscr{H}$ and $\mathscr{H}'$ are isometrically isomorphic. In particular, there is a unique $f = (\alpha, \beta) \in \mathscr{H}$ such that $X(\Gamma) = \langle f, \Gamma \rangle$. To simplify presentation, we shall identify X with f whenever confusion can be avoided, thus writing $X(\Gamma) = \langle X, \Gamma \rangle$. Similarly, we do not distinguish between the inner product on $\mathscr{H}$ or $\mathscr{H}'$, i.e., $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{\mathscr{H}} = \langle \cdot, \cdot \rangle_{\mathscr{H}'}$.

For a given observable $X \in \mathscr{H}'$, let us define the *trajectory* of X as the family of observables $\{X_t \in \mathscr{H}' : t \geq 0\}$ such that

$$X_t : \Gamma \to X(\Gamma_t) = X(\Phi_t(\Gamma)).$$

Then for $X = (\alpha, \beta)$ we have

$$\begin{aligned} X_t(\Gamma) &= \langle X, \Gamma_t \rangle \\ &= \int_0^\infty x \begin{bmatrix} \alpha(x) & \beta(x) \end{bmatrix} \begin{bmatrix} \cos(xt) & \sin(xt) \\ -\sin(xt) & \cos(xt) \end{bmatrix} \begin{bmatrix} \mathsf{q}_0(x) \\ \mathsf{p}_0(x) \end{bmatrix} \mathrm{d}x \\ &= \langle \Phi_{-t}(X), \Gamma_0 \rangle, \end{aligned}$$

and

$$\begin{aligned}\langle X_{u+t}, X_u\rangle &= \langle \Phi_{-u-t}(X), \Phi_{-u}(X)\rangle \\ &= \int_0^\infty x \cdot \begin{bmatrix}\alpha(x) & \beta(x)\end{bmatrix} \begin{bmatrix}\cos(xt) & -\sin(xt) \\ \sin(xt) & \cos(xt)\end{bmatrix} \begin{bmatrix}\alpha(x) \\ \beta(x)\end{bmatrix} \mathrm{d}x \\ &= \int_0^\infty \cos(xt) \cdot \big[x\alpha^2(x) + x\beta^2(x)\big]\mathrm{d}x,\end{aligned}$$

such that $\langle X_{u+t}, X_u\rangle$ does not depend on u. In fact, let $\mathcal{C}(t)$ denote the autocorrelation function (ACF) of any continuous stationary process. Then there exists (e.g., Khintchine, 1934; Itô, 1954) a non-negative symmetric finite measure $\mu(\omega)$ such that

$$\mathcal{C}(t) = 2\int_0^\infty \cos(\omega t)\mathrm{d}\mu(\omega).$$

If $\mathcal{C}(t)$ arises from a continuous stationary process which is purely nondeterministic in the sense of its Wold decomposition (Cramér, 1961), then $\mu(\omega)$ is an absolutely continuous measure with $\mathrm{d}\mu(\omega) = \mathcal{S}(\omega)\mathrm{d}\omega$ (Cramér, 1967). Thus, let $\alpha(x) = \beta(x) = (\mathcal{S}(x)/x)^{1/2}$. Then for $X = (\alpha, \beta)$ we have

$$\begin{aligned}\langle X_{u+t}, X_u\rangle &= \int_0^\infty \cos(xt) \cdot \big[x\alpha^2(x) + x\beta^2(x)\big]\mathrm{d}x \\ &= 2\int_0^\infty \cos(xt) \cdot \mathcal{S}(x)\mathrm{d}x = \mathcal{C}(t).\end{aligned} \tag{2.8}$$

Thus, any autocorrelation function $\mathcal{C}(t)$ arising for a power spectral density (PSD) $\mathcal{S}(\omega)$ can be expressed as $\mathcal{C}(t) = \langle X_{u+t}, X_u\rangle$ for a suitably chosen observable $X \in \mathscr{H}'$.

2.2.1 Hamiltonians and C_0-Semigroups

In light of the connection between ACFs and observables established in the motivating example above, we pose the following questions:

Q1. The representation result (2.8) is for the specific Hamiltonian system given by (2.6). In the same spirit, we can show the result also holds for some other well-known Hamiltonians (see Appendix A.2). Our question then becomes: To what extent does the result generalize to other Hamiltonian systems?

Q2. The result only reveals a connection between Hamiltonians and ACFs. How can this be extended to stochastic processes in Hamiltonian systems?

In order to address Q1, we now give the precise definition of Hamiltonian systems on infinite-dimensional separable Hilbert spaces established by Chernoff and Marsden (1975).

Definition 2.2.1. *Let $\{\Phi_t : t \geq 0\}$ be a family of bounded linear operators defined on a Banach space $\mathscr{B}$ (real or complex) satisfying the following conditions:*

(i) $\Phi_0 = \mathcal{I}$, *the identity operator.*

(ii) $\Phi_u \circ \Phi_t = \Phi_{u+t}$.

(iii) Φ_t *is strongly continuous, i.e.,* $\lim_{t\to 0^+} \|\Phi_t x - x\| = 0$ *for every* $x \in \mathscr{B}$.

Then Φ_t is called a C_0-semigroup *on $\mathscr{B}$. Furthermore, consider the (possibly unbounded) linear operator $\mathcal{L}$ defined by*

$$\mathcal{L}x = \lim_{h\to 0^+} \frac{1}{h}(\Phi_h x - x) \tag{2.9}$$

with domain $\mathrm{dom}(\mathcal{L})$ *consisting of all $x \in \mathscr{B}$ for which the limit* (2.9) *exists. Then $\mathcal{L}$ is called the* infinitesimal generator *of Φ_t.*

The generators of C_0-semigroups are characterized by the Hille–Yoshida theorem (Engel and Nagel, 2000). Of particular interest to us is when $\mathcal{L}$ is a skew-adjoint operator on a complex Hilbert space $\mathscr{H}$, in which case there is a one-to-one correspondence between such $\mathcal{L}$ and unitary C_0-semigroups, i.e., those for which $\langle \Phi_t x, \Phi_t y\rangle = \langle x, y\rangle$, due to the celebrated Stone's theorem (Stone, 1930, 1932; von Neumann, 1932; Reed and Simon, 1981; Lax, 2002) We return to this point in Section 2.3. We shall also eventually require a number of results about self-adjoint and skew-adjoint operators, which are detailed in Appendix A.1.

The notion of C_0-semigroups allows us to give a firm theoretical foundation for the so-called abstract Cauchy problem (ACP), i.e., for the solution of abstract differential equations of the form

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\Gamma_t &= \mathcal{L}\Gamma_t, \qquad t \geq 0 \\ \Gamma_0 &= x, \qquad x \in \mathscr{B}. \end{aligned} \tag{2.10}$$

Indeed, the following is a well-known result in semigroup theory:

Theorem 2.2.1 (Arendt et al., 2011, Theorem 3.1.12)**.** *Let $\mathcal{L}$ be the generator of a C_0-semigroup Φ_t. Then for $x \in \mathrm{dom}(\mathcal{L})$, $\Gamma_t = \Phi_t x$ is differentiable with respect to t and satisfies the ACP* (2.10)*, i.e., $\Gamma_0 = x$ and*

$$\lim_{h \to 0^+} \| \tfrac{1}{h}(\Gamma_{t+h} - \Gamma_t) - \mathcal{L}\Gamma_t \| = 0, \qquad t \geq 0.$$

Due to this result, the C_0-semigroup Φ_t with generator $\mathcal{L}$ is typically denoted $e^{t\mathcal{L}}$, a convention we follow for the remainder of this Chapter.

Definition 2.2.2. *Consider a Hilbert space $\mathscr{H}$ over the field $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ equipped with a mapping $\omega : \mathscr{H} \times \mathscr{H} \to \mathbb{F}$ having the following properties:*

(i) ω *is a bilinear form.*

(ii) $\omega(\Gamma, \Gamma) = 0$ *for all* $\Gamma \in \mathscr{H}$*.*

(iii) $\omega(\Gamma_1, \Gamma_2) = 0$ *for all* $\Gamma_2 \in \mathscr{H}$ *implies that* $\Gamma_1 = 0$*.*

Then $\omega(\cdot, \cdot)$ is called a symplectic form*.*

Symplectic forms and C_0-semigroups are the two main ingredients for characterizing "linear" Hamiltonian systems on infinite-dimensional separable Hilbert spaces.

Theorem 2.2.2 (Chernoff and Marsden, 1975, Theorem 1)**.** *Let $\mathscr{H}$ be an infinite-dimensional separable Hilbert space over $\mathbb{R}$ with symplectic form $\omega(\cdot, \cdot)$, and let $\mathcal{L}$ be the generator of a C_0-semigroup on $\mathscr{H}$. If $\mathcal{L}$ is skew-symmetric with respect to ω,*

$$\omega(\mathcal{L}\Gamma_1, \Gamma_2) = -\omega(\Gamma_1, \mathcal{L}\Gamma_2), \qquad \Gamma_1, \Gamma_2 \in \mathrm{dom}(\mathcal{L}),$$

then the ACP (2.10) *solved by $\Gamma_t = e^{t\mathcal{L}} x$ describes the dynamics of a Hamiltonian system whenever $x \in \mathrm{dom}(\mathcal{L})$, with the Hamiltonian energy function given by*

$$\mathcal{H}(\Gamma) = \frac{1}{2}\omega(\mathcal{L}\Gamma, \Gamma).$$

For the remainder of the paper, we shall refer to linear Hamiltonian systems on real Hilbert spaces as constructed above simply as Hamiltonian systems.

As in the motivating example, the observables of these more general Hamiltonian systems are the elements of the continuous dual space $\mathscr{H}'$ of $\mathscr{H}$, and the trajectory of an observable $X \in \mathscr{H}'$, is the family of observables $\{X_t \in \mathscr{H}' : t \geq 0\}$ such that

$$X_t : \Gamma \to X(\Gamma_t) = X(e^{t\mathcal{L}}\Gamma).$$

When $\mathcal{L}$ is skew-adjoint, since its generated C_0-semigroup is unitary (e.g. Hall, 2013), by the property of semigroup, it immediately follows that $\left(e^{t\mathcal{L}}\right)^\dagger = \left(e^{t\mathcal{L}}\right)^{-1} = e^{-t\mathcal{L}}$ on $\mathscr{H}$ where $\dagger$ denotes adjoint as explained in Appendix A.1, we may further have

$$X_t(\Gamma) = \langle X, e^{t\mathcal{L}}\Gamma\rangle = \langle e^{-t\mathcal{L}}X, \Gamma\rangle,$$

and thus $X_t = e^{-t\mathcal{L}}X$. But when $\mathcal{L}$ is unbounded in general, we may only define the adjoint semigroup $\left(e^{t\mathcal{L}}\right)^\dagger$ on the adjoint dual space of $\mathscr{H}$. For details, see van Neerven (1992).

Moreover, we have the following:

Proposition 2.2.1. *Consider an observable $X \in \mathscr{H}'$ with corresponding trajectory $\{X_t = e^{-t\mathcal{L}} : t \geq 0$. Then $\langle X_{u+t}, X_u\rangle$ is independent of u. In fact, $\mathcal{C}(t) = \langle X_t, X_0\rangle$ is continuous, symmetric, and positive-definite, i.e.,*

$$\sum_{i=1}^{N} \mathcal{C}(t_i - t_j) z_i \overline{z_j} \geq 0$$

for any choice of $t_1, \ldots, t_N \in \mathbb{R}$ and $z_1, \ldots, z_N \in \mathbb{C}$. Thus, $\mathcal{C}(t)$ is an autocorrelation function.

The proof is given in Appendix A.3.1.

2.3 Hamiltonian Representation of Autocorrelation Functions

Let $\mathscr{H}$ denote the real infinite-dimensional separable Hilbert space underlying the Hamiltonian system defined by the symplectic form $\omega(\cdot, \cdot)$ and the Liouville operator $\mathcal{L}$. For the results to follow, we restrict our attention to Hamiltonian systems in which $\mathscr{H}$ can be embedded into a complex Hilbert space $\mathscr{H}_C$, in the sense that $x \in \mathscr{H} \iff x \in \mathscr{H}_C$, in such a way that $\mathcal{L}$ is a skew-adjoint operator of $\mathscr{H}_C$.

That is, let $\mathscr{H}_C = \{V_C, \langle\cdot,\cdot\rangle_C\}$ be a complex separable Hilbert space with inner product $\langle\cdot,\cdot\rangle_C$. Then $\mathscr{H}_C$ can be constructed from a "base" real Hilbert space $\mathscr{H}_0 = \{V_0, \langle\cdot,\cdot\rangle_0\}$, with $V_C = V_0 \times V_0$ and

$$\begin{aligned}\langle\Gamma_1,\Gamma_2\rangle_C &= \langle(\mathsf{q}_1,\mathsf{p}_1),(\mathsf{q}_2,\mathsf{p}_2)\rangle_C \\ &= \langle\mathsf{q}_1,\mathsf{q}_2\rangle_0 + \langle\mathsf{p}_1,\mathsf{p}_2\rangle_0 + i\cdot\big(\langle\mathsf{q}_1,\mathsf{p}_2\rangle_0 - \langle\mathsf{q}_2,\mathsf{p}_1\rangle_0\big) \\ &= \langle\Gamma_1,\Gamma_2\rangle_R + i\cdot\omega(\Gamma_1,\Gamma_2),\end{aligned}$$

where $\langle\cdot,\cdot\rangle_R$ is the inner product on the direct sum (real) Hilbert space $\mathscr{H}_R = \mathscr{H}_0 \oplus \mathscr{H}_0$, and $\omega(\cdot,\cdot)$ is a symplectic form.

Now suppose that $\mathcal{L}$ is a skew-adjoint operator on $\mathscr{H}_C$. Then by Stone's theorem $e^{t\mathcal{L}}$ is a C_0-semigroup on $\mathscr{H}_C$, and it is straightforward to check that it is also a C_0-semigroup on $\mathscr{H}_R$. Moreover, by the definition of skew-adjointness we have

$$\omega(\mathcal{L}\Gamma_1,\Gamma_2) = -\omega(\Gamma_1,\mathcal{L}\Gamma_2), \quad \forall\ \Gamma_1,\Gamma_2 \in \mathrm{dom}(\mathcal{L}).$$

Thus, $e^{t\mathcal{L}}$ generates the Hamiltonian equations of motion on the real Hilbert space $\mathscr{H}_R$ by Chernoff and Marsden (1975) Theorem 1.

Example 1. *The Hamiltonian system in the motivating example can be recast into the form above with*

$$\mathscr{H}_0 = \left\{ f : \mathbb{R}^+ \to \mathbb{R} : \int_0^\infty x|f(x)|^2 \mathrm{d}x < \infty \right\}$$

and

$$\mathcal{L} = \begin{bmatrix} 0 & x \\ -x & 0 \end{bmatrix}, \qquad \mathrm{dom}(\mathcal{L}) = \left\{ (\mathsf{q},\mathsf{p}) \in \mathscr{H}_0 \oplus \mathscr{H}_0 : \int_0^\infty x^2[\mathsf{q}(x)^2 + \mathsf{p}(x)^2]\mathrm{d}x < \infty \right\}.$$

Example 2. *A famous example of an infinite-dimensional Hamiltonian system is defined as follows (Chernoff and Marsden, 1975). Let H^k denote the Sobolev space with square-integrable k-th order derivatives (details in Appendix A.2.1 and A.2.2). The Hilbert space underlying the Hamiltonian system is*

$$\mathscr{H} = H^1 \oplus L^2,$$

the symplectic form is

$$\omega((\mathsf{q}_1,\mathsf{p}_1),(\mathsf{q}_2,\mathsf{p}_2)) = \langle\mathsf{q}_1,\mathsf{p}_2\rangle - \langle\mathsf{q}_2,\mathsf{p}_1\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the L^2 inner product, and the Liouville operator is

$$\mathcal{L} = \begin{bmatrix} 0 & I \\ -m^2 I + \partial_x^2 & 0 \end{bmatrix}, \qquad \mathrm{dom}(\mathcal{L}) = H^2 \oplus H^1,$$

where ∂_x^2 is the second order differential operator. The Hamiltonian equations of motion are

$$\begin{aligned} \dot{\mathsf{q}}_t &= \mathsf{p}_t \\ \dot{\mathsf{p}}_t &= \partial_x^2 \mathsf{q}_t - m^2 \mathsf{q}_t, \end{aligned}$$

or upon rearranging (and with slightly different notation),

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} + m^2 \right) \mathsf{q}(x,t) = 0.$$

This is known as the Klein–Gordon equation in physics. For $m = 0$, it reduces to the wave equation. In either case, the Hamiltonian system cannot be embedded as desired into a complex Hilbert space $\mathscr{H}_C$.

In Proposition 2.2.1, we showed that for any observable $X \in \mathscr{H}'$, $\langle X_u, X_t \rangle$ is the covariance function of a stationary process with autocorrelation function $\mathcal{C}(t) = \langle X_t, X_0 \rangle$. By Bochner's theorem (e.g., Lax, 2002, Theorem 8, Chapter 14), any autocorrelation function $\mathcal{C}(t)$ is the Fourier transform of a nonnegative, symmetric measure ν, such that

$$\mathcal{C}(t) = \int_{\mathbb{R}} e^{it\omega} \mathrm{d}\nu(\omega).$$

It turns out that a skew-adjoint operator $\mathcal{L}$ on a complex Hilbert space $\mathscr{H}_C$ can also be associated with a nonnegative measure μ on $\mathbb{R}$, as a consequence of the spectral theorem (Appendix A.1.1 and A.1.2). This leads us to the answer of Q1, i.e., to characterize exactly when an autocorrelation function can be expressed as the inner product between observables along a trajectory $\{X_t \in \mathscr{H}'\}$.

Theorem 2.3.1. *Consider the real Hamiltonian system $\{\mathscr{H}, \omega(\cdot,\cdot), \mathcal{L}\}$, and suppose it can be embedded in the complex Hilbert space $\mathscr{H}_C$ on which $\mathcal{L}$ is skew-adjoint. Then an autocorrelation function*

$$\mathcal{C}(t) = \int_{\mathbb{R}} e^{it\omega} \mathrm{d}\nu(\omega)$$

can be represented as $\langle e^{-t\mathcal{L}} X, X \rangle$ for an observable $X \in \mathscr{H}'$ if and only if ν is absolutely

continuous with respect to the measure μ associated with $\{\mathscr{H}_C, \mathcal{L}\}$ via the spectral theorem.

The proof is given in Appendix A.3.2.

We close this Section by noting that a wide variety of ACFs can be represented by the wave equation and Klein–Gordon Hamiltonian systems of Example 2 as shown in Appendix A.2. That being said, for the remainder of this Chapter we assume that whenever we have a Hamiltonian system $\{\mathscr{H}, \omega(\cdot,\cdot), \mathcal{L}\}$, the real Hilbert space $\mathscr{H}$ can be embedded in a complex Hilbert space $\mathscr{H}_C$ on which $\mathcal{L}$ is skew-adjoint.

2.4 Stochastic Processes on Hamiltonian Systems

The basic idea for constructing a stochastic process on a Hamiltonian system is merely to sample the initial phase-space variable Γ_0 from some distribution $p(\Gamma_0)$. Then for any observable $X \in \mathscr{H}'$,

$$X_t(\Gamma) = \langle e^{-t\mathcal{L}} X, \Gamma \rangle$$

is a well-defined stochastic process.

Ideally, we would like to do this in such a way that

$$\operatorname{cov}(X_s, X_t) = \langle X_s, X_t \rangle,$$

i.e., such that the covariance function of the stochastic process coincides with the inner product of the corresponding observable trajectory. However, this turns out to be impossible even in the Gaussian case (e.g., Gross, 1967). Rather, one must proceed via the construction of abstract Wiener measures (Gross, 1967). Such a construction guarantees that there exists a space $\mathscr{H}^* \supsetneq \mathscr{H}$ along with a bilinear form $\langle \cdot, \cdot \rangle^* : \mathscr{H} \times \mathscr{H}^* \to \mathbb{C}$ such that

$$\langle \cdot, f \rangle^* = \langle \cdot, f \rangle, \qquad f \in \mathscr{H}, \tag{2.11}$$

and a probability distribution $p(\Gamma^*)$ on $\mathscr{H}^*$ such that for any $X_1, \ldots, X_N \in \mathscr{H}$, sampling $\Gamma^* \sim p(\Gamma^*)$ results in

$$(\langle X_1, \Gamma^* \rangle^*, \ldots, \langle X_N, \Gamma^* \rangle^*) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \qquad \mathbf{\Sigma}_{ij} = \langle X_i, X_j \rangle.$$

Consequently, $\langle e^{-t\mathcal{L}} X, \Gamma^* \rangle^*$ is a Gaussian stochastic process with autocorrelation $\langle e^{-t\mathcal{L}} X, X \rangle$.

In Appendix A.3.3, we generalize this idea to address Q2:

Theorem 2.4.1. *Consider the real Hamiltonian system* $\{\mathscr{H}, \omega(\cdot,\cdot), \mathcal{L}\}$*, and suppose it can be embedded into the complex Hilbert space* $\mathscr{H}_C$ *on which* $\mathcal{L}$ *is skew-adjoint. Let*

$$\mathcal{C}(t) = \langle e^{-t\mathcal{L}} X, X\rangle$$

be an autocorrelation function obtained from a given $X \in \mathscr{H}'$*. Then for any continuous stationary process* X_t *having autocorrelation function* $\mathcal{C}(t)$*, there exists a space* $\mathscr{H}^* \supsetneq \mathscr{H}$*, a bilinear form* $\langle\cdot,\cdot\rangle^* : \mathscr{H} \times \mathscr{H}^* \to \mathbb{C}$ *satisfying* (2.11)*, and a probability distribution* $p(\Gamma^*)$ *on* $\mathscr{H}^*$ *such that*

$$X_t = \langle e^{-t\mathcal{L}} X, \Gamma^*\rangle^*, \qquad \Gamma^* \sim p(\Gamma^*).$$

2.5 Mori–Zwanzig Projections

The GLE for Hamiltonian systems is obtained by a procedure known as the Mori–Zwanzig projection technique (Mori, 1965b; Zwanzig, 1973, 2001; Nordholm and Zwanzig, 1975). The results in this section, having been derived in the references above and below for the N-particle setting in irreversible statistical mechanics (Zwanzig, 1960, 1961), are extended here to the setting of infinite-dimensional (separable) Hilbert spaces.

Definition 2.5.1. *Consider the observables* $\mathcal{A}_1, \ldots, \mathcal{A}_d \in \mathscr{H}'$*. The projection operator* $\mathcal{P}$ *is the operator on* $\mathscr{H}'$ *defined by*

$$\mathcal{P}\mathcal{B} = \langle \mathcal{B}, \boldsymbol{\mathcal{A}}\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}}^{-1} \boldsymbol{\mathcal{A}}^\dagger\,, \quad \mathcal{B} \in \mathscr{H}'$$

where

$$\begin{aligned} \boldsymbol{\mathcal{A}} &= [\mathcal{A}_1, \ldots, \mathcal{A}_d]_{1\times d}\,, \\ \langle \mathcal{B}, \boldsymbol{\mathcal{A}}\rangle &= (\langle \mathcal{B}, \mathcal{A}_1\rangle, \ldots, \langle \mathcal{B}, \mathcal{A}_d\rangle)_{1\times d} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}\ i,j} &= \langle \mathcal{A}_i, \mathcal{A}_j\rangle, \quad i,j = 1,\ldots,d. \end{aligned}$$

Projection operators have the following elementary properties (proved in Appendix A.3.4 and A.3.5.):

Proposition 2.5.1. *The projection operator* $\mathcal{P}$ *and* $\mathcal{Q} = \mathcal{I} - \mathcal{P}$ *are both self-adjoint.*

Proposition 2.5.2. *The projection operator* $\mathcal{P}$ *and* $\mathcal{P}\mathcal{L}$ *are bounded. In particular, we have* $\|\mathcal{P}\| = 1$*.*

By the bounded perturbation theorem (Engel and Nagel, 2000, III.1.3), $\mathcal{QL} = \mathcal{L} - \mathcal{PL}$ is the generator of a strongly continuous semigroup $e^{t\mathcal{QL}}$. Therefore, we can well establish the exponential operators $e^{t\mathcal{PL}}$ and $e^{t\mathcal{QL}}$ as strongly continuous semigroups. $e^{t\mathcal{QL}}$ is also known as the orthogonal dynamics whose existence and uniqueness for classical dynamic systems under Mori's projection operator has been proved by Givon et al. (2004). Indeed, for the development to follow we require the operator identity

$$e^{t\mathcal{L}} = e^{t(\mathcal{I}-\mathcal{P})\mathcal{L}} + \int_0^t e^{(t-s)\mathcal{L}}\mathcal{PL}e^{s(\mathcal{I}-\mathcal{P})\mathcal{L}}\mathrm{d}s, \tag{2.12}$$

which is valid on $\mathrm{dom}(\mathcal{L})$ provided that $\boldsymbol{\mathcal{A}} \in \mathrm{dom}(\mathcal{L})$. Identity (2.12) is a type of Nakajima–Zwanzig equation (Nakajima, 1958; Zwanzig, 1960), and is also called a Dyson operator identity (Zhu et al., 2018). Indeed, it can be transformed into the Dyson-Phillips series using the Volterra operator (e.g., Engel and Nagel, 2000, III.1.8-10). Hereinafter, we will just call it the Zwanzig operator identity.

It is rigorously proved in Appendix A.3.6.

2.5.1 Mori–Zwanzig Formalism

Consider the observables $\boldsymbol{\mathcal{A}} = (\mathcal{A}_1, \ldots, \mathcal{A}_d), \mathcal{B} \in \mathrm{dom}(L) \subset \mathscr{H}'$ and suppose that $\mathcal{LB} \in \mathrm{dom}(\mathcal{L})$.

Let us start with the abstract Cauchy problem,

$$\dot{\mathcal{B}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{B}(t) = \mathcal{LB}(t), \qquad \mathcal{B}(0) = \mathcal{B}. \tag{2.13}$$

By Theorem 2.2.1, since $\mathcal{L}$ generates a strongly continuous unitary group, (2.13) has a unique classical solution

$$\mathcal{B}(t) = e^{t\mathcal{L}}\mathcal{B}(0), \quad \mathcal{B}(t) \in \mathrm{dom}(\mathcal{L}). \tag{2.14}$$

Substituting (2.14) into (2.13) and noting that $e^{t\mathcal{L}}$ and $\mathcal{L}$ commute, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{B}(t) = e^{t\mathcal{L}}\mathcal{LB}(0) = e^{t\mathcal{L}}\left(\mathcal{P}+\mathcal{I}-\mathcal{P}\right)\mathcal{LB}(0) = e^{t\mathcal{L}}\mathcal{PLB}(0) + e^{t\mathcal{L}}\left(\mathcal{I}-\mathcal{P}\right)\mathcal{LB}(0). \tag{2.15}$$

First, we focus on the first term on the right side, i.e.

$$\begin{aligned} e^{t\mathcal{L}}\mathcal{P}\mathcal{L}\mathcal{B}(0) &= e^{t\mathcal{L}}\mathcal{P}\left(\mathcal{L}\mathcal{B}(0)\right) = e^{t\mathcal{L}}\langle \mathcal{L}B(0), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1}\boldsymbol{\mathcal{A}}^{\dagger}(0) \\ &= \langle \mathcal{L}\mathcal{B}(0), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1} e^{t\mathcal{L}}\boldsymbol{\mathcal{A}}^{\dagger}(0) \\ &= \boldsymbol{\Omega}\boldsymbol{\mathcal{A}}^{\dagger}(t), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\Omega} &= \langle \mathcal{L}\mathcal{B}(0), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1}, \\ \boldsymbol{\mathcal{A}}(t) = e^{t\mathcal{L}}\boldsymbol{\mathcal{A}}(0) &= \left[e^{t\mathcal{L}}\mathcal{A}_1(0), \ldots, e^{t\mathcal{L}}\mathcal{A}_d(0)\right] = \left[\mathcal{A}_1(t), \ldots, \mathcal{A}_d(t)\right]. \end{aligned}$$

Now let us turn our attention to the second term on the right side of (2.15). Note that $(\mathcal{I}-\mathcal{P})\mathcal{L}\mathcal{B}(0) \in \mathrm{dom}(L)$, and thus applying the operator identity (2.12) to it gives

$$e^{t\mathcal{L}}(\mathcal{I}-\mathcal{P})\mathcal{L}\mathcal{B}(0) = F(t) + \int_0^t e^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}F(s)\mathrm{d}s$$

where

$$F(t) = e^{t(\mathcal{I}-\mathcal{P})\mathcal{L}}(\mathcal{I}-\mathcal{P})\mathcal{L}\mathcal{B}(0). \tag{2.16}$$

Moreover,

$$\begin{aligned} \int_0^t e^{(t-s)\mathcal{L}}\mathcal{P}\mathcal{L}F(s)\mathrm{d}s &= \int_0^t e^{(t-s)\mathcal{L}}\langle \mathcal{L}F(s), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1}\boldsymbol{\mathcal{A}}^{\dagger}(0)\mathrm{d}s \\ &= \int_0^t \langle \mathcal{L}F(s), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1} e^{(t-s)\mathcal{L}}\boldsymbol{\mathcal{A}}^{\dagger}(0)\mathrm{d}s \\ &= \int_0^t \langle \mathcal{L}F(s), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1}\boldsymbol{\mathcal{A}}^{\dagger}(t-s)\mathrm{d}s \\ &= -\int_0^t \boldsymbol{K}(s)\boldsymbol{\mathcal{A}}^{\dagger}(t-s)\mathrm{d}s \\ &= -\int_0^t \boldsymbol{K}(t-s)\boldsymbol{\mathcal{A}}^{\dagger}(s)\mathrm{d}s, \end{aligned}$$

where we define the $1 \times d$ matrix

$$\boldsymbol{K}(t) = -\langle \mathcal{L}F(t), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1} = \langle F(t), \mathcal{L}\boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{A}}(0)}^{-1}.$$

Finally, we combine all the pieces together to rewrite (2.15) as the Mori–Zwanzig projection

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{B}(t) = \boldsymbol{\Omega}\boldsymbol{\mathcal{A}}^\dagger(t) - \int_0^t \boldsymbol{K}(t-s)\boldsymbol{\mathcal{A}}^\dagger(s)\mathrm{d}s + F(t).$$

The following properties of the force term $F(t)$, proved in Appendix A.3.7, shall become useful momentarily.

Proposition 2.5.3. *The random force $F(t)$ is second-order stationary in the sense that $\langle F(s+t), F(s)\rangle = \langle F(t), F(0)\rangle$ is an ACF. Moreover, we have $\langle F(t), \boldsymbol{\mathcal{A}}(0)\rangle = \mathbf{0}$.*

2.6 The GLE Representation Theorem

Consider a Hamiltonian system $\{\mathscr{H}, \omega(\cdot,\cdot), \mathcal{L}\}$ on which $\mathcal{L}$ is skew-adjoint. Let $X \in \mathscr{H}'$ denote an observable of which the corresponding autocorrelation function

$$\mathcal{C}_X(t) = \langle e^{-t\mathcal{L}}X, X\rangle$$

is that of a twice continuously differentiable stationary process.

Lemma 2.6.1. *For the Hamiltonian system above, let $X \in \mathscr{H}'$ such that $\langle e^{-t\mathcal{L}}X, X\rangle$ is the autocorrelation of a k-times continuously differentiable stationary process. Then $\mathcal{L}^m X \in \mathrm{dom}(\mathcal{L})$ for $m = 0, \ldots, k-1$.*

This is proved in Appendix A.3.8. With Lemma 2.6.1 in hand, we may derive the GLE of Theorem 2.1.1 for the trajectory X_t by simply applying the Mori–Zwanzig formalism of Section 2.5.1 with $\boldsymbol{\mathcal{A}} = [X, \dot{X}]$ and $\mathcal{B} = \dot{X}$, where $\dot{X} = \mathcal{L}X$. Indeed, it is verified in Appendix A.3.9 that

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{\langle X, \ddot{X}\rangle}{\langle X, X\rangle}, & 0 \end{pmatrix}, \qquad \boldsymbol{\Omega}\boldsymbol{\mathcal{A}}^\dagger(t) = -\frac{\langle \dot{X}, \dot{X}\rangle}{\langle X, X\rangle}X(t),$$
$$\boldsymbol{K}(t) = \begin{pmatrix} 0, & \frac{\langle F(t), \ddot{X}\rangle}{\langle \dot{X}, \dot{X}\rangle} \end{pmatrix}, \quad \boldsymbol{K}(s)\boldsymbol{\mathcal{A}}^\dagger(t-s) = \frac{\mathcal{C}_F(s)}{\langle \dot{X}, \dot{X}\rangle}\dot{X}(t-s),$$

where $\ddot{X} = \mathcal{L}\dot{X}$ and $\mathcal{C}_F(t) = \langle F(t), F(0)\rangle$. Putting it all together we have

$$\ddot{X}(t) = -\theta X(t) - \int_0^t \gamma(t-s)\dot{X}(s)\mathrm{d}s + F(t), \tag{2.17}$$

where $\theta = \langle \dot{X}, \dot{X} \rangle / \langle X, X \rangle$ and $\gamma(t) = \mathcal{C}_F(t) / \langle \dot{X}, \dot{X} \rangle$. If we multiply both sides by $m = k_B T / \langle \dot{X}, \dot{X} \rangle$, we obtain the standard GLE (2.4)

$$m\ddot{X}(t) = -\kappa X(t) - \int_0^t K(t-s)\dot{X}(s)\mathrm{d}s + \tilde{F}(t), \tag{2.18}$$

where

$$\kappa = \frac{k_B T}{\langle X, X \rangle}, \qquad \tilde{F}(t) = mF(t), \qquad K(t) = k_B T \cdot \frac{\mathcal{C}_F(t)}{\langle \dot{X}, \dot{X} \rangle^2},$$

and we may verify the fluctuation-dissipation theorem (FDT) (2.2):

$$\mathcal{C}_{\tilde{F}}(t) = \langle \tilde{F}(t), \tilde{F}(0) \rangle = m^2 \langle F(t), F(0) \rangle = \left(\frac{k_B T}{\langle \dot{X}, \dot{X} \rangle} \right)^2 \mathcal{C}_F(t) = k_B T \cdot K(t).$$

2.6.1 Memory Kernel in the Laplace Domain

So far in the GLE of Theorem 2.1.1 we have obtained the explicit formulas for m and κ, but not for the memory kernel $K(t)$. We now produce such a result by passing to the Laplace domain.

Taking the inner product of (2.18) with $X(0)$ we obtain

$$m\langle \ddot{X}(t), X(0) \rangle = -\kappa \langle X(t), X(0) \rangle - \int_0^t K(t-s)\langle \dot{X}(s), X(0) \rangle \mathrm{d}s, \tag{2.19}$$

where we have eliminated the force term via $\langle \tilde{F}(t), X(0) \rangle = m\langle F(t), X(0) \rangle = 0$ as shown in Proposition 2.5.3.

Next, we make use of the following lemma proved in Appendix A.3.10:

Lemma 2.6.2. *For $\psi, \phi \in \mathscr{H}'$ let $\mathcal{C}_{\psi,\phi}(t) = \langle e^{-t\mathcal{L}}\psi, \phi \rangle$. If $\psi \in \mathrm{dom}(\mathcal{L})$, then $\mathcal{C}_{\psi,\phi}(t)$ is continuously differentiable and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}_{\psi,\phi}(t) = \langle e^{-t\mathcal{L}}\mathcal{L}\psi, \phi \rangle.$$

Thus, we may rewrite (2.19) as

$$m\ddot{\mathcal{C}}_X(t) = -\kappa\mathcal{C}_X(t) - \int_0^t K(t-s)\dot{\mathcal{C}}_X(s)\mathrm{d}s, \tag{2.20}$$

where $\dot{\mathcal{C}}_X(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}_X(t)$ and $\ddot{\mathcal{C}}_X(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{C}_X(t)$.

We now wish to take the Laplace transform of (2.20), and justify why this can be done term by term. First, we note that $K(t)$ is an ACF. Thus, we have the spectral representation

$$K(t) = \int_{\mathbb{R}} e^{-it\lambda}\mathrm{d}\mu(\lambda),$$

and for $z = a + ib \in \mathbb{C}$, note that

$$\begin{aligned}\left|\int_0^\infty e^{-zt}K(t)\mathrm{d}t\right| &= \left|\int_0^\infty e^{-zt}\int_{\mathbb{R}} e^{-it\lambda}\mathrm{d}\mu(\lambda)\mathrm{d}t\right| \\ &\leq \int_0^\infty e^{-|a|t}\int \left|e^{-it(b+\lambda)}\right|\mathrm{d}\mu(\lambda)\mathrm{d}t \leq \int_0^\infty e^{-|a|t}M_0\mathrm{d}t\end{aligned}$$

where the finite constant $M_0 < \infty$ just depends on $\mu(\lambda)$. Therefore, the region of (absolute) convergence (ROC) of the Laplace transform

$$\mathfrak{L}\{K(t)\}(z) = \int_0^\infty e^{-zt}K(t)\mathrm{d}t$$

is the right half plane $\{z \in \mathbb{C} : \Re\{z\} > 0\}$ where $\check{K}(z) = \mathfrak{L}\{K(t)\}(z)$ should be analytic. A slightly different argument is used to take the Laplace transform of $\mathcal{C}_X(t)$, $\dot{\mathcal{C}}_X(t)$ and $\ddot{\mathcal{C}}_X(t)$ in Appendix A.3.11.

Let $\check{K}(z) = \mathfrak{L}\{K(t)\}(z)$ and $\check{\mathcal{C}}_X(z) = \mathfrak{L}\{\mathcal{C}_X(t)\}(z)$. Then applying the Laplace transform to (2.20) and using the properties of the Laplace transform for derivatives and convolutions yields

$$z^2\check{\mathcal{C}}_X(z) - z\mathcal{C}_X(0) - \dot{\mathcal{C}}_X(0) = -\theta\check{\mathcal{C}}_X(z) - \check{K}(z)\cdot\left(z\check{\mathcal{C}}_X(z) - \mathcal{C}_X(0)\right),$$

which we rearrange to obtain

$$\check{K}(z) = \frac{-(z^2+\theta)\check{\mathcal{C}}_X(z) + z\mathcal{C}_X(0)}{z\check{\mathcal{C}}_X(z) - \mathcal{C}_X(0)}. \tag{2.21}$$

Note that $\theta\mathcal{C}_X(0) = \langle \dot{X}(0), \dot{X}(0) \rangle = -\ddot{\mathcal{C}}_X(0)$ (see Appendix A.3.12), $\dot{\mathcal{C}}_X(0) = 0$ and the denominator $z\check{\mathcal{C}}_X(z) - \mathcal{C}_X(0) \neq 0$ for $\{z \in \mathbb{C} : \Re\{z\} > 0\}$ as explained in Appendix A.3.11.

2.7 Numerical Computation

In the previous section, we derived the GLE for a twice-continuously differentiable stationary process $X(t)$ and obtained closed-form expression for the term in (2.17).

While the mere existence theorem would be sufficient to use any twice continuously differentiable process as a valid physical model, these formulas can be used to provide important physical insights via interpretation and analysis of the GLE parameters and memory kernel.

In this respect, the Laplace transform representation of the memory kernel is of limited practical applicability due to the difficulty of numerical inversion to recover the memory kernel $K(t)$ in the time domain. To this end, we now provide a representation of the memory kernel $K(t)$ in the Fourier domain, for which numerical inversion is readily accomplished by means of the fast Fourier transform (FFT).

Let $\mathcal{C}(t)$ be a general ACF and suppose $\mathcal{C}(t) \in L^1(\mathbb{R})$, its Fourier transform is then defined as

$$\mathcal{F}\{\mathcal{C}(t)\}(\omega) = \int_{-\infty}^{\infty} e^{-2\pi i t\omega} \mathcal{C}(t) \mathrm{d}t.$$

The Laplace transform $\mathfrak{L}\{\mathcal{C}(t)\}(z)$ is then well-defined in the ROC $\{z : \Re\{z\} \geq 0\}$, and a simple calculation gives

$$\mathcal{F}\{\mathcal{C}(t)\}(\omega) = 2 \cdot \Re\{\check{\mathcal{C}}(2\pi i\omega)\}, \qquad \check{\mathcal{C}}(z) = \mathfrak{L}\{\mathcal{C}(t)\}(z)$$

where $z = 2\pi i\omega$ and $\Re\{\cdot\}$ denotes the real part of a complex number. This relation helps us convert the explicit formula (2.21) in the Laplace domain to the Fourier domain (see Theorem 2.7.1 below). However, the assumption that $\mathcal{C}(t)$ be integrable is quite restrictive. If instead we only assume that $\mathcal{C}(t) \in L^2(\mathbb{R})$ (this is indeed less restrictive than $L^1(\mathbb{R})$ since $\mathcal{C}(t)$ is continuous), then $\mathcal{F}\{\mathcal{C}(t)\}(\omega)$ is well defined as the L^2 limit of $\mathcal{F}\{\mathcal{C}(t)\mathbb{1}(|t| \leq N)\}(\omega)$ where $\mathbb{1}(E)$ is the indicator function of set E. Furthermore, we can establish the following limit

$$\check{\mathcal{C}}(2\pi i\omega) = \lim_{a \to 0^+} \check{\mathcal{C}}(a + 2\pi i\omega) = \mathcal{F}\{\mathbb{1}(t)\mathcal{C}(t)\}$$

in either the L^2 sense or in the pointwise (a.e. for every ω) sense where $\mathbb{1}(t)$ is the Heaviside

function, i.e.

$$\mathbb{1}(t) = \begin{cases} 1, & x > 0 \\ \frac{1}{2}, & x = 0 \\ 0, & x < 0 \end{cases}.$$

Please see Appendix A.3.13 for a complete discussion.

So for numerical purposes, let us assume for the given process $x(t)$, the ACF $\mathcal{C}(t)$, its time derivative $\dot{\mathcal{C}}(t) = \mathrm{d}\mathcal{C}(t)/\mathrm{d}t$, and the memory kernel $K(t)$ are all in $L^2(\mathbb{R})$. Then we can safely carry out all the calculations in Appendix A.3.14 to derive another explicit representation formula of the memory kernel $K(t)$ for Theorem 2.1.1, i.e.

Theorem 2.7.1 (Fourier version of Theorem 2.1.1)**.** *Let $x(t)$ be any twice differentiable stationary process. Suppose its ACF $\mathcal{C}_x(t), \dot{\mathcal{C}}_x(t) \in L^2(\mathbb{R})$ and the memory kernel $K(t) \in L^2(\mathbb{R})$. Then there exists an interacting particle system for which $x(t)$ is one of the observables, and $x(t)$ satisfies the linear GLE* (2.4) *with parameters explicitly given by*

$$\kappa = \frac{k_B T}{\mathcal{C}_x(0)}, \quad m = -\frac{k_B T}{\ddot{\mathcal{C}}_x(0)}$$

$$K(\omega) = \mathcal{F}\{K(t)\}(\omega) = \frac{1}{\pi\omega} \Im \left\{ \frac{1}{\hat{\psi}(\omega)} \right\},$$

where

$$\hat{\psi}(\omega) = -\frac{1}{k_B T} \mathcal{F}\{\mathbb{1}(t) \cdot \dot{\mathcal{C}}_x(t)\}.$$

2.7.1 Example: Gaussian Autocorrelation Function

To illustrate the computational result of Theorem 2.7.1, we consider the example of a stationary process $x(t)$ with ACF given by

$$\mathcal{C}_x(t) = \sigma^2 e^{-\frac{t^2}{2\eta^2}}, \quad \sigma, \eta > 0. \tag{2.22}$$

By taking its Fourier transform $\mathcal{F}\{\cdot\}$, we can get the analytic PSD which is again a Gaussian kernel

$$\mathcal{S}_x(\omega) = \mathcal{F}\left\{\mathcal{C}_x(t)\right\}(\omega) = \sqrt{2\pi}\sigma^2 \eta e^{-2\pi^2 \eta^2 \omega^2}.$$

Our goal is to estimate the GLE parameters m, κ, and $K(t)$ using either $\mathcal{C}_x(t)$ or $\mathcal{S}_x(\omega)$, which is often available in practice.

Analytically, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{C}_x(t) = -\sigma^2\frac{t}{\eta^2}e^{-\frac{t^2}{2\eta^2}},$$
$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{C}_x(t) = -\frac{\sigma^2}{\eta^2}\frac{\mathrm{d}}{\mathrm{d}t}\left(te^{-\frac{t^2}{2\eta^2}}\right) = -\frac{\sigma^2}{\eta^2}\left(1-\frac{t^2}{\eta^2}\right)e^{-\frac{t^2}{2\eta^2}}.$$

Thus, we can obtain $\kappa = k_BT/\mathcal{C}_x(0) = k_BT/\sigma^2$ and $m = -k_BT/\ddot{\mathcal{C}}_x(0) = k_BT\eta^2/\sigma^2$ analytically. In practice, if $\mathcal{C}_x(t)$ and its derivatives are not analytically available (whereas $\mathcal{S}_x(\omega)$ is known to us), then we could obtain them using numerical integration as follows

$$\kappa = \frac{k_BT}{\int_{-\infty}^{\infty}\mathcal{S}_x(\omega)\mathrm{d}\omega}$$
$$m = \frac{k_BT}{\int_{-\infty}^{\infty}(2\pi\omega)^2\mathcal{S}_x(\omega)\mathrm{d}\omega}$$

where the denominators, which are derived from the fact that $\dot{\mathcal{C}}_x(t) = \mathcal{F}^{-1}\{2\pi i\omega\mathcal{S}_x(\omega)\}$ and $\ddot{\mathcal{C}}_x(t) = \mathcal{F}^{-1}\{(2\pi i\omega)^2\mathcal{S}_x(\omega)\}$, can be approximated by Riemann sums numerically (or by some other more advanced numerical methods).

As for the memory kernel

$$\mathcal{F}\{K(t)\}(\omega) = \frac{1}{\pi\omega}\Im\left\{\frac{1}{\hat{\psi}(\omega)}\right\},$$

For the Gaussian ACF (2.22) we have

$$\hat{\psi}(\omega) = \int_0^{\infty} e^{-2\pi i\omega t}\frac{\sigma^2 t}{k_BT\eta^2}e^{-\frac{t^2}{2\eta^2}}\mathrm{d}t, \tag{2.23}$$

which can be obtained analytically using the Dawson function and associated error functions (Ng and Geller, 1969). In general, we can estimate $K(\omega)$ numerically from either $\dot{\mathcal{C}}_x(t)$ or $\mathcal{S}_x(\omega)$ as described below.

For an arbitrary infinite sequence $\{a_n : n \in \mathbb{Z}\}$, let $a_{[N]} = (a_0, \ldots, a_N)$ and $a_{[\pm N]} = (a_0, \ldots, a_N, a_{-(N-1)}, \ldots, a_{-1})$. Let `FFT` denote the discrete Fourier transform operator,

$$y_{[N-1]} = \texttt{FFT}(x_{[N-1]}) \qquad \Longleftrightarrow \qquad y_k = \sum_{n=0}^{N-1} x_n e^{-2\pi ikn/N},$$

and iFFT denote its inverse transform,

$$x_{[N-1]} = \texttt{iFFT}(y_{[N-1]}) \qquad \Longleftrightarrow \qquad x_k = \frac{1}{N}\sum_{n=0}^{N-1} y_n e^{2\pi i k n/N}.$$

For a given sample size N and sampling frequency $f_s = 1/\Delta t$, we can estimate $K(t)$ at time points $t_{[N]} = (0, \Delta t, \ldots, N\Delta t)$ by following Algorithm 2.1.

Algorithm 2.1 Numerical algorithm to recover memory kernel $K(t)$.

1: **if** $\dot{\mathcal{C}}_x(t)$ is analytically available **then**
2: $\quad \psi_{[\pm N]} \leftarrow -(k_B T)^{-1}\dot{\mathcal{C}}_x(t_{[\pm N]})$, where $t_n = n\Delta t$.
3: **else**
4: $\quad \psi_{[\pm N]} \leftarrow -(k_B T)^{-1} \cdot \texttt{iFFT}\{2\pi i f_{[\pm N]} \cdot \mathcal{S}_x(f_{[\pm N]})\}$, where $f_n = \frac{n}{N}\frac{f_s}{2}$.
5: **end if**
6: $\hat{\psi}_{[\pm N]} \leftarrow \texttt{FFT}\{\mathbb{1}(t_{[\pm N]} \cdot \psi_{[\pm N]}\}$
7: $K_{[\pm N]} \leftarrow \texttt{iFFT}\{(\pi f_{[\pm N]})^{-1} \cdot \Im\{1/\hat{\psi}_{[\pm N]}\}\}$
8: Keep the symmetric half $K_{[N]}$ as the approximation to $K(t)$ at time points $t_{[N]}$.

In order to assess the accuracy of Algorithm 2.1, we verify that the estimated GLE parameters recover the original PSD $\mathcal{S}_x(\omega)$ by

$$\mathcal{S}_x(\omega) = \frac{k_B T \cdot K(\omega)}{\left|\kappa + 2\pi i\omega \cdot \hat{K}(\omega) - 4\pi^2\omega^2\right|^2}$$

where $K(\omega)$ (i.e., the memory kernel in the Fourier domain $K(f_{[\pm N]})$) has already been obtained in Algorithm 2.1, and $\hat{K}(\omega) = \mathcal{F}\{\mathbb{1}(t)\cdot K(t)\}(\omega)$ can be numerically approximated by using FFT and iFFT, i.e.,

$$\hat{K}(f_{[\pm N]}) = \texttt{FFT}\{\mathbb{1}_{[\pm N]} K_{[\pm N]}\} = \texttt{FFT}\{\mathbb{1}_{[\pm N]} \cdot \texttt{iFFT}\{K(f_{[\pm N]})\}\}.$$

Figure 2.1 displays estimates of $\hat{K}(f)$ and $\mathcal{S}_x(f)$ computed for $T = 298\,\mathrm{K}$, $\sigma = 0.01\,\mathrm{m}$, $\eta = 0.0001\,\mathrm{sec}$, with $N = 1\,000\,000$ samples and sampling frequency $f_s = 100\,000\,\mathrm{Hz}$. If we approximate (2.23) numerically using the FFT, we can only recover it in the frequency domain up to about $10\,000\,\mathrm{Hz}$, the result may suffer severe numerical errors for the ultra-high frequency (due to the condition number (Trefethen and Bau, 1997) inflated by $1/\hat{\psi}(\omega)$), as shown in Figure 2.1a. But note that in the ultra-high frequency range over $10\,000\,\mathrm{Hz}$, the

PSD decays dramatically and is essentially too small to be numerically significant. If we truncate the frequency up to about 10 000 Hz, we can see the recovered PSD is perfectly matched up with the theoretical one, as shown by Figure 2.1b.

If we calculate (2.23) analytically using Dawson function and replace $\hat{\psi}(\omega)$ obtained at the step 4 in Algorithm 2.1 with this analytic result, then the recovered $K(\omega)$ is exact for all the frequencies $\omega \in \mathbb{R}$, as shown in Figure 2.1c. We can also compare the recovered PSD using the analytic GLE parameters with the theoretical PSD shown by Figure 2.1d.

2.8 Discussion

We rigorously showed that *any* twice differentiable stationary process $x(t)$ can be used to model nanoscopic phenomena without violating the fundamental laws of physics. Inference about $x(t)$ can be performed using any statistical method of choice, with the GLE parameters of scientific interest being recoverable post-inference using the formulas presented. It might sound "trivial" at first glance, but in a deeper sense, our result establishes a fundamental connection between the dynamics of motion in an abstract infinite dimensional Hamiltonian space and any continuous stationary process defined on a probability measure space, which allows us to model the single-particle dynamics without too much restriction.

One might question however if our result is still valid for non-stationary processes. Indeed, we believe our result can be extended to any continuous increment stationary process. However, a mathematically rigorous and complete proof is the subject of future research.

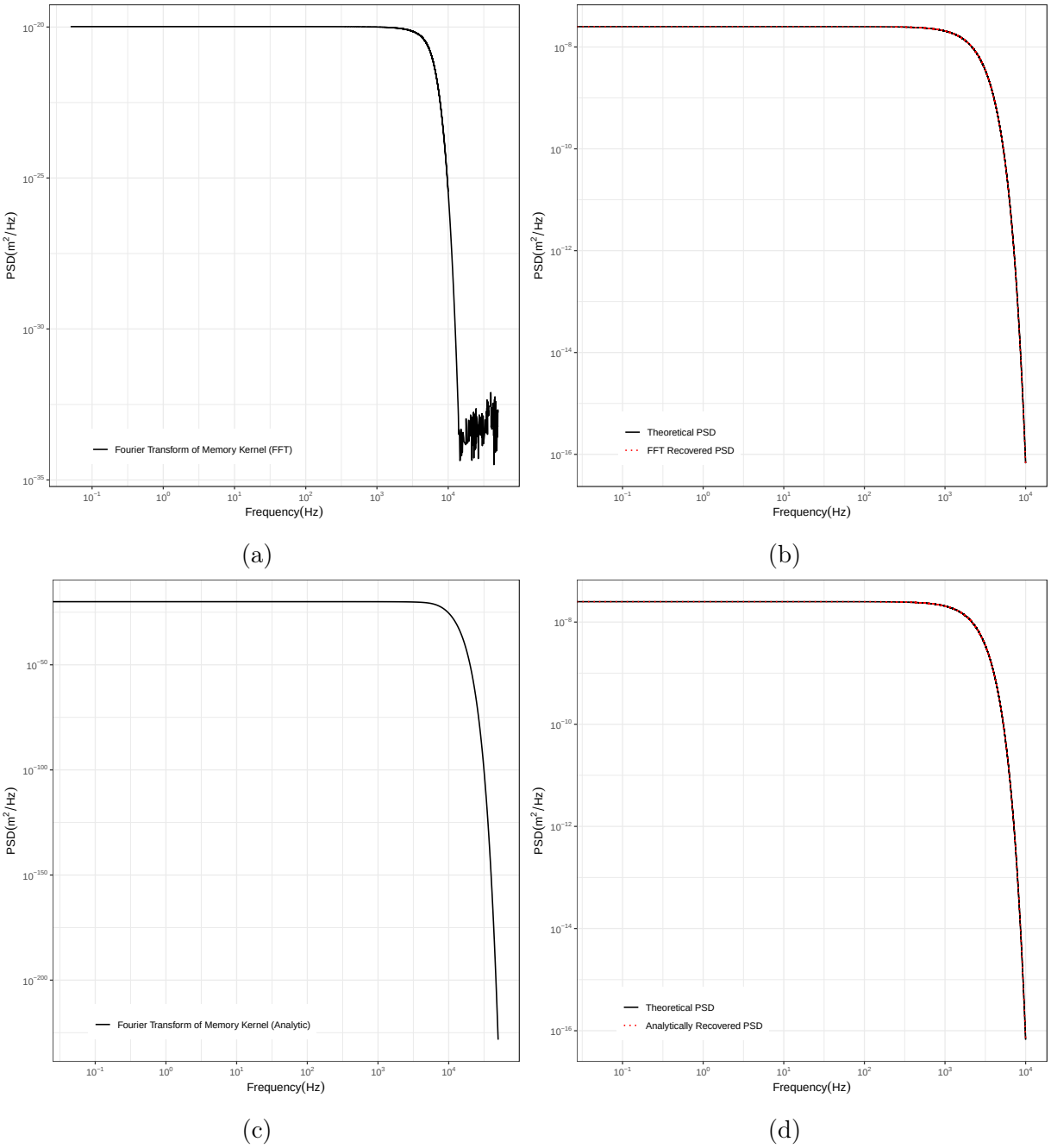

Figure 2.1: (a) Memory kernel numerically recovered by FFT. (b) Theoretical PSD v.s. the recovered PSD using numerical parameters. (c) Memory kernel analytically recovered using Dawson function. (d) Theoretical PSD v.s. the recovered PSD using analytic parameters. All the figures are displayed on the log-log scale.

Chapter 3

Parametric Spectral Density Estimation for High-Throughput Data

3.1 Introduction

3.1.1 Background

Over the last two decades, development of scientific instruments has opened a new era in biological research, allowing scientists to observe and measure biophysical dynamics with extremely short time interval as well as ultra-precision. For example, an atomic force microscope (AFM) can be used to produce ultra-high resolution measurements of nanoscopic objects and forces (Binnig et al., 1986; Kirmizis and Logothetidis, 2010) which now becomes an indispensable tool for various scientific studies in biophysics, biomedical engineering, chemical engineering and molecular biology, etc. (Radmacher, 1997; Hoffmann et al., 2001; Evans and Calderwood, 2007; García et al., 2007; Sugimoto et al., 2007; Alsteens et al., 2008; Yu et al., 2017). A huge amount of time series data can be recorded in those advanced experiments at high-frequency and for extended periods of time. Such "high-throughput" (HTP) data give us unprecedented opportunities to probe the molecular mechanisms of many biological processes, yet leading statisticians to theoretical difficulties in estimation and inference.

As such "high-throughput" (HTP) data typically span timescales over several orders of magnitude, often it is convenient to analyze them in the frequency domain. That is, suppose the phenomenon of interest X_t is a continuous-time mean-zero stationary Gaussian process with autocorrelation $\mathcal{C}(t) = \mathrm{cov}(X_s, X_{s+t})$. Then a frequency domain analysis involves the

Fourier transform of the autocorrelation, namely the *power spectral density* (PSD) function of X_t,

$$\mathcal{S}(f) = \int_{-\infty}^{\infty} e^{-2\pi i t f} \mathcal{C}(t) \mathrm{d}t = \mathcal{F}\{\mathcal{C}(t)\}. \tag{3.1}$$

While the large size of HTP datasets can be dealt with by some nonparametric PSD estimation techniques (e.g., Bartlett, 1950; Welch, 1967; Thomson, 1982; Gasser et al., 1985; Nason and Silverman, 1995), in many scientific applications a parametric spectral model is still preferred. Many such models are members of the continuous-time autoregressive fractionally integrated moving average (CARFIMA) family (Tsai and Chan, 2005), for which the PSD is

$$\mathcal{S}(f) = \frac{\sigma^2}{2\pi} \Gamma(2H+1) \sin(\pi H) \times |f|^{1-2H} \frac{|1 + \sum_{k=1}^{q} \beta_k (if)^k|^2}{|(if)^p - \sum_{m=1}^{p} \alpha_m (if)^{m-1}|^2}, \qquad 0 < H < 1.$$

Some CARFIMA processes have a direct physical interpretation. A widely used model for HTP data is the simple harmonic oscillator (SHO) model as discussed later in Section 3.5. Others are able to capture diverse stochastic characteristics, such as long-range dependence ($H \neq \frac{1}{2}$), and the shape or size of multiple spectral modes. For narrowband features such as the width and location of sharp spectral peaks, parametric methods can offer lower estimation uncertainty than nonparametric alternatives. Therefore, CARFIMA and other parametric models have found numerous applications in HTP data analysis, for example: mRNA molecule tracking in live cells (Burnecki, 2012); identification of chemical bond lengths and angles (Tsekov, 2016); electroencephalogram (EEG) signal processing (Sanei and Chambers, 2013); persistence analysis of geophysical processes (Witt and Malamud, 2013); turbulence characterization of oceanographic surface drifters (Sykulski et al., 2016); and modeling of X-ray emissions from neutron stars (Barret and Vaughan, 2012).

3.1.2 Statistical Challenges

While parametric spectral density estimation serves a wide range of scientific applications, HTP data present two outstanding challenges as pointed out by Lysy et al. (2022). The first relates to the theoretical foundation of statistical inference for the HTP data.

(i) The large size of HTP data can render likelihood-based estimation method prohibitively expensive due to the intractability of time-domain maximum likelihood estimation, i.e. the computationally-intensive extraction of $\mathcal{C}(t)$ from the PSD model (3.1).

(ii) The existing asymptotic results in the field of time series are mostly based on fixed time interval (i.e. with fixed sampling frequency). In recent literature, Fasen and Fuchs (2013) studied the limiting behavior of the periodogram of high-frequency sampled stable continuous-time autoregressive moving average (CARMA) processes. However, the asymptotic properties of many important estimators as the sampling frequency goes to infinity are still unknown.

The second challenge to PSD estimation is that HTP recordings are often contaminated by various sources of instrumental noise. While in many cases such noise can be incorporated into the PSD model itself, the periodic noise inexorably emitted by electronic circuitry cannot, and has been shown to severely affect PSD estimates (e.g., Lees, 1995; Kast et al., 2003; Yang et al., 2009; Littenberg and Cornish, 2015).

To solve the second issue (which will not be covered in this thesis), Lysy et al. (2022) proposed a two-stage noise elimination approach. The basic idea goes as follows:

(i) Adapt the variance-stabilized least-squares estimator of Moulines and Soulier (1999); Fay et al. (2002) to the HTP regime, proving its consistency and asymptotic efficiency.

(ii) Eliminate electronic noise outliers via a novel statistical testing framework for hidden periodicities, for which the Type-I error and false discovery rate can be controlled. The two-stage procedure refits the least-squares estimator after the outliers have been removed.

For details, please refer to Lysy et al. (2022) where extensive simulations and experimental results are included for a popular spectral model for HTP data, a substantial reduction in parameter mean squared error can be achieved. All the methods discussed and applied in (Lysy et al., 2022) are implemented efficiently in the R package `realPSD` developed by Zhu and Lysy (2021).

3.1.3 Our Contribution

In this Chapter, we establish the theoretical foundations of parametric statistical inference for HTP data, then propose a robust and efficient parametric spectral density estimation method based on the Whittle likelihood (Whittle, 1953a,b, 1957), which provides a good approximation to the likelihood of a stationary Gaussian process and allows us to perform parameter estimation directly in the frequency domain.

Under mild regularity conditions, the Whittle estimator is consistent and asymptotically efficient (Fox and Taqqu, 1986; Dahlhaus, 1989). However, for discrete-time data recorded at sampling frequency f_s, the Whittle estimator involves the discrete-time PSD

$$\mathcal{S}_{f_s}(f) = \frac{1}{f_s} \sum_{n=-\infty}^{\infty} e^{-2\pi i n f/f_s} \mathcal{C}(n/f_s) = \sum_{n=-\infty}^{\infty} \mathcal{S}(f + n f_s), \tag{3.2}$$

which typically is not available in closed form. Even if it were, the large volume of HTP data can render the Whittle estimator prohibitively expensive (Nørrelykke and Flyvbjerg, 2010). Thus, practitioners routinely employ a much faster, nonlinear least-squares (NLS) estimation method, at the cost of substantial reduction in statistical efficiency to ease the computational burden. In order to avoid the efficiency loss incurred by using the NLS method, another least-squares estimator based on the log-periodogram (LP) is proposed. We will rigorously verify and discuss the asymptotic properties of all these estimators for HTP data.

3.1.4 Outline

The remaining Chapter is organized as follows: In Section 3.2, we discuss some basic settings of HTP data by introducing continuous processes and its sampled processes, revealing the Poisson summation formula (3.2), a key relationship between the spectral densities of the two processes. Besides, we also introduce the basic characterization of different processes by using different types of memory, and discuss two different ways of describing the high-frequency asymptotics and explain the rationale behind our choice of HTP settings motivated by real experiments. Our main results will be shown in Section 3.3 and 3.4 where we establish the consistency and asymptotic normality of the (bandlimited) Whittle estimator and the log-periodogram estimator. In Section 3.5, we verify the theoretical results based on a simulation study of a ubiquitous model in scientific research and a real-world application of AFM calibration.

3.2 Sampling Continuous-Time Processes

In this chapter, we only discuss stationary processes which can always be represented as the GLE according to our remarkable representation theorem revealed in Chapter 2. Inference for nonstationary processes is beyond the scope of our current work. Gaussian processes are often desirable enough in many applications and it can be fully characterized by its mean

and autocovariance structure. In the following sections, all the processes under discussion are implicitly assumed to be stationary and Gaussian.

3.2.1 Continuous Time

Consider the continuous-time stationary (zero-mean) Gaussian process $\{X(t) : t \in \mathbb{R}\}$ with autocorrelation function $\mathcal{C}(t) = \mathrm{Cov}(X(s), X(s+t))$. The power spectral density (PSD) of $X(t)$ is defined as the unique nonnegative symmetric integrable function $\mathcal{S}(f)$ satisfying

$$\mathcal{C}(t) = \int_{-\infty}^{\infty} e^{2\pi i t f} \mathcal{S}(f) \, \mathrm{d}f.$$

Here, frequency f is measured in Hertz. When $\mathcal{C}(t)$ is integrable (e.g., when $\mathcal{S}(f)$ is continuous), we have the Fourier inversion theorem

$$\mathcal{S}(f) = \int_{-\infty}^{\infty} e^{-2\pi i t f} \mathcal{C}(t) \, \mathrm{d}t.$$

3.2.2 Discrete Time

Suppose that $X(t)$ is sampled at frequency $f_s = 1/\Delta t$ to produce the discrete-time process $\{X_n = X(n\Delta t) : n \in \mathbb{Z}\}$. The discrete-time PSD $\mathcal{S}_{f_s}(f)$ is defined as the unique nonnegative symmetric integrable function on $f \in (-f_s/2, f_s/2]$ such that

$$\mathcal{C}(n\Delta t) = \mathcal{C}(n/f_s) = \frac{1}{f_s} \int_{-f_s/2}^{f_s/2} e^{2\pi i n f/f_s} \mathcal{S}_{f_s}(f) \, \mathrm{d}f.$$

Here f is also measured in Hertz. When $\{\mathcal{C}(n\Delta t)\}$ is absolutely summable, we have the Fourier inversion formula

$$\mathcal{S}_{f_s}(f) = \frac{1}{f_s} \sum_{n=-\infty}^{\infty} e^{-2\pi i n f/f_s} \mathcal{C}(n/f_s).$$

Moreover, under appropriate conditions we have the *Poisson summation formula*

$$\mathcal{S}_{f_s}(f) = \sum_{n=-\infty}^{\infty} \mathcal{S}(f + n f_s).$$

3.2.3 Finite Samples

Now suppose we have a finite sample of observations $\boldsymbol{X} = (X_0, \ldots, X_{N-1})$ and $N = 2K+1$. The discrete Fourier transform of $\boldsymbol{X}$ is defined as the vector $\tilde{\boldsymbol{X}} = (\tilde{X}_{-K}, \ldots, \tilde{X}_0, \ldots, \tilde{X}_K)$ such that

$$\tilde{X}_k = \sum_{n=0}^{N-1} e^{-2\pi i k n/N} X_n. \tag{3.3}$$

As with each observation X_n we associate a time point $t_n = n \cdot \Delta t$, with each Fourier observation $\tilde{X}_k$ we associate a frequency $f_k = k/N \cdot f_s$. Moreover, if $\texttt{FFT}(\cdot)$ denotes the fast Fourier transform (FFT) algorithm function in R or MATLAB, then

$$\texttt{FFT}(\boldsymbol{X}) = (\tilde{X}_0, \ldots, \tilde{X}_K, \tilde{X}_{-K}, \ldots \tilde{X}_{-1}).$$

3.2.4 Different Types of Memory

Time series data observed in experiments or real applications are rarely *i.i.d.*, that is, there are always some kind of (weak or strong) dependence between two observations at different time points. There are many ways to generally characterize the dependence involved in a stationary processes, e.g. *α-mixing* or *m-dependence*. But since we are discussing stationary Gaussian processes, a more natural way of describing and measuring dependence is by using second-order autocovariance structure (time-domain) or spectral information (frequency-domain). Therefore, we have the following concepts:

Definition 3.2.1. *A stationary process $\{X_n\}$ with autocovariance function $\gamma(k)$ is said to have*

short memory (short-range dependence) if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \qquad \textit{and} \qquad \sum_{k\in\mathbb{Z}} \gamma(k) > 0;$$

long memory (long-range dependence) if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| = \infty;$$

negative memory (antipersistence) if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \qquad \textit{and} \qquad \sum_{k\in\mathbb{Z}} \gamma(k) = 0.$$

Definition 3.2.2. *Suppose the spectral density of a stationary process $\{X_n\}$ exists, satisfying*

$$\mathcal{S}_{f_s}(f) = |f|^{-2d} g(f), \quad f \in [-f_s/2, f_s/2], \quad |d| < \frac{1}{2}$$

where the sampling frequency f_s is fixed and $g(f)$ is a continuous function bounded away from zero. The process is said to have negative memory, short memory or long memory, if $d \in (-\frac{1}{2}, 0)$, $d = 0$ or $d \in (0, \frac{1}{2})$, respectively.

Note that for a long-memory process, the spectral density, provided it exists, is unbounded at the origin.

These concepts are adapted from two great resources on this subject, Beran et al. (2013) and Giraitis et al. (2012). We only introduce them in a minimal way necessary for our discussions in the following sections. For further details, please check out the aforementioned references.

3.2.5 High-Frequency Asymptotics

High throughput data are collected in a situation where we have a very large (and also flexible) sampling frequency (i.e., extremely short observation time interval) with an increasing amount of observations, as shown in Figure 3.1. The limiting behavior of various statistics in this situation is generally called "high-frequency asymptotics". This topic is rarely seen in the current literature. Generally, there are two ways to approach this type of asymptotics:

(1) Assume the sampling frequency f_s is a function of the total number of observations N, i.e. $N \to \infty$ and $f_s \to \infty$ happen simultaneously. Usually $N\Delta t = N/f_s \sim N^\alpha$ with some rate α, as $N \to \infty$. Fasen and Fuchs (2013) adopted this kind of approach. The advantage is that it allows us to view the asymptotic process in a straightforward way. However, it usually incurs more mathematical complexity.[1] Moreover, this type

[1]We may need to modify the standard triangular array argument to account for non-independent row observations to get the asymptotic distribution. Besides, analysis on manifolds may be an indispensable tool to get a rigorous result. Please see Fasen and Fuchs (2013).

of asymptotics is not realistic at all in many physical experiments in which people always need to set a sampling frequency first and then collect more and more data. It is impossible to alter the sampling frequency after the experiment begins. This kind of restriction in real experiments leads us to another procedure of asymptotics.

(2) Assume the sampling frequency f_s is always chosen (can be arbitrarily large) beforehand, and then we allow the observation number N to increase. Even though both f_s and N can increase to infinity, we always require taking $N \to \infty$ under a given fixed f_s. This order indeed makes sense from a practitioner's perspective because in experiments we can only first determine the sampling frequency (no matter how large it is) based on the error bound we would like to control and then collect more and more observations.

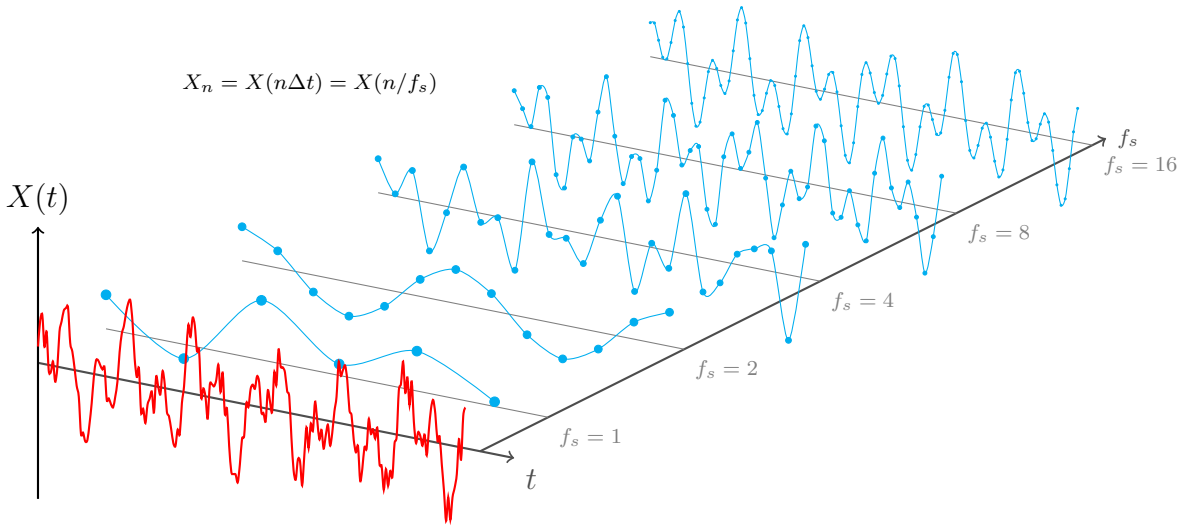

Figure 3.1: Illustration of HTP data recorded from observing a continuous stationary process $X(t)$, with an increasing sampling frequency $f_s = 1/\Delta t$ for extended periods of time.

In mathematical language, the consistency of a statistical estimator $\hat{\boldsymbol{\theta}}_{f_s,N}$ to the true parameter $\boldsymbol{\theta}_0$ can be described as: for any $\epsilon > 0$, there exists a large sampling frequency f_0 such that for $f_s > f_0$,

$$\lim_{N\to\infty} \left| \hat{\boldsymbol{\theta}}_{f_s,N} - \boldsymbol{\theta}_0 \right| < \epsilon.$$

More generally, given any random variable $\{X_{f_s,N}\}$, the limiting process is

$$\lim_{f_s\to\infty}\lim_{N\to\infty} X_{f_s,N} := \lim_{f_s\to\infty}(\lim_{N\to\infty} X_{f_s,N}) = \lim_{f_s\to\infty} Y_{f_s}$$

where

$$Y_{f_s} := X_{f_s,\infty} := \lim_{N\to\infty} X_{f_s,N}.$$

This procedure is referred to as the *high-frequency asymptotics* later on.

3.3 Parametric Spectral Density Estimation

Ideally one would perform parameter estimation[2] with the log-likelihood $\ell(\boldsymbol{\theta}\|\boldsymbol{X})$ directly in the time domain. However, this approach is typically inviable in practice due to the computational intractability of the autocorrelation of $\boldsymbol{X}$. Instead, most spectral estimators make use of the following result.

Proposition 1. *Let $N = 2K+1$ and define the finite Fourier transform as in ((3.3)). For each $\tilde{X}_k$, let $f_k = kf_s/N$ denote the corresponding frequency and let $Y_k = \left|\tilde{X}_k\right|^2/(Nf_s)$. Then under suitable conditions on the discrete-time PSD $\mathcal{S}_{f_s}(f)$ defined in (3.2), as $N\to\infty$ we have*

$$Y_k \overset{\text{ind}}{\sim} \mathcal{S}_{f_s}(f_k)\times \text{Expo}(1), \qquad 0<k<K.$$

A precise meaning of the convergence in Proposition 1 as the number of variables $K\to\infty$ is given by e.g., Proposition 4.5.2 of Brockwell and Davis (2006), or more generally, by Theorem 2 of Moulines and Soulier (1999) which we have summarized in Appendix B.2.3. For ease of exposition, we shall employ Proposition 1 heuristically to motivate our estimators, providing rigorous derivations of their theoretical properties in the Appendix.

[2] A underlying model or an estimation method is said to be *parametric* if its spectral density or autocovariance structure is known up to an unknown Euclidean parameter, e.g. CARFIMA model. If the spectral density is known to depend on an Euclidean parameter and an unknown infinite-dimensional parameter, it is called *semiparametric*, e.g. $S(u) = |u|^{-d}g(u)$ where $|d| < \frac{1}{2}$ and g is some unspecified function. For details please see Giraitis et al. (2012). The distinctions between parametric and semiparametric methods are not the focus of our work here. We just mention these concepts for a better organization of our results.

3.3.1 Whittle log-Likelihood Estimator

The Whittle likelihood (first introduced by Whittle (1953a)) is an approximation to the exact Gaussian likelihood which is usually intractable for complex models. Usually it can be defined in both a continuous form and a discrete form as used in Hannan (1973); Fox and Taqqu (1986); Dahlhaus (1989). The continuous form Whittle likelihood $\ell_N^W(\boldsymbol{\theta}|\boldsymbol{X})$ can be written as

$$\ell_N^W(\boldsymbol{\theta}) := \bar{\ell}_N^W(\boldsymbol{\theta}|\boldsymbol{X}) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log \mathcal{S}_{f_s}(f;\boldsymbol{\theta}) + \frac{I_N(f)}{\mathcal{S}_{f_s}(f;\boldsymbol{\theta})}\right)\mathrm{d}f$$

where $I_N(f) := \frac{1}{2\pi N}\left|\sum_{n=0}^{N-1} e^{-i\cdot f\cdot n}X_n\right|^2$ for $f \in [-\pi,\pi]$.

For computational purposes, we also need its discretized form $\tilde{\ell}_N^W(\boldsymbol{\theta}|\boldsymbol{X})$, i.e.

$$\tilde{\ell}_N^W(\boldsymbol{\theta}) := \tilde{\ell}_N^W(\boldsymbol{\theta}|\boldsymbol{X}) = -\frac{1}{2N}\sum_{j\in\Pi_N}\left(\log \mathcal{S}_{f_s}(f_j;\boldsymbol{\theta}) + \frac{I_N(f_j)}{\mathcal{S}_{f_s}(f_j;\boldsymbol{\theta})}\right). \tag{3.4}$$

where $j \in \Pi_N := \{-\lceil N/2\rceil + 1,\ldots,-1,0,1,\ldots,\lfloor N/2\rfloor\}$, $f_j = j/N\cdot f_s$ and $I_N(f_j)$ is similarly defined.

Since the periodogram $I_N = (I_N(f_1),\ldots,I_N(f_j),\ldots,I_N(f_K))$ can be computed in $\mathcal{O}(N\log N)$ time using the FFT, maximization of the Whittle log-likelihood (3.4) is considerably easier than that of the original likelihood $\ell(\boldsymbol{\theta}\|\boldsymbol{X})$.

Motivated by Proposition 1, the discretized Whittle log-likelihood function (Whittle, 1953a,b, 1957) can be written as

$$\ell(\boldsymbol{\theta}\|\boldsymbol{Y}) = -\sum_{k=1}^{K}\left[\log \mathcal{S}_{f_s}(f_k,\boldsymbol{\theta}) + \frac{Y_k}{\mathcal{S}_{f_s}(f_k,\boldsymbol{\theta})}\right]. \tag{3.5}$$

For the same reason due to FFT, maximization of the Whittle log-likelihood (3.5) is considerably easier than that of the likelihood $\ell(\boldsymbol{\theta}\|\boldsymbol{X})$ in the original time domain.

For fixed interval (i.e. f_s fixed) stationary processes, the asymptotic behavior of the Whittle MLE $\hat{\boldsymbol{\theta}}_N^W = \arg\max_{\boldsymbol{\theta}} \ell_N^W(\boldsymbol{\theta}\|\boldsymbol{X})$ has been widely studied in the literature. Consistency and asymptotic efficiency of the Whittle MLE $\tilde{\boldsymbol{\theta}}_{\mathrm{W}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}\|\boldsymbol{Y})$ have been established by Fox and Taqqu (1986); Dahlhaus (1989). Since for HTP applications $\mathcal{S}_{f_s}(f,\boldsymbol{\theta})$ is not available in closed form, a natural approximation is to truncate (3.2) at a finite

number of terms

$$\mathcal{S}_{f_s}(f, \boldsymbol{\theta}) \approx \sum_{|n|<M} \mathcal{S}(f, \boldsymbol{\theta}), \tag{3.6}$$

with $M = 100$ being commonly recommended in practice (e.g. Comte, 1996; Tsai and Chan, 2005). However, $f_s \to \infty$ requires that $M \to \infty$ in order to maintain the same order of accuracy (Comte, 1996). Moreover, each evaluation of the Whittle log-likelihood (3.5) with the truncated approximation (3.6) requires KM evaluations of $\mathcal{S}(f, \boldsymbol{\theta})$, which for HTP datasets with K in the millions would bear a massive computational burden. In contrast, we show in Lemma B.1.2 of Appendix B that under mild conditions we have

$$\lim_{f_s \to \infty} \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sup_{|f| < f_{\max}} |\mathcal{S}_{f_s}(f, \boldsymbol{\theta}) - \mathcal{S}(f, \boldsymbol{\theta})| = 0, \tag{3.7}$$

for a compact set $\boldsymbol{\Theta}$ and predetermined frequency $f_{\max} < f_s$. In other words, approximation (3.6) becomes arbitrarily accurate even with a single term $M = 1$ over a predetermined bandwidth, leading to the following extension of the Whittle estimator to the HTP setting:

Theorem 3.3.1. *Let $\boldsymbol{\theta}_0$ denote the true parameter value, and consider the bandlimited Whittle likelihood*

$$\ell_{f_{\max}}(\boldsymbol{\theta} \mid \boldsymbol{Y}) = -\sum_{k=1}^{K_{\max}} \left[\log \mathcal{S}(f_k, \boldsymbol{\theta}) + \frac{Y_k}{\mathcal{S}(f_k, \boldsymbol{\theta})} \right],$$

where $K_{\max} = \lfloor f_{\max} N / f_s \rfloor$, and let $\hat{\boldsymbol{\theta}}_{\mathrm{W}} = \arg\max \ell_{f_{\max}}(\boldsymbol{\theta} \mid \boldsymbol{Y})$. Then for fixed f_s and under Assumptions 3.3.1 – 3.3.5, there exists $\boldsymbol{\theta}_{f_s}$ and $\mathcal{I}_{f_s}$ such that for $N \to \infty$ we have

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_{f_s}) \to \mathcal{N}(\mathbf{0}, \mathcal{I}_{f_s}^{-1}),$$

and as $f_s \to \infty$ we have $\boldsymbol{\theta}_{f_s} \to \boldsymbol{\theta}_0$ and $\mathcal{I}_{f_s} \to \mathcal{I}_{f_{\max}}$, where

$$\mathcal{I}_{f_{\max}} = \frac{1}{4 f_{\max}} \int_{-f_{\max}}^{f_{\max}} \left[\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}(f, \boldsymbol{\theta}_0) \right] \left[\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}(f, \boldsymbol{\theta}_0) \right]' \mathrm{d}f. \tag{3.8}$$

The estimator $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ defined in Theorem 3.3.1 is bandlimited optimal, in the sense that

$$\check{\boldsymbol{\theta}}_{\mathrm{W}} = \arg\max_{\boldsymbol{\theta}} \sum_{k=1}^{K_{\max}} \left[\frac{-Y_k}{\mathcal{S}_{f_s}(f_k, \boldsymbol{\theta})} - \log \mathcal{S}_{f_s}(f_k, \boldsymbol{\theta}) \right],$$

the bandlimited Whittle estimator with the exact discrete-time PSD, has Fisher information converging to $\mathcal{I}_{f_{\max}}$ as $f_s \to \infty$. In contrast, the asymptotic variance of the full Whittle estimator defined by (3.5) has Fisher information

$$\mathcal{I}_{f_s}^{\mathrm{W}} = \frac{1}{4f_s} \int_{-f_s}^{f_s} \left[\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}_{f_s}(f, \boldsymbol{\theta}_0)\right] \left[\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}_{f_s}(f, \boldsymbol{\theta}_0)\right]' \mathrm{d}f.$$

If $\lim_{f_{\max}\to\infty} \mathcal{I}_{f_{\max}} = \mathcal{I}_\infty$ exists, then (3.7) implies that $\lim_{f_s\to\infty} \mathcal{I}_{f_s}^{\mathrm{W}} = \mathcal{I}_\infty$ as well, such that the loss of efficiency due to bandlimiting can be made arbitrarily small. However, for the CARFIMA model

$$\mathcal{S}(f, H) = \frac{|f|^{1-2H}}{f^2 + 1},$$

as $f_{\max} \to \infty$ we have

$$\mathcal{I}_{f_{\max}} = \frac{2}{f_{\max}} \int_0^{f_{\max}} (\log f)^2 \, \mathrm{d}f = 2(\log f_{\max})^2 - 4 \log f_{\max} + 4 \to \infty.$$

Since $\mathcal{I}_{f_s}^{\mathrm{W}} > \mathcal{I}_{f_{\max}}$ for sufficiently large f_s, bandlimiting incurs an arbitrarily large penalty in asymptotic efficiency relative to the full Whittle estimator, for fixed $f_{\max}$ as $f_s \to \infty$.

Remark 3.3.1. *If the tail of the spectral density $\mathcal{S}(f, \boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$, e.g. $\mathcal{S}(f, \boldsymbol{\theta}) = \mathcal{O}(1/|f|^{1+\delta(\boldsymbol{\theta})})$ for $|f| \to \infty$, we should get new information as we increase the frequency bandwidth $2f_{\max}$. Then the asymptotic variance of our bandlimited estimator $\hat{\boldsymbol{\theta}}_{\mathrm{W}}$ cannot be arbitrarily close to that of the full Whittle estimator given by Dahlhaus (1989), since the amount of information accumulated can never converge. This reveals the motivation for the assumption in Lemma B.1.2, i.e., $\mathcal{S}(f, \boldsymbol{\theta}) \sim \mathcal{O}(1/|f|^{1+\delta})$ as $|f| \to \infty$ by which we can show $\mathcal{I}_{f_{\max}}$ converges to the Fisher information matrix of the full Whittle estimator given by Dahlhaus (1989), as $f_{\max} \to \infty$ and thus is optimal. See Appendix B.1.1 for further details.*

3.3.2 Central Limit Theorem for the Whittle MLE

We now provide a general central limit theorem for estimating equations of Whittle type. This theorem may be of independent interest for the study of PSD estimation under model misspecification.

Theorem 3.3.2. *For fixed f_s, let $\mathcal{S}_{f_s}(f)$ denote the discrete-time PSD of $\{X_n : n \in \mathbb{Z}\}$,*

and consider the integrated periodogram

$$I_N(f) = \frac{1}{Nf_s}\left|\sum_{n=0}^{N-1} e^{-2\pi i n f/f_s} X_n\right|^2.$$

Let $g(f,\boldsymbol{\theta}) : \mathbb{R}\otimes\boldsymbol{\Theta} \to \mathbb{R}^+$ *where* $\boldsymbol{\Theta}$ *is a compact* p*-dimensional parameter space, and let*

$$\hat{\boldsymbol{\theta}}_N = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\, \ell_N(\boldsymbol{\theta}), \quad \ell_N(\boldsymbol{\theta}) = \frac{1}{2f_s}\int_{-f_s/2}^{f_s/2} \log g(f,\boldsymbol{\theta}) + \frac{I_N(f)}{g(f,\boldsymbol{\theta})}\,\mathrm{d}f,$$
$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\, \ell_0(\boldsymbol{\theta}), \quad \ell_0(\boldsymbol{\theta}) = \frac{1}{2f_s}\int_{-f_s/2}^{f_s/2} \log g(f,\boldsymbol{\theta}) + \frac{\mathcal{S}_{f_s}(f)}{g(f,\boldsymbol{\theta})}\,\mathrm{d}f,$$

Now suppose there exists a compact ball $\boldsymbol{\Theta}_0 \subset \boldsymbol{\Theta}$ *containing* $\boldsymbol{\theta}_0$ *such that* $g(f,\boldsymbol{\theta})$ *satisfies the following assumptions:*

Assumption 3.3.1. $\boldsymbol{\theta}_0$ *lies in the interior of* $\boldsymbol{\Theta}$*. If* $\boldsymbol{\theta} \neq \boldsymbol{\theta}' \in \boldsymbol{\Theta}$*, the set* $\{f : g(f,\boldsymbol{\theta}) \neq g(f,\boldsymbol{\theta}')\}$ *has positive Lebesgue measure.*

Assumption 3.3.2. $\frac{1}{2f_s}\int_{-f_s/2}^{f_s/2} \log g(f,\boldsymbol{\theta})\mathrm{d}f$ *is twice differentiable in* $\boldsymbol{\theta}$ *under the integral sign.*

Assumption 3.3.3. *The functions* $g(f,\boldsymbol{\theta}) > 0$*,* $g^{-1}(f,\boldsymbol{\theta}) = 1/g(f,\boldsymbol{\theta})$*,* $\frac{\partial}{\partial\boldsymbol{\theta}} g^{-1}(f,\boldsymbol{\theta})$*, and* $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} g^{-1}(f,\boldsymbol{\theta})$ *are continuous at all* $(f,\boldsymbol{\theta}) \in \{\mathbb{R}\setminus 0\}\otimes\boldsymbol{\Theta}_0$*.*

Assumption 3.3.4. *There exists* $-1 < \alpha, \beta < 1$ *such that* $\alpha+\beta < \frac{1}{2}$*,* $g(f,\boldsymbol{\theta}_0) \leq C|f|^{-\alpha}$*,* $|\frac{\partial}{\partial\boldsymbol{\theta}} g(f,\boldsymbol{\theta}_0)| \leq C|f|^{-\alpha-1}$ *and* $|\frac{\partial}{\partial\boldsymbol{\theta}} g^{-1}(f,\boldsymbol{\theta}_0)| \leq C|f|^{-\beta}$*, as* $f \to 0$*.*

Assumption 3.3.5. *If* $\alpha \leq 0$*, then there exists* $s(f) \in L_1(\mathbb{R})$ *such that* $|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} g^{-1}(f,\boldsymbol{\theta})| \leq s(f)$ *as* $f \to 0$*. If* $\alpha > 0$*, then* $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} g^{-1}(f,\boldsymbol{\theta})$ *is continuous in* $(f,\boldsymbol{\theta}) \in \mathbb{R}\otimes\boldsymbol{\Theta}_0$*.*

Then as $N \to \infty$*,* $\hat{\boldsymbol{\theta}}_N \to \boldsymbol{\theta}_0$ *almost surely, and*

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\ \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\right),$$

where

$$\boldsymbol{A} = \tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_0(\boldsymbol{\theta}_0), \qquad \boldsymbol{B} = \frac{1}{2f_s}\int_{-f_s/2}^{f_s/2} \left[\tfrac{\partial}{\partial\boldsymbol{\theta}} g(f,\boldsymbol{\theta}_0)^{-1}\right]\left[\tfrac{\partial}{\partial\boldsymbol{\theta}} g(f,\boldsymbol{\theta}_0)^{-1}\right]' (\mathcal{S}_{f_s}(f))^2\,\mathrm{d}f.$$

Remark 3.3.2. *The assumptions given above are adapted from Section 2 of Fox and Taqqu (1986) and Section 8.3 of Giraitis et al. (2012) which are compatible with short-memory, long-memory (persistent) and negative memory (anti-persistent) processes.*

Remark 3.3.3. *As pointed out by Dahlhaus (1989), $\hat{\boldsymbol{\theta}}_N$ may equivalently be defined in terms of the discretized Whittle-type likelihood*

$$\ell_N^{\dagger}(\boldsymbol{\theta}) = \sum_{k=1}^{K} -\log g(f_k, \boldsymbol{\theta}) - \frac{I_N(f)}{g(f_k, \boldsymbol{\theta})}$$

without affecting the result. We proceed with the integrated likelihood $I_N(f)$ in order to simplify the proof.

All the proofs are given in Appendix B.2.

3.4 Semiparametric Methods

Despite its advantages relative to time-domain methods, calculating the Whittle estimator $\tilde{\boldsymbol{\theta}}_{\mathrm{W}}$ in the HTP setting still often remains a computational challenge, even when $\mathcal{S}_{f_s}(f, \boldsymbol{\theta})$ is analytically tractable. Each evaluation of $\ell(\boldsymbol{\theta} \mid \boldsymbol{Y})$ can require millions of evaluations of $\mathcal{S}_{f_s}(f, \boldsymbol{\theta})$. This is compounded with the fact that numerical optimization of $\ell(\boldsymbol{\theta} \mid \boldsymbol{Y})$ can be unstable, with many iterative algorithms requiring a large number of steps to achieve convergence.

3.4.1 Nonlinear Least-Squares Estimator

To reduce the computational burden, a common technique to overcome these issues is to compress the periodogram frequencies into consecutive and non-overlapping bins (e.g., Daniell (1946), (Brockwell and Davis, 2006, Section 10.4), (Fay et al., 2002, Section 2)). That is, assume that $K = B \cdot N_B$ is a multiple of the bin size B, and consider the average periodogram value in bin m,

$$\bar{Y}_m = \frac{1}{B} \sum_{k \in I_m} Y_k, \qquad I_m = \{k : (m-1)B < k \leq mB\}.$$

It then follows from Proposition 1 that if $\mathcal{S}_{f_s}(f_k)$ is relatively constant within bins, the distribution of $\bar{\boldsymbol{Y}} = (\bar{Y}_1, \ldots, \bar{Y}_{N_B})$ can be well-approximated by

$$\bar{Y}_m \overset{\text{ind}}{\sim} \mathcal{S}_{f_s}(\bar{f}_m) \times \text{Gamma}(B, B), \tag{3.9}$$

where $\bar{f}_m = (1/B)\sum_{k \in I_m} f_k$, and $\text{Gamma}(B, B)$ is a Gamma distribution with mean 1 and variance $1/B$. This leads to a popular nonlinear least-squares (NLS) estimator

$$\tilde{\boldsymbol{\theta}}_{\text{NLS}} = \arg\min_{\boldsymbol{\theta}} \sum_{m=1}^{N_B} \left(\bar{Y}_m - \mathcal{S}_{f_s}(\bar{f}_m, \boldsymbol{\theta})\right)^2, \tag{3.10}$$

which produces a consistent estimator of $\boldsymbol{\theta}$ (Nørrelykke and Flyvbjerg, 2010).

3.4.2 Log-Periodogram (LP) Estimator

Not only does the sum-of-squares criterion (3.10) reduce the number of calls to $\mathcal{S}_{f_s}(f, \boldsymbol{\theta})$ by a factor of B, it can also be minimized using specialized algorithms such as Levenberg–Marquardt (Levenberg, 1944; Marquardt, 1963), thus rendering the calculation of $\tilde{\boldsymbol{\theta}}_{\text{NLS}}$ considerably faster than that of $\tilde{\boldsymbol{\theta}}_W$. The drawback of the NLS estimator is that it often incurs a significant loss in statistical precision, since (3.10) does not account for the parameter-dependent variance term $\text{var}(\bar{Y}_m) = \mathcal{S}_{f_s}(\bar{f}_m, \boldsymbol{\theta})/B$. However, taking the logarithm of the compressed periodogram converts model (3.9) to that of constant-variance additive errors, such that

$$Z_m = \log \bar{Y}_m = \log \mathcal{S}_{f_s}(\bar{f}_m) + \epsilon_m, \tag{3.11}$$

with $\exp(\epsilon_m) \overset{\text{iid}}{\sim} \text{Gamma}(B, B)$. Noting that $\mathbf{E}[\epsilon_m] = C_B = \psi(B) - \log B$, where $\psi(\cdot)$ is the digamma function, the least-squares estimator corresponding to (3.11) is the log-periodogram (LP) estimator

$$\tilde{\boldsymbol{\theta}}_{\text{LP}} = \arg\min_{\boldsymbol{\theta}} \sum_{m=1}^{N_B} \left(Z_m - C_B - \log \mathcal{S}_{f_s}(\bar{f}_m, \boldsymbol{\theta})\right)^2.$$

The LP estimator was pioneered by Moulines and Soulier (1999) for the purpose of long-range dependence modeling, i.e., for estimating H in a PSD model of the form $\mathcal{S}_{f_s}(f) \sim |f|^{1-2H}$ as $|f| \to 0$. In this context and with bin size $B = 1$, the LP estimator is akin to that of Geweke and Porter-Hudak (1983) (GPH). However, as B increases, the LP estimator

dominates that of GPH both in terms of convergence rate (Hurvich et al., 1998; Moulines and Soulier, 1999) and asymptotic efficiency (Robinson, 1995).

The LP estimator was originally developed to model long-range dependence in a semi-parametric fashion, and has since been extensively used in this context (e.g., Andrews and Guggenberger, 2003; Sun and Phillips, 2003; Moulines and Soulier, 2003; Andrews and Sun, 2004; Dalla et al., 2006). In the fully-parametric setting, asymptotic properties of the LP estimator have been derived by Fay et al. (2002), and compared favorably therein to the family of efficient estimators proposed by Taniguchi (1987). However, the theory of Fay et al. (2002) applies only to PSD models with short-range dependence. The following Theorem both circumvents this limitation and extends the LP estimator to the HTP setting.

Theorem 3.4.1. *Let $\boldsymbol{\theta}_0$ denote the true parameter value, and for arbitrary frequencies $f_{\min} < f_{\max}$, consider the bandlimited* LP *estimator objective function*

$$\hat{\boldsymbol{\theta}}_{\text{LP}} = \arg\min_{\boldsymbol{\theta}} \sum_{m=N_{\min}+1}^{N_{\max}} \big(Z_m - C_B - \log \mathcal{S}(\bar{f}_m, \boldsymbol{\theta})\big)^2,$$

where $N_{\max} = \lfloor f_{\max} N / f_s \rfloor$ and $N_{\min} = \lfloor f_{\min} N^{\delta} / f_s \rfloor$ for any $\frac{1}{2} < \delta < 1$. Then for fixed f_s, suppose both $\mathcal{S}_{f_s}(f, \boldsymbol{\theta})$ and $\mathcal{S}(f, \boldsymbol{\theta})$ satisfy Assumptions 3.3.1 – 3.3.5, as well as B.2.1 – 3.4.2, there exists $\boldsymbol{\theta}_{f_s}$ and $\mathcal{I}_{f_s}$ such that for $N \to \infty$ we have

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\text{LP}} - \boldsymbol{\theta}_{f_s}) \to \mathcal{N}(\mathbf{0}, D_B \times \mathcal{I}_{f_s}^{-1}),$$

where $D_B = B\psi'(B)$. Moreover, as $f \to \infty$ we have $\boldsymbol{\theta}_{f_s} \to \boldsymbol{\theta}_0$ and $\mathcal{I}_{f_s} \to \mathcal{I}_{f_{\max}}$, the bandlimited optimal Fisher information defined by (3.8).

As noted by Fay et al. (2002), the loss of asymptotic efficiency due to binning is typically negligible, since D_B converges very quickly to 1. For example, $D_{10} \approx 1.05$ and for $B = 100$, $D_{100} \approx 1.005$. Also note that $\mathcal{I}_{f_s}$ here is different from the one in Theorem 3.3.1 for Whittle MLE, which will be clear in the proof.

The proof of Theorem 3.4.1 is provided in Appendix B.2.5 which relies upon the theoretical setup introduced in the following section.

3.4.3 Central Limit Theorem for the LP Estimator

Based on the results of Moulines and Soulier (1999) (summarized in Appendix B.2.3), with two classical assumptions (adapted from Jennrich (1969)), an analog of Theorem 3.3.2 for

log-periodogram regression is provided below. Once again this may be of independent interest for parametric PSD estimation under a misspecified model.

Assumption 3.4.1. *Suppose the true data-generating process is the discrete-time process characterized by $\mathcal{S}_{f_s}(f)$ and $g(f,\boldsymbol{\theta})$ is some proposed spectral density function, let*

$$D(\boldsymbol{\theta}) = \int_0^{f_s} \left(\log g(f,\boldsymbol{\theta}) - \log \mathcal{S}_{f_s}(f)\right)^2 \mathrm{d}f.$$

Then we assume $D(\boldsymbol{\theta})$ is a continuous function in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and $D(\boldsymbol{\theta})$ has a unique minimum at

$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\, D(\boldsymbol{\theta}) \tag{3.12}$$

where $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$.

Assumption 3.4.2. *Suppose $\frac{\partial}{\partial\boldsymbol{\theta}} g(f,\boldsymbol{\theta})$ and $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} g(f,\boldsymbol{\theta})$ exist and are bounded for all $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$ and*

$$\mathcal{I}(\boldsymbol{\theta}_0) = \frac{1}{2f_s}\int_0^{f_s} [\tfrac{\partial}{\partial\boldsymbol{\theta}} \log g(f,\boldsymbol{\theta}_0)][\tfrac{\partial}{\partial\boldsymbol{\theta}} \log g(f,\boldsymbol{\theta}_0)]' \mathrm{d}f \tag{3.13}$$

is positive definite.

Besides, we also need the following technical assumptions to complete our later proof.

Assumption 3.4.3. *The function $g^{-1}(f,\boldsymbol{\theta})$ is continuous at all $(f,\boldsymbol{\theta}) \in \{\mathbb{R}\setminus 0\}\otimes\boldsymbol{\Theta}_0$ and $g^{-1}(f,\boldsymbol{\theta}_0) \le C|f|^\alpha$, where $\boldsymbol{\Theta}_0$, C and α are the same as in Assumptions 3.3.3 and 3.3.4.*

Based on the above assumptions, we obtain the following result.

Theorem 3.4.2. *For fixed f_s, let $\mathcal{S}_{f_s}(f)$ denote the discrete-time PSD of $\{X_n : n \in \mathbb{Z}\}$, such that $\mathcal{S}_{f_s}(f)$ satisfies Assumption B.2.1. Let $g(f,\boldsymbol{\theta}) : \mathbb{R}\otimes\boldsymbol{\Theta} \to \mathbb{R}^+$ satisfy Assumptions 3.4.1 – 3.4.3 as well as Assumptions 3.3.1 – 3.3.5 in Theorem 3.3.2. Now consider the least-squares estimator $\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}} Q_N(\boldsymbol{\theta})$, where*

$$Q_N(\boldsymbol{\theta}) = \frac{1}{N_B}\sum_{m=1}^{N_B} \left(Z_m - C_B - \log g(\bar{f}_m,\boldsymbol{\theta})\right)^2, \qquad N_B = [N/2B].$$

Then $\hat{\boldsymbol{\theta}}_N \to \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is defined by (3.12)*, and*

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\ (2\,\mathcal{I}(\boldsymbol{\theta}_0) - \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1}\cdot 4B\psi'(B)\mathcal{I}(\boldsymbol{\theta}_0)\cdot(2\,\mathcal{I}(\boldsymbol{\theta}_0) - \boldsymbol{A}(\boldsymbol{\theta}_0))^{-1}\right),$$

where $\mathcal{I}(\boldsymbol{\theta}_0)$ *is defined in* (3.13) *and*

$$\boldsymbol{A}(\boldsymbol{\theta}_0) = \frac{1}{f_s}\int_0^{f_s} (\log \mathcal{S}_{f_s}(f) - \log g(f;\boldsymbol{\theta}_0)) \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \log g(f,\boldsymbol{\theta}_0)\mathrm{d}f.$$

The proof is given in Appendix B.2.4.

3.5 Numerical Study

In order to visually reveal the statistical properties proposed in Theorem 3.3.1 and Theorem 3.4.1, we introduce the following simulation study that is based on a ubiquitous model for HTP data known as the Simple Harmonic Oscillator (SHO).

3.5.1 Simple Harmonic Oscillator

The SHO model is a stochastic extension of the deterministic mass-spring system, describing a type of periodic motion where movement from rest state entails a restoring force directly proportional to the displacement. Physics of oscillators have been employed to model ocean surface drifter trajectories (Sykulski et al., 2016) and AFM cantilever calibration (Chon et al., 2000). That is, the time-dependent position X_t of a particle of interest satisfies the differential equation

$$m\ddot{X}_t = -kX_t - \varsigma\dot{X}_t + F_t, \tag{3.14}$$

where $\dot{X}_t$ and $\ddot{X}_t$ are the particle's velocity and acceleration, m is its mass, k is the spring-like restoring force, ς is the friction coefficient, and F_t is the thermal force driving the particle's motion. It is modeled as a mean-zero white noise process with $\mathrm{cov}(F_s, F_{s+t}) = 2k_BT\varsigma\delta(t)$, where $\delta(\cdot)$ is the Dirac delta function, T is absolute temperature in Kelvin and k_B is Boltzmann's constant.

In fact, equation (3.14) can be regarded as a degenerate GLE influenced by a harmonic potential $U(x_t) = -\frac{1}{2}kx_t^2$ with the memory kernel being $\delta(t)$, then the noise term is white or Markovian. A detailed construction of such a simple harmonic oscillator heat bath is given in Zwanzig (2001, Section 1.6).

Applying Fourier transform simply gives us the spectral density of X_t

$$\mathcal{S}(f) = \frac{k_B T/(k \cdot \pi f_0 Q)}{\left[(f/f_0)^2 - 1\right]^2 + \left[f/(f_0 Q)\right]^2}, \tag{3.15}$$

where $f_0 = \sqrt{k/m}/(2\pi)$ is SHO resonance frequency and $Q = \sqrt{km}/\varsigma$ is its "quality factor", which measures the width of the spectral peak around f_0. The SHO is used pervasively in the natural sciences, e.g., describing the random oscillations of a particle trapped in an optical tweezer (Nørrelykke and Flyvbjerg, 2010), chemical bond lengths and angles (Tsekov, 2016), X-ray emissions from neutron stars (Barret and Vaughan, 2012), and oceanographic surface drifters (Sykulski et al., 2016).

3.5.2 Simulation Setup

In our simulation study, the true underlying data were generated using a standard FFT-based algorithm (e.g., Labuda et al., 2012). The parameters used in all simulations are displayed in Table 3.1. Each dataset consists of a 5 s time series sampled at 10 MHz ($N = 5 \times 10^7$ data points) from the SHO model (3.15).

When we fix all parameters except $Q \in \{1, 10, 100, 500\}$, the corresponding SHO spectra displayed in Figure 3.2 represent a broad range of stochastic oscillatory phenomena.

For the estimation comparison, we added pure white noise to the original data such that the noisy observations follow the following PSD.

$$\mathcal{S}(f, \boldsymbol{\theta}) = A_{\mathrm{w}} + \frac{k_B T/(k \cdot \pi f_0 Q)}{\left[(f/f_0)^2 - 1\right]^2 + \left[f/(f_0 Q)\right]^2},$$

for which A_{w} represents the additional noise component in the frequency domain, and the parameters to be estimated are $\boldsymbol{\theta} = (k, f_0, Q, A_{\mathrm{w}})$.

For each simulation setting $M = 1000$ datasets were generated, and for each dataset we calculated the three estimators $\hat{\boldsymbol{\theta}}_W$, $\hat{\boldsymbol{\theta}}_{\mathrm{NLS}}$, and $\hat{\boldsymbol{\theta}}_{\mathrm{LP}}$. Bandlimiting was applied by using only the $N_{\mathrm{band}} = 325000$ periodogram frequencies between $f_{\mathrm{min}} = 15$ kHz and $f_{\mathrm{max}} = 80$ kHz, outside of which the remaining frequencies add little information about $\boldsymbol{\theta}$. In practice, determining a suitable frequency bandwidth may require a combination of good domain knowledge and a preliminary nonparametric estimation. For the NLS and LP estimators, the bin size was set to $B = 100$. For all estimators, the optimization problem was reduced from four to three parameters by a profile likelihood procedure (Lysy et al.,

Table 3.1: SHO parameters in baseline environment.

SHO Parameter	Value
Temperature	$T = 298\,\mathrm{K}$
Noise Magnitude	$A_\mathrm{w} = 19\,000\,\mathrm{fm}^2\,\mathrm{Hz}^{-1}$
Cantilever Stiffness	$k = 0.172\,\mathrm{N\,m}^{-1}$
Resonance Frequency	$f_0 = 33.533\,\mathrm{kHz}$
Quality Factor	$Q \in \{1, 10, 100, 500\}$

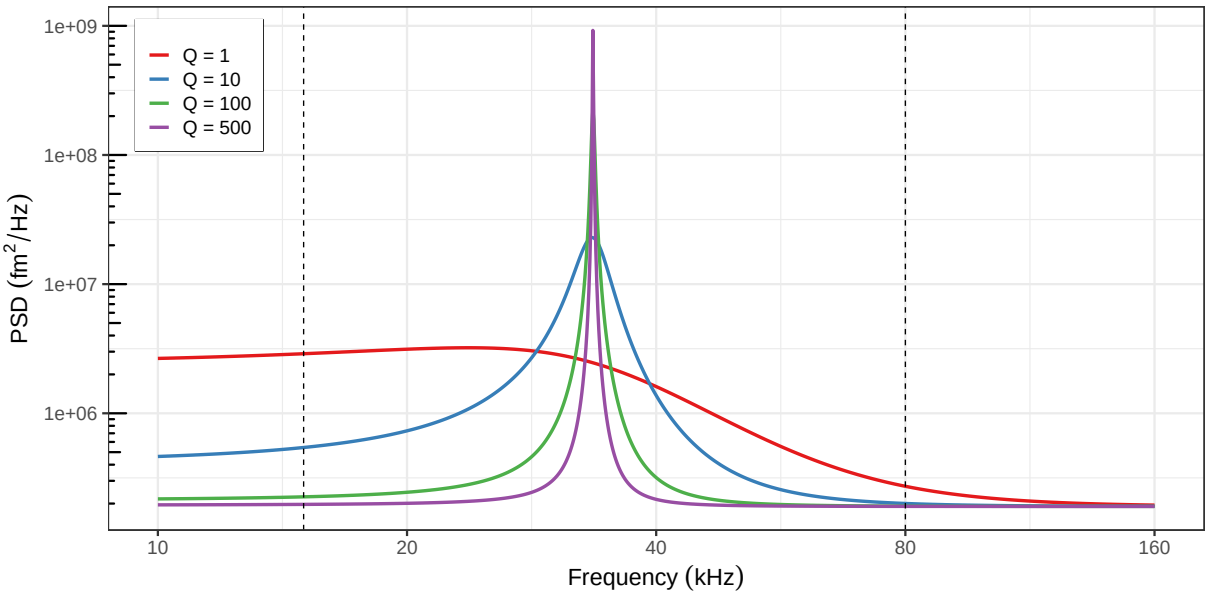

Figure 3.2: Baseline PSDs over a range of quality factors (Q). The dashed vertical lines indicate the frequency bandwidth used for parameter estimation.

2022, Appendix A). The profile likelihood was maximized using the Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) for $\hat{\boldsymbol{\theta}}_\mathrm{LP}$ and $\hat{\boldsymbol{\theta}}_\mathrm{NLS}$, whereas for the Whittle estimator the more general BFGS algorithm (Fletcher, 1987) was used. Both algorithms require gradient information, for which we have provided an efficient C++ implementation via automatic differentiation in the R/C++ package `realPSD` (Zhu and Lysy, 2021).

3.5.3 Baseline Estimation Comparison

Figure 3.3 displays parameter-wise boxplots for each of the estimator standardized relative to its true value. The numbers on top of each boxplot correspond to the mean squared error (MSE) ratios between each estimator and the Whittle estimator $\hat{\boldsymbol{\theta}}_W$, hereafter referred to as the MLE. That is, for each of the SHO parameters $\varphi \in (f_0, k, Q)$ and estimator $j \in \{\text{NLS}, \text{LP}, \text{MLE}\}$, the corresponding MSE ratio in Figure 3.3 is calculated as

$$\mathcal{R}_j(\varphi) = \frac{\sum_{i=1}^{M}(\hat{\varphi}_j^{(i)} - \varphi_0)^2}{\sum_{i=1}^{M}(\hat{\varphi}_W^{(i)} - \varphi_0)^2},$$

where φ_0 is the true parameter value and $\hat{\varphi}_j^{(i)}$ is its estimate by method j for dataset i.

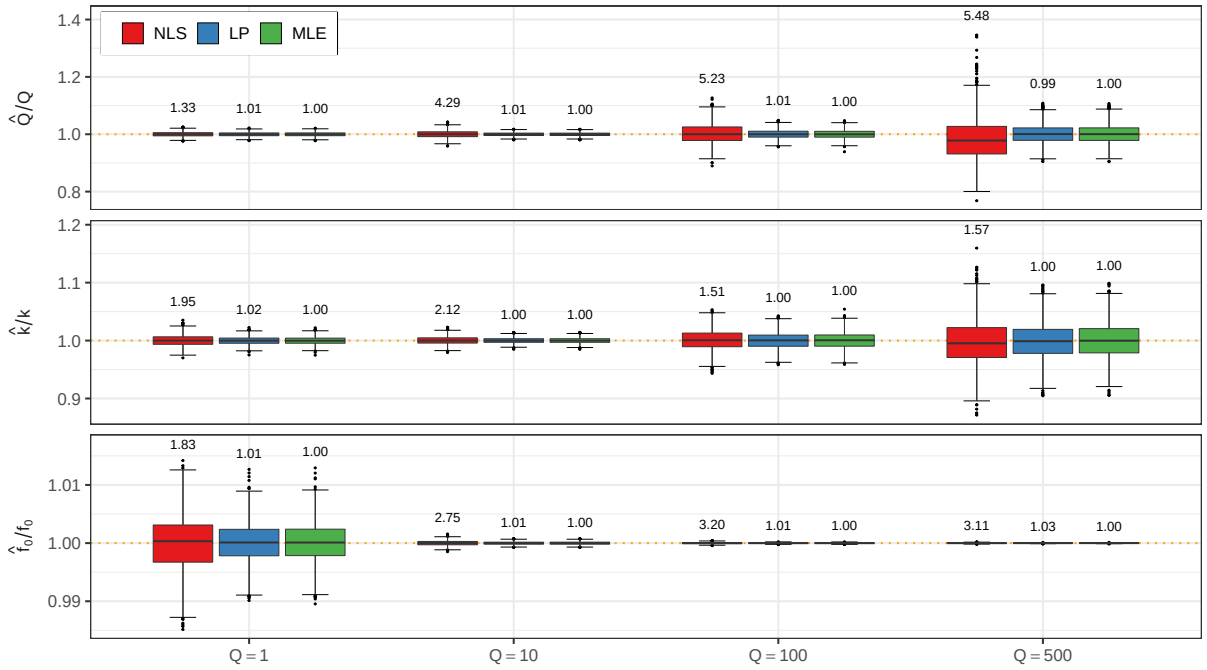

Figure 3.3: Comparison of NLS, LP, and MLE estimators in the baseline simulation environment. Numbers indicate MSE ratios of the corresponding estimators relative to MLE.

For low quality factor $Q = 1$, the NLS method has roughly 1.5-2 times higher MSE than the MLE. For higher values of Q, the MSE of NLS increases to roughly 3-5 times that of MLE. In contrast, the LP estimator achieves virtually the same MSE as the MLE,

but at a 150 to 200 times smaller computational cost. This is due in large part to the bin size $B = 100$ providing must faster log-likelihood evaluations, and also to the Levenberg–Marquardt algorithm typically converging in fewer steps than BFGS.

3.5.4 Electronic Noise Contamination

In many real applications, observations are usually contaminated by some electronic noise. In order to assess the performance of three estimators, we simulated periodic noise component of the following form.

$$g(t) = D \cdot \sin(2\pi\zeta t + \phi)$$

The frequency ζ of each sine wave was drawn from a Normal with mean and standard deviation

$$\mu_{\text{sine}} = f_0 = 33.533\,\text{kHz}, \qquad \sigma_{\text{sine}} = 20\,\text{Hz}.$$

This scenario is particularly difficult for SHO parameter estimation due to the proximity of ζ to the SHO frequency f_0, but it is not particularly uncommon. For details, please refer to Lysy et al. (2022).

Figure 3.4 displays a simulated dataset with electronic noise contamination. The cutoff line was based on the noise-removal procedure proposed by Lysy et al. (2022) which we do not care about here.

Figure 3.5 displays boxplots of each parameter estimate relative to its true value with the existence of electronic noise. To assess the impact of the noise contamination, these estimates do not include the preliminary denoising step. The numbers in the plot correspond to MSE ratios between the estimator with noise contamination, relative to its own performance in the baseline dataset. The ratios formula is given below.

$$\mathcal{R}_j(\varphi) = \frac{\sum_{i=1}^{M}(\hat{\varphi}^{(i)}_{j,\text{noise}} - \varphi_0)^2}{\sum_{i=1}^{M}(\hat{\varphi}^{(i)}_{j,\text{base}} - \varphi_0)^2},$$

where $\hat{\varphi}^{(i)}_{j,\text{base}}$ and $\hat{\varphi}^{(i)}_{j,\text{noise}}$ are parameter estimates with method j for dataset i under baseline and noise-contaminated settings, respectively. At low Q, the MSE ratios are close to one, suggesting that the estimators are relatively insensitive to the electronic noise. However, for high Q the effect of the noise is considerably more detrimental, particularly for NLS. In all cases, the performance of the LP estimator is affected the least, indicating it is naturally

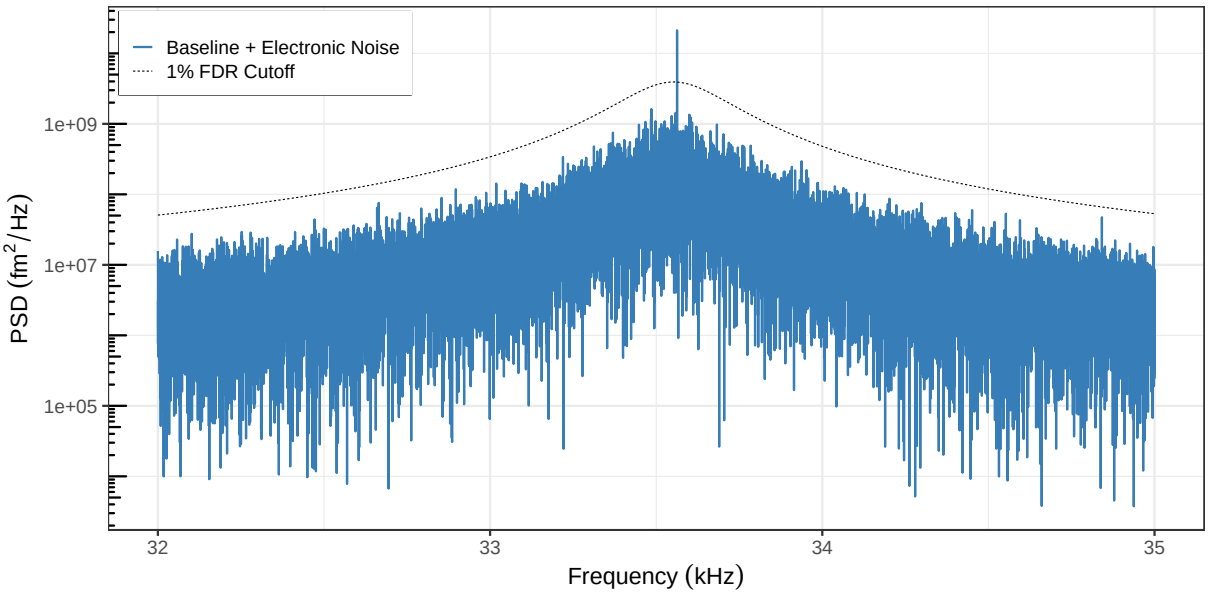

Figure 3.4: Simulated SHO periodogram with $Q = 100$ and added electronic noise, along with the FWER periodogram cutoff corresponding to an FDR level of $\alpha = 1\%$.

more robust than NLS and MLE to periodic noise contamination, even before the denoising technique is applied.

So far, we have gained some basic insights into the performance of three different estimators. Since we are not focused on the two-stage noise removal method, we will not discuss the improvement of estimates after the noise removal. For the details of effectiveness of our noise removal method and further simulation based on severe periodic noise conditions, please refer to Lysy et al. (2022).

3.5.5 Application: Calibration of an AFM

Finally, we introduce a real-world application of fitting the SHO model — the calibration of an atomic force microscopy (AFM), a very-high-resolution type of scanning probe microscopy, with demonstrated resolution on the order of fractions of a nanometer.

In a typical AFM experiment, the cantilever's bending response is measured in opposition to its spring-like restoring force, which requires proper calibration of the cantilever stiffness in order to convert measured displacement readings into force (Cleveland et al.,

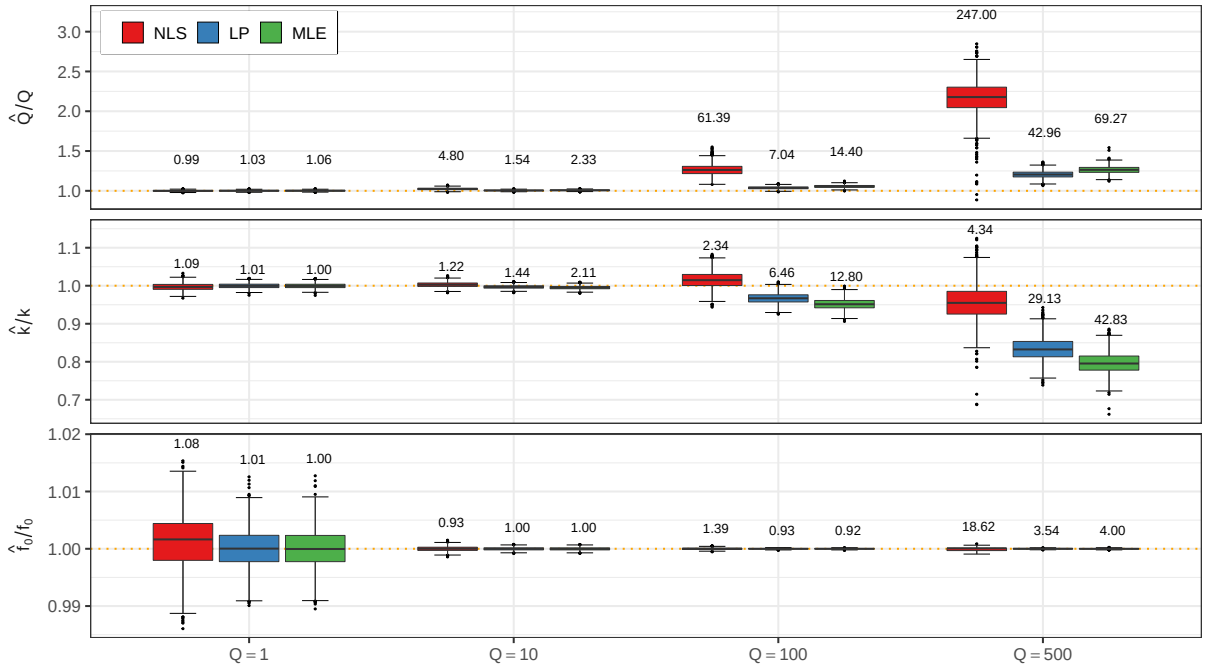

Figure 3.5: Comparison of NLS, LP, and MLE preliminary estimators (i.e., prior to noise removal), in the noise-contaminated environment. Numbers indicate MSE ratios of the corresponding estimators relative to their own performance at baseline.

1993; Burnham et al., 2002; Clarke et al., 2006; Sader et al., 2011). This is accomplished by fitting various parametric models to a baseline spectral density recording, i.e., to a cantilever driven by thermal noise alone.

The problem of calibrating an AFM cantilever mainly focuses on fitting its periodogram which is displayed in Figure 3.6. The data consist of a 5-second time series recorded by a commercial cantilever (TR400-S Olympus) at a sampling frequency of $f_s = 5\,\text{MHz}$ ($N = 2.5 \times 10^7$ observations). The objective is to determine the parameters of the best-fitting model to the first cantilever eigenmode (Figure 3.6b), which is severely contaminated by electronic noise near its peak (around 33.5 kHz).

While the PSDs in the simulations of Section 3.5.2 are dominated at low frequencies by white noise, that of the experimental data in Figure 3.6a exhibits power-law behavior,

$$\mathcal{S}(f) \sim 1/f^{\alpha} \quad \text{as} \quad f \to 0.$$

This phenomenon is referred to as "$1/f$ noise" and features prominently in AFM experi-

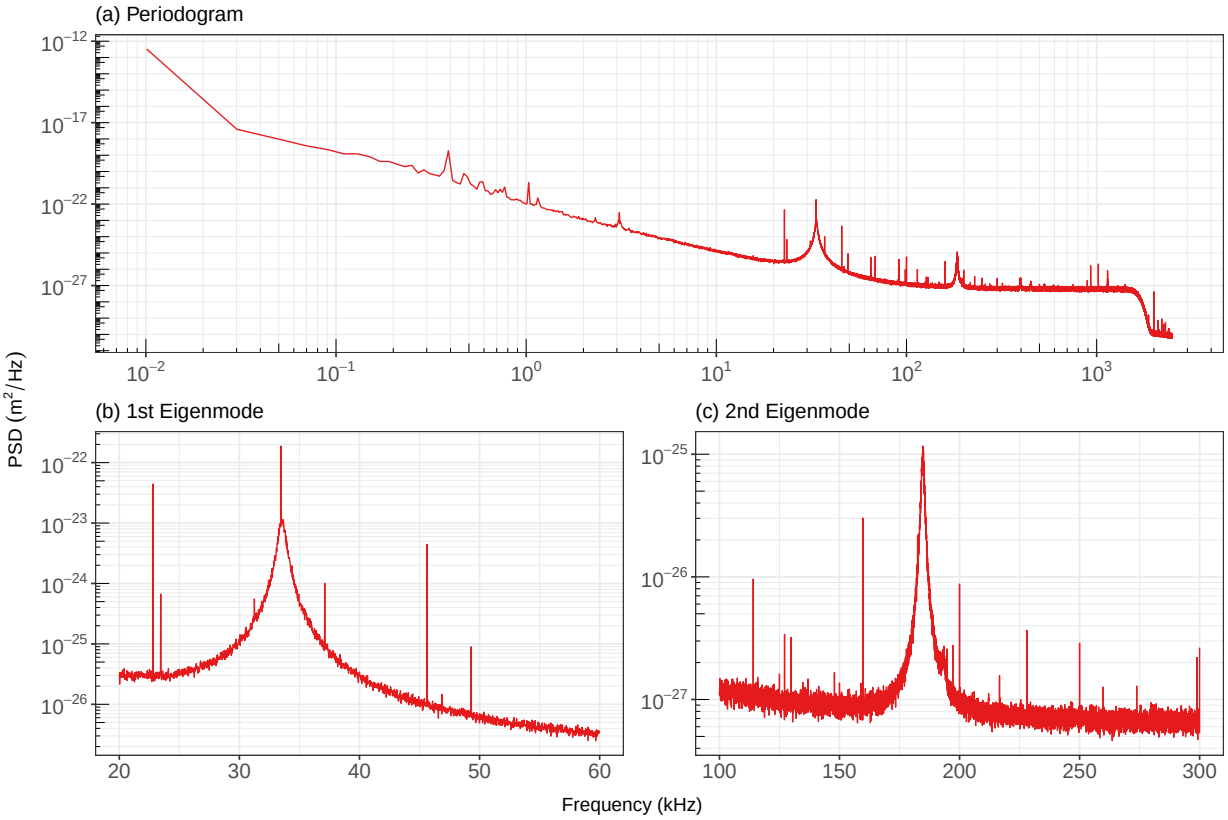

Figure 3.6: (a) Periodogram for a TR400-S Olympus cantilever recorded for 5 s at 5 MHz ($N = 2.5 \times 10^7$ observations). The data have been averaged by bins of size $B = 100$ to enhance visibility. (b-c) Magnified view of first and second eigenmodes.

ments (e.g., Harkey and Kenny, 2000; Giessibl, 2003; Heerema et al., 2015). It is due in this case to slow fluctuations of the measurement sensor. Depending on the value of α, $1/f$ noise induces long-range dependence in the cantilever displacement ($0 < \alpha < 1$), or even lack of stationarity ($\alpha \geq 1$). Failing to account for it can significantly bias SHO parameter estimates. Fortunately, $1/f$ noise can be readily dealt with by adding a correction term to the SHO model (3.15), which becomes

$$\mathcal{S}(f, k, f_0, Q, A_\mathrm{w}, A_\mathrm{f}, \alpha) = A_\mathrm{w} + \frac{A_\mathrm{f}}{f^\alpha} + \frac{k_B T/(k \cdot \pi f_0 Q)}{\left[(f/f_0)^2 - 1\right]^2 + \left[f/(f_0 Q)\right]^2}. \tag{3.16}$$

Model (3.16) is used to calibrate the SHO AFM parameters.

In addition to the first eigenmode near 33.5 kHz, the AFM data contain higher eigenmodes arising from the flexural oscillations of the cantilever beam (Sader, 1998). The first of these higher eigenmodes is displayed in Figure 3.6c.

It is essentially important to calibrate the higher eigenmodes for many bimodal and multifrequency AFM imaging techniques (e.g., Martinez et al., 2008; García and Proksch, 2013; Herruzo et al., 2014; Labuda et al., 2016b). We could do this by fitting separate SHO models to each successive eigenmode. However, as the peak amplitude of these higher modes approaches the noise floor, the accuracy of separate SHO estimators rapidly deteriorates. An alternative avenue is to combine SHO models on the basis of hydrodynamic principles (van Eysden and Sader, 2006; Clark et al., 2010) and other scaling laws (Labuda et al., 2016a) which we do not pursue here.

Figure 3.7 displays the periodogram of the AFM data over the frequency range used for parameter estimation ($N_{\text{band}} = 200000$ periodogram frequencies). The data were fit with the two-stage estimator for the MLE method and for the NLS and LP methods with bin size $B = 100$, for two special cases of the SHO + noise model (3.16). The first fixes $A_{\text{f}} = 0$, i.e., SHO + pure white noise, and the second fixes $A_{\text{w}} = 0$, i.e., SHO + pure $1/f$ noise. We refer to these models as SHOW and SHOF, respectively. Indeed, fitting the full model (3.16) to the data in Figure 3.7 resulted in very large standard errors for A_{w}, A_{f}, and α, suggesting a lack of parameter identifiability. For more detailed analysis of the effects caused by both white noise and $1/f$ noise, please refer to Lysy et al. (2022).

Figure 3.7 displays SHOW and SHOF models fit with the various methods (missing is the SHOW MLE estimator which failed to converge). The SHOW model fits the data considerably worse than SHOF. Out of the SHOF estimators, the fit of NLS is somewhat worse than that of LP and MLE, which are virtually indistinguishable from one another. This is because the NLS estimator penalizes the distance between empirical and fitted PSD on the regular scale, whereas Figure 3.7 plots these on the log scale. Therefore, visual differences at the tails of the plot are downweighted by NLS relative to differences near the peak (around 33.5 kHz). In contrast, the LP estimator penalizes differences on the log scale, thus weighting them uniformly on the visual scale of Figure 3.7. For further discussion about the estimation and noise contamination effects on the standard error estimation, please refer to Lysy et al. (2022, Section 5.1).

The estimation results above were obtained with a bin size of $B = 100$, for which the NLS and LP estimators are two orders of magnitude faster than the MLE. To investigate the impact of bin size, Figure 3.8a plots parameter estimates for the NLS and LP methods for values of $B = 10$ to 1000. The behavior of NLS is considerably more erratic, presumably

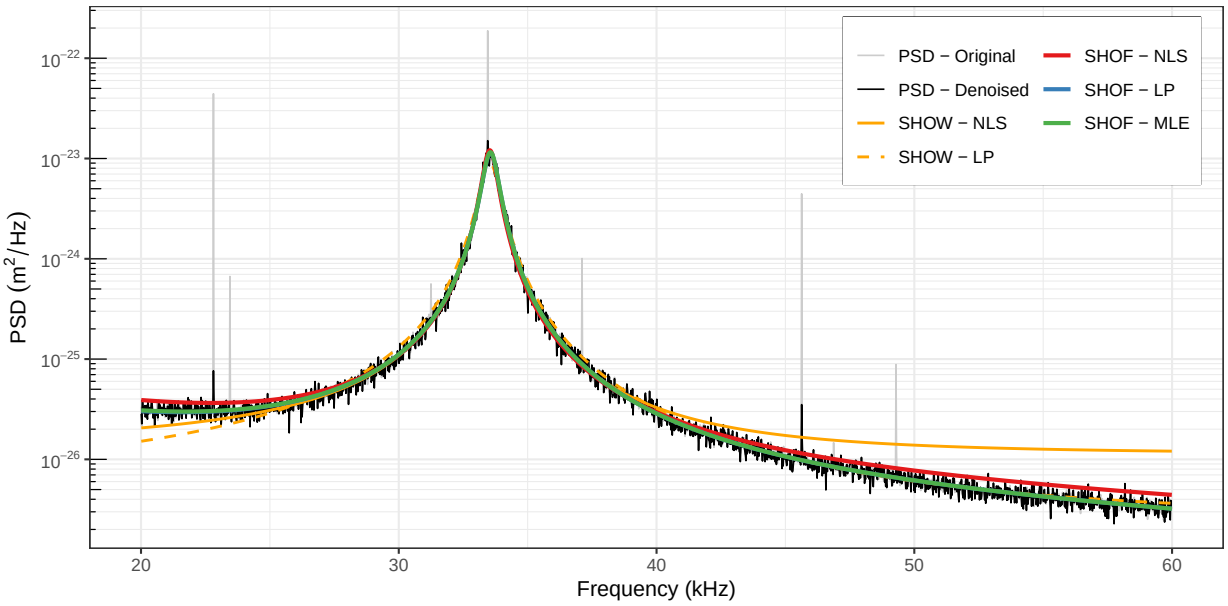

Figure 3.7: Periodogram (averaged by bins of size $B = 100$) and the fitted SHOW and SHOF models using NLS, LP and MLE estimators (not shown is SHOW - MLE which failed to converge). The large spikes indicate frequencies at which there is electronic noise contamination.

due to small changes in the bin end points having larger impact on $\mathcal{S}(\bar{f}_m, \boldsymbol{\theta})$ than on its logarithm. A similar pattern is seen for standard errors in Figure 3.8b. The NLS parameter estimates and standard errors appear to be negatively correlated, resulting in fluctuation of the coefficient of variation (CV) in Figure 3.8c by a factor of two. In contrast, the effect of bin size on the LP standard error and CV is imperceptible.

3.6 Conclusion

We laid the theoretical foundations for the asymptotic properties of the Whittle MLE and LP estimators for HTP data upon which Lysy et al. (2022)'s work relies. Our theoretical results may help shed some time on this relatively new direction of time series analysis.

In addition, we introduced two numerical studies to evaluate the theoretical properties under various situations, including (1) simulated SHO model with white noise measurement error; (2) simulated SHO model with periodic electronic noise; (3) real world AFM data

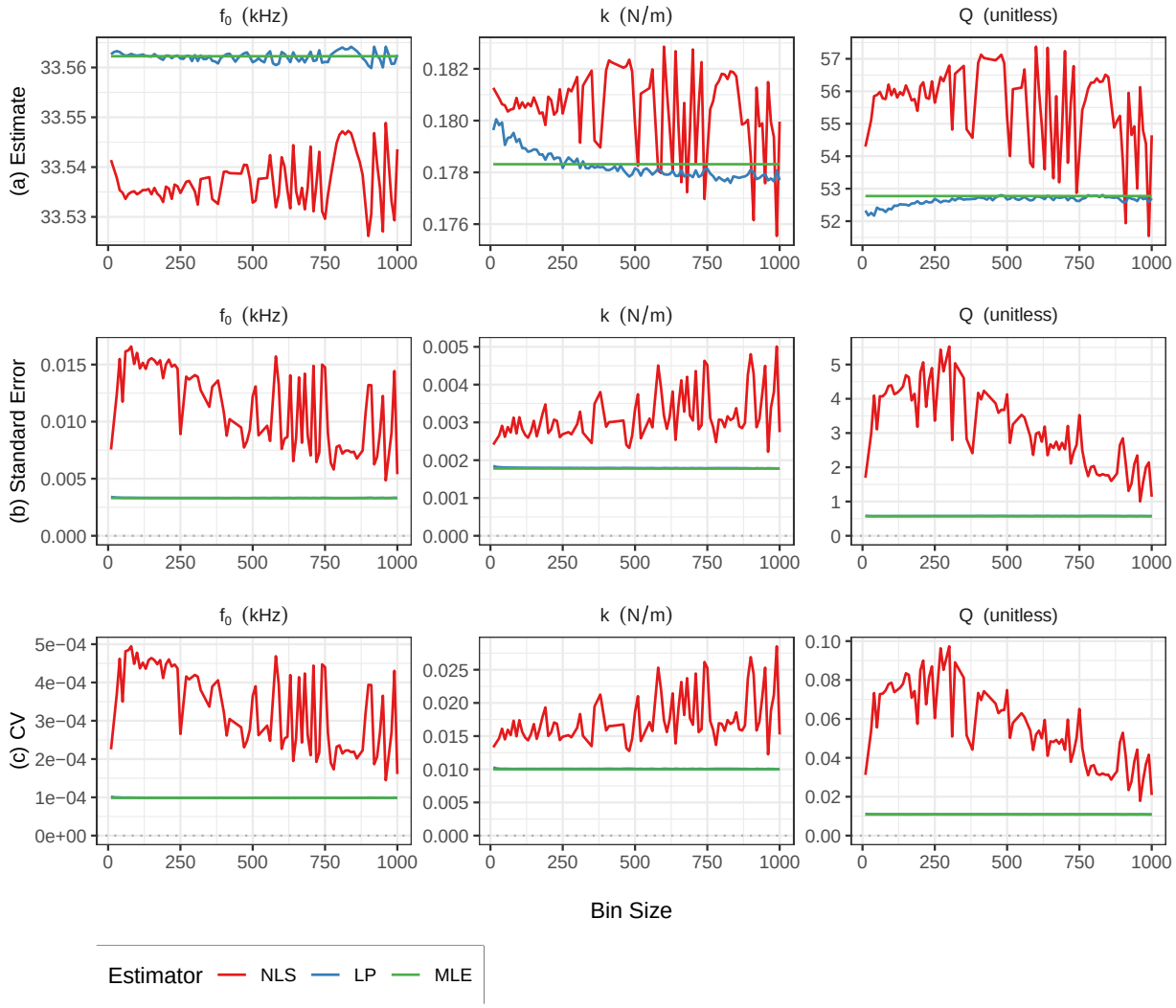

Figure 3.8: NLS, LP, and MLE estimators for bin size $B = 10$ to 1000. (a) Parameter estimates. (b) Standard errors. (c) Coefficient of variation (CV).

fitted by SHO model with noise term. From the numerical results, we can see that the Whittle MLE and the LP estimators are indeed bandlimited optimal as suggested by Theorem 3.3.1 and Theorem 3.4.1, respectively. In particular, the LP estimator may be more robust, and in most cases can achieve a level of statistical efficiency comparable to the MLE at a small fraction of computational costs. This advantage is largely due to the data

compression by binning of periodogram ordinates before the estimation. This is the key ingredient of our successful two-stage noise removal method proposed in the work of Lysy et al. (2022). A potential direction of new research may be to employ adaptive bin size in the data compression, depending on the local shape of spectral density, or to propose some other more effective data reduction techniques.

Moreover, the numerical studies provide great examples of performing parameter inference (in particular spectral estimation) given the massive amount of (possibly noisy) data generated from an underlying GLE model. However, the SHO model is just a degenerate GLE with the memory kernel being a Dirac delta function. For a general GLE with non-trivial memory kernel and nonlinear potential force term, the convolution term will make statistical inference much more difficult, as we shall discuss in Chapter 4.

Chapter 4

Quasi-Markovian Approximation of the Generalized Langevin Equation

4.1 Introduction

Recall the GLE from Chapter 2:

$$\ddot{X}(t) = -U'(X(t)) - \int_0^t K(t-s)\dot{X}(s)\mathrm{d}s + F(t), \tag{4.1}$$

where we have assumed for simplicity that the particle mass $m = 1$, and where the force ACF $\mathcal{C}_F(t) = \beta^{-1}K(t)$, with $\beta = 1/(k_B T)$ being a constant assumed to be known (T is the temperature preset in an experiment). In this Chapter, we are interested in parametrizing the potential energy $U_{\boldsymbol{\phi}}(x)$ and memory kernel $K_{\boldsymbol{\eta}}(t)$ by unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\eta})$ and estimate these using discrete observations of the GLE $\boldsymbol{X} = (X_0, \ldots, X_N)$, where $X_n = X(n\Delta t)$ for $n = 1, \ldots, N$.

4.1.1 Statistical Challenges

However, there are two main challenges.

First, analytically the easy solutions to GLEs have largely been restricted to the linear case where the potential is harmonic, i.e. $U(x) = \frac{1}{2}\kappa x^2$. However, linear GLEs can only be useful for systems near equilibrium. In a nonlinear system ("nonlinear" here means the

potential energy contains higher powers than quadratic), the application of linear GLE is limited. For example, it cannot be used to treat nonlinear transport processes (Zwanzig, 2001, Chapter 9). Due to the complex nature of nonlinear problems, solving nonlinear GLEs analytically is difficult, let alone obtaining its analytic likelihood function, which impedes the likelihood-based inference.

Second, the convolution term $\int_0^t K(t-s)\dot{X}(s)\mathrm{d}s$ reveals that the trajectory $X(t)$ is non-Markovian since at time t it depends on the entire past trajectory. Therefore, numerically, we still lack a proper and general way to approximate the exact GLE. By "proper", we mean if we have a theoretical autocovariance structure or spectral density, we would like to fully recover the covariance information either in the time domain or in the spectral domain (via the Fourier transform) by using the discretized GLE. By "general", we mean the discretization method should be suitable for a general class of or hopefully any nonlinear GLEs. If this could be done, it would provide a framework for fast parameter fitting of GLE and thus propel the development of many single-particle experiments.

A promising approach to GLE inference is that *certain* GLE models can be recast as stochastic differential equations (SDEs). We say that a d-dimensional stochastic process $\boldsymbol{X}(t) = (X_1(t), \ldots, X_d(t))$ is an SDE when it satisfies the differential equation (Oksendal, 2013)

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{\Lambda}(\boldsymbol{X}(t))\mathrm{d}t + \boldsymbol{\Sigma}(\boldsymbol{X}(t))^{1/2}\mathrm{d}\boldsymbol{B}(t), \tag{4.2}$$

with some stochastic noise driven by Brownian motion $\boldsymbol{B}(t)$ where $\boldsymbol{\Lambda}(\boldsymbol{X})$ is a d-dimensional drift function and $\boldsymbol{\Sigma}(\boldsymbol{X})$ is a $d \times d$ positive-definite diffusion matrix. As for modeling GLEs using discrete observations, if the memory kernel $K(t)$ is a sum of exponential functions (Fricks et al., 2009; McKinley et al., 2009), then we can characterize the GLE by adding a finite number of auxiliary variables latent variables. That is, we have the following *quasi-Markovian approximation* of the GLE (qmGLE).

Theorem 4.1.1 (Pavliotis, 2014, Proposition 8.1)**.** *Consider the GLE* (4.1) *for which the memory kernel* $K(t)$ *may be written as*

$$K(t) = \langle e^{-\boldsymbol{A}t}\boldsymbol{\lambda}, \boldsymbol{\lambda}\rangle,$$

where $\boldsymbol{\lambda} \in \mathbb{R}^d$ *and* $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ *is a positive definite matrix. Then the law of the GLE process* $X(t)$ *in* (4.1) *is identical to that arising from the system of SDEs*

$$\begin{aligned}
\mathrm{d}X(t) &= V(t)\mathrm{d}t \\
\mathrm{d}V(t) &= -U'_{\boldsymbol{\phi}}(X(t))\mathrm{d}t + \langle \boldsymbol{\lambda}, \boldsymbol{z}(t)\rangle \mathrm{d}t \\
\mathrm{d}\boldsymbol{z}(t) &= -\left(V(t)\boldsymbol{\lambda} + \boldsymbol{A}\boldsymbol{z}(t)\right)\mathrm{d}t + \boldsymbol{\Sigma}\mathrm{d}\boldsymbol{B}(t),
\end{aligned}$$

where $\boldsymbol{z}(t) \in \mathbb{R}^{d\times 1}$, $\boldsymbol{z}(0) \sim \mathcal{N}(\boldsymbol{0}, \beta^{-2}\boldsymbol{I})$ *($\beta > 0$) and the diffusion matrix* $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ *satisfies*

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\dagger} = \beta^{-2}(\boldsymbol{A} + \boldsymbol{A}^{\dagger})$$

where $\dagger$ *denotes the conjugate transpose, and* $\boldsymbol{B}(t)$ *is d-dimensional Brownian motion.*

The above quasi-Markovian scheme (or a similar one) was first proposed by theoretical physicists and mathematicians, for example Eckmann et al. (1999); Kupferman (2004), and later was carefully studied from a mathematician's perspective by Ottobre and Pavliotis (2010). However, the main focus was on the solution or infinitesimal generator, the ergodic properties and some other analytic properties instead of the parameter inference.

In contrast to GLEs, there is an enormous amount of studies on likelihood-based parameter inference for the SDE with (full or partial observations). One direction is to consider estimators alternative to the MLE, for example indirect inference (Gourieroux et al., 1993), efficient methods of moments (Gallant and Long, 1997), estimating functions (Bibby et al., 2004), and higher-order Taylor approximations to the intractable likelihood function (Aït-Sahalia, 2002, 2008). However, arguably the most widely-used approach is Bayesian data augmentation via an Euler–Maruyama numerical discretization scheme (or simply Euler scheme, as explained later in Section 4.2), for which sampling from the joint posterior distribution of parameters and latent variables is achieved by using a variety of Markov chain Monte Carlo (MCMC) methods (e.g., Elerian et al., 2001; Eraker, 2001; Roberts and Stramer, 2001; Golightly and Wilkinson, 2008; Kou et al., 2012; Whitaker et al., 2017).

In the context of the data augmentation approach, the complete data is generated from a so-called state-space model (defined in Section 4.2.1), for which sampling from the latent variables can be done efficiently using a particle filter (Andrieu et al., 2010). The simplest particle filter is the so-called bootstrap filter (Gordon et al., 1993) which proposes latent variables from their forward distribution. As the bootstrap particle filter works very poorly for SDE problems, a bridge proposal developed by Durham and Gallant (2002) specifically for SDEs has been successfully used instead (e.g., Chib et al., 2004; Golightly and Wilkinson, 2008; Whitaker et al., 2017; Picchini and Forman, 2019; Botha et al., 2021). However, the qmGLE is a hypoelliptic diffusion (Ottobre and Pavliotis, 2010), for which using the Euler discretization for parameter inference must be done with care (Pokern et al., 2009; Ditlevsen and Samson, 2019). In particular, it precludes a direct application of the SDE bridge proposal, as we shall show in Section 4.3.2.

4.1.2 Our Contribution

In this Chapter, we propose to use a non-degenerate discretization scheme for the qmGLE via the Itô-Taylor (IT) expansion presented in Pokern et al. (2009). These authors show that the IT scheme produces biased estimates of SDE drift parameters when used in a Gibbs sampler which alternately updates parameters and latent variables conditioned on each other and the observed data. However, we argue in Section 4.3.4 that this problem does not affect MCMC (or other inference methods) based on particle filters which integrates the latent variables out of the problem, rather than sampling from them along with the parameters. A different discretization scheme for hypoelliptic diffusions is given by Ditlevsen and Samson (2019), which ostensibly does not break down in the Gibbs sampling scheme described above. However, our IT scheme is much simpler, and indeed the main contribution of this chapter is to extend the SDE bridge proposal of Durham and Gallant (2002), originally developed for the Euler scheme, to the IT scheme as well. We believe this will lead to an efficient particle filtering method for the qmGLE and for hypoelliptic diffusions in general. That being said, the work in this Chapter is merely conceptual, as implementation of the proposed particle filter is the subject of future work.

4.1.3 Outline

The remainder of this chapter is organized as follows. In Section 4.2 we introduce our model settings and present the particle filter algorithm along with the bridge proposal tailored to SDEs. In Section 4.3 we demonstrate why this approach breaks down for the qmGLE. Then, we introduce the modified discretization scheme and build the corresponding bridge proposal for general particle filters. Concluding remarks are offered in Section 4.4.

4.2 SDE Inference with Particle Filters

Before we discuss the parameter inference for the qmGLE, let us introduce some basic knowledge of particle filters and SDEs.

Consider the SDE (4.2) in the context of parameter inference,

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{\Lambda}_{\boldsymbol{\theta}}(\boldsymbol{X}(t))\mathrm{d}t + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{X}(t))^{1/2}\mathrm{d}\boldsymbol{B}(t)\,,$$

where $\boldsymbol{\Lambda}_{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ now depend on some unknown parameters $\boldsymbol{\theta}$ to be estimated from the data. In many scientific experiments, $\boldsymbol{X}(t)$ is measured with error at times $t_n = n\Delta t$ such

that the observed data is $\boldsymbol{Y}_{0:N} = (\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_N)$ with

$$\boldsymbol{Y}_n \overset{\text{ind}}{\sim} g(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{\theta}),$$

where $\boldsymbol{X}_n = \boldsymbol{X}(t_n)$. Assume that the initial value of the SDE at time $t_0 = 0$ has the prior distribution $\boldsymbol{X}_0 \sim \pi(\boldsymbol{X}_0 \mid \boldsymbol{\theta})$. Then the likelihood function is given by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) &= \int p\left(\boldsymbol{Y}_{0:N}, \boldsymbol{X}_{0:N} | \boldsymbol{\theta}\right) \mathrm{d}\boldsymbol{X}_{0:N} \\ &= \int \left[\pi(\boldsymbol{X}_0 \mid \boldsymbol{\theta}) \cdot \prod_{n=0}^{N} g(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{\theta}) \cdot \prod_{n=1}^{N} p(\boldsymbol{X}_n \mid \boldsymbol{X}_{n-1}, \boldsymbol{\theta}) \right] \mathrm{d}\boldsymbol{X}_{0:N}. \end{aligned} \tag{4.3}$$

However, this expression requires the SDE transition density $p(\boldsymbol{X}_n \mid \boldsymbol{X}_{n-1}, \boldsymbol{\theta})$, which is rarely available in closed form. Instead, the likelihood function is often approximated by the so-called Euler–Maruyama discretization method (Maruyama, 1955; Pedersen, 1995a,b). Namely, for $m \geq 1$, let $\boldsymbol{X}_n^{(m)}$ denote the value of the SDE at time $n\Delta t/m$, such that $\boldsymbol{X}_{mn}^{(m)} = \boldsymbol{X}_n = \boldsymbol{X}(t_n)$. As $m \to \infty$, $\Delta t/m \to 0$, the normal approximation

$$\boldsymbol{X}_n^{(m)} \sim \mathcal{N}(\boldsymbol{X}_{n-1}^{(m)} + \boldsymbol{\Lambda}_{\boldsymbol{\theta}}(\boldsymbol{X}_{n-1}^{(m)})\Delta t/m, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{X}_{n-1}^{(m)})\Delta t/m) \tag{4.4}$$

becomes increasingly accurate (Kloeden and Platen, 1999), at the expense of greater computational effort. Thus, the Euler approximation of resolution m (Kou et al., 2012) to the likelihood function is

$$\hat{\mathcal{L}}_m(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) = \int \left[\pi(\boldsymbol{X}_0^{(m)} \mid \boldsymbol{\theta}) \cdot \prod_{n=0}^{N} g(\boldsymbol{Y}_n \mid \boldsymbol{X}_{nm}^{(m)}, \boldsymbol{\theta}) \cdot \prod_{n=1}^{Nm} \varphi(\boldsymbol{X}_n^{(m)} \mid \boldsymbol{X}_{n-1}^{(m)}, \boldsymbol{\theta}) \right] \mathrm{d}\boldsymbol{X}_{0:Nm}^{(m)},$$

where $\phi(\boldsymbol{X}_n^{(m)} \mid \boldsymbol{X}_{n-1}^{(m)}, \boldsymbol{\theta})$ is the PDF of the normal distribution in (4.4), and we have $\hat{\mathcal{L}}_m(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) \to \mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N})$ as $m \to \infty$ (Pedersen, 1995b, Theorem 4).

4.2.1 State-Space Model

The above SDEs can be treated under a general model framework known as state-space models (e.g., Murphy, 2012, Chapter 18). A general state-space model with static parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d$ consists of a hidden state process $\{\boldsymbol{S}_n\}_{n \geq 0}$ and an observation process $\{\boldsymbol{Y}_n\}_{n \geq 0}$, which can be illustrated by Figure 4.1.

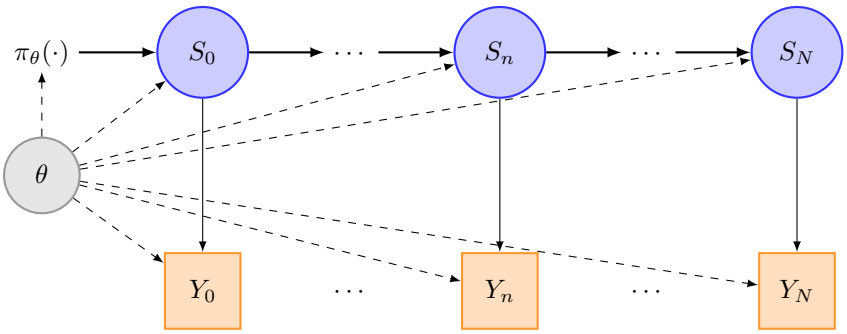

Figure 4.1: Illustration of the state-space model.

The hidden state process is characterized by its initial state density $\boldsymbol{S}_0 \sim \pi_{\boldsymbol{\theta}}(\cdot)$ and state transition density is assumed to be Markovian

$$p(\boldsymbol{S}_{n+1} \mid \boldsymbol{S}_{0:n}, \boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(\boldsymbol{S}_{n+1} \mid \boldsymbol{S}_n), \quad n \geq 0$$

where $\boldsymbol{S}_{0:n} = \boldsymbol{S}_0, \ldots, \boldsymbol{S}_n$. The state process $\{\boldsymbol{S}_n\}_{n\geq 0}$ is observed, not directly, but through another process $\{\boldsymbol{Y}_n\}_{n\geq 0}$. The observation $\boldsymbol{Y}_n$ is assumed to be conditionally independent given $\boldsymbol{S}_n$ and its marginal probability density is given by

$$p(\boldsymbol{Y}_n \mid \boldsymbol{Y}_{1:n-1}, \boldsymbol{S}_{0:n}, \boldsymbol{\theta}) = p(\boldsymbol{Y}_n \mid \boldsymbol{S}_n, \boldsymbol{\theta}) = g_{\boldsymbol{\theta}}(\boldsymbol{Y}_n \mid \boldsymbol{S}_n), \quad n \geq 1.$$

When $\boldsymbol{\theta}$ is known and the current time is $T = N\Delta t$, Bayesian inference relies on the posterior density $p_{\boldsymbol{\theta}}(\boldsymbol{S}_{0:n} \mid \boldsymbol{Y}_{0:n}) \propto p_{\boldsymbol{\theta}}(\boldsymbol{S}_{0:n}, \boldsymbol{Y}_{0:n})$ where the joint likelihood can be written as

$$p(\boldsymbol{S}_{0:N}, \boldsymbol{Y}_{0:N} \mid \boldsymbol{\theta}) = \pi(\boldsymbol{S}_0 \mid \boldsymbol{\theta}) \prod_{n=1}^{N} f_{\boldsymbol{\theta}}(\boldsymbol{S}_n \mid \boldsymbol{S}_{0:n-1}) \prod_{n=0}^{N} g_{\boldsymbol{\theta}}(\boldsymbol{Y}_n \mid \boldsymbol{S}_n)$$

If $\boldsymbol{\theta}$ is unknown, we ascribe a prior density $\pi(\boldsymbol{\theta})$ to it. Then the joint density is

$$p(\boldsymbol{\theta}, \boldsymbol{S}_{0:N}, \boldsymbol{Y}_{0:N}) \propto p(\boldsymbol{S}_{0:N}, \boldsymbol{Y}_{0:N} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}).$$

Alternatively, under a hierarchical sampling framework, the SSM can be expressed as

$$\begin{aligned} \boldsymbol{S}_0 &\sim \pi(\boldsymbol{S}_0 \mid \boldsymbol{\theta}) \\ \boldsymbol{S}_n &\sim f_{\boldsymbol{\theta}}(\boldsymbol{S}_n \mid \boldsymbol{S}_{n-1}) \\ \boldsymbol{Y}_n &\sim g_{\boldsymbol{\theta}}(\boldsymbol{S}_n) \end{aligned}$$

In our SDE problem, at time n, the observation is still $\boldsymbol{Y}_n$, while the hidden state is

$$\boldsymbol{S}_n = \boldsymbol{X}^{(m)}_{(n-1)m+1:nm}$$

The prior on $\boldsymbol{S}_0 = (\boldsymbol{S}^{(m)}_{-m+1}, \ldots, \boldsymbol{S}^{(m)}_0)$, is given by

$$\begin{aligned} \boldsymbol{S}^{(m)}_0 &\sim \pi(\boldsymbol{S}^{(m)}_0 \mid \boldsymbol{\theta}), \\ \boldsymbol{S}^{(m)}_n &= 0, \qquad -m+1 \leq n < 0. \end{aligned}$$

In other words, we use the SSM prior for $\boldsymbol{X}^{(m)}_0$, and set the dummy variables $\boldsymbol{X}^{(m)}_n = 0$ for $n = -m+1, \ldots, -1$.

The task of Bayesian inference is to condition this joint distribution on some particular data observations, $\boldsymbol{Y}_{0:N}$ to obtain the posterior distribution $p(\boldsymbol{S}_{0:N}, \boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N})$, which can be written as

$$p(\boldsymbol{S}_{0:N}, \boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) = p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) \cdot p(\boldsymbol{S}_{0:N} \mid \boldsymbol{\theta}, \boldsymbol{Y}_{0:N}).$$

To obtain the first factor $p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N})$ is parameter estimation, while to get the second factor $p(\boldsymbol{S}_{0:N} \mid \boldsymbol{\theta}, \boldsymbol{Y}_{0:N})$ conditioning on some $\boldsymbol{\theta}$ drawn from the first is state estimation or filtering problem (Murray, 2013). For non-linear non-Gaussian SSMs, $p(\boldsymbol{\theta}, \boldsymbol{S}_{0:N}, \boldsymbol{Y}_{0:N})$ does not have closed form expressions, rendering analytic inference difficult in practice. Therefore, it is necessary for us to resort to approximations for which particle filters are designed.

4.2.2 Particle Filters

The particle filter is used to construct an efficient Monte Carlo approximation of the intractable likelihood function (4.3) by sampling Bayesian posterior distributions recursively, which is a particularly effective method to deal with the SSM.

There are numerous variants of particle filters, a standard algorithm of particle filter is given by Algorithm 4.1. For a complete description under the coherent framework of sequential Monte Carlo (Liu, 2004; Robert and Casella, 2004), please refer to Doucet et al. (2001, 2009); Johansen (2009).

Suppose we set J particles for our problem, then we initialize the particle filter by sampling J random variables from our prior distribution $\pi(\boldsymbol{S}_0 \mid \boldsymbol{\theta})$ and assigning each an equal weight $1/J$. These random variables are referred to as *particles*. The particle filter

algorithm runs sequentially as new observations are coming in through three main steps: *propagation*, *weighting* and *resampling*.

(i) Propagation: Suppose at the end of step $n-1$ we have J particles $\bar{\boldsymbol{S}}_{n-1}^1, \ldots, \bar{\boldsymbol{S}}_{n-1}^J$ (the "bar" notation will be justified momentarily). Each particle is advanced to the next observation time n according to a proposal distribution $\boldsymbol{S}_n^j \sim q(\boldsymbol{S}_n \mid \bar{\boldsymbol{S}}_{n-1}^j, \boldsymbol{Y}_n, \boldsymbol{\theta})$.

(2) Weighting: Each particle is weighted according to the likelihood of the new observation, i.e., the incremental weight for particle j is calculated as

$$w_n^j = \frac{g_{\boldsymbol{\theta}}(\boldsymbol{Y}_n \mid \boldsymbol{S}_n^j) f_{\boldsymbol{\theta}}(\boldsymbol{S}_n^j \mid \bar{\boldsymbol{S}}_{n-1}^j)}{q(\boldsymbol{S}_n^j \mid \boldsymbol{Y}_n, \bar{\boldsymbol{S}}_{n-1}^j, \boldsymbol{\theta})}. \tag{4.5}$$

There are many ways to define the weight by using different proposal distributions, as we shall see in Section 4.2.3 and Section 4.3.5.

(ii) Resampling: We sample the new particles $\boldsymbol{S}_n^1, \ldots, \boldsymbol{S}_n^J$ with replacement with probability proportional to w_n^j to obtain the at the end of step n the set of particles $\bar{\boldsymbol{S}}_n^1, \ldots \bar{\boldsymbol{S}}_n^J$. Simply discarding particles does not work. Instead, we usually follow some well-designed resampling mechanisms, one of which is to draw the number of offsprings of the current particle (known as the parent particle) from a multinomial distribution such that parent particles with larger weights tend to produce more equally-weighted new offsprings to move forward to the next observation time/generation whereas the one with small weights tend to be eliminated. This strategy is designed to overcome the well-known particle degeneracy problem that the particle weights $\{w_n^j\}_{j=1}^J$ tend to be more and more skewed and concentrated such that eventually only one particle has non-zero weight (Doucet et al., 2009).

In fact, in the current literature, there are two equivalent ways of describing the multinomial resampling mechanism. One is to sample a number of offsprings for each parent particle from a multinomial distribution according to their particle weights. Let $(O_n^1, \ldots, O_n^J)$ be the offspring vector at time n, it is drawn from a multinomial distribution, i.e.

$$(O_n^1, \ldots, O_n^J) \sim \text{Multinomial}(J; \bar{w}_n^1, \ldots, \bar{w}_n^J)$$
$$\mathbf{P}(O_n^1 = x_1, \ldots, O_n^J = x_J) = \frac{J!}{x_1! \ldots x_J!} \prod_{j=1}^J (\bar{w}_n^j)^{x_j}, \quad \sum_{j=1}^J x_j = J.$$

Algorithm 4.1 Particle filter.

At time $n = 0$
for each particle j in $1 : J$ **do**
    Initialize particle $\boldsymbol{S}_0^j \sim q_0(\boldsymbol{S}_0 \mid \boldsymbol{\theta})$.
    Initialize weight $w_0^j \leftarrow \pi(\boldsymbol{S}_0^j \mid \boldsymbol{\theta}) / q_0(\boldsymbol{S}_0^j \mid \boldsymbol{\theta})$.
end for

for $n = 1, \ldots, N$ **do**
    Normalize weights $\bar{w}_{n-1}^j = w_{n-1}^j / \sum_{k=1}^J w_{n-1}^k$.
    for each particle j in $1 : J$ **do**
        Sample ancestral index $A_n^j \sim \text{Categorical} \left\{ \bar{w}_{n-1}^k \right\}_{k=1}^J$.
        Propagate particle $\boldsymbol{S}_n^j \sim q(\boldsymbol{S}_n \mid \bar{\boldsymbol{S}}_{n-1}^{A_n^j}, \boldsymbol{Y}_n, \boldsymbol{\theta})$.
        Weight $w_n^j \leftarrow g_{\boldsymbol{\theta}}(\boldsymbol{Y}_n \mid \boldsymbol{S}_n^j) f_{\boldsymbol{\theta}}(\boldsymbol{S}_n^j \mid \bar{\boldsymbol{S}}_{n-1}^{A_n^j}) / q(\boldsymbol{S}_n^j \mid \bar{\boldsymbol{S}}_{n-1}^{A_n^j}, \boldsymbol{Y}_n, \boldsymbol{\theta})$.
    end for
end for
Return marginal likelihood: $\hat{\mathcal{L}}(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) \leftarrow \prod_{n=0}^N \left(\frac{1}{J} \sum_{j=1}^J w_n^j \right)$.

Another way is to sample the ancestor index A_n^j of particle j at step n as shown in Algorithm 4.1. Let $(A_n^1, \ldots, A_n^J)$ be the vector of ancestor indices, then it follows the categorical distribution (with J categories) defined as follows

$$
\begin{aligned}
&A_n^j \overset{\text{iid}}{\sim} \text{Categorical}(\bar{w}_{n-1}^1, \ldots, \bar{w}_{n-1}^J), \quad j = 1, \ldots, J \\
&\mathbf{P}(A_n^j = k) = \bar{w}_{n-1}^k, \quad k \in \{1, \ldots, J\}
\end{aligned}
$$

For more different resampling strategies, please refer to Doucet et al. (2009) and also Hol et al. (2006); Kuptametee and Aunsri (2022).

In the Bayesian context, Algorithm 4.1 can be used to estimate the posterior parameter distribution $p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N})$ via the particle filter approximation

$$
\hat{p}(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) \propto \hat{\mathcal{L}}(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N}) \cdot \pi(\boldsymbol{\theta}). \tag{4.6}
$$

The (approximate) posterior (4.6) can be sampled from using the family of particle marginal Metropolis-Hastings (PMMH) algorithms proposed in Andrieu et al. (2010). With minor modifications, Algorithm 4.1 can also be used to return a draw from $p(\boldsymbol{S}_{0:N} \mid \boldsymbol{\theta}, \boldsymbol{Y}_{0:N})$, which could in turn be used in a Gibbs sampler which alternates between draws from $p(\boldsymbol{S}_{0:N} \mid \boldsymbol{\theta}, \boldsymbol{Y}_{0:N})$ and $p(\boldsymbol{\theta} \mid \boldsymbol{S}_{0:N}, \boldsymbol{Y}_{0:N})$. However, for hypoelliptic diffusions this so-called

particle Gibbs sampler (Andrieu et al., 2010) must be implemented with extreme care to approximate the posterior distribution correctly (Pokern et al., 2009). We shall return to this point in Section 4.3.4.

4.2.3 SDE Bridge Proposal

We now return to the SDE inference problem described at the beginning of section 4.2, and assume henceforth that we have a normal measurement model

$$\boldsymbol{Y}_n \overset{\text{ind}}{\sim} \mathcal{N}\left(\boldsymbol{A}\boldsymbol{S}_n, \boldsymbol{\Omega}\right) \tag{4.7}$$

In this setting, inference for noise-free SDEs can be achieved by setting $\boldsymbol{\Omega} = \mathbf{0}$ or for numerical stability $\boldsymbol{\Omega} = \sigma_\epsilon^2 \mathcal{I}$ for some small constant $\sigma_\epsilon > 0$.

The particle filter described by Algorithm 4.1 requires us to specify a proposal distribution in (4.5). If we set $q_{\boldsymbol{\theta}}(\boldsymbol{S}_n^j \mid \boldsymbol{Y}_n, \boldsymbol{S}_{n-1}^{A_n^j}) = f_{\boldsymbol{\theta}}(\boldsymbol{S}_n^j \mid \boldsymbol{S}_{n-1}^{A_n^j})$ which is the forward transition density of the hidden states, then $w_n^j = g_{\boldsymbol{\theta}}(\boldsymbol{Y}_n \mid \boldsymbol{S}_n^j)$, which gives us the bootstrap particle filter.

The bootstrap filter is widely used for many applications and easy to implement, but it works poorly when there is little noise in the measurement model $\boldsymbol{Y}_n \sim g_{\boldsymbol{\theta}}(\boldsymbol{Y}_n \mid \boldsymbol{S}_n)$ since the incremental weights would then be calculated according to a deterministic model. For this reason, a proposal tailored to SDEs has been developed by Durham and Gallant (2002); Chib et al. (2004); Golightly and Wilkinson (2008).

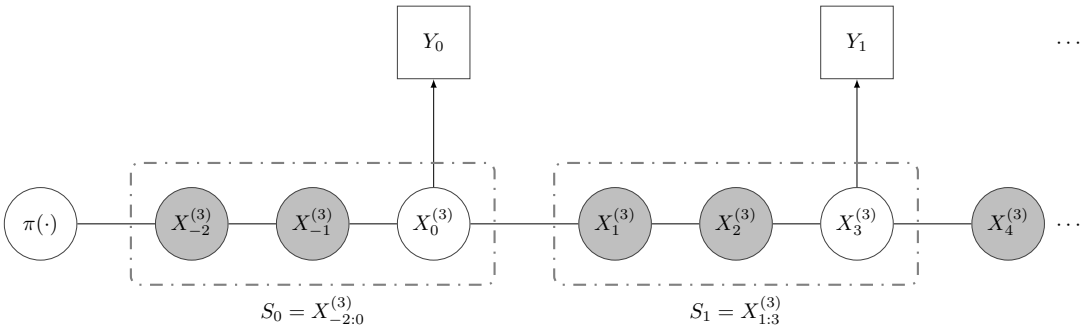

Figure 4.2: Illustration of the bridge proposal with resolution $m = 3$.

The idea of the bridge proposal goes as follows. When we are at time $t_n = (n-1)\Delta t$ for $n \geq 1$ and want to move forward, we need a proposal distribution to sample $\boldsymbol{X}_{(n-1)m+1}^{(m)}$

which is one of the intermediate particles unobserved. Given the information obtained up to time $t_{n-1} = (n-1)\Delta t$ and the fact that when $t = n\Delta t$, we must have $\boldsymbol{X}(t) = X_{nm}^{(m)}$ and $\boldsymbol{Y}(t) = \boldsymbol{Y}_n$, we would like to find the conditional distribution of $\boldsymbol{X}_{(n-1)m+1}^{(m)}$ given $\boldsymbol{X}_{0:(n-1)m}^{(m)}$ (and the prior initialized $X_{-m+1}^{(m)}, \ldots, X_{-1}^{(m)}$) with the required ending point $\left(\boldsymbol{Y}_n, \boldsymbol{X}_{nm}^{(m)}\right)$ at time $n\Delta t$, which is the idea behind our bridge proposal. Then we can recursively sample $\boldsymbol{X}_{(n-1)m+k}^{(m)}$ for $0 < k \leq m$. In this way, $\boldsymbol{Y}_n$ is the observation that corresponds to state variables $\boldsymbol{S}_n = \boldsymbol{X}_{(n-1)m+1:nm}^{(m)}$ under the framework of SSMs. Figure 4.2 illustrates this idea with $m = 3$ as an example. As long as our SDEs are driven by Brownian motions, the random noise within the small time interval $\Delta t/m$ follows a normal distribution. Therefore, the diffusion bridge is analytically tractable. For normal distributions, we have the following result (proved in Appendix C.2).

Lemma 4.2.1. *Suppose that we have*

$$\begin{aligned} \boldsymbol{W} &\sim \mathcal{N}(\boldsymbol{\mu}_W, \boldsymbol{\Sigma}_W) \\ \boldsymbol{X} \mid \boldsymbol{W} &\sim \mathcal{N}(\boldsymbol{W} + \boldsymbol{\mu}_{X|W}, \boldsymbol{\Sigma}_{X|W}) \\ \boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{W} &\sim \mathcal{N}(\boldsymbol{A}\boldsymbol{X}, \boldsymbol{\Omega}). \end{aligned}$$

then

$$\begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{X} \\ \boldsymbol{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_W \\ \boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W} \\ \boldsymbol{\mu}_Y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_W & \boldsymbol{\Sigma}_W & \boldsymbol{\Sigma}_W \boldsymbol{A}' \\ \boldsymbol{\Sigma}_W & \boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W} & (\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W})\boldsymbol{A}' \\ \boldsymbol{A}\boldsymbol{\Sigma}_W & \boldsymbol{A}(\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W}) & \boldsymbol{\Sigma}_Y \end{bmatrix} \right), \tag{4.8}$$

where $\boldsymbol{\mu}_Y = \boldsymbol{A}[\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W}]$ *and* $\boldsymbol{\Sigma}_Y = \boldsymbol{A}(\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W})\boldsymbol{A}' + \boldsymbol{\Omega}$*, such that*

$$\boldsymbol{W} \mid \boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{\mu}_W + \boldsymbol{\Sigma}_W \boldsymbol{A}' \boldsymbol{\Sigma}_Y^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y),\ \boldsymbol{\Sigma}_W - \boldsymbol{\Sigma}_W \boldsymbol{A}' \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{A} \boldsymbol{\Sigma}_W\right). \tag{4.9}$$

The bridge proposal for the state variable $\boldsymbol{S}_t$ then proceeds recursively as follows:

(i) Without loss of generality, assume we are at time $t_n = \Delta t$, i.e. $n = 1$, and that for fixed $0 < k \leq m$ the proposal values until $\boldsymbol{X}_{(n-1)m+k}^{(m)} = \boldsymbol{X}_k^{(m)}$ are given. Now suppose we wish to specify the proposal distribution for $\boldsymbol{X}_{k+1}^{(m)}$ given our observation $\boldsymbol{Y}(t_n) = \boldsymbol{Y}_1$.

(ii) Approximate $\boldsymbol{X}(t)$ as a Brownian motion with constant drift for $t \in (\frac{k}{m}\Delta t, \Delta t)$:

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{\Lambda}_k \mathrm{d}t + \boldsymbol{\Sigma}_k^{1/2} \mathrm{d}\boldsymbol{B}(t),$$

where $\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_{\boldsymbol{\theta}}(\boldsymbol{X}_k^{(m)})$ and $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\boldsymbol{X}_k^{(m)})$. This means that for any $\frac{k}{m}\Delta t \leq s \leq s+u \leq \Delta t$ we have

$$\boldsymbol{X}(s+u) \mid \boldsymbol{X}(s) \sim \mathcal{N}(\boldsymbol{X}(s) + u\boldsymbol{\Lambda}_k, u\boldsymbol{\Sigma}_k),$$

To obtain the proposal for $\boldsymbol{X}_{k+1}^{(m)}$, we apply the formula to $(\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{X}_{k+1}^{(m)}, \boldsymbol{X}_m^{(m)}, \boldsymbol{Y}_1)$, for which we have

$$\begin{aligned}
\boldsymbol{X}_{k+1}^{(m)} &\sim \mathcal{N}\left(\boldsymbol{X}_n^{(m)} + \frac{\Delta t}{m}\boldsymbol{\Lambda}_k,\ \frac{\Delta t}{m}\boldsymbol{\Sigma}_k\right) \\
\boldsymbol{X}_m^{(m)} \mid \boldsymbol{X}_{k+1}^{(m)} &\sim \mathcal{N}\left(\boldsymbol{X}_{k+1}^{(m)} + (m-k-1)\frac{\Delta t}{m}\boldsymbol{\Lambda}_k,\ (m-k-1)\frac{\Delta t}{m}\boldsymbol{\Sigma}_k\right) \\
\boldsymbol{Y}_1 \mid \boldsymbol{X}_m^{(m)}, \boldsymbol{X}_{k+1}^{(m)} &\sim \mathcal{N}\left(\boldsymbol{A}\boldsymbol{X}_m^{(m)},\ \boldsymbol{\Omega}\right),
\end{aligned}$$

such that in the formula (4.9) we have

$$\begin{aligned}
\boldsymbol{\mu}_W &= \boldsymbol{X}_k^{(m)} + \boldsymbol{\Lambda}_k\frac{\Delta t}{m}, & \boldsymbol{\Sigma}_W &= \frac{\Delta t}{m}\boldsymbol{\Sigma}_k, \\
\boldsymbol{\mu}_Y &= \boldsymbol{A}\left[\boldsymbol{X}_k^{(m)} + (m-k-1)\frac{\Delta t}{m}\boldsymbol{\Lambda}_k\right], & \boldsymbol{\Sigma}_Y &= (m-k-1)\frac{\Delta t}{m}\boldsymbol{A}\boldsymbol{\Sigma}_k\boldsymbol{A}' + \boldsymbol{\Omega}.
\end{aligned}$$

4.3 Parameter Estimation of the Quasi-Markovian GLE

Now we are ready to focus on the parameter inference for the qmGLE.

4.3.1 Reparameterization

Clearly there is an identifiability issue if we try to estimate the parameter $\boldsymbol{\lambda}, \boldsymbol{A}$ (and implicitly $\boldsymbol{\Sigma}$) given by Theorem 4.1.1. The identifiability issue motivates us to assume that $\boldsymbol{A}$ is symmetric such that we can diagonalize $\boldsymbol{A} = \mathcal{P}\mathcal{D}\mathcal{P}^\dagger$ where $\mathcal{P} \in \mathbb{R}^{d\times d}$ is an orthogonal matrix, i.e. $\mathcal{P}\mathcal{P}^\dagger = \mathcal{P}^\dagger\mathcal{P} = \boldsymbol{I}$ and $\mathcal{D} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with diagonal components $\boldsymbol{\rho} = \{\rho_1, \ldots, \rho_d\}$. We notice that

$$K(t) = \langle e^{-\boldsymbol{A}t}\boldsymbol{\lambda}, \boldsymbol{\lambda}\rangle = \langle \mathcal{P}e^{-\mathcal{D}t}\mathcal{P}^\dagger\boldsymbol{\lambda}, \boldsymbol{\lambda}\rangle = \sum_{i=1}^{d}\lambda_i^2 e^{-\rho_i t}$$

where we used the orthogonality of $\mathcal{P}$. Since $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\dagger = 2\beta^{-1}\boldsymbol{A}$, and the eigenvalues ρ_i, $i = 1, \ldots, d$ are all positive (since $\boldsymbol{A}$ is assumed to be positive definite), we can further require $\boldsymbol{\Sigma} = \sqrt{2\boldsymbol{A}}/\beta = \sqrt{2}/\beta\mathcal{P}\mathcal{D}^{\frac{1}{2}}\mathcal{P}^\dagger$. With these restrictions, we can eliminate the identifiability issue, then the original SDEs become

$$\begin{aligned}
\mathrm{d}X(t) &= V(t)\mathrm{d}t\,,\\
\mathrm{d}V(t) &= -U'_{\boldsymbol{\phi}}(X(t))\mathrm{d}t + \langle\boldsymbol{\lambda}, \boldsymbol{z}(t)\rangle\mathrm{d}t\,,\\
\mathrm{d}\boldsymbol{z}(t) &= -\left(V(t)\boldsymbol{\lambda} + \mathcal{P}\mathcal{D}\mathcal{P}^\dagger\boldsymbol{z}(t)\right)\mathrm{d}t + \sqrt{2}/\beta\mathcal{P}\mathcal{D}^{\frac{1}{2}}\mathcal{P}^\dagger\mathrm{d}\boldsymbol{B}(t)\,,
\end{aligned}$$

Now apply an orthogonal transformation $\mathcal{P}^\dagger$ onto the third equation, i.e. let

$$\boldsymbol{Z}(t) = \mathcal{P}^\dagger\boldsymbol{z}(t)$$

and accordingly,

$$\begin{aligned}
\boldsymbol{\mu} &= \mathcal{P}^\dagger\boldsymbol{\lambda},\\
\tilde{\boldsymbol{B}}(t) &= \mathcal{P}^\dagger\boldsymbol{B}(t),
\end{aligned}$$

we have

$$\mathrm{d}\boldsymbol{Z}(t) = -\left(V(t)\boldsymbol{\mu} + \mathcal{D}\boldsymbol{Z}(t)\right)\mathrm{d}t + \mathcal{D}^{\frac{1}{2}}\mathrm{d}\tilde{\boldsymbol{B}}(t).$$

Since $\langle\boldsymbol{\lambda}, \boldsymbol{z}(t)\rangle = \sum_{i=1}^d \lambda_i z_i(t) = \langle\mathcal{P}^\dagger\boldsymbol{\lambda}, \mathcal{P}^\dagger\boldsymbol{Z}(t)\rangle = \sum_{i=1}^d \mu_i Z_i(t)$, and (multidimensional) Brownian motion is invariant under orthogonal transformation, i.e. $\tilde{\boldsymbol{B}}(t)$ is again a d-dimensional standard Brownian motion (each component is a one-dimensional Brownian motion and different components are independent), the original SDE can be expressed as

$$\begin{aligned}
\mathrm{d}X(t) &= V(t)\mathrm{d}t\,,\\
\mathrm{d}V(t) &= -U'_{\boldsymbol{\phi}}(X(t))\mathrm{d}t + \sum_{i=1}^d \mu_i Z_i(t)\mathrm{d}t\,,\\
\mathrm{d}Z_i(t) &= -\left(V(t)\mu_i + \rho_i Z_i(t)\right)\mathrm{d}t + \sqrt{2\rho_i}/\beta\mathrm{d}\tilde{B}_i(t)\,, \quad i = 1, \ldots, d.
\end{aligned}$$

Therefore, if we can estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\rho}, \boldsymbol{\phi})$ from the above SDE, we can recover our original GLE and the memory kernel $K(t) = \langle e^{-\mathcal{D}t}\mathcal{P}^\dagger\boldsymbol{\lambda}, \mathcal{P}^\dagger\boldsymbol{\lambda}\rangle = \sum_{i=1}^d \mu_i^2 e^{-\rho_i t}$. Note that d is predetermined in experiments.

4.3.2 Degenerate Euler Scheme

In order to perform simulation, a natural approach is to apply an Euler–Maruyama scheme to discretize the SDE. We wish to simulate date at time intervals of Δt by sampling at an internal frequency $\Delta t_m = \Delta t/m$ for $m > 0$ using a simple procedure to be described below, and keeping only every m-th observation. In other words, let $X_n^{(m)}$, $n = 0, 1, \ldots$ be the value of $X(t)$ at time $t = n\Delta t_m$ such that $X_n = X_{mn}^{(m)}$. Then the Euler scheme for the qmGLE is

$$\begin{aligned}
X_{n+1}^{(m)} - X_n^{(m)} &= V_n^{(m)} \Delta t/m \,, \\
V_{n+1}^{(m)} - V_n^{(m)} &= -U'_{\boldsymbol{\phi}}(X_n^{(m)})\Delta t/m + \sum_{i=1}^{d} \mu_i Z_{i,n}^{(m)} \Delta t/m \,, \\
Z_{i,n+1}^{(m)} - Z_{i,n}^{(m)} &= -\left(V_n^{(m)}\mu_i + \rho_i Z_{i,n}^{(m)}\right)\Delta t/m + \sqrt{2\rho_i}/\beta \Delta \tilde{B}_{i,n}^{(m)}
\end{aligned}$$

where $\Delta \tilde{B}_{i,n}^{m} = \tilde{B}_{i,n+1}^{(m)} - \tilde{B}_{i,n}^{(m)} \overset{iid}{\sim} \mathcal{N}(0, \Delta t/m)$, $i = 1, \ldots, d$.

The Euler discretized qmGLE can be treated as an SSM. At time $t_n = n\Delta t$, the hidden state vector is

$$\boldsymbol{S}_n = (X_{m(n-1)+1:mn}^{(m)}, V_{m(n-1)+1:mn}^{(m)}, \boldsymbol{Z}_{m(n-1)+1:mn}^{(m)})$$

where $\boldsymbol{Z}_t^{(m)} = (Z_{1,t}^{(m)}, \ldots, Z_{d,t}^{(m)})$. It is clear that $\boldsymbol{S}_n$ only depends on the last state $\boldsymbol{S}_{n-1}$ and thus is Markovian. At each observation time, there are $d+2$ number of state variables whereas the randomness only comes from d independent Brownian motions. Denote $\{Y_n = X_{mn}^{(m)}\}$ as the observation sequence, then the measurement model (4.7) becomes

$$Y_n \sim \mathcal{N}\left(X_{mn}^{(m)}, \sigma_\epsilon^2\right) \tag{4.10}$$

which is just a special case of Lemma 4.2.1 with $\boldsymbol{A} = \left[\begin{smallmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{smallmatrix}\right]$ and $\boldsymbol{\Omega} = \sigma_\epsilon^2 \boldsymbol{A}$.

Then we can see that it would be meaningless to directly apply the bridge proposal under the degenerate Euler scheme. Note that the system of SDEs becomes

$$\boldsymbol{S}_{nm+1} - \boldsymbol{S}_{nm} = \begin{pmatrix} V_{mn}^{(m)} \\ -U'_{\boldsymbol{\phi}}(X_{mn}^{(m)}) + \boldsymbol{\mu}' \boldsymbol{Z}_{mn}^{(m)} \\ -V_{mn}^{(m)} \boldsymbol{\mu} - \boldsymbol{\rho} \boldsymbol{Z}_{mn}^{(m)} \end{pmatrix} \frac{\Delta t}{m} + \begin{pmatrix} 1 & 0 & \mathbf{0} \\ 0 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\sqrt{2\boldsymbol{\rho}}}{\beta} \mathbf{1}_{d\times d} \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \tilde{\boldsymbol{B}}_{nm}^{(m)} \end{pmatrix}.$$

If we blindly apply the bridge proposal formula (4.9), the result degenerates to (4.10) with $\sigma = 0$, which is equivalent to the bootstrap particle filter. The intuition behind is indeed

simple in that the d-dimensional Brownian motion (randomness) is essentially independent of the "deterministic" variables of dimension 2, resulting in failure of constructing a meaningful diffusion bridge. However, there is a deep root behind the degeneracy problem, as we shall discuss next.

4.3.3 Hypoelliptic Diffusions

A deeper reason behind the failure of the Euler scheme lies in the fact that qmGLE belongs to a class of hypoelliptic diffusion processes where a smooth density solution exists (Hörmander, 1961), but the diffusion matrix is not of full rank (Pokern et al., 2009; Ditlevsen and Samson, 2019). In such cases, the ergodicity of the (degenerate) Euler discretization scheme may fail since the Lyapunov condition given in Mattingly et al. (2002) is not satisfied.

Moreover, for SDEs with partial observations, under the degenerate Euler scheme, the unobserved components (e.g. $V(t)$) are directly computed via the time differentiation from the observed coordinates (e.g. $X(t)$) in order to be compatible with the model, which leads to estimating missing paths with incorrect quadratic variations.

From an abstract mathematical point of view, the generator of the qmGLE solution is actually a degenerate second-order hypoelliptic differential operator (Ottobre et al., 2012) which can be written in Hömander's "sum of squares" form (Hörmander, 1967). This hypoelliptic structure leads to the failure of estimation based on the degenerate Euler discretization scheme as we previously pointed out. Without modifying the Euler scheme, even the bridge proposal tailored to SDEs would certainly fail. We are then motivated to propose a non-degenerate discretization scheme by propagating the noise from the "rough" components $\boldsymbol{Z}(t)$ to the "smooth" components $X(t), V(t)$, which we shall provide in the next section.

4.3.4 Modified Discretization Scheme

For our convenience of the following exposition, let us restate the qmGLE of interest for parameter inference,

$$\begin{aligned} \mathrm{d}X(t) &= V(t)\mathrm{d}t, \\ \mathrm{d}V(t) &= -U'_{\boldsymbol{\phi}}(X(t))\mathrm{d}t + \sum_{i=1}^{d} \mu_i Z_i(t)\mathrm{d}t, \\ \mathrm{d}Z_i(t) &= -\left(\mu_i V(t) + \rho_i Z_i(t)\right)\mathrm{d}t + \sigma_i \mathrm{d}\tilde{B}_i(t) \end{aligned} \tag{4.11}$$

where $\sigma_i = \sqrt{2\rho_i}/\beta$. Moreover, let

$$\boldsymbol{Z}(t) = (Z_1(t), \ldots, Z_d(t))$$

and

$$\boldsymbol{W}(t) = (X(t), V(t), \boldsymbol{Z}(t)).$$

To propagate the randomness to the full qmGLE model, we design a modified Euler discretization scheme by using the q-times integrated Wiener process (Schober et al., 2019) with $q = 2$.

When $t_n = n\Delta t$, let $X_n = X(t_n)$, $V_n = V(t_n)$, $\boldsymbol{Z}_n = \boldsymbol{Z}(t_n)$, $\Delta X_n(s) = X(t_n+s) - X(t_n)$ for $s \in \mathbb{R}$, and similarly for $\Delta V_n(s)$, $\Delta\tilde{B}_{i,n}(s)$, $\Delta\boldsymbol{W}_n(s)$. Then for small $s \geq 0$, the SDE (4.11) may be approximated on the time interval $t \in (t_n, t_n + s)$ by

$$\begin{aligned} \Delta X_n(s) &= \int_0^s \left(V_n + \Delta V_n(h)\right)\mathrm{d}h \\ \Delta V_n(s) &= -U'_{\boldsymbol{\phi}}(X_n)s + \sum_{i=1}^{d} \mu_i \int_0^s \left(Z_{i,n} + \Delta Z_{i,n}(h)\right)\mathrm{d}h \\ \Delta Z_{i,n}(s) &= -\left(\mu_i V_n + \rho_i Z_{i,n}\right)s + \sigma_i \Delta\tilde{B}_{i,n}(s), \quad i = 1, \ldots, d. \end{aligned}$$

Let $U_{i,n}^{(j)}(s)$ for $j \in \{0,1,2\}$ be defined as

$$\begin{aligned} U_{i,n}^{(2)}(s) &= \sigma_i \Delta \tilde{B}_{i,n}(s), \\ U_{i,n}^{(1)}(s) &= \int_0^s U_{i,n}^{(2)}(h)\mathrm{d}h, \\ U_{i,n}^{(0)}(s) &= \int_0^s U_{i,n}^{(1)}(h)\mathrm{d}h \end{aligned}$$

such that $U_{i,n}^{(j)}(s)$ is the j-th order derivative of $U_{i,n}^{(0)}(s)$ at time s.

Let $\boldsymbol{U}_n^{(j)}(t) = (U_{1,n}^{(j)}(t), \ldots, U_{d,n}^{(j)}(t))$ and consider the $3d$-dimensional process

$$\boldsymbol{U}_n(t) = (\boldsymbol{U}_n^{(0)}(t), \boldsymbol{U}_n^{(1)}(t), \boldsymbol{U}_n^{(2)}(t)).$$

Then upon letting $A_{i,n} = \mu_i V_n + \rho_i Z_{i,n}$ and $\boldsymbol{A}_n = (A_{1,n}, \ldots, A_{d,n})$, we have

$$\Delta \boldsymbol{W}_n(s) = \boldsymbol{\lambda}(\boldsymbol{W}_n, s) + \tilde{\boldsymbol{Z}}_n(s),$$

where each of the three terms above is a $(d+2)$-dimensional process with

$$\boldsymbol{\lambda}(\boldsymbol{W}_n, s) = \begin{bmatrix} V_n - \frac{1}{2}U_{\boldsymbol{\phi}}'(X_n)s^2 + \boldsymbol{\mu}'\left(\frac{1}{2}\boldsymbol{Z}_n s^2 - \frac{1}{6}\boldsymbol{A}_n s^3\right) \\ -U_{\boldsymbol{\phi}}'(X_n)s + \boldsymbol{\mu}'\left(\boldsymbol{Z}_n s - \frac{1}{2}\boldsymbol{A}_n s^2\right) \\ -\boldsymbol{A}_n s \end{bmatrix}, \quad \tilde{\boldsymbol{Z}}_n(s) = \begin{bmatrix} \boldsymbol{\mu}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mu}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_{d\times d} \end{bmatrix} \boldsymbol{U}_n(s).$$

Note that $\boldsymbol{U}_n(s)$ is a Markov process, with $\boldsymbol{U}_n(0) = \mathbf{0}$ and

$$\boldsymbol{U}_n(s+h) \mid \boldsymbol{U}_n(s) \sim \mathcal{N}\left\{(\boldsymbol{R}(h) \otimes \mathbf{1}_{d\times d})\,\boldsymbol{U}_n(s), \boldsymbol{\Sigma}(h) \otimes \mathrm{diag}(\boldsymbol{\sigma}^2)\right\},$$

where $\otimes$ is the Kronecker matrix product, and

$$\boldsymbol{R}(h) = \begin{bmatrix} 1 & h & \frac{1}{2}h^2 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}(h) = \begin{bmatrix} \frac{1}{20}h^5 & \frac{1}{8}h^4 & \frac{1}{6}h^3 \\ \frac{1}{8}h^4 & \frac{1}{3}h^3 & \frac{1}{2}h^2 \\ \frac{1}{6}h^3 & \frac{1}{2}h^2 & h \end{bmatrix},$$

Note $\boldsymbol{\Sigma}(h)$ can be derived using (Schober et al., 2019, Eq (10) for $q = 2$). In Appendix C.1, we also give a clear and direct derivation of this covariance matrix without referring to the aforementioned result (which is a more general one for our purposes).

It follows that $\tilde{\boldsymbol{Z}}_n(s)$ is also a Markov process with $\tilde{\boldsymbol{Z}}_n(0) = \mathbf{0}$ and

$$\tilde{\boldsymbol{Z}}_n(s+h) \mid \tilde{\boldsymbol{Z}}_n(s) \sim \mathcal{N}\left(\tilde{\boldsymbol{R}}(h)\tilde{\boldsymbol{Z}}_n(s), \tilde{\boldsymbol{\Sigma}}(h)\right),$$

where

$$\tilde{\boldsymbol{R}}(h) = \begin{bmatrix} 1 & h & \frac{1}{2}h^2\boldsymbol{\mu}' \\ 0 & 1 & h\boldsymbol{\mu}' \\ 0 & 0 & \mathbf{1}_{d\times d} \end{bmatrix}, \quad \tilde{\boldsymbol{\Sigma}}(h) = \begin{bmatrix} \frac{1}{20}\gamma h^5 & \frac{1}{8}\gamma h^4 & \frac{1}{6}h^3\boldsymbol{\mu}'\operatorname{diag}(\boldsymbol{\sigma}^2) \\ \frac{1}{8}\gamma h^4 & \frac{1}{3}\gamma h^3 & \frac{1}{2}h^2\boldsymbol{\mu}'\operatorname{diag}(\boldsymbol{\sigma}^2) \\ \frac{1}{6}h^3\operatorname{diag}(\boldsymbol{\sigma}^2)\boldsymbol{\mu} & \frac{1}{2}h^2\operatorname{diag}(\boldsymbol{\sigma}^2)\boldsymbol{\mu} & h\operatorname{diag}(\boldsymbol{\sigma}^2) \end{bmatrix},$$

and where $\gamma = \sum_{i=1}^{d} \mu_i^2\sigma_i^2$. Then we have the following algorithm to simulate $\boldsymbol{W}_0, \ldots, \boldsymbol{W}_N$,

(i) Fix the value of $\boldsymbol{W}_0$, or draw it from its stationary distribution.

(ii) Given $\boldsymbol{W}_n$, generate $\boldsymbol{W}_{n+1}$ via

$$\boldsymbol{W}_{n+1} = \boldsymbol{W}_n + \boldsymbol{\lambda}(\boldsymbol{W}_n, \Delta t) + \tilde{\boldsymbol{Z}}_n(\Delta t),$$

where $\tilde{\boldsymbol{Z}}_n(\Delta t) \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}(\Delta t))$.

Essentially speaking, this propagation of noise is equivalent to adding the first non-zero Brownian noise terms arising in the Itô–Taylor (IT) expansion (Kloeden and Platen, 1999, e.g. Chapter 5, p. 182) of the "smooth" components $X(t)$ and $V(t)$ in $\boldsymbol{W}(t)$, which is similar to the idea suggested by Pokern et al. (2009), leading to the exact covariance matrix obtained from the integrated Wiener process (IWP). The modified discretization scheme is then in agreement with the truncated Itô–Taylor expansion up to error terms of order $O(\Delta t^{1.5})$ in the rows that correspond to $\boldsymbol{Z}(t)$, resulting in a weak order 1.5 convergence scheme (Kloeden and Platen, 1999).

Pokern et al. (2009) showed that the IT scheme obtained above can be used to estimate the diffusion coefficients consistently, as well as to infer the unobserved "rough" components of the path. However, it produces a bias of order $O(\Delta t)$ in estimates of SDE drift parameters when used in a Gibbs sampler which alternately updates parameters and latent variables conditioned on each other and the observed data. This result warns us that we should be very careful about applying the parameter inference method even with a non-degenerate discretization scheme. They also showed that the simple Euler scheme cannot be used to estimate the diffusion coefficients consistently. This complements a result of Roberts and Stramer (2001) that Gibbs samplers for regular SDEs have arbitrarily poor mixing time in the diffusion parameters as the resolution number m increases.

However, it is important to note that these negative results do not imply that the IT (or Euler) scheme is in and of itself invalid. Indeed, Kou et al. (2012) prove that the Euler scheme converges in distribution to the true posterior $p(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N})$ as $m \to \infty$. It seems that the proof can be adapted for hypoelliptic diffusions with the IT (or Euler) scheme though we have not pursued this here. The point is that it is not the discretization scheme which is at fault, but rather the algorithm used to sample from it. The failures of the IT scheme for hypoelliptic diffusions are demonstrated only for the Gibbs sampler which alternates between parameter and latent variable updates. It is our hope that the PMMH algorithms which operate on the marginal posterior $\hat{p}(\boldsymbol{\theta} \mid \boldsymbol{Y}_{0:N})$ directly, i.e., which integrate out the latent variables for each value of $\boldsymbol{\theta}$, would not suffer for the estimation bias reported by Pokern et al. (2009). Ditlevsen and Samson (2019) proposed an order 1.5 strong IT approximation scheme to solve the issue of estimation bias for the drift parameters using the non-degenerate scheme in Pokern et al. (2009). However, their scheme applies only to one-dimensional smooth variables, whereas we have both $X(t)$ and $V(t)$. More importantly, their scheme is more complicated than ours, such that Ditlevsen and Samson (2019) have to integrate out the latent variables using a bootstrap particle filter. In contrast, the simplicity of our scheme allows us to extend the bridge proposal of Durham and Gallant (2002) to the IT setting. This is the heart of the novel contribution of this Chapter, which we develop in the following Section.

4.3.5 Particle Filtering with Bridge Proposal

For resolution number $m \geq 1$, let $\boldsymbol{W}_n^{(m)}$ denote the value of the SDE at time $t_n = n\Delta t$, such that $\boldsymbol{W}_{mn}^{(m)} = \boldsymbol{W}_n = \boldsymbol{W}(t_n)$. Then we establish the SSM with hidden states

$$\{\boldsymbol{S}_n = \boldsymbol{W}_{(n-1)m+1:nm}^{(m)}\}_{n\geq 1}$$

and observations $\{\boldsymbol{Y}_n\}_{n\geq 1}$.

A bridge proposal for $\boldsymbol{S}_n$ is constructed as follows ($0 \leq k \leq m-1$):

(i) Without loss of generality, suppose $n = 1$, we have $\boldsymbol{W}_{(n-1)m+k}^{(m)} = \boldsymbol{W}_k^{(m)} = \boldsymbol{W}(k\Delta t/m)$ is given, and we wish to draw the proposal for $\boldsymbol{W}_{k+1}^{(m)} = \boldsymbol{W}((k+1)\Delta t/m)$.

(ii) For $0 \leq s \leq (m-k)\Delta t/m$, let $\boldsymbol{W}_k(s) = \boldsymbol{W}(k\Delta t/m + s)$. Then consider the approximation

$$\boldsymbol{W}(k\Delta t/m + s) = \boldsymbol{\mu}_k(s) + \tilde{\boldsymbol{Z}}_k(s),$$

where $\boldsymbol{\mu}_k(s) = \boldsymbol{W}_k^{(m)} + \boldsymbol{\lambda}(\boldsymbol{W}_k^{(m)}, s)$ and $\tilde{\boldsymbol{Z}}_k(s)$ is a Markov process as defined above. Under this approximation we have the model

$$\begin{aligned}
\boldsymbol{W}_{k+1}^{(m)} \mid \boldsymbol{W}_k^{(m)} &\sim \mathcal{N}(\boldsymbol{\mu}_k(\Delta t/m), \tilde{\boldsymbol{\Sigma}}(\Delta t/m)) \\
\boldsymbol{W}_m^{(m)} \mid \boldsymbol{W}_{k+1}^{(m)}, \boldsymbol{W}_k^{(m)} &\sim \mathcal{N}(\boldsymbol{\mu}_k(\Delta t) + \tilde{\boldsymbol{R}}(\Delta t_k)(\boldsymbol{W}_{k+1}^{(m)} - \boldsymbol{\mu}_k(\frac{\Delta t}{m})), \tilde{\boldsymbol{\Sigma}}(\Delta t_k)) \\
\boldsymbol{Y}_1 \mid \boldsymbol{W}_m^{(m)}, \boldsymbol{W}_{k+1}^{(m)}, \boldsymbol{W}_k^{(m)} &\sim \mathcal{N}\left(\boldsymbol{A}\boldsymbol{W}_m^{(m)}, \boldsymbol{\Omega}\right),
\end{aligned}$$

where $\Delta t_k = (m-k-1)\Delta t/m$. We may now apply Lemma 4.2.1 and formulas for conditional distributions of the multivariate normal to sample

$$\boldsymbol{W}_{k+1}^{(m)} \sim p(\boldsymbol{W}_{k+1}^{(m)} \mid \boldsymbol{W}_k^{(m)}, \boldsymbol{Y}_1),$$

namely

$$\boldsymbol{W}_{k+1}^{(m)} \mid \boldsymbol{W}_k^{(m)}, \boldsymbol{Y}_1 \sim \mathcal{N}(\boldsymbol{\lambda}_{k+1}, \boldsymbol{\Omega}_{k+1}).$$

In fact, we can derive the joint distribution of $(\boldsymbol{W}_{k+1}^{(m)}, \boldsymbol{Y}_1)$ as

$$\begin{bmatrix} \boldsymbol{W}_{k+1}^{(m)} \\ \boldsymbol{Y}_1 \end{bmatrix} \mid \boldsymbol{W}_k^{(m)} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_k(\Delta t/m) \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}(\frac{\Delta t}{m}) & \tilde{\boldsymbol{\Sigma}}(\frac{\Delta t}{m})\tilde{\boldsymbol{R}}(\Delta t_k)'\boldsymbol{A}' \\ \boldsymbol{A}\tilde{\boldsymbol{R}}(\Delta t_k)\tilde{\boldsymbol{\Sigma}}(\frac{\Delta t}{m})' & \boldsymbol{\Pi}_{m,k} \end{bmatrix}\right).$$

where $\boldsymbol{\Pi}_{m,k} = \boldsymbol{A}\tilde{\boldsymbol{R}}(\Delta t_k)\tilde{\boldsymbol{\Sigma}}(\frac{\Delta t}{m})\tilde{\boldsymbol{R}}(\Delta t_k)'\boldsymbol{A}' + \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}(\Delta t_k)\boldsymbol{A}' + \boldsymbol{\Omega}$. Note $\boldsymbol{\Pi}_{m,k}$ is the sum of three symmetric, positive definite matrices, and thus is invertible. Then by applying the formula for the conditional distribution of a multivariate normal distribution, we get find the formulae for $\boldsymbol{\lambda}_{k+1}$ and $\boldsymbol{\Omega}_{k+1}$

$$\begin{aligned}
\boldsymbol{\lambda}_{k+1} &= \boldsymbol{\mu}_k(\Delta t/m) + \tilde{\boldsymbol{\Sigma}}(\Delta t/m)\tilde{\boldsymbol{R}}(\Delta t_k)'\boldsymbol{A}'\boldsymbol{\Pi}_{m,k}^{-1}\boldsymbol{Y}_1 \\
\boldsymbol{\Omega}_{k+1} &= \tilde{\boldsymbol{\Sigma}}(\Delta t/m) - \tilde{\boldsymbol{\Sigma}}(\Delta t/m)\tilde{\boldsymbol{R}}(\Delta t_k)'\boldsymbol{A}'\boldsymbol{\Pi}_{m,k}^{-1}\boldsymbol{A}\tilde{\boldsymbol{R}}(\Delta t_k)\tilde{\boldsymbol{\Sigma}}(\Delta t/m)'
\end{aligned}$$

A general result regarding the joint distribution of $(\boldsymbol{W}_m^{(m)}, \boldsymbol{W}_{k+1}^{(m)}, \boldsymbol{Y}_1)$ and the conditional distribution $(\boldsymbol{W}_m^{(m)} \mid \boldsymbol{W}_{k+1}^{(m)}, \boldsymbol{Y}_1)$ is given in Appendix C.3. We can now embed this bridge proposal into the particle filter described in Algorithm 4.1.

4.4 Concluding Remarks

In order to at least partially solve the parameter inference problem for the general nonlinear GLE, we favored a type of quasi-Markovian scheme (denoted as qmGLE) to convert the non-Markovian GLE to a Markovian system of SDEs with the same solution process (Pavliotis, 2014; Ottobre and Pavliotis, 2010). Based on the qmGLE, we analyzed the limitations of default particle filter techniques (e.g. bootstrap particle filter) and carefully derived the bridge proposal for particle filters. Furthermore, we revealed that the failure lies in the degenerate Euler scheme such that the bridge proposal is essentially as bad as the bootstrap filter due to a lack of randomness, which is an intrinsic problem for hypoelliptic SDEs. Then we were motivated to propose a modified scheme based on the integrated Wiener process to propagate randomness into the full qmGLE system. By applying the particle filter with the bridge proposal, we present a promising method to perform particle filtering and parameter inference. We have discussed parameter estimation conceptually in Section 4.2.2. The implementation of a careful designed estimation algorithm and extensive numerical verification will be the subject of future work.

Our proposed method avoids the degeneracy caused by hypoelliptic diffusions. The non-degenerate discretization scheme with data augmentation is guaranteed to converge in distribution to the true posterior, as shown by Kou et al. (2012). However, a badly designed parameter inference method could easily ruin this. That basically explains some reasons behind the failure of the Gibbs sampler used by Pokern et al. (2009) and the success of the particle filter method proposed by Ditlevsen and Samson (2019), aside from their different discretization schemes, for the parameter estimation of hypoelliptic SDEs. We believe our proposed method, i.e. the particle filter with bridge proposal built upon a non-degenerate Itô-Taylor approximation scheme for the qmGLE can avoid the parameter estimation issue as long as a well-designed inferential algorithm is implemented such that the parameter estimation is not intertwined with the updating of latent variables. Some promising methods are available in the current literature, e.g. either the particle MCMC (Andrieu et al., 2010) or simulated maximum likelihood (e.g. Cappé and Moulines, 2005; Poyiadjis et al., 2011; Corenflos et al., 2021). But the implementation of parameter estimation with careful numerical checks will be left for future research.

We realized there is another way to simulate the diffusion process, using the exact algorithm (EA) proposed and studied by Beskos and Roberts (2005); Beskos et al. (2006a,b, 2009) and Fearnhead et al. (2008). However, the EA method requires the diffusion process to have some good analytic properties. In contrast, we still prefer a non-degenerate discretization scheme which allows us to build more advanced sampling mechanism for Bayesian parameter estimation.

One last thing to note is that in our parameterization, we have assumed $\boldsymbol{A}$ to be symmetric such that it can be easily diagonalized. In general, it may not be true. For a general positive definite $\boldsymbol{A}$, a better way to overcome the identifiability issue may be needed. But this will lead us to a challenging field of memory kernel approximation for the GLE, which will also be the subject of future work, as discussed in Chapter 5.

Chapter 5

Conclusion and Future Work

In Chapter 2, we established a representation theorem which reveals the fundamental connections between the GLE and any continuous weakly stationary process. This result greatly helps and facilitates the process of statistical modeling based on complex data generated from an underlying GLE dynamics, which was illustrated by the example of parametric spectral density estimation for the HTP data in Chapter 3. Moreover, we also proved in Chapter 3 the theoretical high-frequency asymptotics of three main statistical estimators of interest, the Whittle-type MLE, LP estimator, and NLS estimator, laying the foundation for a very efficient and robust noise removal procedure proposed by Lysy et al. (2022). Finally, in Chapter 4 we made an attempt with thorough theoretical consideration to solve the inference problem for the nonlinear (and also non-Markovian) GLE with extensive applications in more complex thermodynamic environment. By designing a non-degenerate discretization scheme with additional randomness introduced from the integrated Wiener process, we propose a particle filter technique with a bridge proposal to fit the quasi-Markovian SDE approximation of the GLE. We emphasize the critical point that the failure of designing a suitable inferential algorithm would easily destroy the convergence of the discretization scheme (or the data-generating model). To properly deal with the parameter estimation for the hypoelliptic SDEs, we suggest performing the task solely in the parameter space without sampling problematic latent variables along the way. Particle filter based methods, which can be used to integrate latent variables, are promising for this purpose.

However, there are still some very interesting and promising topics left unsolved or for future research.

For the fundamental theorem we established in Chapter 2, it is likely that there is

a non-trivial extension to the continuous increment stationary process. But to prove the conjecture in a rigorous mathematical language (mainly based on unbounded operators and semigroup theory) is not easy, especially in the abstract infinite dimensional Hamiltonian Hilbert space. That being said, this direction is also very promising in that once proved completely, it would certainly give much deeper insights to our understanding of complex thermodynamics and its relationship with various stochastic processes in statistics.

For asymptotic properties of likelihood-based estimators under HTP data, a theoretically intriguing topic may be the case where the sampling frequency f_s and the number of sample size N can approach infinity simultaneously. Even though we have already argued that from an experimental perspective, this case may be of no practical interest, at least in scientific experiments. But mathematically, there is certainly a huge gap between the high-frequency (and even high-dimensional at the same time) asymptotics and the traditional large sample asymptotics in time series, which may be worth filling.

Finally, for the nonlinear GLE, an efficient and general implementation of our proposed method is the subject of future work. As for the methodology per se, we actually assumed that the matrix $\boldsymbol{A}$ in the memory kernel $\gamma(t) = \langle e^{-\boldsymbol{A}t}\boldsymbol{\lambda}, \boldsymbol{\lambda}\rangle$ is symmetric and thus easily diagonalizable. Otherwise, we would certainly face some identifiability issues. A long-lasting research interest in the realm of GLEs is about the memory kernel approximation. Due to the mathematical intricacy caused by the convolution integral in the GLE, this task is non-trivial. Over the years, many techniques have been proposed to address this problem, either numerically or analytically. In summary, there are two main directions. The first one is data-driven methods, i.e., to recover the memory kernel based on data. For example, Lei et al. (2016) developed rational function approximation using Laplace transform; Brennan and Venturi (2018) designed a conditional expectation technique. The second direction is to approximate the memory kernel from first-principles based on the internal nonlinear structure without using simulation. A classic and well-known example is the continued-fraction expansion proposed by Mori (1965a), or using recurrence relations (Lee, 1982a,b). For further discussions on this, please refer to Zhu and Lei (2021) and the references therein. This topic remains to be the subject of future work.

Overall, better statistical models and inference methods are indispensable in nanoscopic systems for us to verify, investigate and understand the hidden dynamics of individual molecules or nanoparticles from enormous amount of noisy observations, as scientists zoom in on the intrinsically stochastic world. In retrospect of what Richard Feynman said over sixty years ago, the same is so true for statistics today. This thesis only serves as a small step forward to fill some theoretical gaps in this realm, hoping we or other research colleagues could make a further leap in the near future.

References

Adelman, S. A. and Doll, J. D. (1974). Generalized Langevin equation approach for atom/solid-surface scattering: Collinear atom/harmonic chain model. *The Journal of Chemical Physics*, 61(10):4242–4245.

Adelman, S. A. and Doll, J. D. (1976). Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *The Journal of Chemical Physics*, 64(6):2375–2388.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1):223–262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, 36(2).

Albers, J., M, D. J., and Oppenheim, I. (1971). Generalized Langevin equations. *Journal of Chemical Physics*, 54:3541–3546.

Alsteens, D., Verbelen, C., Dague, E., Raze, D., Baulard, A. R., and Dufrêne, Y. F. (2008). Organization of the mycobacterial cell wall: A nanoscale view. *Pflügers Archiv-European Journal of Physiology*, 456(1):117–125.

Andrews, D. W. K. and Guggenberger, P. (2003). A bias–reduced log–periodogram regression estimator for the long–memory parameter. *Econometrica*, 71(2):675–712.

Andrews, D. W. K. and Sun, Y. (2004). Adaptive local polynomial whittle estimation of long-range dependence. *Econometrica*, 72(2):569–614.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342.

Arendt, W., Batty, C. J., Hieber, M., and Neubrander, F. (2011). *Vector-Valued Laplace Transforms and Cauchy Problems*. Birkhäuser, Basel, Switzerland, second edition.

Ariel, G. and Vanden-Eijnden, E. (2009). A strong limit theorem in the Kac–Zwanzig model. *Nonlinearity*, 22(1):145–162.

Barret, D. and Vaughan, S. (2012). Maximum likelihood fitting of X-ray power density spectra: Application to high-frequency quasi-periodic oscillations from the neutron star X-ray binary 4U1608-522. *The Astrophysical Journal*, 746(2):131.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, 37(1/2):1–16.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. SpringerLink : Bücher. Springer, Berlin, Heidelberg.

Beskos, A., Papaspiliopoulos, O., and Roberts, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *The Annals of Statistics*, 37(1).

Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2006a). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6).

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006b). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. *The Annals of Applied Probability*, 15(4).

Bibby, B. M., Jacobsen, M., and Sørensen, M. (2004). Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics: Tools and Techniques*, pages 203–268. Elsevier.

Binnig, G., Quate, C. F., and Gerber, C. (1986). Atomic force microscope. *Physical Review Letters*, 56(9):930.

Botha, I., Kohn, R., and Drovandi, C. (2021). Particle methods for stochastic differential equation mixed effects models. *Bayesian Analysis*, 16(2).

Brennan, C. and Venturi, D. (2018). Data-driven closures for stochastic dynamical systems. *Journal of Computational Physics*, 372:281–298.

Brockwell, P. J. and Davis, R. A. (2006). *Time Series: Theory and Methods*. Springer Science & Business Media, New York, NY, second edition.

Burnecki, K. (2012). FARIMA processes with application to biophysical data. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(05):P05015.

Burnham, N. A., Chen, X., Hodges, C. S., Matei, G. A., Thoreson, E. J., Roberts, C. J., Davies, M. C., and Tendler, S. J. B. (2002). Comparison of calibration methods for atomic-force microscopy cantilevers. *Nanotechnology*, 14(1):1.

Cappé, O. and Moulines, E. (2005). On the use of particle filtering for maximum likelihood parameter estimation. In *2005 13th European Signal Processing Conference*, pages 1–4.

Celli, M., Colognesi, D., and Zoppi, M. (2002). Direct experimental access to microscopic dynamics in liquid hydrogen. *Physical Review E*, 66(2):021202.

Ceriotti, M., Manolopoulos, D. E., and Parrinello, M. (2011). Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *The Journal of Chemical Physics*, 134(8):084104.

Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York, NY.

Chernoff, P. R. and Marsden, J. E. (1975). Some basic properties of infinite dimensional Hamiltonian systems. In *Géométrie Symplectique et Physique Mathématique*, No.237. Éditions, pages 313–330. du C.N.R.S., Paris.

Chib, S., Pitt, M., and Shephard, N. (2004). Likelihood based inference for diffusion driven models. *Nuffield College (University of Oxford) Series: Economics working papers*.

Chon, J. W., Mulvaney, P., and Sader, J. E. (2000). Experimental validation of theoretical models for the frequency response of atomic force microscope cantilever beams immersed in fluids. *Journal of applied physics*, 87(8):3978–3988.

Chui, C. K. (1971). Concerning rates of convergence of Riemann sums. *Journal of Approximation Theory*, 4(3):279–287.

Ciccotti, G. and Ryckaert, J.-P. (1981). On the derivation of the generalized Langevin equation for interacting Brownian particles. *Journal of Statistical Physics*, 26(1):73–82.

Clark, M. T., Sader, J. E., Cleveland, J. P., and Paul, M. R. (2010). Spectral properties of microcantilevers in viscous fluid. *Physical Review E*, 81:045306 1–10.

Clarke, R. J., Jensen, O. E., Billingham, J., Pearson, A. P., and Williams, P. M. (2006). Stochastic elastohydrodynamics of a microcantilever oscillating near a wall. *Physical Review Letters*, 96(5):050801.

Cleveland, J. P., Manne, S., Bocek, D., and Hansma, P. K. (1993). A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy. *Review of Scientific Instruments*, 64(2):403–405.

Comte, F. (1996). Simulation and estimation of long memory continuous time models. *Journal of Time Series Analysis*, 17(1):19–36.

Corenflos, A., Thornton, J., Deligiannidis, G., and Doucet, A. (2021). Differentiable particle filtering via entropy-regularized optimal transport. In *Proceedings of the 38th International Conference on Machine Learning*, pages 2100–2111. PMLR.

Cortés, E., West, B. J., and Lindenberg, K. (1985). On the generalized Langevin equation: Classical and quantum mechanical. *The Journal of Chemical Physics*, 82(6):2708–2717.

Cramér, H. (1961). On some classes of nonstationary stochastic processes. In Neyman, J., editor, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory*, volume 4, pages 57–78.

Cramér, H. (1967). A contribution to the multiplicity theory of stochastic processes. In Le Cam, L. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1*, pages 215–221.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4):1749–1766.

Dalla, V., Giraitis, L., and Hidalgo, J. (2006). Consistent estimation of the memory parameter for nonlinear time series. *Journal of Time Series Analysis*, 27(2):211–251.

Daniell, P. J. (1946). Discussion on the paper by M.S. Bartlett "On the theoretical specification and sampling properties of auto-correlated time series". *Supplement to Journal of the Royal Statistical Society*, 8(1):88–90.

de Oliveira, M. J. (2020). Quantum Langevin equation. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(2):023106.

Démery, V., Bénichou, O., and Jacquin, H. (2014). Generalized Langevin equations for a driven tracer in dense soft colloids: Construction and applications. *New Journal of Physics*, 16(5):053032.

Didier, G., McKinley, S. A., Hill, D., and Fricks, J. (2012). Statistical challenges in microrheology. *ERN: Other Econometrics: Single Equation Models (Topic)*.

Ditlevsen, S. and Samson, A. (2019). Hypoelliptic diffusions: Filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(2):361–384.

Doll, J. D. and Dion, D. R. (1976). Generalized Langevin equation approach for atom/solid–surface scattering: Numerical techniques for Gaussian generalized Langevin dynamics. *The Journal of Chemical Physics*, 65(9):3762–3766.

Doll, J. D., Myers, L. E., and Adelman, S. A. (1975). Generalized Langevin equation approach for Atom/Solid-surface scattering: Inelastic studies. *The Journal of Chemical Physics*, 63(11):4908–4914.

Doucet, A., Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York, NY.

Doucet, A., Johansen, A. M., et al. (2009). A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3.

Du, C. and Kou, S. C. (2019). Supplement: Statistical methodology in single-molecule experiments.

Du, C. and Kou, S. C. (2020). Statistical methodology in single-molecule experiments. *Statistical Science*, 35(1).

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–338.

Ebbinghaus, S., Kim, S. J., Heyden, M., Yu, X., Heugen, U., Gruebele, M., Leitner, D. M., and Havenith, M. (2007). An extended dynamical hydration shell around proteins. *Proceedings of the National Academy of Sciences*, 104:20749–20752.

Eckmann, J.-P., Pillet, C.-A., and Rey-Bellet, L. (1999). Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Communications in Mathematical Physics*, 201(3):657–697.

Einstein, A. (1905). über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen (On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat). *Annalen der Physik (in German)*, 322(8):549–560.

Einstein, A. (1956). *Investigations on the Theory of the Brownian Movement*. Dover Books on Physics Series. Dover Publications (Original work published in 1926, edited with notes by R. Fürth, translated by A. D. Cowper), New York, NY.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4):959–993.

Engel, K. and Nagel, R. (2000). *One-Parameter Semigroups for Linear Evolution Equations*. Graduate Texts in Mathematics. Springer-Verlag, New York, NY.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, 19(2):177–191.

Evans, E. A. and Calderwood, D. A. (2007). Forces and bond dynamics in cell adhesion. *Science*, 316(5828):1148–1153.

Fasen, V. and Fuchs, F. (2013). On the limit behavior of the periodogram of high-frequency sampled stable CARMA processes. *Stochastic Processes and their Applications*, 123(1):229–273.

Fay, G., Moulines, E., and Soulier, P. (2002). Nonlinear functionals of the periodogram. *Journal of Time Series Analysis*, 23(5):523–553.

Fearnhead, P., Papaspiliopoulos, O., and Roberts, G. O. (2008). Particle filters for partially observed diffusions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(4):755–777.

Fletcher, R. (1987). *Practical Methods of Optimization*. John Wiley & Sons, Inc., New York, NY.

Ford, G. W. and Kac, M. (1987). On the quantum Langevin equation. *Journal of Statistical Physics*, 46(5):803–810.

Ford, G. W., Kac, M., and Mazur, P. (1965). Statistical mechanics of assemblies of coupled oscillators. *Journal of Mathematical Physics*, 6(4):504–515.

Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 14(2):517–532.

Fricks, J., Yao, L., Elston, T. C., and Forest, M. G. (2009). Time-domain methods for diffusive transport in soft matter. *SIAM Journal on Applied Mathematics*, 69(5):1277–1308.

Gallant, A. and Long, J. (1997). Estimating stochastic differential equations efficiently by minimum Chi-squared. *Biometrika*, 84(1):125–141.

García, R., Magerle, R., and Perez, R. (2007). Nanoscale compositional mapping with gentle forces. *Nature Materials*, 6(6):405–411.

García, R. and Proksch, R. (2013). Nanomechanical mapping of soft matter by bimodal force microscopy. *European Polymer Journal*, 49(8):1897–1906.

Gasser, T., Muller, H.-G., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, pages 238–252.

Gerardi, C., Cory, D., Buongiorno, J., Hu, L., and McKrell, T. (2009). Nuclear magnetic resonance-based study of orderd layering on the surface of alumina nanoparticles in water. *Applied Physics Letters*, 95:253104 1–3.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

Giessibl, F. J. (2003). Advances in atomic force microscopy. *Reviews of Modern Physics*, 75(3):949.

Giraitis, L., Koul, H. L., and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. World Scientific Publishing Company, Singapore.

Givon, D., Kupferman, R., and Stuart, A. (2004). Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity*, 17(6):R55–R127.

Go, J., Nair, P. R., and Alam, M. A. (2012). Theory of signal and noise in double-gated nanoscale electronic pH sensors. *Journal of Applied Physics*, 112:034516 1–10.

Goldstein, H., Poole, C., and Safko, J. (2002). *Classical Mechanics*. Addison Wesley, New York, NY, 3rd edition.

Golightly, A. and Wilkinson, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, 52(3):1674–1693.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140(2):107–113(6).

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:S85–S118.

Gritsun, A. and Branstator, G. (2007). Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *Journal of the Atmospheric Sciences*, 64(7):2558–2575.

Gross, L. (1967). Abstract wiener spaces. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1*, volume 5, pages 31–43. University of California Press.

Hall, B. (2013). *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics. Springer Science & Business Media, New York, NY.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10(1):130–145.

Harkey, J. A. and Kenny, T. W. (2000). 1/f noise considerations for the design and process optimization of piezoresistive cantilevers. *Journal of Microelectromechanical Systems*, 9(2):226–235.

Heerema, S. J., Schneider, G. F., Rozemuller, M., Vicarelli, L., Zandbergen, H. W., and Dekker, C. (2015). 1/f noise in graphene nanopores. *Nanotechnology*, 26(7):074001.

Herruzo, E. T., Perrino, A. P., and Garcia, R. (2014). Fast nanomechanical spectroscopy of soft matter. *Nature Communications*, 5.

Hille, E. and Phillips, R. S. (1957). *Functional Analysis and Semigroups*. American Mathematical Society, New York, NY, revised and expanded edition.

Hoffmann, P. M., Oral, A., Grimble, R. A., Jeffery, S., and Pethica, J. B. (2001). Direct measurement of interatomic force gradients using an ultra-low-amplitude atomic force microscope. *Proceedings of the Royal Society of London, Series A*, 457(2009):1161–1174.

Hohenegger, C. and McKinley, S. A. (2017). Fluid–particle dynamics for passive tracers advected by a thermally fluctuating viscoelastic medium. *Journal of Computational Physics*, 340:688–711.

Hol, J. D., Schon, T. B., and Gustafsson, F. (2006). On resampling algorithms for particle filters. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 79–82.

Hörmander, L. (1961). Hypoelliptic differential operators. *Annales de l'Institut Fourier*, 11:477–492.

Hörmander, L. (1967). Hypoelliptic second order differential equations. *Acta Mathematica*, 119:147–171.

Hurvich, C. M., Deo, R., and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19(1):19–46.

Itô, K. (1954). Stationary random distributions. *Memoirs of the College of Science, University of Kyoto. Series A: Mathematics*, 28(3):209–223.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643.

Johansen, A. M. (2009). SMCTC: Sequential monte carlo in C++. *Journal of Statistical Software*, 30(6):1–41.

Kantorovich, L. (2008). Generalized Langevin equation for solids. I. Rigorous derivation and main properties. *Physical Review B*, 78(9):094304.

Kantorovich, L. and Rompotis, N. (2008). Generalized Langevin equation for solids. II. Stochastic boundary conditions for nonequilibrium molecular dynamics. *Physical Review E*, 78:094305 1–8.

Kast, J., Gentzel, M., Wilm, M., and Richardson, K. (2003). Noise filtering techniques for electrospray quadrupole time of flight mass spectra. *Journal of the American Society for Mass Spectrometry*, 14(7):766–776.

Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, 109:604–615.

Kirmizis, D. and Logothetidis, S. (2010). Atomic force microscopy probing in the measurement of cell mechanics. *International Journal of Nanomedicine*, 5:137.

Kloeden, P. E. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, Heidelberg, 3rd corrected printing of 1992 edition.

Kneller, G. R. (2011). Generalized Kubo relations and conditions for anomalous diffusion: Physical insights from a mathematical theorem. *The Journal of Chemical Physics*, 134(22):224106.

Kou, S. and Xie, X. (2004). Generalized Langevin equation with fractional Gaussian noise: Subdiffusion within a single protein molecule. *Physical Review Letters*, 93(18):180603.

Kou, S. C. (2008). Stochastic networks in nanoscale biophysics: Modeling enzymatic reaction of a single protein. *Journal of the American Statistical Association*, 103:961–975.

Kou, S. C., Olding, B. P., Lysy, M., and Liu, J. S. (2012). A multiresolution method for parameter estimation of diffusion processes. *Journal of the American Statistical Association*, 107(500):1558–1574.

Krommes, J. A. (2018a). Projection-operator methods for classical transport in magnetized plasmas. Part 1. Linear response, the Braginskii equations and fluctuating hydrodynamics. *Journal of Plasma Physics*, 84(4):925840401.

Krommes, J. A. (2018b). Projection-operator methods for classical transport in magnetized plasmas. Part 2. Nonlinear response and the Burnett equations. *Journal of Plasma Physics*, 84(6):905840601.

Kubo, R. (1966). The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29:255–284.

Kupferman, R. (2004). Fractional kinetics in Kac–Zwanzig heat bath models. *Journal of Statistical Physics*, 114(1/2):291–326.

Kuptametee, C. and Aunsri, N. (2022). A review of resampling techniques in particle filtering framework. *Measurement*, 193:110836.

Labuda, A., Kocuń, M., Lysy, M., Walsh, T., Meinhold, J., Proksch, T., Meinhold, W., Anderson, C., and Proksch, R. (2016a). Calibration of higher eigenmodes of cantilevers. *Review of Scientific Instruments*, 87(7):073705.

Labuda, A., Kocuń, M., Meinhold, W., Walters, D., and Proksch, R. (2016b). Generalized Hertz model for bimodal nanomechanical mapping. *Beilstein Journal of Nanotechnology*, 7(1):970–982.

Labuda, A., Lysy, M., and Grütter, P. (2012). Stochastic simulation of tip-sample interactions in atomic force microscopy. *Applied Physics Letters*, 101:113105 1–4.

Lange, O. F. and Grubmüller, H. (2006). Collective Langevin dynamics of conformational motions in proteins. *Journal of Chemical Physics*, 124(21):214903.

Lax, P. D. (2002). *Functional Analysis*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons, Inc., New York, NY.

Lee, M. H. (1982a). Orthogonalization process by recurrence relations. *Physical Review Letter*, 49(15):1072–1075.

Lee, M. H. (1982b). Solutions of the generalized Langevin equation by a method of recurrence relations. *Physical Review B*, 26(5):2547–2551.

Lees, J. M. (1995). Reshaping spectrum estimates by removing periodic noise: Application to seismic spectral ratios. *Geophysical research letters*, 22(4):513–516.

Lei, H., Baker, N. A., and Li, X. (2016). Data-driven parameterization of the generalized Langevin equation. *Proceedings of the National Academy of Sciences*, 113(50):14183–14188.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168.

Li, J. F., Huang, Y. F., Ding, Y., Yang, Z. L., Li, S. B., Zhou, X. S., Fan, F. R., Zhang, W., Zhou, Z. Y., Ren, B., et al. (2010). Shell-isolated nanoparticle-enhanced Raman spectroscopy. *Nature*, 464(7287):392–395.

Li, Z., Bian, X., Li, X., and Karniadakis, G. E. (2015). Incorporation of memory effects in coarse-grained modeling via the Mori–Zwanzig formalism. *The Journal of Chemical Physics*, 143(24):243128.

Littenberg, T. B. and Cornish, N. J. (2015). Bayesian inference for spectral estimation of gravitational wave detector noise. *Physical Review D*, 91(8):084034.

Liu, J. S. (2004). *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer, New York, NY.

Loudet, A. and Burgess, K. (2007). BODIPY dyes and their derivatives: Syntheses and spectroscopic properties. *Chemical Reviews*, 107(11):4891–4932.

Lysy, M., Pillai, N. S., Hill, D. B., Forest, M. G., Mellnik, J. W. R., Vasquez, P. A., and McKinley, S. A. (2016). Model comparison and assessment for single particle tracking in biological fluids. *Journal of the American Statistical Association*, 111(516):1413–1426.

Lysy, M., Zhu, F., Yates, B., and Labuda, A. (2022). Robust and efficient parametric spectral density estimation for high-throughput data. *Technometrics*, 64(1):30–51.

Marchetto, D., Rota, A., Calabri, L., Gazzadi, G. G., Menozzi, C., and Valeri, S. (2008). AFM investigation of tribological properties of nano-patterned silicon surface. *Wear*, 265:577–582.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

Martinez, N. F., Lozano, J. R., Herruzo, E. T., Garcia, F., Richter, C., Sulzbach, T., and Garcia, R. (2008). Bimodal atomic force microscopy imaging of isolated antibodies in air and liquids. *Nanotechnology*, 19(38):384011.

Maruyama, G. (1955). Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48–90.

Mason, T. G., Gang, H., and Weitz, D. A. (1997). Diffusing-wave-spectroscopy measurements of viscoelasticity of complex fluids. *Journal of the Optical Society of America A*, 14(1):139–149.

Mason, T. G. and Weitz, D. (1995). Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids. *Physical Review Letters*, 74(7):1250.

Mattingly, J. C., Stuart, A. M., and Higham, D. J. (2002). Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185–232.

Mazur, P. and Oppenheim, I. (1970). Molecular theory of Brownian motion. *Physica*, 50(2):241–258.

McDowell, H. K. (2000). Quantum generalized Langevin equation: Explicit inclusion of nonlinear system dynamics. *The Journal of Chemical Physics*, 112(16):6971–6982.

McKinley, S. A. and Nguyen, H. D. (2018). Anomalous diffusion and the generalized Langevin equation. *SIAM Journal on Mathematical Analysis*, 50(5):5119–5160.

McKinley, S. A., Yao, L., and Forest, M. G. (2009). Transient anomalous diffusion of tracer particles in soft matter. *Journal of Rheology*, 53:1487–1506.

Moerner, W. E. (2002). A dozen years of single-molecule spectroscopy in physics, chemistry, and biophysics. *The Journal of Physical Chemistry B*, 106(5):910–927.

Morgado, R., Oliveira, F. A., Batrouni, G. G., and Hansen, A. (2002). Relation between anomalous and normal diffusion in systems with memory. *Physical Review Letter*, 89(10):100601.

Mori, H. (1965a). A continued-fraction representation of the time-correlation functions. *Progress of Theoretical Physics*, 34(3):399–416.

Mori, H. (1965b). Transport, collective motion, and Brownian motion*). *Progress of Theoretical Physics*, 33(3):423–455.

Moulines, E. and Soulier, P. (1999). Broadband log-periodogram regression of time series with long-range dependence. *The Annals of Statistics*, 27(4):1415–1439.

Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In Doukhan, P., Oppenheim, G., and Taqqu, M., editors, *Theory and Applications of Long-Range Dependence*, pages 251–301. Birkhauser, Boston, MA.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.

Murray, L. M. (2013). Bayesian state-space modelling on high-performance hardware using LibBi. *arXiv:1306.3277*.

Nakajima, S. (1958). On quantum theory of transport phenomena: Steady diffusion. *Progress of Theoretical Physics*, 20(6):948–959.

Nason, G. and Silverman, B. (1995). The stationary wavelet transform and some statistical applications. In Antoniadis, A. and Oppenheim, G., editors, *Wavelets and Statistics*, pages 281–299. Springer-Verlag, New York, NY.

Nelson, E. (2001). *Dynamical Theories of Brownian Motion*. Princeton University Press, Princeton, NJ, second edition.

Ng, E. W. and Geller, M. (1969). A table of integrals of the error functions. *Journal of Research of the National Bureau of Standards B*, 73(1):1–20.

Nie, S. and Zare, R. N. (1997). Optical detection of single molecules. *Annual Review of Biophysics and Biomolecular Structure*, 26(1):567–596.

Nordholm, S. and Zwanzig, R. (1975). A systematic derivation of exact generalized Brownian motion theory. *Journal of Statistical Physics*, 13(4):347–371.

Nørrelykke, S. F. and Flyvbjerg, H. (2010). Power spectrum analysis with least-squares fitting: Amplitude bias and its elimination, with application to optical tweezers and atomic force microscope cantilevers. *Review of Scientific Instruments*, 81(7):075103.

Okabe, Y. (1986). KMO-Langevin equation and fluctuation-dissipation theorem (II). *Hokkaido Mathematical Journal*, 15(3):317–355.

Oksendal, B. (2013). *Stochastic Differential Equations: An Introduction with Applications*. Springer-Verlag, Berlin, Heidelberg.

Ottobre, M. and Pavliotis, G. (2010). Asymptotic analysis for the generalized Langevin equation. *Nonlinearity*, 24:1629–1653.

Ottobre, M., Pavliotis, G., and Pravda-Starov, K. (2012). Exponential return to equilibrium for hypoelliptic quadratic systems. *Journal of Functional Analysis*, 262(9):4000–4039.

Panja, D. (2010). Generalized Langevin equation formulation for anomalous polymer dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(2):L02001 1–8.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, volume 60. Springer, New York, NY.

Pavliotis, G. A. and Stuart, A. M. (2007). Parameter estimation for multiscale diffusions. *Journal of Statistical Physics*, 127(4):741–781.

Pedersen, A. R. (1995a). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, 1(3):257–279.

Pedersen, A. R. (1995b). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71.

Picchini, U. and Forman, J. L. (2019). Bayesian inference for stochastic differential equation mixed effects models of a tumour xenography study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(4):887–913.

Pokern, Y., Stuart, A. M., and Wiberg, P. (2009). Parameter estimation for partially observed hypoelliptic diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):49–73.

Poyiadjis, G., Doucet, A., and Singh, S. S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98(1):65–80.

Radmacher, M. (1997). Measuring the elastic properties of biological samples with the AFM. *Ieee Engineering in Medicine and Biology Magazine*, 16(2):47–57.

Reed, M. and Simon, B. (1981). *I: Functional Analysis*. Methods of Modern Mathematical Physics. Academic Press, San Diego, CA, revised and enlarged edition.

Rey-Bellet, L. (2006). Open classical systems. In Attal, S., Joye, A., and Pillet, C.-A., editors, *Open Quantum Systems II: The Markovian Approach*, pages 41–78. Springer, Berlin, Heidelberg.

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York, NY.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika*, 88(3):603–621.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072.

Rudin, W. (1973). *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, New York, NY, second edition.

Sader, J. E. (1998). Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope. *Journal of Applied Physics*, 84(1):64–76.

Sader, J. E., Sanelli, J., Hughes, B. D., Monty, J. P., and Bieske, E. J. (2011). Distortion in the thermal noise spectrum and quality factor of nanomechanical devices due to finite frequency resolution with applications to the atomic force microscope. *Review of Scientific Instruments*, 82(9):095104.

Sanei, S. and Chambers, J. A. (2013). *EEG Signal Processing*. John Wiley & Sons, Inc., Chichester, UK.

Schlick, T. (2010). *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer Science & Business Media, New York, NY, second edition.

Schober, M., Särkkä, S., and Hennig, P. (2019). A probabilistic model for the numerical solution of initial value problems. *Statistics and Computing*, 29(1):99–122.

Seisenberger, G., Ried, M. U., Endreß, T., Büning, H., Hallek, M., and Bräuchle, C. (2001). Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science*, 294(5548):1929–1932.

Sikula, J. and Levinshtein, M., editors (2004). *Advanced Experimental Methods for Noise Research in Nanoscale Electronic Devices*. NATO Science Series. Klewer Academic Publishers, Norwell.

Solomon, M. J. and Lu, Q. (2001). Rheology and dynamics of particles in viscoelastic media. *Current Opinion in Colloid & Interface Science*, 6(5):430–437.

Squires, T. and Mason, T. (2010). Fluid mechanics of microrheology. *Annual Review of Fluid Mechanics*, 42(1):413–438.

Stein, E. M. and Shakarchi, R. (2005). *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton Lectures in Analysis. Princeton University Press, Princeton, NJ.

Stone, M. H. (1930). Linear transformations in Hilbert space. *Proceedings of the National Academy of Sciences*, 16(2):172–175.

Stone, M. H. (1932). On one-parameter unitary groups in Hilbert space. *Annals of Mathematics*, 33(3):643–648.

Sugimoto, Y., Pou, P., Abe, M., Jelinek, P., Pérez, R., Morita, S., and Custance, O. (2007). Chemical identification of individual surface atoms by atomic force microscopy. *Letters to Nature*, 446:64–67.

Sun, Y. and Phillips, P. C. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, 115(2):355–389.

Sykulski, A. M., Olhede, S. C., Lilly, J. M., and Danioux, E. (2016). Lagrangian time series models for ocean surface drifter trajectories. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(1):29–50.

Tamarat, P., Maali, A., Lounis, B., and Orrit, M. (2000). Ten years of single-molecule spectroscopy. *The Journal of Physical Chemistry A*, 104(1):1–16.

Taniguchi, M. (1987). Minimum contrast estimation for spectral densities of stationary processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 49(3):315–325.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096.

Trefethen, L. N. and Bau, D. I. (1997). *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics.

Tsai, H. and Chan, K.-S. (2005). Quasi-maximum likelihood estimation for a class of continuous-time long-memory processes. *Journal of Time Series Analysis*, 26(5):691–713.

Tsekov, R. (2016). Brownian Emitters. *Fluctuation and Noise Letters*, 15(04):1650022.

Tuckerman, M. (2010). *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, New York, NY.

Valentine, M. T., Kaplan, P. D., Thota, D., Crocker, J. C., Gisler, T., Prud'homme, R. K., Beck, M., and Weitz, D. A. (2001). Investigating the microenvironments of inhomogeneous soft materials with multiple particle tracking. *Physical Review E*, 64(6):061506.

van Eysden, C. A. and Sader, J. E. (2006). Resonant frequencies of a rectangular cantilever beam immersed in a fluid. *Journal of Applied Physics*, 100(11):114916.

van Neerven, J. (1992). *The Adjoint of a Semigroup of Linear Operators*. Number 1529 in Lecture Notes in Mathematics. Springer, Berlin Heidelberg.

von Neumann, J. (1932). Uber einen Satz von Herrn M. H. Stone. *Annals of Mathematics*, 33(3):567–573.

Weiss, S. (1999). Fluorescence spectroscopy of single biomolecules. *Science*, 283(5408):1676–1683.

Weiss, S. (2000). Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature Structural Biology*, 7:724–9.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73.

Whitaker, G. A., Golightly, A., Boys, R. J., and Sherlock, C. (2017). Improved bridge constructs for stochastic differential equations. *Statistics and Computing*, 27(4):885–900.

Whittle, P. (1953a). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 15(1):125–139.

Whittle, P. (1953b). Estimation and information in stationary time series. *Arkiv för Matematik*, 2(5):423–434.

Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 19(1):38–63.

Witt, A. and Malamud, B. D. (2013). Quantification of long-range persistence in geophysical time series: Conventional and benchmark-based improvement techniques. *Surveys in Geophysics*, 34(5):541–651.

Xiang, T.-x., Liu, F., and Grant, D. M. (1991). Generalized Langevin equations for molecular dynamics in solution. *The Journal of Chemical Physics*, 94(6):4463–4471.

Xie, X. S. and Lu, H. P. (1999). Single-molecule Enzymology. *Journal of Biological Chemistry*, 274(23):15967–15970.

Xie, X. S. and Trautman, J. K. (1998). Optical studies of single molecules at room temperature. *Annual Review of Physical Chemistry*, 49(1):441–480.

Xing, J. and Kim, K. S. (2011). Application of the projection operator formalism to non-Hamiltonian dynamics. *The Journal of Chemical Physics*, 134(4):044132.

Yang, C., He, Z., and Yu, W. (2009). Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 10(1):4.

Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T.-M., Rech, I., Cova, S., Xun, L., and Xie, X. S. (2003). Protein conformational dynamics probed by single-molecule electron transfer. *Science*, 302(5643):262–266.

Yosida, K. (1995). *Functional Analysis*. Springer-Verlag, Berlin, Heidelberg, reprint of the 6th 1980 edition.

Yu, H., Siewny, M. G., Edwards, D. T., Sanders, A. W., and Perkins, T. T. (2017). Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science*, 355(6328):945–950.

Zhang, L., Wang, L., Kao, Y. T., Qiu, W., Yang, Y., Okobiah, O., and Zhong, D. (2007). Mapping hydration dynamics around a protein surface. *Proceedings of the National Academy of Sciences*, 104(47):18461–18466.

Zhu, F. and Lysy, M. (2021). *realPSD: Robust and Efficient Calibration of Parametric Power Spectral Density Models*. https://github.com/mlysy/realPSD.

Zhu, Y., Dominy, J. M., and Venturi, D. (2018). On the estimation of the Mori-Zwanzig memory integral. *Journal of Mathematical Physics*, 59(10):103501.

Zhu, Y. and Lei, H. (2021). Effective Mori-Zwanzig equation for the reduced-order modeling of stochastic systems. *arXiv:2102.01377*.

Zwanzig, R. (1960). Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5):1338–1341.

Zwanzig, R. (1961). Memory effects in irreversible thermodynamics. *Physical Review*, 124(4):983.

Zwanzig, R. (1973). Nonlinear generalized Langevin equations. *Journal of Statistical Physics*, 9(3):215–220.

Zwanzig, R. (1980). Problems in nonlinear transport theory. In Garrido, L., editor, *Systems Far from Equilibrium*, pages 198–225, Berlin, Heidelberg. Springer.

Zwanzig, R. (2001). *Nonequilibrium Statistical Mechanics*. Oxford University Press, New York, NY.

APPENDICES

Appendix A

Appendix for Chapter 2

In this Appendix, there are three main sections. Section A.1 presents some basic concepts related to self-adjoint operators and then introduce the functional calculus as well as the spectral theorems. Section A.2 shows in detail some other examples of Hamiltonian systems. Section A.3 gives detailed proofs of all the results in the paper.

A.1 Self-Adjoint Operators

In this section, we present some important concepts related to self-adjoint operators that are crucial for readers to understand the mathematical tools applied in our construction of Hamiltonians and later proofs. These materials in this section are adapted from classic references Hille and Phillips (1957); Rudin (1973); Reed and Simon (1981); Yosida (1995); Engel and Nagel (2000); Lax (2002); Hall (2013). Some properties are given directly or stated with only succinct proof. For the remainder of this section, we will work with a complex separable Hilbert space $\mathscr{H}$, which can be finite or infinite.

Definition A.1.1 (Adjoint operator)**.** *Let $\mathcal{A}$ be a linear (possibly unbounded) operator defined on a dense subspace* $\mathrm{dom}(\mathcal{A}) \subseteq \mathscr{H}$*. The* adjoint of $\mathcal{A}$ *is an operator $\mathcal{A}^\dagger$ whose domain* $\mathrm{dom}(\mathcal{A}^\dagger)$ *consists of all vectors $v \in \mathscr{H}$ for which there is a vector $z \in \mathscr{H}$ such that*

$$\langle \mathcal{A}u, v\rangle = \langle u, z\rangle \tag{A.1}$$

holds for all $u \in \mathrm{dom}(\mathcal{A})$. For such $v, z \in \mathscr{H}$ we write $\mathcal{A}^\dagger v = z$.

The reason behind our focus on a subspace $\mathscr{D}$ is that if A is defined everywhere on $\mathscr{H}$ and satisfies (A.1), then it can be shown that A must be bounded due to the closed graph theorem. Thus, unbounded linear operators can only be defined on a subspace of $\mathscr{H}$.

Usually the subspace $\mathrm{dom}(\mathcal{A})$ is assumed to be dense. This assumption is motivated by the fact that the denseness of the domain is necessary and sufficient for the existence of the adjoint operator $\mathcal{A}^\dagger$. Since $u \mapsto \langle \mathcal{A}u, v\rangle$ is a continuous linear map on the domain $\mathrm{dom}(\mathcal{A})$ of $\mathcal{A}$ which can be extended to the whole space $\mathscr{H}$ by the Hahn-Banach theorem, the Riesz theorem tells us that there is a unique $w \in \mathscr{H}$ satisfying

$$\langle \mathcal{A}u, v\rangle = \langle u, w\rangle, \quad \forall u \in \mathrm{dom}(\mathcal{A})$$

where w is uniquely determined by v if and only if the linear map $u \mapsto \langle \mathcal{A}u, v\rangle$ is densely defined (which is equivalent to $\mathcal{A}$ being densely defined). Finally, we can set $\mathcal{A}^\dagger v = w$ to define $\mathcal{A}^\dagger$. The upshot is that as long as $\mathrm{dom}(\mathcal{A})$ is dense, for any given v, $\mathcal{A}^\dagger v$ is unique if exists.

Definition A.1.2. *For a linear operator $\mathcal{A}$ on $\mathscr{H}$, the set of values*

$$\sigma(\mathcal{A}) = \{z \in \mathbb{C} : (zI - \mathcal{A}) \text{ has a bounded inverse}\}$$

is called the spectrum of $\mathcal{A}$.

Definition A.1.3. *An operator $\mathcal{A}$ is called* self-adjoint *if* $\mathrm{dom}(\mathcal{A}^\dagger) = \mathrm{dom}(\mathcal{A})$ *and* $\mathcal{A}^\dagger = \mathcal{A}$, *and* skew-adjoint *if* $\mathrm{dom}(\mathcal{A}^\dagger) = \mathcal{A}^\dagger$ *and* $\mathcal{A}^\dagger = -\mathcal{A}$.

It turns out that if $\mathcal{A}$ is self-adjoint, then $\sigma(\mathcal{A}) \subseteq \mathbb{R}$ (Hall, 2013, Theorem 9.17). More generally, self-adjoint and skew-adjoint operators admit some degree of theoretical tractability via the so-called Spectral Theorem which we introduce momentarily. By way of motivation we first present a consequence of this result.

Theorem A.1.1 (Functional Calculus, Hall, 2013, Definition 10.5)**.** *Let $\mathcal{A}$ be a self-adjoint operator on $\mathscr{H}$. Then for any measurable function $f : \sigma(\mathcal{A}) \to \mathbb{C}$, there exists a unique operator $f(\mathcal{A})$ such that for any $x \in \mathscr{H}$, there exists a unique finite measure μ_x (independent of f) with $\mu_x(\mathbb{R} \setminus (\sigma(\mathcal{A})) = 0$ such that*

$$\langle f(\mathcal{A})x, x\rangle = \int_{\mathbb{R}} f(\lambda)\mathrm{d}\mu_x(\lambda), \tag{A.2}$$

with

$$\mathrm{dom}(f(\mathcal{A})) = \left\{x \in \mathscr{H} : \int_{\mathbb{R}} |f(\lambda)|\mu_x(\lambda) < \infty\right\}.$$

A few consequences of the functional calculus are:

(i) Addition and multiplication operations on regular functions extend directly to their operator versions, i.e., $(f+g)(\mathcal{A}) = f(\mathcal{A}) + g(\mathcal{A})$ and $(fg)(\mathcal{A}) = f(A)g(A)$.

(ii) For given $f(\mathcal{A})$, its adjoint $f(\mathcal{A})^\dagger$ is such that

$$\langle f(\mathcal{A})^\dagger x, x\rangle = \langle x, f(\mathcal{A})x\rangle = \overline{\langle f(\mathcal{A})x, x\rangle}$$

for any $x \in \mathrm{dom}(f(\mathcal{A}))$. On the other hand, (A.2) implies that

$$\overline{\langle f(\mathcal{A})x, x\rangle} = \int_{\mathbb{R}} \overline{f(\lambda)} \mathrm{d}\mu_x(\lambda),$$

such that $f(\mathcal{A})^\dagger = \overline{f}(\mathcal{A})$.

(iii) The operator $\mathcal{L} = i\mathcal{A}$ is skew-adjoint, and the operator $e^{t\mathcal{L}} = e^{it\mathcal{A}}$ defined by functional calculus precisely agrees with the corresponding C_0-semigroup defined by Definition 2.2.1.

A.1.1 Spectral Theorem: Direct Integral Form

For our purposes we shall require the spectral theorem in two equivalent forms. The first of these hinges on the concept of a *direct integral* Hilbert space (Hall, 2013, Definition 7.18).

Suppose we have a partition of the real line $\mathbb{R} = \amalg_{n=0}^{\infty} S_n$, and consider the family of Hilbert spaces $\{\mathscr{H}_\lambda, \lambda \in \mathbb{R}\}$ such that

$$\mathscr{H}_\lambda = \begin{cases} \mathbb{C}^n, & \lambda \in S_n, n > 0 \\ \ell^2, & \lambda \in S_0, n = 0, \end{cases}$$

where ℓ^2 is the Hilbert space of square-summable (complex) infinite sequences. Let $\langle \cdot, \cdot \rangle_\lambda$ and $|\cdot|_\lambda$ denote respectively the inner product and norm on $\mathscr{H}_\lambda$.

Now, let μ be a σ-finite countably additive measure on $\mathbb{R}$, and consider functions $s(\cdot)$ defined on $\mathbb{R}$ such that:

(i) $s(\lambda) \in \mathscr{H}_\lambda$.

(ii) $s(\lambda)$ is square-integrable with respect to μ:

$$\int_{\mathbb{R}} |s(\lambda)|_\lambda^2 \mathrm{d}\mu(\lambda) = \sum_{n=0}^{\infty} \int_{S_n} |s(\lambda)|_n^2 \mathrm{d}\mu(\lambda) < \infty.$$

The set of such functions forms a (complex, separable, finite or infinite) Hilbert space called a *direct integral* which we denote by $\mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n, \mu\}$, on which the inner product is defined as

$$\langle s, v \rangle = \int_{\mathbb{R}} \langle s(\lambda), v(\lambda) \rangle_\lambda \mathrm{d}\mu(\lambda) = \sum_{n=0}^{\infty} \int_{S_n} \langle s(\lambda), v(\lambda) \rangle_n \mathrm{d}\mu(\lambda).$$

Direct integrals provide the fundamental tools to define the spectral decomposition of a self-adjoint operator $\mathcal{A}$:

Theorem A.1.2 (Hall, 2013, Theorem 10.9)**.** *Let $\mathcal{A}$ be a self-adjoint operator on $\mathscr{H}$. Then $\mathscr{H}$ is isomorphic to a direct integral $\mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n, \mu\}$ with $\mu(\mathbb{R} \setminus \sigma(\mathcal{A})) = 0$ on which $\mathcal{A}$ is a multiplication operator:*

$$\mathcal{A} : s(\lambda) \to \lambda s(\lambda).$$

Moreover, the functional calculus on the direct integral is given by

$$f(\mathcal{A}) : s(\lambda) \to f(\lambda) s(\lambda),$$

with

$$\mathrm{dom}(f(\mathcal{A})) = \left\{ s(\lambda) : \int_{\mathbb{R}} |f(\lambda) s(\lambda)|_\lambda^2 \mathrm{d}\mu(\lambda) < \infty \right\}$$

The direct integral $\mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n, \mu\}$ in Theorem A.1.2 is not unique. However, if $\mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n^{(1)}, \mu_1\}$ and $\mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n^{(2)}, \mu_2\}$ are both candidates for the result of Theorem A.1.2, then $S_n^{(1)} = S_n^{(2)}$ and μ_1 and μ_2 are absolutely continuous with respect to each other (Hall, 2013, Proposition 7.22).

A.1.2 Spectral Theorem: Projection-Valued Measure Form

The second form of the spectral theorem requires the following definition.

Definition A.1.4. *Let $\mathscr{H}$ be a complex separable Hilbert space and $\mathscr{B}(\mathbb{R})$ the Borel sets on the real line. A function μ from $\mathscr{B}(\mathbb{R})$ to the bounded linear operators on $\mathscr{H}$ is called a* projection-valued measure *if it satisfies the following:*

(i) *For each $E \in \mathscr{B}(\mathbb{R})$, $\mu(E)$ is an orthogonal projection on a closed subspace $V_E \subseteq \mathscr{H}$. That is, we have $\mu(E)x = x$ for $x \in V_E$ and $\mu(E)x = 0$ for $x \in V_E^{\perp}$, the orthogonal complement of V_E.*

(ii) *$\mu(\emptyset) = 0$ and $\mu(\mathbb{R}) = I$.*

(iii) *For disjoint sets $E_1, E_2, \ldots \in \mathscr{B}(\mathbb{R})$, for all $v \in \mathscr{H}_C$ we have*

$$\mu\left(\cup_{j=1}^{\infty} E_j\right) v = \sum_{j=1}^{\infty} \mu(E_j) v.$$

(iv) *For all $E, F \in \mathscr{B}(\mathbb{R})$, we have $\mu(E \cap F) = \mu(E)\mu(F)$.*

If μ is a projection-valued measure and $\phi \in \mathscr{H}$, then

$$\mu_{\phi,\phi} : \mathscr{B}(\mathbb{R}) \to \mathbb{R}^+, \qquad \mu_{\phi,\phi}(E) = \langle \phi, \mu(E)\phi \rangle$$

is a positive finite measure (Hall, 2013, Proposition 7.11). This can be generalized to the following lemma:

Lemma A.1.1. *If μ is a projection-valued measure and $\psi, \phi \in \mathscr{H}$, then*

$$\mu_{\psi,\phi} : \mathscr{B}(\mathbb{R}) \to \mathbb{C}, \qquad \mu_{\psi,\phi}(E) = \langle \psi, \mu(E)\phi \rangle$$

is a complex measure.

Proof. By direct calculation note that

$$\mu_{\psi,\phi} = \frac{1}{2}\left[\mu_{\psi+\phi,\psi+\phi} - (\mu_{\psi,\psi} + \mu_{\phi,\phi})\right] - \frac{i}{2}\left[\mu_{\psi+i\phi,\psi+i\phi} - (\mu_{\psi,\psi} + \mu_{i\phi,i\phi})\right],$$

where each of the terms above is a positive finite measure, such that the result follows from the definition of a complex measure. □

Thus for measurable functions $f : \mathbb{R} \to \mathbb{C}$ we can define operators $\mu(f) = \int_{\mathbb{R}} f(\lambda) \mathrm{d}\mu(\lambda)$ on $\mathscr{H}$ via

$$\langle \psi, \mu(f)\phi \rangle = \int_{\mathbb{R}} f(\lambda) \mathrm{d}\mu_{\psi,\phi}(\lambda), \qquad \psi, \phi \in \mathscr{H}.$$

We are now ready to present the spectral theorem in terms of projection-valued measures:

Theorem A.1.3 (Hall, 2013, Theorem 10.4)**.** *Let $\mathcal{A}$ be a self-adjoint operator on $\mathscr{H}$. Then there exists a unique projection-valued measure $\mu_{\mathcal{A}}$ with $\mu_{\mathcal{A}}(\mathbb{R} \setminus \sigma(\mathcal{A})) = 0$ such that*

$$\mathcal{A} = \int_{\mathbb{R}} \lambda \mathrm{d}\mu_{\mathcal{A}}(\lambda).$$

Moreover, the functional calculus is obtained via

$$f(\mathcal{A}) = \int_{\sigma(\mathcal{A})} f(\lambda) \mathrm{d}\mu_{\mathcal{A}}(\lambda),$$

with

$$\mathrm{dom}(f(\mathcal{A})) = \left\{ \phi \in \mathscr{H} : \int_{\sigma(\mathcal{A})} |f(\lambda)| \mathrm{d}\mu_{\mathcal{A}_{\phi,\phi}} < \infty \right\}.$$

A.2 ACF Representation with Hamiltonians of Interest

For the following examples we introduce the notion of Sobolev spaces.

Definition A.2.1. *Let $u(x)$ and $v(x)$ be locally integrable function on $\mathbb{R}$. Then $v(x)$ is called the* k-th weak derivative *of $u(x)$ if*

$$\int u(x) \frac{\mathrm{d}^k}{\mathrm{d}x^k} \phi(x) \mathrm{d}x = (-1)^k \int v(x) \phi(x) \mathrm{d}x$$

for all smooth functions $\phi(x)$ with compact support.

Weak derivatives generalize regular derivatives in that the two coincide when the latter exists, and the former also obeys the formula for integration by parts.

Definition A.2.2. *The* Sobolev space $W^{k,p}$ *consists of the subset of functions $f \in L^p$ such that f and its weak derivatives of order up to k are all in L^p.*

The Sobolev space $W^{k,p}$ is a Banach space with respect to the norm

$$\|f\|_{k,p} = \sum_{i=0}^{k} \left(\int_{\mathbb{R}} |f^{(i)}(x)|^p \mathrm{d}x \right)^{1/p}.$$

The Sobolev spaces $H^k = W^{k,2}$ are Hilbert spaces with inner product

$$\langle u, v \rangle = \sum_{i=0}^{k} \int_{\mathbb{R}} u^{(i)}(x) v^{(i)}(x) \mathrm{d}x.$$

A.2.1 Wave Equation Hamiltonian

Let us consider the heat-bath Hamiltonian given in Pavliotis (2014). The infinite-dimensional system is characterized by the one-dimensional wave equation $\partial_t^2 \mathsf{q} = \partial_x^2 \mathsf{q}$ where $\mathsf{q} = \mathsf{q}(x,t)$ is the position and $\mathsf{p} = \mathsf{p}(x,t) = \partial_t \mathsf{q}$ is the conjugate momentum. The Hamiltonian is given by

$$\mathcal{H}(\Gamma) = \frac{1}{2} \int_{\mathbb{R}} \left(|\mathsf{p}|^2 + |\partial_x \mathsf{q}|^2 \right) \mathrm{d}x.$$

The resulting Hamiltonian equations of motion are just the wave equations,

$$\partial_t \mathsf{q} = \frac{\delta \mathcal{H}}{\delta \mathsf{p}} = \frac{\partial \mathcal{H}}{\partial \mathsf{p}} - \frac{\partial}{\partial_x} \frac{\partial \mathcal{H}}{\partial (\partial_x \mathsf{p})} = \mathsf{p}, \tag{A.3}$$

$$\partial_t \mathsf{p} = -\frac{\delta \mathcal{H}}{\delta \mathsf{q}} = -\frac{\partial \mathcal{H}}{\partial \mathsf{q}} + \frac{\partial}{\partial_x} \frac{\partial \mathcal{H}}{\partial (\partial_x \mathsf{q})} = \partial_x^2 \mathsf{q}. \tag{A.4}$$

For $\Gamma = (\mathsf{q}, \mathsf{p})$, the Hamiltonian can be written as

$$\mathcal{H}(\Gamma) = \frac{1}{2} \|\Gamma\|^2$$

The corresponding Hilbert space $\mathscr{H}_E$ is equipped with an inner product as follows

$$\langle \Gamma_1, \Gamma_2 \rangle = \int_{\mathbb{R}} \left(\partial_x \mathsf{q}_1 \overline{\partial_x \mathsf{q}_2} + \mathsf{p}_1 \overline{\mathsf{p}_2} \right) \mathrm{d}x.$$

Define an operator $\mathscr{A}$ as

$$\mathscr{A} = \begin{bmatrix} 0 & 1 \\ \partial_x^2 & 0 \end{bmatrix}.$$

Then the Hamiltonian equations of motion (A.3) and (A.4) can be expressed as

$$\partial_t \Gamma = \mathscr{A} \Gamma$$

for $\Gamma = (\mathsf{q}, \mathsf{p})$. It is an ODE in terms of t which has the solution $\Gamma_t = e^{\mathscr{A} t} \Gamma_0$.

Consider an observable $X_t = \langle f, \Gamma_t \rangle$ in the dual space of $\mathscr{H}_E$ where $f = (\alpha, \beta) \in \mathscr{H}_E$. Since $\mathscr{A}$ is skew adjoint, we have

$$X_t = \langle f, \Gamma_t \rangle = \langle e^{-\mathscr{A}t} f, \Gamma_0 \rangle$$

where $e^{-t\mathscr{A}}$ is a strong continuous one-parameter unitary group. Thus,

$$\langle X_u, X_{u+t} \rangle_{\mathscr{H}'} = \langle e^{-\mathscr{A}u} f, e^{-\mathscr{A}(u+t)} f \rangle = \langle e^{\mathscr{A}t} f, f \rangle$$

is in fact an ACF of a Gaussian process. In the following subsection, we can explicitly establish the construction.

The d'Alembert solution is not as useful as the Fourier solution. Instead, we can turn to the Fourier solution of the wave equation

$$\frac{\partial^2}{\partial x^2} \mathsf{q}(x,t) = \frac{\partial^2}{\partial t^2} \mathsf{q}(x,t). \tag{A.5}$$

Let $\mathsf{q}(x,0) = f(x)$ and $\frac{\partial}{\partial t}\mathsf{q}(x,0) = g(x)$, by taking the Fourier transform (w.r.t. x) of (A.5) and recalling that differentiation w.r.t. t commutes with the Fourier transform, we obtain

$$-4\pi^2 |\xi|^2 \hat{\mathsf{q}}(\xi,t) = \frac{\partial^2}{\partial t^2} \hat{\mathsf{q}}(\xi,t) \tag{A.6}$$

where $\hat{\mathsf{q}}(\xi,t) = \mathcal{F}\{\mathsf{q}(x,t)\}$. For each ξ, (A.6) is an ODE in t whose solution is given by

$$\hat{\mathsf{q}}(\xi,t) = C_1(\xi)\cos(2\pi|\xi|t) + C_2(\xi)\sin(2\pi|\xi|t) \tag{A.7}$$

where $C_1(\xi)$ and $C_2(\xi)$ are constants determined by the initial conditions

$$\mathcal{F}\{\mathsf{q}(x,0)\} = \hat{\mathsf{q}}(\xi,0) = \hat{f}(\xi) \qquad \text{and} \qquad \mathcal{F}\left\{\frac{\partial}{\partial t}\mathsf{q}(x,0)\right\} = \frac{\partial}{\partial t}\hat{\mathsf{q}}(\xi,0) = \hat{g}(\xi)$$

By setting $t = 0$ in (A.7), we find

$$C_1(\xi) = \hat{f}(\xi) \qquad \text{and} \qquad 2\pi|\xi| C_2(\xi) = \hat{g}(\xi).$$

Thus, the solution to (A.5) is given by

$$\mathsf{q}(x,t) = \int_{\mathbb{R}} \left[\hat{f}(\xi)\cos(2\pi|\xi|t) + \hat{g}(\xi)\frac{\sin(2\pi|\xi|t)}{2\pi|\xi|}\right] e^{2\pi i x\xi} \mathrm{d}\xi.$$

The covariance of a stationary random variable can be written as (suppose here $f = (\mathsf{q}, \mathsf{p}) \in \mathscr{H}_E$)

$$\begin{aligned}\langle X_u, X_{u+t}\rangle &= \langle e^{-\mathscr{A}u} f, e^{-\mathscr{A}(u+t)} f\rangle = \langle e^{\mathscr{A}t} f, f\rangle \\ &= \frac{1}{2}\int_{\mathbb{R}} \left(\partial_x \mathsf{q}(x,t)\overline{\partial_x \mathsf{q}(x,0)} + \mathsf{p}(x,t)\overline{\mathsf{p}(x,0)}\right) \mathrm{d}x.\end{aligned}$$

By the Plancherel's theorem, we have

$$\begin{aligned}\int \partial_x \mathsf{q}(x,t)\overline{\partial_x \mathsf{q}(x,0)}\mathrm{d}x &= \int_{\mathbb{R}} \left[2\pi|\xi|\hat{f}(\xi)\cos(2\pi|\xi|t) + \hat{g}(\xi)\sin(2\pi|\xi|t)\right] 2\pi|\xi|\overline{\hat{f}(\xi)}\mathrm{d}\xi \\ \int \mathsf{p}(x,t)\overline{\mathsf{p}(x,0)}\mathrm{d}x &= \int_{\mathbb{R}} \left[-2\pi|\xi|\hat{f}(\xi)\sin(2\pi|\xi|t) + \hat{g}(\xi)\cos(2\pi|\xi|t)\right] \overline{\hat{g}(\xi)}\mathrm{d}\xi\end{aligned}$$

Let $a = 2\pi|\xi|\hat{f}(\xi)$, $b = \hat{g}(\xi)$ and $\theta = 2\pi|\xi|$, we have

$$\langle X_u, X_{u+t}\rangle_{\mathscr{H}'} = \frac{1}{2}\int_{\mathbb{R}} (a^2+b^2)\cos(\theta t)\mathrm{d}\xi = \frac{1}{2}\int_{\mathbb{R}} \left(4\pi^2\xi^2|\hat{f}(\xi)|^2 + |\hat{g}(\xi)|^2\right)\cos(2\pi|\xi|t)\mathrm{d}\xi.$$

Thus, we can connect the inner product with the autocovariance (or equivalently the PSD) of any given continuous and stationary process as we previously did.

A.2.2 The Klein–Gordon Hamiltonian

Let us consider the Klein–Gordon Hamiltonian (e.g. Goldstein et al., 2002, Section 13.6)

$$\mathcal{H}(\mathsf{q}, \mathsf{p}) = \frac{1}{2}\int_{\mathbb{R}} \left(\mathsf{p}^2 + (\nabla \mathsf{q})^2 + m^2\mathsf{q}^2\right) \mathrm{d}x$$

where $\mathsf{q} = \mathsf{q}(x,t)$, $\mathsf{p} = \mathsf{p}(x,t)$, $m \in \mathbb{R}$ and the corresponding inner product can be defined as

$$\langle \Gamma_1, \Gamma_2\rangle = \int_{\mathbb{R}} \left(\mathsf{p}\bar{\mathsf{p}} + (\nabla \mathsf{q})\cdot(\overline{\nabla \mathsf{q}}) + m^2\mathsf{q}\bar{\mathsf{q}}\right) \mathrm{d}x, \qquad \Gamma_i = (\mathsf{q}_i, \mathsf{p}_i),\ i = 1, 2.$$

The Hamiltonian equations of motion can be calculated using Poisson brackets (or functional derivatives)

$$\dot{\mathsf{q}} = \partial_t \mathsf{q} = \frac{\delta \mathcal{H}}{\delta \mathsf{p}} = \{\mathsf{q}, \mathcal{H}\} = \mathsf{p}$$
$$\dot{\mathsf{p}} = \partial_t \mathsf{p} = -\frac{\delta \mathcal{H}}{\delta \mathsf{q}} = \{\mathsf{p}, \mathcal{H}\} = \nabla^2 \mathsf{q} - m^2 \mathsf{q}$$

which gives us the famous Klein–Gordan equation

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + m^2\right) \mathsf{q} = 0.$$

Now, let us apply the Fourier transform trick again. We have the Klein–Gordan equation in the frequency domain as follows

$$\frac{\partial^2}{\partial t^2} \hat{\mathsf{q}}(\xi, t) = -\left(4\pi^2 |\xi|^2 + m^2\right) \hat{\mathsf{q}}(\xi, t)$$

which is a second-order homogeneous ODE in t for each fixed ξ with the solution written as

$$\hat{\mathsf{q}}(\xi, t) = C_1(\xi) \cos(\theta t) + C_2(\xi) \sin(\theta t)$$

where $\theta = \sqrt{4\pi^2 |\xi|^2 + m^2}$ and $C_1(\xi) = \hat{f}(\xi)$, $C_2(\xi) = \hat{g}(\xi)/\theta$ determined by the initial conditions

$$\hat{\mathsf{q}}(\xi, 0) = \hat{f}(\xi) \qquad \text{and} \qquad \frac{\partial}{\partial t} \hat{\mathsf{q}}(\xi, 0) = \hat{g}(\xi)$$

where $\hat{\mathsf{q}}(\xi, t) = \mathcal{F}\{\mathsf{q}(x, t)\}$. Therefore, the solution to the Klein–Gordan equation in the time domain can be written as the inverse Fourier transform

$$\mathsf{q}(x, t) = \int_{\mathbb{R}} \left[\hat{f}(\xi) \cos(\theta t) + \hat{g}(\xi) \frac{\sin(\theta t)}{\theta}\right] e^{2\pi i x \xi} \mathrm{d}\xi.$$

The inner product between two observables can be written as

$$\begin{aligned} \langle X_u, X_{u+t} \rangle_{\mathscr{H}'} &= \langle e^{-\mathscr{A} u} f, e^{-\mathscr{A}(u+t)} f \rangle = \langle e^{\mathscr{A} t} f, f \rangle \qquad f = (\mathsf{q}, \mathsf{p}) \in \mathscr{H} \\ &= \frac{1}{2} \int_{\mathbb{R}} \left(\mathsf{p}(x, t) \overline{\mathsf{p}(x, 0)} + \nabla \mathsf{q}(x, t) \overline{\nabla \mathsf{q}(x, 0)} + m^2 \mathsf{q}(x, t) \overline{\mathsf{q}(x, 0)}\right) \mathrm{d}x. \end{aligned}$$

By the Plancherel's theorem, we have

$$\int_{\mathbb{R}} \mathsf{p}(x,t)\overline{\mathsf{p}(x,0)}\mathrm{d}x = \int_{\mathbb{R}} \left(-\theta\hat{f}(\xi)\sin(\theta t) + \hat{g}(\xi)\cos(\theta t)\right)\overline{\hat{g}(\xi)}\mathrm{d}\xi$$
$$\int_{\mathbb{R}} \nabla\mathsf{q}(x,t)\overline{\nabla\mathsf{q}(x,0)}\mathrm{d}x = 4\pi^2|\xi|^2\int_{\mathbb{R}} \left(\hat{f}(\xi)\cos(\theta t) + \hat{g}(\xi)\frac{\sin(\theta t)}{\theta}\right)\overline{\hat{f}(\xi)}\mathrm{d}\xi$$
$$\int_{\mathbb{R}} m^2\mathsf{q}(x,t)\overline{\mathsf{q}(x,0)}\mathrm{d}x = m^2\int_{\mathbb{R}} \left(\hat{f}(\xi)\cos(\theta t) + \hat{g}(\xi)\frac{\sin(\theta t)}{\theta}\right)\overline{\hat{f}(\xi)}\mathrm{d}\xi$$

Plug the above results into the inner product expression and notice that $\theta = \sqrt{4\pi^2|\xi|^2 + m^2}$, we get

$$\langle X_u, X_{u+t}\rangle_{\mathscr{H}'} = \frac{1}{2}\int_{\mathbb{R}} \left[\left(4\pi^2|\xi|^2 + m^2\right)|\hat{f}(\xi)|^2 + |\hat{g}(\xi)|^2\right]\cos(\theta t)\,\mathrm{d}\xi$$

which corresponds to the Fourier transform of the PSD of a stationary process.

A.3 Proofs of Technical Results

A.3.1 Proof of Proposition 2.2.1

Proof. First, let $\Phi(t) = e^{t\mathcal{L}}$, we have

$$|\mathcal{C}(t) - \mathcal{C}(0)| = |\langle X_0, (\Phi(t) - \mathcal{I})X_0\rangle| \leq \|\Phi(t) - \mathcal{I})\|\|X_0\|^2 \to 0$$

as $t \to 0$ since $\|X_0\|$ is just a constant and $\Phi(t)$ is strongly continuous at $t = 0$. Then by using the group property, we have the fact that $\Phi(-t) = \Phi(t)^{-1} = \Phi(t)^\dagger$, we can show the positive definiteness,

$$\begin{aligned}\sum \mathcal{C}(t_i - t_j)z_i\bar{z}_j &= \sum\langle\Phi(t_i - t_j)f, f\rangle z_i\bar{z}_j = \sum\langle\Phi(t_j)^{-1}\Phi(t_i)f, f\rangle z_i\bar{z}_j \\ &= \sum\langle\Phi(t_i)f, \Phi(t_j)f\rangle z_i\bar{z}_j = \langle\sum z_i\Phi(t_i)f, z_j\Phi(t_j)f\rangle \\ &= \|\sum z_i\Phi(t_i)f\|^2 \geq 0.\end{aligned}$$

Finally, it is indeed (conjugate) symmetric

$$\overline{\mathcal{C}(t)} = \overline{\langle\Phi(-t)X_0, X_0\rangle} = \langle X_0, \Phi(-t)X_0\rangle = \mathcal{C}(-t).$$

But since all the ACFs we are discussing are essentially real-valued functions, this just lead to the time symmetry of an ACF. □

A.3.2 Proof of Theorem 2.3.1

Proof. Since $\mathcal{L}$ is a skew-adjoint operator on $\mathscr{H}_C$, $i\mathcal{L}$ is self-adjoint. Using the direct integral version of the spectral theorem (Theorem A.1.2), $\mathscr{H}_C$ is isometrically isomorphic to a direct integral $\mathscr{H}^{\oplus} = \mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n, \mu\}$ under which we have

$$f(\mathcal{L}) = f(-i \cdot (i\mathcal{L})) : s(\lambda) \mapsto f(-i\lambda).s(\lambda)$$

Thus, for $X \in \mathscr{H}'$ we have

$$\langle e^{-t\mathcal{L}} X, X\rangle = \int_{\mathbb{R}} e^{it\lambda} \langle s(\lambda), s(\lambda)\rangle_\lambda \mathrm{d}\mu(\lambda) = \int_{\mathbb{R}} e^{it\lambda} |s(\lambda)|_\lambda^2 \mathrm{d}\mu(\lambda),$$

where $s(\lambda)$ is the element of $\mathscr{H}^{\oplus}$ identified with $X \in \mathscr{H}'$ via Theorem A.1.2. Let us now consider the measure ν for which the Radon–Nikodym derivative with respect to μ is given by $\mathrm{d}\nu(\lambda)/\mathrm{d}\mu(\lambda) = |s(\lambda)|_\lambda^2$. Due to the square-integrability assumption on $s(\lambda)$, ν is a finite positive measure for which we have

$$\langle e^{-t\mathcal{L}} X, X\rangle = \int_{\mathbb{R}} e^{it\lambda} \mathrm{d}\nu(\lambda),$$

which shows that $\langle e^{-t\mathcal{L}} X, X\rangle$ is an ACF.

Conversely, suppose that $\mathcal{C}(t)$ is an ACF with spectral decomposition

$$\mathcal{C}(t) = \int_{\mathbb{R}} e^{it\lambda} \mathrm{d}\nu(\lambda),$$

and that ν is absolutely continuous with respect to μ, such that we have the Radon–Nikodym derivative $\mathrm{d}\nu(\lambda)/\mathrm{d}\mu(\lambda) = a(\lambda) \geq 0$. Now consider the function

$$s(\lambda) = (a(\lambda)^{1/2}, 0, \ldots).$$

It then follows that

$$\mathcal{C}(t) = \int_{\mathbb{R}} e^{it\lambda} |s(\lambda)|_\lambda^2 \mathrm{d}\mu(\lambda) = \int_{\mathbb{R}} e^{it\lambda} a(\lambda) \mathrm{d}\mu(\lambda),$$

which completes the proof.

□

A.3.3 Proof of Theorem 2.4.1

We first introduce the characterization of continuous stationary processes introduced by Itô (1954).

Definition A.3.1. *Let μ be a non-negative finite measure on $\mathscr{B}(\mathbb{R})$, the Borel sets on the real line. A* random measure *$\mathscr{M}$ is a function from $\mathscr{B}(\mathbb{R})$ to the complex-valued random variables such that for any $E, F \in \mathscr{B}(\mathbb{R})$ we have*

$$\operatorname{cov}(\mathscr{M}(E), \mathscr{M}(F)) = \mu(E \cap F).$$

As a consequence of Definition A.3.1, we also get

$$\begin{aligned} &\|\mathscr{M}(E)\|^2 = \mu(E), \quad E \in \mathscr{B}(\mathbb{R}), \\ &\quad \mathscr{M}(E_1) \perp \mathscr{M}(E_2) \quad \text{if } E_1 \cap E_2 = \emptyset, \quad E_1, E_2 \in \mathscr{B}(\mathbb{R}), \\ &\mathscr{M}(\sum_{n=1}^{\infty} E_n) = \sum_{n=0}^{\infty} \mathscr{M}(E_n) \quad \text{for disjoint sets } E_1, E_2, \ldots \in \mathscr{B}(\mathbb{R}). \end{aligned}$$

Thus for $f \in L^2_\mu(\mathbb{R})$ we can define random variables of the form

$$\mathscr{M}(f) = \int_{\mathbb{R}} f(\lambda) \mathrm{d}\mathscr{M}(\lambda)$$

by Lebesgue integration. That is, for simple functions $f(x) = \sum_{i=1}^n c_i \mathbf{1}(x \in E_i)$ we have

$$\mathscr{M}(f) = \sum_{i=1}^{n} c_i \mathscr{M}(E_i),$$

and for arbitrary $f \in L^2_\mu(\mathbb{R})$, $\mathscr{M}(f)$ is the mean square limit of $\mathscr{M}(f_n)$, where the f_n are simple functions converging in mean square to f.

As a consequence, we have

$$\operatorname{cov}(\mathscr{M}(f_1), \mathscr{M}(f_2)) = \int_{\mathbb{R}} f_1(\lambda) \overline{f_2(\lambda)} \mathrm{d}\mu(\lambda).$$

Ito's characterization of continuous stationary processes is then as follows:

Theorem A.3.1 (Itô, 1954, Theorem 4.1)**.** *Let X_t be a continuous stationary process with mean zero and ACF*

$$\operatorname{cov}(X_t, X_0) = \int e^{it\lambda} \mathrm{d}\mu(\lambda). \tag{A.8}$$

Then there is a unique $\mathscr{M}(\lambda)$ with measure μ such that

$$X_t = \int_{\mathbb{R}} e^{it\lambda} \mathrm{d}\mathscr{M}(\lambda). \tag{A.9}$$

Conversely, any random measure $\mathscr{M}(\lambda)$ with respect to μ defines a continuous stationary process X_t with ACF given by (A.8)*.*

This motivates us to construct such a random measure in the context of Hamiltonian systems as given in Section 2.2. To realize this, we use the projection-valued measure form of the spectral theorem (Theorem A.1.3) to obtain the unique projection-valued measure $\mu_{\mathcal{L}}$ for which the functional calculus is given by

$$f(\mathcal{L}) = \int_{\mathbb{R}} f(-i\lambda) \mathrm{d}\mu_{\mathcal{L}}(\lambda).$$

(Note that $\mu_{\mathcal{L}}$ is actually $\mu_{i\mathcal{L}}$ in the statement of Theorem A.1.3 since it is $i\mathcal{L}$ which is self-adjoint.)

Lemma A.3.1. *For $\psi, \phi \in \mathscr{H}$ the function $\mu_{\psi,\phi} : \mathscr{B}(\mathbb{R}) \mapsto \mathbb{C}$ defined by*

$$\mu_{\psi,\phi}(E) = \langle \mu_{\mathcal{L}}(E)\psi, \phi \rangle.$$

is a complex measure, and for a measurable function f we have

$$\langle \int_{\mathbb{R}} f(\lambda) \mathrm{d}\mu_{\mathcal{L}}(\lambda)\psi, \phi \rangle = \int_{\mathbb{R}} f(\lambda) \mathrm{d}\mu_{\psi,\phi}(\lambda).$$

Proof. When $\phi = \psi$ the result is given by Hall (2013), Proposition 7.11. Otherwise, a direct calculation shows that

$$\mu_{\psi,\phi} = \frac{1}{2}\left[\mu_{\psi+\phi,\psi+\phi} - \mu_{\psi,\psi} - \mu_{\phi,\phi}\right] - \frac{i}{2}\left[\mu_{\psi+i\phi,\psi+i\phi} - \mu_{\psi,\psi} - \mu_{i\phi,i\phi}\right],$$

and a similar calculation holds for $\langle \int_{\mathbb{R}} f(\lambda) \mathrm{d}\mu_{\mathcal{L}}(\lambda)\psi, \phi \rangle$. □

Now consider the Hamiltonian system $\{\mathscr{H}, \omega(\cdot,\cdot), \mathcal{L}\}$ embedded in the complex Hilbert space $\mathscr{H}_C$ on which $\mathcal{L}$ is skew-adjoint. For the remainder of the proof we shall only refer to $\mathscr{H}_C$, and thus for easy of notation let us denote it by $\mathscr{H}$. Thus, let $X \in \mathscr{H}'$ be an observable of the system. Let $\{X_t, t \geq 0\}$ denote the trajectory of X. Then

$$\begin{aligned}\langle X_t, X_0\rangle = \langle e^{-t\mathcal{L}}X, X\rangle &= \langle \int e^{it\lambda} \mathrm{d}\mu_{\mathcal{L}}(\lambda) X, X\rangle \\ &= \int e^{it\lambda}\mathrm{d}\mu_{X,X}(\lambda)\end{aligned}$$

and similarly

$$X_t(\Gamma) = \int e^{it\lambda}\mathrm{d}\mu_{X,\Gamma}(\lambda).$$

Given a random measure $\mathscr{M}$ on $\mu_{X,X}$, the idea now is to construct a probability distribution $p(\Gamma)$ for Γ in such a way that $\mu_{X,\Gamma} = \mathscr{M}$. In this manner, $\{X_t(\Gamma) : t \geq 0, \Gamma \sim p(\Gamma)\}$ is the (mean-zero) continuous stationary process characterized by $\mathscr{M}$ and $\mu = \mu_{X,X}$ in (A.8) and (A.9) of Theorem A.3.1.

Theorem A.3.2. *Let $X \in \mathscr{H}'$ and $\mathscr{M}$ be a random measure on $\mu_{X,X}$. Then there exists a vector space $\mathscr{H}^* \supsetneq \mathscr{H}$, a bilinear map $\langle\cdot,\cdot\rangle^* : \mathscr{H} \times \mathscr{H}^*$, and a probability distribution $p(\Gamma)$ on $\mathscr{H}^*$ with the following properties:*

(i) The bilinear form $\langle\cdot,\cdot\rangle^*$ is an extension of the inner product $\langle\cdot,\cdot\rangle$ on $\mathscr{H}$, i.e., for any $\phi, \Gamma \in \mathscr{H}$ we have $\langle\phi,\Gamma\rangle^* = \langle\phi,\Gamma\rangle$.

(ii) For any $\phi \in \mathscr{H}$ the random variable

$$Z(\phi) = Z(\phi;\Gamma) = \langle\phi,\Gamma\rangle^*, \qquad \Gamma \sim p(\Gamma)$$

is well defined.

(iii) For $E \in \mathscr{B}(\mathbb{R})$,

$$Z(\mu_{\mathcal{L}}(E)X) = \mathscr{M}(E).$$

The proof closely follows the construction of Gaussian measures on infinite-dimensional separable Hilbert spaces (Gross, 1967). As therein, the probability distribution $p(\Gamma)$ is such that $\Pr(\Gamma \in \mathscr{H}) = 0$. However, the extension of the inner product $\langle\cdot,\cdot\rangle$ from $\mathscr{H} \times \mathscr{H}$ to the bilinear operator $\langle\cdot,\cdot\rangle^*$ on $\mathscr{H}^* \times \mathscr{H}$ is all that is required to construct the stochastic

process X_t, or more generally, the stochastic process $\{Z(\phi) : \phi \in \mathscr{H}\}$. Finally, note that for $\Gamma \in \mathscr{H}$,

$$Z(\mu_{\mathcal{L}}(E)X;\Gamma) = \langle \mu_{\mathcal{L}}(E)X, \Gamma\rangle^* = \langle \mu_{\mathcal{L}}(E)X, \Gamma\rangle = \mu_{X,\Gamma}(E),$$

which completes the proof of Theorem 2.4.1.

Proof. Let $\mu = \mu_{X,X}$, and for any $f \in L^2_\mu(\mathbb{R})$ define

$$v(f) = \mu_{\mathcal{L}}(f)X,$$

such that

$$\begin{aligned}\langle v(f_1), v(f_2)\rangle &= \langle \mu_{\mathcal{L}}(f_1)X, \mu_{\mathcal{L}}(f_2)X\rangle \\ &= \langle X, \mu_{\mathcal{L}}(\overline{f_1})\mu_{\mathcal{L}}(f_2)X\rangle \\ &= \langle X, \mu_{\mathcal{L}}(\overline{f_1}f_2)X\rangle = \int f_1(\lambda)\overline{f_2}(\lambda)\mathrm{d}\mu(\lambda).\end{aligned}$$

Therefore, if $\{f_n\}_{n=1}^\infty$ is an orthonormal basis for $L^2_\mu(\mathbb{R})$ [1], then $\{v_n\}_{n=1}^\infty$ is a set of orthonormal vectors in $\mathscr{H}$. Let $V = \operatorname{span}\{v_n : n \geq 1\}$ and $V^\perp = \{v \in \mathscr{H} : \langle v, v'\rangle = 0 \text{ for all } v' \in \mathscr{H}\}$ be the orthogonal complement of V. Then $V^\perp$ is also a Hilbert subspace with orthonormal basis $\{v_n\}_{n=-\alpha}^0$, where $\alpha - 1 \in \{0, \mathbb{N}, \infty\}$ is the cardinality of W. Thus, any element $\phi \in \mathscr{H}$ can be uniquely expressed as

$$\phi = \sum_{n=-\alpha}^{\infty} a_n v_n,$$

where the coefficients $a_n \in \mathbb{C}$ are such that $\sum_{n=-\alpha}^\infty |a_n|^2 < \infty$. Indeed, this construction provides an isomorphism between $\mathscr{H}$ and the square-summable complex sequences ℓ^2.

Let us now define random variables $\{X_n\}_{n=-\alpha}^\infty$ such that $X_n = \mathscr{M}(f_n)$ for $n \geq 1$ and independently $X_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ for $n \leq 0$. By construction, the X_n have mean zero, variance one and are uncorrelated with each other. Moreover, for $\phi \in \mathscr{H}$ let $\{a_n(\phi)\}_{n=-\alpha}^\infty$ denote the corresponding element of ℓ^2 defined by the isomorphism above, and consider the random

[1] Since $L^2(\mathbb{R})$ is separable, $\{f_n\}_{n=1}^\infty$ is a finite or countably infinite collection of elements in $L^2(\mathbb{R})$ such that their linear combinations are dense in $L^2(\mathbb{R})$. This can be extended to $L^2(\mathbb{R}^d)$ (Stein and Shakarchi, 2005, Theorem 1.3 in Chapter 4).

variable

$$Z_N(\phi) = \sum_{n=-\alpha}^{N} a_n(\phi) X_n.$$

Then for $M < N$ we have

$$\mathbf{E}(Z_N(\phi) - Z_M(\phi))^2 = \sum_{n=M+1}^{N} |a_n(\phi)|^2,$$

which shows that $Z_1(\phi), Z_2(\phi), \ldots$ is a Cauchy sequence, and thus converges in the Hilbert space consisting of all random variables with finite second moment. In other words, for any $\phi \in \mathscr{H}$ we have a well-defined random variable

$$Z(\phi) = \sum_{n=-\alpha}^{\infty} a_n(\phi) X_n.$$

Indeed, identifying $\mathscr{H}^*$ with the support of the random sequences $\{X_n\}_{n=-\alpha}^{\infty}$, and a specific $\Gamma \in \mathscr{H}^*$ with a specific instance of the sequence $\{X_n(\Gamma)\}_{n=-\alpha}^{\infty}$, by defining

$$\langle \phi, \Gamma \rangle^* = \sum_{n=-\alpha}^{\infty} a_n(\phi) X_n(\Gamma)$$

we have proved parts (i) and (ii).

To prove part (iii), for $f \in L^2_\mu(\mathbb{R})$ let

$$Z(f) = Z(\mu_{\mathcal{L}}(f) X).$$

Then for f_n in the orthonormal basis $\{f_n\}_{n=1}^{\infty}$ of $L^2_\mu(\mathbb{R})$, by construction we have $Z(f_n) = \mathscr{M}(f_n)$. For arbitrary $f = \sum_{n=1}^{\infty} a_n f_n \in L^2_\mu(\mathbb{R})$, let $S_n = \sum_{k=1}^{n} a_k f_k$. Then by linearity of both $Z(f)$ and $\mathscr{M}(f)$ we have $Z(S_n) = \mathscr{M}(S_n)$. Finally, for $M \leq N$ we have

$$|Z(S_N) - Z(S_M)|^2 = |\mathscr{M}(S_N) - \mathscr{M}(S_M)|^2 = \int |S_N(\lambda) - S_M(\lambda)|^2 \mathrm{d}\mu(\lambda)$$

which shows that $Z(S_n) = \mathscr{M}(S_n)$ is a Cauchy sequence, and thus both converge in mean square to the same random variable. That each converges to $Z(f) = \mathscr{M}(f)$ is obtained by an identical argument. We now have the proof of (iii) by setting $f(\lambda) = \mathtt{1}(\lambda \in E)$.

□

A.3.4 Proof of Proposition 2.5.1

Proof. Decompose the Hilbert space $\mathscr{H}$ as the direct sum of a closed subspace G and its orthogonal complement $G^{\perp}$, i.e. $\mathscr{H} = G \bigoplus G^{\perp}$, for any $x \in \mathscr{H}$, $x = y + z$ where $y \in G$, $z \in G^{\perp}$. The projection operator $\mathcal{P} : \mathscr{H} \to G$, then since $\langle y_1, z_2 \rangle = \langle y_2, z_1 \rangle = 0$,

$$\langle \mathcal{P}x_1, x_2 \rangle = \langle y_1, y_2 + z_2 \rangle = \langle y_1, y_2 \rangle = \langle y_1 + z_1, y_2 \rangle = \langle x_1, \mathcal{P}x_2 \rangle$$

which implies $\mathcal{P} = \mathcal{P}^*$. It is obvious $\mathrm{dom}(\mathcal{P}) = \mathrm{dom}(\mathcal{P}^*)$, therefore, $\mathcal{P}$ is self-adjoint. Similarly, $\mathcal{Q}$ is also self-adjoint. □

A.3.5 Proof of Proposition 2.5.2

Proof. For any $x \in \mathscr{H}$ and if $\mathcal{P}x \neq 0$, by the Cauchy–Schwartz inequality, we have

$$\|\mathcal{P}x\| = \frac{\langle \mathcal{P}x, \mathcal{P}x \rangle}{\|\mathcal{P}x\|} = \frac{\langle \mathcal{P}x, \mathcal{P}^2 x \rangle}{\|\mathcal{P}x\|} = \frac{\langle x, \mathcal{P}x \rangle}{\|\mathcal{P}x\|} \leq \|x\|$$

which implies that $\|\mathcal{P}\| \leq 1$. For nonzero $\mathcal{P}$, i.e. $\mathcal{P} \neq 0$, then $\exists x \in \mathscr{H}$ such that $\mathcal{P}x \neq 0$, and $\|\mathcal{P}(\mathcal{P}x)\| = \|\mathcal{P}x\|$ implies $\|\mathcal{P}\| \geq 1$. Combing the two results gives us $\|\mathcal{P}\| = 1$ if it is nonzero.

Similarly, $\mathcal{P}\mathcal{L}$ is also bounded, i.e.

$$\frac{\|\mathcal{P}\mathcal{L}x\|}{\|x\|} = \frac{\|\langle \mathcal{L}x, A \rangle \langle A, A \rangle^{-1} A\|}{\|x\|} = \frac{|\langle x, \mathcal{L}A \rangle| \cdot \|A\|}{\|A\|^2 \|x\|} \leq \frac{\|x\| \|\mathcal{L}A\| \cdot \|A\|}{\|A\|^2 \|x\|} = \frac{\|\mathcal{L}A\|}{\|A\|}$$

which is just a constant. □

A.3.6 Proof of the Zwanzig Operator Identity

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be two unbounded operators (we do not require $\mathcal{A}$ and $\mathcal{B}$ commute with each other), and assume that $\mathcal{A}$, $-\mathcal{A}$, $\mathcal{B}$, and $\mathcal{A} + \mathcal{B}$ are well-defined generators of strongly continuous semigroups. Let us define

$$\mathcal{K}(t) = e^{-\mathcal{A}t} e^{(\mathcal{A}+\mathcal{B})t}, \qquad \mathrm{dom}(\mathcal{K}) = \mathrm{dom}(\mathcal{A}) \cap \mathrm{dom}(\mathcal{B}). \tag{A.10}$$

Applying the product rule to take the derivative (w.r.t. t), we have (notice that A and $e^{-\mathcal{A}t}$ commute) for $x \in \mathrm{dom}(\mathcal{A}) \cap \mathrm{dom}(\mathcal{B})$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}(t)x = -\mathcal{A}e^{-\mathcal{A}t}e^{(\mathcal{A}+\mathcal{B})t}x + e^{-\mathcal{A}t}(\mathcal{A}+\mathcal{B})e^{(\mathcal{A}+\mathcal{B})t}x = e^{-\mathcal{A}t}\mathcal{B}\cdot e^{(\mathcal{A}+\mathcal{B})t}x.$$

Using the definition (A.10) of $\mathcal{K}(t)$ to replace $e^{(\mathcal{A}+\mathcal{B})t}$ gives us

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}(t)x = e^{-\mathcal{A}t}\mathcal{B}e^{\mathcal{A}t}\mathcal{K}(t)x. \tag{A.11}$$

By the semigroup property, we know that $e^{-\mathcal{A}t}$ and $e^{(\mathcal{A}+\mathcal{B})t}$ are both continuous in t (since $\mathcal{A}+\mathcal{B}$ is also self-adjoint or skew-adjoint) such that $K(t)$ must be continuous in t, and that $\mathcal{K}(0) = \mathcal{I}$. Taking the integration of (A.11) yields

$$\mathcal{K}(t)x = \mathcal{I}x + \int_0^t e^{-\mathcal{A}s}\mathcal{B}e^{\mathcal{A}s}\mathcal{K}(s)x\mathrm{d}s.$$

Applying $e^{\mathcal{A}t}$ from left on both sides and then using the definition (A.10) of $\mathcal{K}(t)$, we get

$$e^{(\mathcal{A}+\mathcal{B})t}x = e^{\mathcal{A}t}x + \int_0^t e^{\mathcal{A}(t-s)}\mathcal{B}e^{(\mathcal{A}+\mathcal{B})s}x\mathrm{d}s.$$

By setting $\mathcal{A} = \mathcal{L}$ and $\mathcal{B} = -\mathcal{P}\mathcal{L}$, we have $\mathcal{A}+\mathcal{B} = (\mathcal{I}-\mathcal{P})\mathcal{L}$ and $\mathrm{dom}(\mathcal{L}) \cap \mathrm{dom}(\mathcal{P}\mathcal{L}) = \mathrm{dom}(\mathcal{L})$. For any $x \in \mathrm{dom}(\mathcal{L})$ and $\mathcal{L}x \in \mathrm{dom}(\mathcal{L})$

$$e^{(\mathcal{I}-\mathcal{P})\mathcal{L}t}x = e^{\mathcal{L}t}x - \int_0^t e^{\mathcal{L}(t-s)}\mathcal{P}\mathcal{L}e^{(\mathcal{I}-\mathcal{P})\mathcal{L}s}x\mathrm{d}s.$$

Rearranging the above equation (moving the integral term to the LHS) gives us (2.12). □

Remark A.3.1. *It should be noted that for the proof to be valid $\mathcal{B}$ can be unbounded as long as $\mathcal{A}+\mathcal{B}$ can generate a strongly continuous semigroup and* $\mathrm{dom}(\mathcal{A}) \cap \mathrm{dom}(\mathcal{B})$ *is nonempty. Another proof based on the bounded perturbation theorem (i.e., $\mathcal{B}$ is assumed to be bounded, then $\mathcal{A}+\mathcal{B}$ is consequently a well-defined generator of a strong continuous semigroup) is implied by Corollary III.1.7 in Engel and Nagel (2000).*

A.3.7 Proof of Proposition 2.5.3

Proof. Let $\mathcal{Q} = \mathcal{I} - \mathcal{P}$, we have already shown that $\mathcal{P}$ and $\mathcal{Q}$ are self-adjoint. Since $\mathcal{L}$ is skew-adjoint, we have

$$\langle \mathcal{P}\mathcal{L}x, y\rangle = \langle \mathcal{L}x, \mathcal{P}y\rangle = \langle x, -\mathcal{L}\mathcal{P}y\rangle$$

which implies that $(\mathcal{P}\mathcal{L})^* = -\mathcal{L}\mathcal{P}$. Additionally, we have $\mathrm{dom}(\mathcal{P}\mathcal{L}) = \mathrm{dom}(-\mathcal{L}\mathcal{P}) = \mathrm{dom}(\mathcal{L})$. Similarly, $(\mathcal{Q}\mathcal{L})^* = -\mathcal{L}\mathcal{Q}$. By the property of semigroups we have $(e^{t\mathcal{Q}\mathcal{L}})^* = e^{-t\mathcal{L}\mathcal{Q}}$.

Now consider the autocorrelation of the random force $F(t)$.

According to (2.16), we have

$$\begin{aligned}
\langle F(t), F(t+u)\rangle &= \langle e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0),\ e^{(t+u)\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \\
&= \langle e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0),\ e^{t\mathcal{Q}\mathcal{L}}e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \qquad \text{(since } t\mathcal{Q}\mathcal{L} \text{ and } u\mathcal{Q}\mathcal{L} \text{ commute)} \\
&= \langle e^{-t\mathcal{L}\mathcal{Q}}e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle
\end{aligned}$$

where $e^{-t\mathcal{L}\mathcal{Q}}e^{t\mathcal{Q}\mathcal{L}} \neq e^{-t\mathcal{L}\mathcal{Q}+t\mathcal{Q}\mathcal{L}}$ since $-t\mathcal{L}\mathcal{Q}$ and $t\mathcal{Q}\mathcal{L}$ do not commute.

Since $e^{t\mathcal{L}\mathcal{Q}}$ and $e^{t\mathcal{Q}\mathcal{L}}$ are bounded, by the *dominated convergence theorem* we can take the time derivative of $\langle F(t), F(t+u)\rangle$ inside the inner product (i.e., to interchange the derivative and the integral), by the product rule we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle F(t), F(t+u)\rangle &= \langle \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-t\mathcal{L}\mathcal{Q}}e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\right),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \\
&= \langle \left(-\mathcal{L}\mathcal{Q}e^{-t\mathcal{L}\mathcal{Q}}e^{t\mathcal{Q}\mathcal{L}} + e^{-t\mathcal{L}\mathcal{Q}}\mathcal{Q}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\right)\mathcal{Q}\mathcal{L}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \\
&= \langle e^{-t\mathcal{L}\mathcal{Q}}(\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q})e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \\
&= \langle e^{-t\mathcal{L}\mathcal{Q}}(\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q})\mathcal{Q}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \\
&= \langle e^{-t\mathcal{L}\mathcal{Q}}(\mathcal{Q}\mathcal{L}\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q}\mathcal{L})e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle \\
&= \langle e^{-t\mathcal{L}\mathcal{Q}}(\mathcal{Q}\mathcal{L}\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q}\mathcal{L})e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ \mathcal{Q}\mathcal{L}e^{u\mathcal{Q}\mathcal{L}}\mathcal{B}(0)\rangle \\
&= \langle -\mathcal{L}\mathcal{Q}e^{-t\mathcal{L}\mathcal{Q}}(\mathcal{Q}\mathcal{L}\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q}\mathcal{L})e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{B}(0)\rangle \\
&= \langle -e^{-t\mathcal{L}\mathcal{Q}}\mathcal{L}\mathcal{Q}(\mathcal{Q}\mathcal{L}\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q}\mathcal{L})e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{B}(0)\rangle \\
&= \langle -e^{-t\mathcal{L}\mathcal{Q}}(\mathcal{L}\mathcal{Q}\mathcal{L}\mathcal{Q}\mathcal{L} - \mathcal{L}\mathcal{Q}\mathcal{L}\mathcal{Q}\mathcal{L})e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ e^{u\mathcal{Q}\mathcal{L}}\mathcal{B}(0)\rangle = 0
\end{aligned}$$

which implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle F(t), F(t+u)\rangle = 0 \quad \Longrightarrow \quad \langle F(t), F(t+u)\rangle = \langle F(0), F(u)\rangle = \langle \mathcal{Q}\mathcal{L}\mathcal{B}(0), e^{u\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0)\rangle$$

That is, the function $t \mapsto \langle F(t), F(t+u)\rangle$ is independent of t. We can then denote $\mathcal{C}_F(t) = \langle F(s), F(s+t)\rangle$. To see it is a valid ACF, we still need to verify its positive semi-definiteness and symmetry.

According to Proposition 2.2.1, it immediately follows that $\mathcal{C}_F(t)$ is positive definite. To check its (conjugate) symmetry, we need to show $\mathcal{C}_F(t) = \overline{\mathcal{C}_F(-t)}$. Indeed,

$$\begin{aligned}\overline{\mathcal{C}_F(-t)} &= \overline{\langle \mathcal{QLB}(0),\ e^{-t\mathcal{QL}}\mathcal{QLB}(0)\rangle} = \langle e^{-t\mathcal{QL}}\mathcal{QLB}(0),\ \mathcal{QLB}(0)\rangle \\ &= \langle e^{-t\mathcal{LQ}}\mathcal{QLB}(0),\ \mathcal{QLB}(0)\rangle = \langle \mathcal{QLB}(0),\ e^{t\mathcal{QL}}\mathcal{QLB}(0)\rangle = \mathcal{C}_F(t).\end{aligned}$$

□

A.3.8 Proof of Lemma 2.6.1

Proof. If

$$\mathcal{C}(t) = \langle e^{t\mathcal{L}}X, X\rangle = \int_{\mathbb{R}} e^{it\lambda}\mathrm{d}\nu(\lambda)$$

is the autocorrelation of a k-times continuously differentiable stationary process, then by Itô (1954), Theorem 5.2 we have

$$\int_{\mathbb{R}} \lambda^{2m}\mathrm{d}\nu(\lambda) < \infty.$$

for $m = 0, \ldots, k$.

On the other hand, using the direct integral version of the spectral theorem (Theorem A.1.2), the embedding Hilbert space $\mathscr{H}_C$ is isometric to $\mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n, \mu\}$, wherein $f(\mathcal{L}) : s(\lambda) \mapsto f(-i\lambda)s(\lambda)$. Furthermore, in the proof of Theorem 2.3.1 (Appendix A.3.2) we showed that the element $s(\lambda) \in \mathscr{H}^{\oplus}\{\amalg_{n=0}^{\infty} S_n, \mu\}$ corresponding to X is such that $\mathrm{d}\nu(\lambda)/\mathrm{d}\mu(\lambda) = |s(\lambda)|_\lambda^2$. Thus we have

$$\begin{aligned}\int_{\mathbb{R}} \lambda^{2m}\mathrm{d}\nu(\lambda) &= \int_{\mathbb{R}} \lambda^{2m}|s(\lambda)|_\lambda^2\mathrm{d}\mu(\lambda) \\ &= \int_{\mathbb{R}} |\lambda^m s(\lambda)|_\lambda^2\mathrm{d}\mu(\lambda) < \infty,\end{aligned}$$

which by Theorem A.1.2 shows that $s(\lambda) \in \mathrm{dom}(\mathcal{L}^m)$ for $m = 0, \ldots, k$. □

A.3.9 Some Calculations for the Acceleration Form GLE

Given $\boldsymbol{\mathcal{A}} = [X, \dot{X}]$ and $\mathcal{B} = \dot{X}$ where X is the displacement of interest and $\dot{X} = \mathcal{L}X$ is the velocity, $\ddot{X} = \mathcal{L}\dot{X}$ is the acceleration rate. We have

$$\begin{aligned}
\boldsymbol{\Omega} &= \left(\langle \mathcal{L}\dot{X}, X\rangle, \langle \mathcal{L}\dot{X}, \dot{X}\rangle\right) \begin{pmatrix} \langle X, X\rangle & \langle X, \dot{X}\rangle \\ \langle \dot{X}, X\rangle & \langle \dot{X}, \dot{X}\rangle \end{pmatrix}^{-1} \\
&= \left(\langle \ddot{X}, X\rangle, 0\right) \begin{pmatrix} \langle X, X\rangle^{-1} & 0 \\ 0 & \langle \dot{X}, \dot{X}\rangle^{-1} \end{pmatrix} = \left(\tfrac{\langle \ddot{X}, X\rangle}{\langle X, X\rangle}, 0\right), \\
\boldsymbol{\Omega}\boldsymbol{\mathcal{A}}^{\dagger}(t) &= \left(\tfrac{\langle \ddot{X}, X\rangle}{\langle X, X\rangle}, 0\right) \begin{pmatrix} X(t) \\ \dot{X}(t) \end{pmatrix} = \frac{\langle \ddot{X}, X\rangle}{\langle X, X\rangle} X(t) = -\frac{\langle \dot{X}, \dot{X}\rangle}{\langle X, X\rangle} X(t), \\
F(0) &= (\mathcal{I} - \mathcal{P})\mathcal{L}\dot{X} = \ddot{X} - \boldsymbol{\Omega}\boldsymbol{\mathcal{A}}^{\dagger}(0) = \ddot{X} + \frac{\langle \dot{X}, \dot{X}\rangle}{\langle X, X\rangle} X, \\
\langle F(0), F(t)\rangle &= \langle \ddot{X}, F(t)\rangle + \frac{\langle \dot{X}, \dot{X}\rangle}{\langle X, X\rangle}\langle X, F(t)\rangle = \langle \ddot{X}, F(t)\rangle \quad \text{since } F(t) \perp \mathrm{col}(\boldsymbol{\mathcal{A}}), \\
\boldsymbol{K}(t) &= \left(\langle F(t), \mathcal{L}X\rangle, \langle F(t), \mathcal{L}\dot{X}\rangle\right) \begin{pmatrix} \langle X, X\rangle^{-1} & 0 \\ 0 & \langle \dot{X}, \dot{X}\rangle^{-1} \end{pmatrix} \\
&= \left(0, \tfrac{\langle F(t), \ddot{X}\rangle}{\langle \dot{X}, \dot{X}\rangle}\right) \qquad \text{since } \langle F(t), \dot{X}\rangle = 0, \\
\boldsymbol{K}(s)\boldsymbol{\mathcal{A}}^{\dagger}(t-s) &= \left(0, \tfrac{\langle F(s), \ddot{X}\rangle}{\langle \dot{X}, \dot{X}\rangle}\right) \begin{pmatrix} X(t-s) \\ \dot{X}(t-s) \end{pmatrix} = \frac{\langle F(s), \ddot{X}\rangle}{\langle \dot{X}, \dot{X}\rangle} \dot{X}(t-s) = \frac{\langle F(s), F(0)\rangle}{\langle \dot{X}, \dot{X}\rangle} \dot{X}(t-s)
\end{aligned}$$

where we used the property that $\mathcal{L}$ is skew-adjoint such that $\langle \mathcal{A}, \mathcal{L}\mathcal{A}\rangle = -\langle \mathcal{L}\mathcal{A}, \mathcal{A}\rangle = 0$.

Similarly as what we did in Appendix A.3.7, we can also verify that $\mathcal{C}(t) = \langle F(0), F(t)\rangle$ is symmetric and positive semi-definite.

Notice that $\mathcal{Q} \perp \mathrm{col}(\boldsymbol{\mathcal{A}}(0))$ by the definition of $\mathcal{P}$, we have

$$F(t) = \mathcal{Q}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\mathcal{B} \perp \mathrm{col}(\boldsymbol{\mathcal{A}}(0)) \quad \Longleftrightarrow \quad \langle F(t), \boldsymbol{\mathcal{A}}(0)\rangle = 0.$$

We can see this for any $1 \le j \le d$,

$$\begin{aligned}
\langle F(t), \mathcal{A}_j(0)\rangle &= \langle e^{t(\mathcal{I}-\mathcal{P})\mathcal{L}}(\mathcal{I} - \mathcal{P})\mathcal{L}\mathcal{B}(0),\ \mathcal{A}_j(0)\rangle = \langle e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\mathcal{B}(0),\ \mathcal{A}_j(0)\rangle \\
&= \langle \mathcal{Q}\mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ \mathcal{A}_j(0)\rangle \\
&= \langle \mathcal{L}e^{t\mathcal{Q}\mathcal{L}}\mathcal{B}(0),\ \mathcal{Q}\mathcal{A}_j(0)\rangle = 0
\end{aligned}$$

where we applied the property that $\mathcal{Q}\mathcal{L}$ and $e^{t\mathcal{Q}\mathcal{L}}$ commute

As a consequence, $F(t)$ is along the direction of $\mathcal{Q}$, then if follows that $\mathcal{Q}F(t) = F(t)$. This can also be proved as follows

$$\mathcal{Q}F(t) = (\mathcal{I} - \mathcal{P})\, F(t) = F(t) - \mathcal{P}F(t) = F(t) - \langle F(t), \boldsymbol{\mathcal{A}}(0)\rangle \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\mathcal{A}}(0)} \boldsymbol{\mathcal{A}}^{\dagger}(0) = F(t).$$

A.3.10 Proof of Lemma 2.6.2

Proof. Using the projection-valued measure form of the spectral theorem (Theorem A.1.3),

$$\langle e^{-t\mathcal{L}}\psi, \phi\rangle = \int_{\mathbb{R}} e^{it\lambda} \mathrm{d}\mu_{\psi,\phi}(\lambda),$$

where $\mu_{\psi,\phi}(E) = \langle \mu_{i\mathcal{L}}(E)\psi, \phi\rangle$ and $\mu_{i\mathcal{L}}$ is the projection-valued measure associated with the self-adjoint $i\mathcal{L}$. The desired result follows from interchange of integral and derivative, which itself follows since $\psi \in \mathrm{dom}(\mathcal{L})$ implies that

$$\langle e^{-t\mathcal{L}}\mathcal{L}\psi, \phi\rangle = \int_{\mathbb{R}} e^{it\lambda}(-i\lambda) \mathrm{d}\mu_{\psi,\phi}(\lambda),$$

hence $\frac{\mathrm{d}}{\mathrm{d}t} e^{it\lambda} = e^{it\lambda}(-i\lambda)$ is integrable with respect to $\mu_{\psi,\phi}$. □

A.3.11 Laplace Transforms of Observables and Correlation Functions

Consider in general an observable $\mathcal{A}(t) \in \mathscr{H}'$ with the associated an element $f \in \mathscr{H}$, i.e., by the Riesz representation,

$$\mathcal{A}(\Gamma_t) = \langle f, \Gamma_t\rangle = \langle f, e^{t\mathcal{L}}\Gamma\rangle = \langle e^{-t\mathcal{L}} f, \Gamma\rangle$$

where $\Gamma_t = e^{\mathcal{L}t}\Gamma$, $\Gamma = (\mathsf{q}, \mathsf{p}) \in \mathscr{H}$ and $\mathcal{L}$ is the Liouville operator determined by the Hamiltonian. The Laplace transform $\mathfrak{L}\{\cdot\}$ of $\mathcal{A}$ is well-defined as

$$\check{\mathcal{A}}(z)(\Gamma) = \mathfrak{L}\{\mathcal{A}(\Gamma_t)\} = \int_0^\infty e^{-zt}\langle f, \Gamma_t\rangle \mathrm{d}t = \int_0^\infty e^{-zt}\langle e^{-t\mathcal{L}} f, \Gamma\rangle \mathrm{d}t$$

since $e^{-t\mathcal{L}}$ is a unitary (semi-)group, $\|e^{-t\mathcal{L}}\| = 1$ such that for any fixed Γ,

$$\|\mathcal{A}(\Gamma_t)\| = \left|\langle e^{-t\mathcal{L}} f, \Gamma\rangle\right| \leq \|f\|\|\Gamma\| < \infty.$$

Moreover, since $C_{\mathcal{A}}(t) = \langle \mathcal{A}(u), \mathcal{A}(u+t)\rangle = \langle f, e^{-t\mathcal{L}} f\rangle$, the Laplace transform of this is well-defined for $\Re(z) > 0$ and given by

$$\check{C}_{\mathcal{A}}(z) = \mathfrak{L}\{C_{\mathcal{A}}(t)\} = \int_0^\infty e^{-zt} \langle f, e^{-t\mathcal{L}} f\rangle \mathrm{d}t$$

where $\left|e^{-zt}\langle f, e^{-t\mathcal{L}} f\rangle\right| \leq e^{-\Re(z)t}\|f\|^2$ which guarantees the decay of the Laplace transform.

If we consider another observable $\mathcal{B}(t) \in \mathscr{H}$ associated with $g \in \mathscr{H}$, the cross-correlation $C_{\mathcal{AB}} = \langle \mathcal{A}(t), \mathcal{B}(s)\rangle = \langle e^{-t\mathcal{L}} f, e^{-s\mathcal{L}} g\rangle$ also has a valid Laplace transform, i.e.

$$\check{C}_{\mathcal{AB}}(z) = \mathfrak{L}\{C_{\mathcal{AB}}\} = \int_0^\infty e^{-zt} \langle e^{-t\mathcal{L}} f, e^{-s\mathcal{L}} g\rangle \mathrm{d}t$$

where $\left|e^{-zt}\langle e^{-t\mathcal{L}} f, e^{-s\mathcal{L}} g\rangle\right| \leq e^{-\Re(z)t}\|f\|\|g\|$ by the Cauchy–Schwartz inequality.

When we assume $\mathcal{A}(t)$ is twice continuously differentiable, we have $\mathcal{A}(t) \in \mathrm{dom}(\mathcal{L})$ and $\mathcal{L}\mathcal{A}(t) \in \mathrm{dom}(\mathcal{L})$. Then it immediately follows that $\dot{C}_{\mathcal{A}}(t) = \langle \mathcal{A}(u), \dot{\mathcal{A}}(u+t)\rangle$ and $\ddot{C}_{\mathcal{A}}(t) = \langle \mathcal{A}(u), \ddot{\mathcal{A}}(u+t)\rangle$ have valid Laplace transforms.

Note that since $\mathfrak{L}\{\dot{C}_X(t)\} = z\check{C}_X(z) - C_X(0)$ is the Fourier transform of $e^{-at}\dot{C}_X(t)\mathbb{1}(t > 0) \in L^1(\mathbb{R})$ for $\Re(z) = a > 0$, by the Riemann-Lebesgue lemma we have

$$\lim_{|z|\to\infty} \mathfrak{L}\{\dot{C}_X(t)\} = 0 \qquad \Longrightarrow \qquad \lim_{|z|\to\infty} z\check{C}_X(z) = C_X(0)$$

If $\mathfrak{L}\{\dot{C}_X(t)\} = z\check{C}_X(z) - C_X(0) = 0$ for $\Re\{z\} > 0$, we would have $\check{C}_X(z) = C_X(0)/z$ which implies the inverse Laplace transform $C_X(t) = C_X(0)$. This is certainly a contradiction. For $\Re\{z\} > 0$, (2.21) will always be valid such that $\check{K}(z)$ is an analytic function in its ROC.

A.3.12 A Useful Lemma About the Time Derivatives of ACFs

Let us show a useful property about the time differentiation of the (twice differentiable) autocovariance function (ACF) of a stationary process.

Lemma A.3.2. *For any stationary process $X(t)$ with twice continuously differentiable autocovariance function $\mathcal{C}(t-s) = \langle X(t), X(s)\rangle$, we have*

$$\langle \dot{X}(t), X(s)\rangle = \dot{\mathcal{C}}(t-s), \quad \langle X(t), \dot{X}(s)\rangle = -\dot{\mathcal{C}}(t-s), \quad \langle \dot{X}(t), \dot{X}(s)\rangle = -\ddot{\mathcal{C}}(t-s).$$

Proof.

$$\begin{aligned}
\langle \dot{X}(t), X(s)\rangle &= \langle \lim_{\Delta t\to 0} \frac{X(t)-X(t-\Delta t)}{\Delta t},\ X(s)\rangle \\
&= \lim_{\Delta t\to 0} \frac{1}{\Delta t}\left[\langle X(t), X(s)\rangle - \langle X(t-\Delta t), X(s)\rangle\right] \\
&= \lim_{\Delta t\to 0} \frac{1}{\Delta t}\left[\mathcal{C}(t-s) - \mathcal{C}(t-s-\Delta t)\right] \\
&= \frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{C}(\tau)|_{\tau=t-s} \triangleq \dot{\mathcal{C}}(t-s).
\end{aligned}$$

where we can verify the interchange of limit and the inner product by the continuity of inner products using the Cauchy–Schwarz inequality. Similarly,

$$\begin{aligned}
\langle X(t), \dot{X}(s)\rangle &= \langle X(t),\ \lim_{\Delta s\to 0} \frac{X(s)-X(s-\Delta s)}{\Delta s}\rangle \\
&= \lim_{\Delta s\to 0} \frac{1}{\Delta s}\left[\langle X(t), X(s)\rangle - \langle X(t), X(s-\Delta s)\rangle\right] \\
&= \lim_{\Delta s\to 0} \frac{1}{\Delta s}\left[\mathcal{C}(t-s) - \mathcal{C}(t-s+\Delta t)\right] \\
&= -\lim_{\Delta s\to 0} \frac{1}{\Delta s}\left[\mathcal{C}(t-s+\Delta s) - \mathcal{C}(t-s)\right] \\
&= -\frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{C}(\tau)|_{\tau=t-s} \triangleq -\dot{\mathcal{C}}(t-s).
\end{aligned}$$

Immediately, we have $\langle \dot{X}(t), \dot{X}(s)\rangle = -\ddot{\mathcal{C}}(t-s)$. □

Remark A.3.2. *If we define $\mathcal{C}(t-s) = \langle X(s), X(t)\rangle$, then we have the same result with a different formulation*

$$\langle \dot{X}(t), X(s)\rangle = -\dot{\mathcal{C}}(s-t), \quad \langle X(t), \dot{X}(s)\rangle = \dot{\mathcal{C}}(s-t), \quad \langle \dot{X}(t), \dot{X}(s)\rangle = -\ddot{\mathcal{C}}(s-t).$$

Since $\mathcal{C}(t)$ is an even function such that $\dot{\mathcal{C}}(t)$ is an odd function and $\ddot{\mathcal{C}}(t)$ is again even, we further have $-\dot{\mathcal{C}}(s-t) = \dot{\mathcal{C}}(t-s)$, $\dot{\mathcal{C}}(s-t) = -\dot{\mathcal{C}}(t-s)$ and $-\ddot{\mathcal{C}}(s-t) = -\ddot{\mathcal{C}}(t-s)$, which coincides with the result of Lemma A.3.2.

A.3.13 Hardy Space and the Boundary Value of the Laplace Transform

Let $z = a + 2\pi i\omega$ where a is a positive constant. Consider an ACF $\mathcal{C}(t) \in L^2(\mathbb{R})$, given $\mathbb{1}(t)$ as the Heaviside function, $\mathbb{1}(t)\mathcal{C}(t) \in L^2(\mathbb{R}_+)$, then the Laplace transform

$$\begin{aligned}\check{\mathcal{C}}(z) = \check{\mathcal{C}}(a + 2\pi i\omega) &= \int_0^\infty e^{-(at+2\pi i\omega t)}\mathcal{C}(t)\mathrm{d}t = \int_0^\infty e^{-2\pi i\omega t}\left(e^{-at}\mathcal{C}(t)\right)\mathrm{d}t \\ &= \int_{-\infty}^\infty e^{-2\pi i\omega t}\left(e^{-at}\mathbb{1}(t)\mathcal{C}(t)\right)\mathrm{d}t = \mathcal{F}\{e^{-at}\mathbb{1}(t)\mathcal{C}(t)\}(\omega)\end{aligned} \tag{A.12}$$

is in the Hardy space $\mathscr{H}^2(\mathbb{C})$ satisfying (Stein and Shakarchi, 2005, pp. 213-221)

(i) $\check{\mathcal{C}}(z)$ is an analytic function in the right half-plane $\Re\{z\} = a > 0$;

(ii) By Plancherel's theorem, it can be shown

$$\sup_{a>0}\int_{-\infty}^\infty \left|\check{\mathcal{C}}(a + 2\pi i\omega)\right|^2 \mathrm{d}a < \infty;$$

(iii) $\check{\mathcal{C}}(2\pi i\omega) = \lim_{a\to 0^+}\check{\mathcal{C}}(a + 2\pi i\omega) = \mathcal{F}\{\mathbb{1}(t)\mathcal{C}(t)\}(\omega)$ in either the $L^2(\mathbb{R})$ sense or in the pointwise (a.e. for every ω) sense. (The proof of a.e. convergence requires the result of $L^2(\mathbb{R})$ convergence together with Poisson integral representation to carry out approximations in $L^2(\mathbb{R})$.)

Similarly, for $\bar{z} = a - 2\pi i\omega$, we also have (note $\mathcal{C}(t)$ is an even function on $\mathbb{R}$, i.e. $\mathcal{C}(t) = \mathcal{C}(-t)$)

$$\begin{aligned}\check{\mathcal{C}}(\bar{z}) &= \int_0^\infty e^{-\bar{z}t}\mathcal{C}(t)\mathrm{d}t = \int_0^\infty e^{2\pi i\omega}e^{-at}\mathcal{C}(t)\mathrm{d}t = \int_{-\infty}^0 e^{-2\pi i\omega}\left(e^{at}\mathcal{C}(t)\right)\mathrm{d}t \\ &= \int_{-\infty}^\infty e^{-2\pi i\omega}\left(e^{at}(1-\mathbb{1}(t))\mathcal{C}(t)\right)\mathrm{d}t\end{aligned}$$

is in the Hardy space $\mathscr{H}^2(\mathbb{C})$ with the limit $\check{\mathcal{C}}(-2\pi i\omega) = \lim_{a\to 0^+}\check{\mathcal{C}}(a - 2\pi i\omega) = \mathcal{F}\{(1 - \mathbb{1}(t))\mathcal{C}(t)\}(\omega)$.

Then we can establish

$$\begin{aligned}\mathcal{F}\{\mathcal{C}(t)\}(\omega) &= \int_0^\infty e^{-2\pi i\omega t}\mathcal{C}(t)\mathrm{d}t + \int_{-\infty}^0 e^{-2\pi i\omega t}\mathcal{C}(t)\mathrm{d}t = \check{\mathcal{C}}(2\pi i\omega) + \check{\mathcal{C}}(-2\pi i\omega) \\ &= 2\cdot\Re\{\check{\mathcal{C}}(2\pi i\omega)\}.\end{aligned} \tag{A.13}$$

A.3.14 Technical Derivations of the Fourier Formulas

(i) First we show how to calculate the PSD $S(\omega)$ of $X(t)$ given the memory kernel $\mathcal{C}(t)$. Rearrange (2.21), we have

$$\begin{aligned}\check{\mathcal{C}}(z) &= \frac{\mathcal{C}(0)(\check{\mathcal{C}}(z)+z)}{z\check{\mathcal{C}}(z)+z^2+\theta} \\ \check{\mathcal{C}}(-z) &= \frac{\mathcal{C}(0)(\check{\mathcal{C}}(-z)-z)}{-z\check{\mathcal{C}}(-z)+(-z)^2+\theta}.\end{aligned}$$

By adding the above two equations together we have

$$\check{\mathcal{C}}(z)+\check{\mathcal{C}}(-z) = \frac{\mathcal{C}(0)\theta(\check{\mathcal{C}}(z)+\check{\mathcal{C}}(-z))}{|z\check{\mathcal{C}}(z)+z^2+\theta|^2}$$

where we have used the property that $\mathcal{C}(-z)$ is the complex conjugate of $\mathcal{C}(z)$ since $\mathcal{C}(t)$ is a real function. Let $z = 2\pi i\omega$, we have

$$\begin{aligned}S(\omega) = \mathcal{F}\{\mathcal{C}(t)\} = \check{\mathcal{C}}(2\pi i\omega)+\check{\mathcal{C}}(-2\pi i\omega) &= \frac{\mathcal{C}(0)\theta\mathcal{F}\{\mathcal{C}(t)\}}{|2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)+\kappa/m-4\pi^2\omega^2|^2} \\ &= \frac{k_BT/m\cdot\mathcal{F}\{\mathcal{C}(t)\}}{|2\pi i\omega\mathcal{F}\{\mathbb{1}(t)\mathcal{C}(t)\}+\kappa/m-4\pi^2\omega^2|^2} \\ &= \frac{k_BT\cdot\mathcal{F}\{K(t)\}}{|2\pi i\omega\mathcal{F}\{\mathbb{1}(t)\cdot K(t)\}+\kappa-4\pi^2\omega^2 m|^2} \\ &= \frac{k_BT\cdot K(\omega)}{|2\pi i\omega\hat{K}(\omega)+\kappa-4\pi^2\omega^2 m|^2}\end{aligned}$$

which is exactly the Fourier transformed GLE equation. Note that $\theta\mathcal{C}(0) = k_BT/m$ and $K(t) = m\mathcal{C}(t)$ is the memory kernel in the GLE when we take mass m as an explicit coefficient. We denote $K(\omega) = \mathcal{F}\{K(t)\}$ and $\hat{K}(\omega) = \mathcal{F}\{\mathbb{1}(t)\cdot K(t)\}$.

(ii) Now we show the recipe to calculate the memory kernel (in the frequency domain) given the PSD of $X(t)$ based on the Laplace transform relation (2.21). Let $z = 2\pi i\omega$ and $z = -2\pi i\omega$, respectively, in (2.21), we have

$$\begin{aligned}\check{\mathcal{C}}(2\pi i\omega) &= \frac{(4\pi^2\omega^2 - \theta)\check{\mathcal{C}}(2\pi i\omega) + 2\pi i\omega\mathcal{C}(0)}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega) - \mathcal{C}(0)} \\ \check{\mathcal{C}}(-2\pi i\omega) &= \frac{(4\pi^2\omega^2 - \theta)\check{\mathcal{C}}(-2\pi i\omega) - 2\pi i\omega\mathcal{C}(0)}{-2\pi i\omega\check{\mathcal{C}}(-2\pi i\omega) - \mathcal{C}(0)}.\end{aligned} \tag{A.14}$$

Adding the two equations together gives us

$$\mathcal{F}\{\mathcal{C}(t)\} = \check{\mathcal{C}}(2\pi i\omega) + \check{\mathcal{C}}(-2\pi i\omega) = \frac{\theta\mathcal{C}(0)\mathcal{F}\{\mathcal{C}(t)\}}{|2\pi i\omega\check{\mathcal{C}}(2\pi i\omega) - \mathcal{C}(0)|^2}$$

which gives

$$\mathcal{F}\{K(t)\} = \mathcal{F}\{m\mathcal{C}(t)\} = \frac{k_BT\mathcal{F}\{\mathcal{C}(t)\}}{|2\pi i\omega\mathcal{F}\{\mathbb{1}(t)\mathcal{C}(t)\} - \mathrm{var}(X(0))|^2} \tag{A.15}$$

where $K(t) = m\mathcal{C}(t)$, $\theta\mathcal{C}(0) = k_BT/m$ and $\mathcal{F}\{\mathbb{1}(t)\mathcal{C}(t)\} = \check{\mathcal{C}}(2\pi i\omega)$.

(iii) We can show an alternative way to get the same result based on a time-dependent Hamiltonian and maximum entropy distribution of the stationary system. If we define

$$\hat{\psi}^{-1}(\omega) = \kappa - 4\pi^2\omega^2 m + 2\pi i\omega\hat{K}(\omega), \tag{A.16}$$

then it follows that

$$\begin{aligned}\frac{k_BT}{\pi\omega}\Im\left\{\frac{1}{\hat{\psi}(\omega)}\right\} &= \frac{k_BT}{\pi\omega}\cdot\Im\{\kappa - 4\pi^2\omega^2 m + 2\pi i\omega\hat{K}(\omega)\} \\ &= \frac{k_BT}{\pi\omega}\cdot(2\pi\omega)\cdot\Re\{\hat{K}(\omega)\} \\ &= k_BT\cdot 2\Re\{\check{K}(2\pi i\omega)\} = k_BT\cdot K(\omega) \qquad \text{(by (A.13))}\end{aligned}$$

which gives

$$K(\omega) = \mathcal{F}\{K(t)\} = \frac{1}{\pi\omega}\Im\left\{\frac{1}{\hat{\psi}(\omega)}\right\}. \tag{A.17}$$

Notice that $\hat{K}(\omega) = \mathcal{F}\{\mathbb{1}(t)\cdot K(t)\} = \check{K}(2\pi i\omega)$ according to (A.12). Based on (A.14),

we further have

$$\begin{aligned}
\check{K}(2\pi i\omega) = m\check{\mathcal{C}}(2\pi i\omega) &= \frac{(4\pi^2\omega^2 m-\kappa)\check{\mathcal{C}}(2\pi i\omega)+2\pi i\omega m\mathcal{C}(0)}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)} \\
&= \frac{\frac{(4\pi^2\omega^2 m-\kappa)}{2\pi i\omega}\cdot 2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)+2\pi i\omega m\mathcal{C}(0)}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)} \\
&= \frac{\frac{(4\pi^2\omega^2 m-\kappa)}{2\pi i\omega}\cdot\left[2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)+\mathcal{C}(0)\right]+2\pi i\omega m\mathcal{C}(0)}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)} \\
&= \frac{(4\pi^2\omega^2 m-\kappa)}{2\pi i\omega}+\left[\frac{(4\pi^2\omega^2 m-\kappa)}{2\pi i\omega}+2\pi i\omega m\right]\frac{\mathcal{C}(0)}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)} \\
&= \frac{(4\pi^2\omega^2 m-\kappa)}{2\pi i\omega}-\frac{\kappa}{2\pi i\omega}\cdot\frac{\mathcal{C}(0)}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)}
\end{aligned}$$

which gives us (note $\kappa\mathcal{C}(0)=k_BT$)

$$2\pi i\omega\check{K}(2\pi i\omega) = (4\pi^2\omega^2 m-\kappa)-k_BT\cdot\frac{1}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)}.$$

According to the definition (A.16) of $\hat{\psi}^{-1}(\omega)$, we have

$$\begin{aligned}
\hat{\psi}^{-1}(\omega) &= \kappa-4\pi^2\omega^2 m+2\pi i\omega\hat{K}(\omega) \\
&= -k_BT\cdot\frac{1}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)}
\end{aligned} \tag{A.18}$$

which implies that

$$\hat{\psi}(\omega) = -\frac{1}{k_BT}\left(2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)\right).$$

It then follows that $\hat{\psi}(\omega)$ is indeed the Laplace transform of a well-defined function in the time domain, i.e.,

$$\hat{\psi}(\omega) = -\frac{1}{k_BT}\cdot\mathfrak{L}\{\dot{\mathcal{C}}(t)\}\Big|_{z=2\pi i\omega} = \mathfrak{L}\{\psi(t)\}\Big|_{z=2\pi i\omega}$$

where $\psi(t)=-\frac{1}{k_BT}\dot{\mathcal{C}}(t)$.

We can also establish the result in terms of the Fourier transform of $\psi(t)$. Notice that

$\dot{\mathcal{C}}(t)$ is an odd function (so is $\psi(t)$), we can obtain

$$\begin{aligned}\mathcal{F}\{\psi(t)\} = \hat{\psi}(\omega) - \hat{\psi}(-\omega) &= -\frac{1}{k_B T} \cdot \left[\mathfrak{L}\{\dot{\mathcal{C}}(t)\}\Big|_{z=2\pi i\omega} - \mathfrak{L}\{\dot{\mathcal{C}}(t)\}\Big|_{z=-2\pi i\omega} \right] \\ &= -\frac{1}{k_B T} \cdot \mathcal{F}\{\dot{\mathcal{C}}(t)\} = -\frac{1}{k_B T} \cdot 2\pi i\omega S(\omega).\end{aligned} \tag{A.19}$$

As we argued in (A.12), $\hat{\psi}(\omega)$ can be computed as

$$\hat{\psi}(\omega) = \mathfrak{L}\{\psi(t)\}\Big|_{z=2\pi i\omega} = \mathcal{F}\{\mathbb{1}(t) \cdot \psi(t)\}. \tag{A.20}$$

This result proves the following claim

Proposition A.3.1. *$\hat{\psi}(\omega) = \hat{\phi}_{AA}(\omega)$ where by definition $\phi_{AA}(t) = -\frac{1}{k_B T}\dot{C}_A(t)$ and $\hat{\phi}_{AA}(\omega) = \mathcal{F}\{\mathbb{1}(t)\phi_{AA}(t)\}$.*

It is worth noting that the above relation (in terms of the Laplace transform) is not a special case just for $z = 2\pi i\omega$. In general, if we define

$$\hat{\psi}^{-1}(z) = \kappa + z^2 m + z\check{K}(z).$$

We still have

$$z\check{K}(z) = -(z^2 m + \kappa) - \kappa\mathcal{C}(0) \cdot \frac{1}{z\check{\mathcal{C}}(z) - \mathcal{C}(0)}$$

which gives us

$$\hat{\psi}^{-1}(z) = -k_B T \cdot \frac{1}{z\check{\mathcal{C}}(z) - \mathcal{C}(0)}.$$

However, to make connections between the Laplace transform and the Fourier transform, we have to take $z = 2\pi i\omega$ and $z = -2\pi i\omega$.

(iv) We have two versions of the same result, one is given by (A.15) whereas the other is given by (A.17), (A.19) and (A.20), both of which are derived from (2.21) and by making use of (A.12) and (A.13) with $z = \pm 2\pi i\omega$. To see they are really the same, it is easier to

derive (A.15) from (A.17) and (A.18), i.e.

$$\begin{aligned}
K(\omega) = \mathcal{F}\{K(t)\} &= \frac{1}{\pi\omega}\Im\left\{\frac{1}{\hat{\psi}(\omega)}\right\} \\
&= \frac{1}{\pi\omega}\Im\left\{-k_BT\cdot\frac{1}{2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)}\right\} \\
&= -\frac{1}{\pi\omega}\cdot\frac{k_BT}{|2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)|^2}\cdot\Im\left\{-2\pi i\omega\check{\mathcal{C}}(-2\pi i\omega)-\mathcal{C}(0)\right\} \\
&= -\frac{1}{\pi\omega}\cdot\frac{k_BT}{|2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)|^2}\cdot(-2\pi\omega)\Re\{\check{\mathcal{C}}(-2\pi i\omega)\} \\
&= \frac{k_BT\cdot 2\Re\{\check{\mathcal{C}}(-2\pi i\omega)\}}{|2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)|^2} \\
&= \frac{k_BT\cdot S(\omega)}{|2\pi i\omega\check{\mathcal{C}}(2\pi i\omega)-\mathcal{C}(0)|^2}
\end{aligned}$$

where we noticed that $\mathcal{C}(t)$ is a real, even function and therefore $\check{\mathcal{C}}(-2\pi i\omega)$ is the complex conjugate of $\check{\mathcal{C}}(2\pi i\omega)$ and $S(\omega) = \mathcal{F}\{\mathcal{C}(t)\} = 2\Re\{\check{\mathcal{C}}(2\pi i\omega)\} = 2\Re\{\check{\mathcal{C}}(-2\pi i\omega)\}$ by (A.13).

Remark A.3.3. *The formula given by* (A.15) *is more straightforward based on our previous derivation whereas the set of formulas given by* (A.17), (A.19) *and* (A.20) *is more artfully constructive and reveals some fundamental structure not directly seen from our projection derivation approach. Now, all the pieces from both perspectives are coherently connected and essentially the same.*

Appendix B

Appendix for Chapter 3

B.1 Useful Lemmas

Let us first prove a useful argument used implicitly by Dahlhaus (1989) in his proof of Theorem 2.1.

Lemma B.1.1. *Suppose we have a sequence of possibly random objective functions* $\{\ell_n(\boldsymbol{\theta})\}_{n=1}^{\infty}$ *and a target function* $\ell(\boldsymbol{\theta})$*, and define*

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell_n(\boldsymbol{\theta}), \qquad \boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell(\boldsymbol{\theta}),$$

where $\boldsymbol{\Theta}$ *is a compact parameter space. Then if*

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_n(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})| = 0,$$

we have $\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$ *as well.*

Proof. Since $\ell_n(\boldsymbol{\theta})$ is minimized by $\hat{\boldsymbol{\theta}}_n$ and $\ell(\boldsymbol{\theta})$ is minimized by $\boldsymbol{\theta}_0$,

$$\begin{aligned}
\ell_n(\hat{\boldsymbol{\theta}}_n) \le \ell_n(\boldsymbol{\theta}_0) \quad &\implies \quad \ell_n(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}_0) \le \ell_n(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta}_0) \\
\ell(\boldsymbol{\theta}_0) \le \ell(\hat{\boldsymbol{\theta}}_n) \quad &\implies \quad \ell(\boldsymbol{\theta}_0) - \ell_n(\hat{\boldsymbol{\theta}}_n) \le \ell(\hat{\boldsymbol{\theta}}_n) - \ell_n(\hat{\boldsymbol{\theta}}_n) \\
&\implies \quad -\Big(\ell(\hat{\boldsymbol{\theta}}_n) - \ell_n(\hat{\boldsymbol{\theta}}_n)\Big) \le \ell_n(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}_0) \le \ell_n(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta}_0).
\end{aligned}$$

Therefore,

$$\begin{aligned}\left|\ell_n(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}_0)\right| &\leq \max\left\{\left|\ell(\hat{\boldsymbol{\theta}}_n) - \ell_n(\hat{\boldsymbol{\theta}}_n)\right|,\ \left|\ell_n(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta}_0)\right|\right\} \\ &\leq \sup_{\boldsymbol{\theta}\in\Theta} |\ell_n(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})| \to 0\end{aligned}$$

as $n \to \infty$. By the triangle inequality, we further have

$$\begin{aligned}\left|\ell(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}_0)\right| &= \left|\ell(\hat{\boldsymbol{\theta}}_n) - \ell_n(\hat{\boldsymbol{\theta}}_n) + \ell_n(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}_0)\right| \\ &\leq \left|\ell(\hat{\boldsymbol{\theta}}_n) - \ell_n(\hat{\boldsymbol{\theta}}_n)\right| + \left|\ell_n(\hat{\boldsymbol{\theta}}_n) - \ell(\boldsymbol{\theta}_0)\right| \\ &\leq 2\sup_{\boldsymbol{\theta}\in\Theta} |\ell_n(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})| \to 0\end{aligned}$$

as $n \to \infty$, such that $\lim_{n\to\infty} \hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_0$ as well. □

We now establish a uniform bound on $|\mathcal{S}_{f_s}(f) - \mathcal{S}(f)|$, a crucial ingredient in the proof of our main results.

Lemma B.1.2. *Suppose that $\mathcal{S}(f) = \mathcal{O}(1/|f|^{1+\delta})$ as $|f| \to \infty$ for $\delta > 0$. Then there exists f_0 such that for $f_s > 2f_0$,*

$$0 \leq \mathcal{S}_{f_s}(f) - \mathcal{S}(f) \leq M\left[\frac{1}{(f_s - |f|)^{1+\delta}} + \frac{1}{\delta f_s (f_s - |f|)^{\delta}}\right] \leq M\left[\frac{1}{(f_s/2)^{1+\delta}} + \frac{1}{\delta f_s (f_s/2)^{\delta}}\right]$$

for any $|f| < f_s/2$. Consequently, we have $\lim_{f_s\to\infty} \sup_{|f|<f_s} |\mathcal{S}_{f_s}(f) - \mathcal{S}(f)| = 0$.

Proof. To establish the upper bound, by the tail assumption $\mathcal{S}(f) = \mathcal{O}(1/|f|^{1+\delta})$ as $|f| \to \infty$, we have $\mathcal{S}(f \pm nf_s) \leq \frac{1}{2}M/|f \pm nf_s|^{1+\delta}$ for $|f \pm nf_s| > f_0$, which is the case for any $|f| < f_s/2$ as long as $f_s/2 > f_0$. Thus,

$$\begin{aligned}0 \leq \mathcal{S}_{f_s}(f) - \mathcal{S}(f) &= \sum_{n=1}^{\infty} \mathcal{S}(f - nf_s) + \mathcal{S}(f + nf_s) \\ &\leq \frac{M}{2}\sum_{n=1}^{\infty} \frac{1}{(nf_s - f)^{1+\delta}} + \frac{1}{(nf_s + f)^{1+\delta}}\end{aligned}$$

for any $|f| < f_s/2$. By the integral convergence test we have

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{(nf_s - f)^{1+\delta}} &\leq \frac{1}{(f_s - f)^{1+\delta}} + \int_1^{\infty} \frac{1}{(xf_s - f)^{1+\delta}} \,\mathrm{d}x \\ &= \frac{1}{(f_s - f)^{1+\delta}} + \frac{1}{\delta f_s (f_s - f)^{\delta}} \\ &\leq \frac{1}{(f_s - |f|)^{1+\delta}} + \frac{1}{\delta f_s (f_s - |f|)^{\delta}},\end{aligned}$$

and similarly

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{(nf_s + f)^{1+\delta}} &\leq \frac{1}{(f_s + f)^{1+\delta}} + \frac{1}{\delta f_s (f_s + f)^{\delta}} \\ &\leq \frac{1}{(f_s - |f|)^{1+\delta}} + \frac{1}{\delta f_s (f_s - |f|)^{\delta}},\end{aligned}$$

which completes the proof. □

Remark B.1.1. *If $\mathcal{S}(f)$ is continuous and strictly positive in $\mathbb{R}$ and $\mathcal{S}(f) \sim C|f|^{-2d(\boldsymbol{\theta})}$, as $f \to 0$ for $-\frac{1}{2} < d(\boldsymbol{\theta}) < \frac{1}{2}$, then there exists f_0 such that for $f_s > 2f_0$,*

$$0 < \frac{\mathcal{S}_{f_s}(f)}{\mathcal{S}(f)} - 1 \leq M/C \left[\frac{|f|^{-2d}}{(f_s - |f|)^{1+\delta}} + \frac{|f|^{-2d}}{\delta f_s (f_s - |f|)^{\delta}} \right] < M' \frac{|f|^{-2d}}{(f_s - |f|)^{1+\delta}}, \quad |f| \to 0.$$

Thus, for arbitrarily small $\epsilon > 0$,

$$\begin{aligned}\int_0^{f_{\max}} \left(\frac{\mathcal{S}_{f_s}(f)}{\mathcal{S}(f)} - 1 \right) \mathrm{d}f &< M' \left[\int_0^{\epsilon} \frac{f^{-2d}}{(f_s - 1)^{1+\delta_1}} \mathrm{d}f + \int_{\epsilon}^{f_{\max}} \frac{1/\mathcal{S}(f)}{(f_s - |f|)^{1+\delta_1}} \mathrm{d}f \right] \\ &= M' \left[\frac{1}{(f_s - 1)^{1+\delta_1}} \cdot \frac{1}{(1-2d)} f^{1-2d} \Big|_0^1 + \frac{1}{(f_s - f_{\max})^{1+\delta_1}} \int_{\epsilon}^{f_{\max}} \frac{1}{\mathcal{S}(f)} \mathrm{d}f \right] \\ &\to 0\end{aligned}$$

as $f_s \to \infty$, for any fixed $f_{\max} > 0$, $-\frac{1}{2} < d = d(\boldsymbol{\theta}) < \frac{1}{2}$, $\delta_1, \delta_2 > 0$.

Similarly, by using the fact $\log(1+x) \leq x$ for all $x > -1$, we can also obtain

$$\int_0^{f_{\max}} \left(\log \frac{\mathcal{S}_{f_s}(f)}{\mathcal{S}(f)} \right)^2 \mathrm{d}f \to 0$$

as $f_s \to \infty$.

Moreover, for short memory ($d = 0$) *and long memory* ($0 < d < \frac{1}{2}$) *processes, since* $|f|^{-2d} \leq 1$ *as* $|f| \to 0$, *we further have*

$$\lim_{f_s \to \infty} \sup_{|f| < f_{\max}} \frac{\mathcal{S}_{f_s}(f)}{\mathcal{S}(f)} = 1.$$

B.1.1 Technical Details in Remark 3.3.1

If $\mathcal{S}(f, \boldsymbol{\theta}) = \mathcal{O}(1/|f|^{1+\delta(\boldsymbol{\theta})})$ as $|f| \to \infty$, for the CARFIMA(p, H, q) model with $p = q$ and H being unknown, we have

$$\begin{aligned}
\mathcal{I}_{f_{\max}} &:= \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}(f, \boldsymbol{\theta}_0)\right) \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}(f, \boldsymbol{\theta}_0)\right)' \mathrm{d}f \\
&= \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \left(\frac{\varphi'(H)^2}{\varphi(H)^2} - 4\frac{\varphi'(H)}{\varphi(H)} \log|f| + 4(\log|f|)^2 \right) \mathrm{d}f \\
&= \frac{1}{2}\frac{\varphi'(H)^2}{\varphi(H)^2} - \frac{1}{f_{\max}}\frac{\varphi'(H)}{\varphi(H)} \int_{-f_{\max}}^{f_{\max}} \log|f| \, \mathrm{d}f + \frac{1}{f_{\max}} \int_{-f_{\max}}^{f_{\max}} (\log|f|)^2 \, \mathrm{d}f \\
&= \frac{1}{2}\frac{\varphi'(H)}{\varphi(H)} - 2\frac{\varphi'(H)}{\varphi(H)} \left(\log(f_{\max}) - 1\right) + 2\left[(\log f_{\max})^2 - 2\log F + 2\right] \\
&\sim 2(\log f_{\max})^2 \to \infty
\end{aligned}$$

as $f_{\max} \to \infty$. In such a case, the asymptotic variance of our bandlimited estimator cannot be arbitrarily close to that of the full Whittle estimator given by Dahlhaus (1989), since the amount of information accumulated can never converge.

If we set the CARFIMA model with $p = q$ with H being known while σ is the only unknown parameter, i.e. $\mathcal{S}(f, \sigma) \sim \sigma \cdot \varphi(H)|f|^{1-2H}$. Thus, we can set a large enough bandwidth $f_{\max}$ such that

$$\begin{aligned}
\mathcal{I}_{f_{\max}} &\to \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}_{f_s}(f, \boldsymbol{\theta}_0)\right) \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}_{f_s}(f, \boldsymbol{\theta}_0)\right)' \mathrm{d}f \quad (\text{as } f_s \to \infty) \\
&\to \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}_{f_s}(x, \boldsymbol{\theta})\right) \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{S}_{f_s}(x, \boldsymbol{\theta})\right)' \mathrm{d}x \quad (\text{as } f_{\max} \to \infty)
\end{aligned}$$

which is the information matrix of the full Whittle estimator given by Dahlhaus (1989).

B.2 Proofs of Theorems

B.2.1 Proof of Theorem 3.3.2

Proof. (i) Under Assumptions 3.3.3 – 3.3.5, by Lemma 1 of Fox and Taqqu (1986) we have

$$\lim_{N\to\infty} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}) - \tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_0(\boldsymbol{\theta})| = 0. \tag{B.1}$$

Note that $\frac{\partial}{\partial\boldsymbol{\theta}}\ell_0(\boldsymbol{\theta}_0) = \mathbf{0}$ is easily established by passing the derivative under the integral sign (Assumption 3.3.2), such that $\lim_{N\to\infty}\boldsymbol{\theta}_N \to \boldsymbol{\theta}_0$ by Lemma B.1.1. Moreover, Since for $N > N_0$, $\hat{\boldsymbol{\theta}}_N$ is in a convex compact neighborhood of $\boldsymbol{\theta}_0$, by the mean value theorem for $N > N_0$ we have

$$\begin{aligned} & \tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\hat{\boldsymbol{\theta}}_N) = \tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}_0) + \tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_N(\hat{\boldsymbol{\theta}}_c)(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \\ \Longrightarrow \quad & \sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) = [\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_N(\hat{\boldsymbol{\theta}}_c)]^{-1}\cdot\sqrt{N}[-\tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}_0)], \end{aligned}$$

where $\hat{\boldsymbol{\theta}}_c = c\hat{\boldsymbol{\theta}}_N + (1-c)\boldsymbol{\theta}_0$ for $0 < c < 1$, and the second line holds since by construction $\frac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\hat{\boldsymbol{\theta}}_N) = \mathbf{0}$. Since (B.1) implies that $\mathbf{E}[\frac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}_0)] \to \frac{\partial}{\partial\boldsymbol{\theta}}\ell_0(\boldsymbol{\theta}_0) = 0$, by Proposition 1 of Fox and Taqqu (1986), we have

$$\sqrt{N}\left(-\tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}_0) - \mathbf{0}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{B}),$$

and since $|\hat{\boldsymbol{\theta}}_c - \boldsymbol{\theta}_0| \le |\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0|$, by Lemma 1 of Fox and Taqqu (1986) we have

$$\lim_{N\to\infty} \tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_N(\hat{\boldsymbol{\theta}}_c) = \tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_0(\boldsymbol{\theta}_0) = \boldsymbol{A},$$

such that

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \to [\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_N(\boldsymbol{\theta}_0)]^{-1}\cdot\sqrt{N}[-\tfrac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}_0)] \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\ \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\right).$$

□

B.2.2 Proof of Theorem 3.3.1

Proof. Let us consider for bounded frequencies $|f| \leq f_{\max}$

$$\ell_0^{\dagger}(\boldsymbol{\theta}) = \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \log \mathcal{S}(f,\boldsymbol{\theta}) + \frac{\mathcal{S}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\,\mathrm{d}f,$$
$$\ell_0^{f_s}(\boldsymbol{\theta}) = \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \log \mathcal{S}(f,\boldsymbol{\theta}) + \frac{\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\,\mathrm{d}f.$$

By Assumption 3.3.2, we may interchange integral and derivatives to obtain $\frac{\partial}{\partial \boldsymbol{\theta}} \ell_0^{\dagger}(\boldsymbol{\theta}_0) = \mathbf{0}$, which implies that $\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} \ell_0^{\dagger}(\boldsymbol{\theta})$.

Let $\boldsymbol{\theta}_{f_s} = \arg\min_{\boldsymbol{\theta}} \ell_0^{f_s}(\boldsymbol{\theta})$ and

$$\hat{\boldsymbol{\theta}}_{\mathrm{W}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \log \mathcal{S}(f,\boldsymbol{\theta}) + \frac{I_N(f)}{\mathcal{S}(f,\boldsymbol{\theta})}\mathrm{d}f$$

such that by Theorem 3.3.2 we have $\lim_{N\to\infty} \hat{\boldsymbol{\theta}}_{\mathrm{W}} = \boldsymbol{\theta}_{f_s}$. Moreover, for fixed $f_{\max}$ we have

$$\left|\ell_0^{f_s}(\boldsymbol{\theta}) - \ell_0^{\dagger}(\boldsymbol{\theta})\right| = \frac{1}{4f_{\max}} \int_{-f_{\max}}^{f_{\max}} \left|\frac{\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0) - \mathcal{S}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right| \mathrm{d}f \to 0$$

by noting the compactness of $\boldsymbol{\Theta}$ to apply Remark B.1.1 following Lemma B.1.2 globally for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By Lemma B.1.1 it now follows that $\lim_{f_s\to\infty} \boldsymbol{\theta}_{f_s} = \boldsymbol{\theta}_0$.

As for the asymptotic covariance, by Theorem 3.3.2 as $N \to \infty$ we have

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\mathrm{W}} - \boldsymbol{\theta}_{f_s}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\ \boldsymbol{A}_{f_s}^{-1}\boldsymbol{B}_{f_s}\boldsymbol{A}_{f_s}^{-1}\right),$$

where $\boldsymbol{A}_{f_s} = \boldsymbol{A}_{f_s}(\boldsymbol{\theta}_{f_s})$, $\boldsymbol{B}_{f_s} = \boldsymbol{B}_{f_s}(\boldsymbol{\theta}_{f_s})$, and

$$\begin{aligned}
\boldsymbol{A}_{f_s} &= \tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_0^{f_s}(\boldsymbol{\theta}_{f_s}) \\
&= \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(1-\frac{\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right)\frac{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathcal{S}(f,\boldsymbol{\theta})}{\mathcal{S}(f,\boldsymbol{\theta})}\mathrm{d}f \\
&\quad - \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(1-\frac{2\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right)\boldsymbol{C}(f,\boldsymbol{\theta})\mathrm{d}f, \\
\boldsymbol{B}_{f_s}(\boldsymbol{\theta}) &= \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta})^{-1}\right)\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta})^{-1}\right)'\mathcal{S}_{f_s}(f,\boldsymbol{\theta})^2\mathrm{d}f \\
&= \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(\frac{\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right)^2\cdot\boldsymbol{C}(f,\boldsymbol{\theta})\mathrm{d}f \\
\boldsymbol{C}(f,\boldsymbol{\theta}) &= \left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta}_0)\right)\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta}_0)\right)'.
\end{aligned}$$

By Lemma B.1.2 and dominated convergence, for any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ we have $\lim_{f_s\to\infty}\boldsymbol{A}_{f_s}(\boldsymbol{\theta}) = \boldsymbol{A}(\boldsymbol{\theta})$ and $\lim_{f_s\to\infty}\boldsymbol{B}_{f_s}(\boldsymbol{\theta}) = \boldsymbol{B}(\boldsymbol{\theta})$, where

$$\begin{aligned}
\boldsymbol{A}(\boldsymbol{\theta}) &= \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(1-\frac{\mathcal{S}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right)\frac{\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathcal{S}(f,\boldsymbol{\theta})}{\mathcal{S}(f,\boldsymbol{\theta})}\mathrm{d}f \\
&\quad - \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(1-\frac{2\mathcal{S}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right)\boldsymbol{C}(f,\boldsymbol{\theta})\mathrm{d}f, \\
BB_{f_s}(\boldsymbol{\theta}) &= \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(\frac{\mathcal{S}(f,\boldsymbol{\theta}_0)}{\mathcal{S}(f,\boldsymbol{\theta})}\right)^2\cdot\boldsymbol{C}(f,\boldsymbol{\theta})\mathrm{d}f.
\end{aligned}$$

Since $\lim_{f_s\to\infty}\boldsymbol{\theta}_{f_s} = \boldsymbol{\theta}_0$, it follows that $lim_{f_s\to\infty}\boldsymbol{A}_{f_s}(\boldsymbol{\theta}) = \boldsymbol{A}(\boldsymbol{\theta}_0)$ and $\lim_{f_s\to\infty}\boldsymbol{B}_{f_s}(\boldsymbol{\theta}) = \boldsymbol{B}(\boldsymbol{\theta}_0)$. Straightforward algebraic manipulation show that $\boldsymbol{A}(\boldsymbol{\theta}_0) = \boldsymbol{B}(\boldsymbol{\theta}_0) = \mathcal{I}_{f_{\max}}$, where

$$\mathcal{I}_{f_{\max}} = \frac{1}{4f_{\max}}\int_{-f_{\max}}^{f_{\max}}\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta}_0)\right)\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta}_0)\right)'\mathrm{d}f.$$

Therefore, $\lim_{f_s\to\infty}\boldsymbol{A}_{f_s}^{-1}\boldsymbol{B}_{f_s}\boldsymbol{A}_{f_s}^{-1} = \mathcal{I}_{f_{\max}}$, which completes the proof. □

B.2.3 Summary of the Main Results in Moulines and Soulier (1999)

Our later proof relies heavily on Theorem 2 and Theorem 3 of Moulines and Soulier (1999) which we summarize here.

Assumption B.2.1. *B0 Suppose* $\mathcal{S}_{f_s}(f) = \left|1 - e^{if}\right|^{-2d} \mathcal{S}^{\star}_{f_s}(f)$ *where* $-\frac{1}{2} < d < \frac{1}{2}$ *and* $\mathcal{S}^{\star}_{f_s}(f, \boldsymbol{\theta})$ *is a* 2π*-periodic positive continuous function, differentiable on* $[-f_s/2, f_s/2] \setminus \{0\}$*, and there exists a finite constant* $C^{\star}$ *such that* $\forall f \in [-f_s/2, f_s/2] \setminus \{0\}$*,*

$$\left| \frac{\mathrm{d}}{\mathrm{d}f} \mathcal{S}^{\star}_{f_s}(f) \right| \leq \frac{C^{\star}}{|f|}.$$

Theorem B.2.1 (Moulines and Soulier, 1999)**.** *If* $X_0, X_1, \ldots$ *is a stationary Gaussian process and B.2.1 holds, then for all* $1 \leq m \leq N_B$ *we have*

$$Z_m = \log \bar{Y}_m = \log \mathcal{S}_{f_s}(\bar{f}_m) + \eta_{N,m} + r_{N,m},$$

where $X_{N,m} = \exp(\eta_{N,m} + C_B) = W'_{N,m} W_{N,m}$*, each* $W_{N,m} = (W^{(1)}_{N,m}, \ldots, W^{(2B)}_{N,m})$ *is a* $2B$*-dimensional vector of i.i.d. standard Gaussian random variables, and*

$$\sup_{1 \leq i,j \leq 2B} |\operatorname{cor}(W^{(i)}_{N,k}, W^{(j)}_{N,m})| \leq c_1 \log(m) k^{-|d|} m^{|d|-1}, \qquad 1 \leq k < m \leq N_B,$$

$$|r_{N,m}| < c_d \log(1+m)/m.$$

Moreover, for two nondecreasing integer sequences $\{v_N\}_{N \geq 0}$ *and* $\{w_N\}_{N \geq 0}$ *such that* $1 \leq v_N \leq w_N \leq N_B$*. Let* $\beta_{N,m}$ *be a triangular array of nonidentically vanishing real numbers, and define*

$$S_N = \sum_{m=v_N}^{w_N} \beta_{N,m} \eta_{N,m} \qquad s_N^2 = \sum_{m=v_N}^{w_N} \beta_{N,m}^2,$$

$$a_N = \sum_{m=v_N}^{w_N} |\beta_{N,m}|, \qquad b_N = \max_{v_N \leq m \leq w_N} |\beta_{N,m}|.$$

Then if (i) $\lim_{N \to \infty} b_N / s_N \to 0$ *and (ii)* $\lim_{N \to \infty} \frac{a_N \log(N)}{s_N v_N} = 0$*, we have the central limit theorem*

$$s_N^{-1} S_N \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \psi'(B)\right),$$

where $\psi'(\cdot)$ *is the derivative of the digamma function* $\psi(\cdot)$*.*

Among other things, it follows from Theorem B.2.1 that $X_{N,m} \sim \chi^2_{(2m)}$ and are asymptotically independent, from which follows Proposition 1.

B.2.4 Proof of Theorem 3.4.2

Proof. The proof of strong consistency is a straightforward application of Jennrich (1969), Theorem 6, which we omit for brevity. The proof of asymptotic normality leverages Theorem 2 and Theorem 3 of Moulines and Soulier (1999) (hereby Theorem B.2.1 above).

By an identical argument to the one in the proof of Theorem 3.3.2, for $N > N_0$ we have

$$\begin{aligned} & \tfrac{\partial}{\partial \boldsymbol{\theta}} Q_N(\hat{\boldsymbol{\theta}}_N) = \tfrac{\partial}{\partial \boldsymbol{\theta}} Q_N(\boldsymbol{\theta}_0) + \tfrac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_N(\hat{\boldsymbol{\theta}}_c) \cdot \left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0 \right) \\ \implies \quad & \sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) = \left[\tfrac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_N(\hat{\boldsymbol{\theta}}_c) \right]^{-1} \cdot \sqrt{N} \left(-\tfrac{\partial}{\partial \boldsymbol{\theta}} Q_N(\boldsymbol{\theta}_0) \right), \end{aligned}$$

where $\hat{\boldsymbol{\theta}}_c = c\hat{\boldsymbol{\theta}}_N + (1-c)\boldsymbol{\theta}_0$ for $0 < c < 1$. The proof will be complete upon (i) establishing a central limit theorem for $-\sqrt{N}\frac{\partial}{\partial \boldsymbol{\theta}} Q_N(\boldsymbol{\theta}_0)$ and (ii) showing that $\lim_{N\to\infty} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_N(\hat{\boldsymbol{\theta}}_c) = 2\mathcal{I}(\boldsymbol{\theta}_0) - \boldsymbol{A}(\boldsymbol{\theta}_0)$.

The Limit of $-\sqrt{N}\frac{\partial}{\partial \boldsymbol{\theta}} Q_N(\boldsymbol{\theta}_0)$

Note that

$$\begin{aligned} -\sqrt{N}\tfrac{\partial}{\partial \boldsymbol{\theta}} Q_N(\boldsymbol{\theta}_0) &= \sqrt{N} \cdot \frac{1}{N_B} \sum_{m=1}^{N_B} \left(\log \bar{Y}_m - C_B - \log g(\bar{f}_m, \boldsymbol{\theta}_0) \right) \cdot \tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0) \\ &= \sqrt{N} \cdot \frac{1}{N_B} \sum_{m=1}^{N_B} \tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0) \cdot e_{N,m}, \end{aligned} \tag{B.2}$$

where by Theorem B.2.1 we have

$$\begin{aligned} e_{N,m} &= \log \bar{Y}_m - C_B - \log g(\bar{f}_m, \boldsymbol{\theta}_0) \\ &= \log \bar{Y}_m - C_B - \log \mathcal{S}_{f_s}(\bar{f}_m) + \log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0) \\ &= \eta_{N,m} + r_{N,m} + \log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0). \end{aligned}$$

Using the notation of Theorem B.2.1, let

$$\beta_{N,m} = \frac{1}{\sqrt{N_B}} \frac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0), \qquad v_N = \lceil f_{\min} N_B^{\delta} / f_s \rceil, \qquad w_N = N_B,$$

where $\frac{1}{2} < \delta < 1$ and $f_{\min} > 0$ is a constant. It follows that the lower frequency bound $v_N f_s / N = f_{\min} N_B^{\delta} / N \to 0$ as $N \to \infty$. We shall obtain the desired result by establishing a central limit theorem for $\sum_{m=v_N}^{w_N} \beta_{N,m} \eta_{N,m}$ and show that the remaining terms in (B.2) are negligible as $N \to \infty$. Thus,

$$\begin{aligned} s_N^2 = \sum_{m=v_N}^{w_N} \beta_{N,m}^2 &= \frac{1}{f_s} \cdot \frac{f_s}{N_B} \sum_{m=v_N}^{w_N} \left[\tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right] \left[\tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right]' \\ &\to \frac{1}{f_s} \int_0^{f_s} [\tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(f, \boldsymbol{\theta}_0)][\tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(f, \boldsymbol{\theta}_0)]' \, \mathrm{d}f = 2\mathcal{I}(\boldsymbol{\theta}_0) < \infty \end{aligned}$$

as $N \to \infty$. Moreover, Assumptions 3.3.3, 3.3.4 and 3.4.3 imply that

$$\begin{aligned} b_N &= \max_{v_N \le m \le w_N} |\beta_{N,m}| \\ &\le C_1 \frac{1}{\sqrt{N_B}} \left| \tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(f_s N^{\delta-1}/(2B)^{\delta}, \boldsymbol{\theta}_0) \right| \\ &= C_1 \frac{1}{\sqrt{N_B}} \left| \tfrac{\partial}{\partial \boldsymbol{\theta}} g^{-1}(f_s N^{\delta-1}/(2B)^{\delta}, \boldsymbol{\theta}_0) \cdot g(f_s N^{\delta-1}/(2B)^{\delta}, \boldsymbol{\theta}_0) \right| \\ &\le \frac{1}{\sqrt{N_B}} C_1 C^2 \left| c(f_s, B, \delta) N^{\delta-1} \right|^{-\alpha-\beta} \\ &= \mathcal{O}\left(\frac{1}{N^{\frac{1}{2}-(1-\delta)(\alpha+\beta)}} \right) \to 0 \end{aligned}$$

as $N \to \infty$ (where C and C_1 are some constants), such that $\lim_{N\to\infty} b_N / s_N = 0$. Similarly, by assumption 3.3.4 and 3.4.3 we have

$$a_N = \frac{\sqrt{N_B}}{f_s} \cdot \frac{f_s}{N_B} \sum_{m=v_N}^{w_N} \left| \tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0) \right| = \mathcal{O}(\sqrt{N_B}),$$

such that

$$0 < \lim_{N\to\infty} \frac{a_N \log N}{s_N v_N} \le \lim_{N\to\infty} C_2 N^{\frac{1}{2}-\delta} \log N = 0.$$

Thus, by the central limit theorem in Theorem B.2.1 we have

$$\sqrt{N} \cdot \left(\frac{1}{N_B} \sum_{m=v_N}^{w_N} \tfrac{\partial}{\partial \boldsymbol{\theta}} \log(g(\bar{f}_m, \boldsymbol{\theta}_0)) \cdot \eta_{N,m} \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, 4B\psi'(B)\mathcal{I}(\boldsymbol{\theta}_0)\right).$$

Now let us decompose

$$\begin{aligned}\sum_{m=1}^{N_B} \beta_{N,m} e_{N,m} = &\left(S_N + \sum_{m=1}^{v_N - 1} \beta_{N,m} \eta_{N,m} \right) \\ &+ \sum_{m=1}^{N_B} \beta_{N,m} r_{N,m} + \sum_{m=1}^{N_B} \beta_{N,m} \left(\log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0) \right).\end{aligned}$$

To establish our central limit theorem for $\sum_{m=1}^{N_B} \beta_{N,m} e_{N,m}$, we need to show that S_N is the dominant part. So first, we note that

$$\begin{aligned}\sum_{m=1}^{N_B} \beta_{N,m} r_{N,m} &\leq \frac{C}{\sqrt{N_B}} \sum_{m=1}^{N_B} \frac{c_d \log(1+m)}{m} \\ &\leq \frac{C}{\sqrt{N_B}} \left(\log(2) + \int_2^{N_B} \frac{\log(1+x)}{x} \mathrm{d}x \right) \qquad \text{(by the integral test)} \\ &\leq \frac{C}{\sqrt{N_B}} \left(\log(2) + \int_2^{N_0} \frac{\log(1+x)}{x} \mathrm{d}x + \int_{N_0}^{N_B} \frac{x^\epsilon}{x} \mathrm{d}x \right) \\ &= \mathcal{O}\left(\frac{1}{\sqrt{N_B}} \int_{N_0}^{N_B} x^{\epsilon - 1} \mathrm{d}x \right) \to 0\end{aligned}$$

as $N_B \to \infty$, where we used the fact that $f(x) = \log(1+x)/x$ is a positive, decreasing function for $x > 0$, C is some constant changing from step to step, and $\epsilon \in (0, \frac{1}{2})$, $N_0 \geq 2$

is a constant integer such that $\log(1+x) < x^\epsilon$ when $x > N_0$. Moreover,

$$\begin{aligned}
&\sum_{m=1}^{N_B} \beta_{N,m} \left[\log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right] \\
=&\frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} \tfrac{\partial}{\partial \boldsymbol{\theta}} \log(g(\bar{f}_m, \boldsymbol{\theta}_0)) \cdot \left[\log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right] \\
=&\frac{N_B}{\sqrt{N_B} f_s} \cdot \frac{f_s}{N_B} \sum_{m=1}^{N_B} \tfrac{\partial}{\partial \boldsymbol{\theta}} \log(g(\bar{f}_m, \boldsymbol{\theta}_0)) \cdot \left[\log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right] \\
=&\frac{\sqrt{N_B}}{f_s} \cdot \sum_{m=1}^{N_B} \tfrac{\partial}{\partial \boldsymbol{\theta}} \log(g(\bar{f}_m, \boldsymbol{\theta}_0)) \cdot \left[\log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right] \Delta f_m \\
\sim&\frac{\sqrt{N_B}}{f_s} \cdot \mathcal{O}(1/N_B) \to 0
\end{aligned}$$

as $N_B \to \infty$, where the last step is because the Riemann sum converges to its limit $\frac{\partial}{\partial \boldsymbol{\theta}} D(\boldsymbol{\theta}_0) = \mathbf{0}$ with a rate faster than $\sqrt{N_B}$ (e.g., Chui, 1971, Theorem 1). Finally, $\mathbf{E}\left(\sum_{m=1}^{v_N} \beta_{N,m}\eta_{N,m}\right)^2 = o(1)$ since by the regularity assumptions on $g(f, \boldsymbol{\theta})$ and Theorem B.2.1,

$$\mathbf{E}\left(\sum_{m=1}^{v_N} \beta_{N,m}\eta_{N,m}\right)^2 = \mathcal{O}\left(v_N/N_B\right) = o(1).$$

The Limit of $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_N(\boldsymbol{\theta}_c)$

Note that

$$\begin{aligned}
\tfrac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} Q_N(\boldsymbol{\theta}_0) = &\frac{1}{N_B} \sum_{m=1}^{N_B} \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right) \left(\tfrac{\partial}{\partial \boldsymbol{\theta}} \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right)' \\
&- \frac{1}{N_B} \sum_{m=1}^{N_B} \left(\eta_{N,m} + r_{N,m} + \log \mathcal{S}_{f_s}(\bar{f}_m) - \log g(\bar{f}_m, \boldsymbol{\theta}_0)\right) \tfrac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log g(\bar{f}_m, \boldsymbol{\theta}_0),
\end{aligned}$$

Under the regularity conditions and by Theorem B.2.1, similarly we can show as $N_B \to \infty$ that

$$\frac{1}{N_B}\sum_{m=1}^{N_B}\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log g(\bar{f}_m,\boldsymbol{\theta}_0)\right)\left(\tfrac{\partial}{\partial\boldsymbol{\theta}}\log g(\bar{f}_m,\boldsymbol{\theta}_0)\right)' \to 2\,\mathcal{I}(\boldsymbol{\theta}_0)$$

$$\frac{1}{N_B}\sum_{m=1}^{N_B}(\eta_{N,m}+r_{N,m})\,\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log g(\bar{f}_m,\boldsymbol{\theta}_0) \to \mathbf{0}$$

$$\frac{1}{N_B}\sum_{m=1}^{N_B}\left(\log\mathcal{S}_{f_s}(\bar{f}_m)-\log g(\bar{f}_m,\boldsymbol{\theta}_0)\right)\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log g(\bar{f}_m,\boldsymbol{\theta}_0) \to \boldsymbol{A}(\boldsymbol{\theta}_0)$$

where

$$\boldsymbol{A}(\boldsymbol{\theta}_0)=\frac{1}{f_s}\int_0^{f_s}(\log\mathcal{S}_{f_s}(f)-\log g(f,\boldsymbol{\theta}_0))\,\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log g(f,\boldsymbol{\theta}_0)\mathrm{d}f.$$

Thus, $\lim_{N\to\infty}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}Q_N(\boldsymbol{\theta}_0)=2\mathcal{I}(\boldsymbol{\theta}_0)-\boldsymbol{A}(\boldsymbol{\theta}_0)$. Finally, since $\lim_{N\to\infty}\hat{\boldsymbol{\theta}}_c=\boldsymbol{\theta}_0$ and $Q_N(\boldsymbol{\theta})$ is continuous about $\boldsymbol{\theta}_0$, we obtain the desired limit $\lim_{N\to\infty}Q_N(\hat{\boldsymbol{\theta}}_c)=2\mathcal{I}(\boldsymbol{\theta}_0)-\boldsymbol{A}(\boldsymbol{\theta}_0)$, which completes the proof of Theorem 3.4.2. □

B.2.5 Proof of Theorem 3.4.1

Proof. Let

$$D_0^{\dagger}(\boldsymbol{\theta})=\frac{1}{f_{\max}}\int_0^{f_{\max}}\left(\log\mathcal{S}(f,\boldsymbol{\theta})-\log\mathcal{S}(f,\boldsymbol{\theta}_0)\right)^2,$$

$$D_0^{f_s}(\boldsymbol{\theta})=\frac{1}{f_{\max}}\int_0^{f_{\max}}\left(\log\mathcal{S}(f,\boldsymbol{\theta})-\log\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0)\right)^2.$$

Then by Assumption 3.3.2 we may pass the derivative under the integral sign, such that $\frac{\partial}{\partial\boldsymbol{\theta}}D_0^{\dagger}(\boldsymbol{\theta}_0)=\mathbf{0}$, which implies that $\boldsymbol{\theta}_0=\arg\min D_0^{\dagger}(\boldsymbol{\theta})$. Moreover, by setting $v_N=\lceil f_{\min}N_B^{\delta}/f_s\rceil$ and $w_N=\lfloor f_{\max}N_B/f_s\rfloor$ for some constant frequencies $0<f_{\min}\le f_{\max}$, we can apply the same proof of Theorem 3.4.2 with $g(f,\boldsymbol{\theta})=\mathcal{S}(f,\boldsymbol{\theta})$ to show a central limit theorem for

$$\sum_{m=1}^{w_N}\beta_{N,m}e_{N,m}.$$

We find that $\lim_{N\to\infty}\hat{\boldsymbol{\theta}}_{\mathrm{LP}} = \boldsymbol{\theta}_{f_s}$, where $\boldsymbol{\theta}_{f_s} = \arg\min_{\boldsymbol{\theta}} D_0^{f_s}(\boldsymbol{\theta})$, and with corresponding asymptotic variance terms

$$\mathcal{I}(\boldsymbol{\theta}_{f_s}) = \frac{1}{2f_{\max}}\int_0^{f_{\max}} [\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta}_{f_s})][\tfrac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{S}(f,\boldsymbol{\theta}_{f_s})]'\mathrm{d}f$$
$$\boldsymbol{A}(\boldsymbol{\theta}_{f_s}) = \frac{1}{f_{\max}}\int_0^{f_{\max}} (\log\mathcal{S}_{f_s}(f,\boldsymbol{\theta}_0) - \log\mathcal{S}(f,\boldsymbol{\theta}_{f_s}))\,\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log\mathcal{S}(f,\boldsymbol{\theta}_{f_s})\mathrm{d}f.$$

Moreover,

$$\left|\sqrt{D_0^{\dagger}(\boldsymbol{\theta})} - \sqrt{D_0^{f_s}(\boldsymbol{\theta})}\right| \le \sqrt{\frac{1}{f_{\max}}\int_0^{f_{\max}} (\log\mathcal{S}_{f_s}(f,\boldsymbol{\theta}) - \log\mathcal{S}(f,\boldsymbol{\theta}))^2\,\mathrm{d}f}$$
$$= \sqrt{\frac{1}{f_{\max}}\int_0^{f_{\max}} \left(\log\frac{\mathcal{S}_{f_s}(f,\boldsymbol{\theta})}{\mathcal{S}(f,\boldsymbol{\theta})}\right)^2\mathrm{d}f} \to 0$$

as $f_s \to 0$, which follows from Remark B.1.1. Therefore, as a consequence of Lemma B.1.1 we have $\lim_{f_s\to\infty}\boldsymbol{\theta}_{f_s} = \boldsymbol{\theta}_0$. Similarly, by repeated application of Remark B.1.1 we find that $\lim_{f_s\to 0}\boldsymbol{A}(\boldsymbol{\theta}_{f_s}) = \mathbf{0}$, and $\lim_{f_s\to 0}\mathcal{I}(\boldsymbol{\theta}_{f_s}) = \mathcal{I}_{f_{\max}}$ as defined by (3.8). □

Appendix C

Appendix for Chapter 4

C.1 Covariance Structure of q-times Integrated Wiener Process

Generally, let $W^{(q)}(t)$ be the q-times integrated Wiener process, we have for $0 \leq t_1 \leq t_2$

$$\begin{aligned}
W^{(1)}(t_2) &= W^{(1)}(t_1) + \int_{t_1}^{t_2} W(s)\mathrm{d}s \\
&= W^{(1)}(t_1) + [sW(s)]\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} s\mathrm{d}W(s) \quad \text{(by Itô's lemma)} \\
&= W^{(1)}(t_1) + W(t_1)(t_2 - t_1) + t_2 \int_{t_1}^{t_2} \mathrm{d}W(s) - \int_{t_1}^{t_2} s\mathrm{d}W(s) \\
&= W^{(1)}(t_1) + W(t_1)(t_2 - t_1) + \int_{t_1}^{t_2} (t_2 - s)\mathrm{d}W(s)
\end{aligned}$$

Iteratively, we can derive the general formula

$$W^{(q)}(t_2) = \sum_{k=0}^{q} W^{(q-k)}(t_1)\frac{(t_2 - t_1)^k}{k!} + \int_{t_1}^{t_2} \frac{(t_2 - s)^q}{q!}\mathrm{d}W(s).$$

The covariance of the increments $\Delta W^{(i)} = W^{(i)}(t_2) - W^{(i)}(t_1)$ and $\Delta W^{(j)} = W^{(j)}(t_2) - W^{(j)}(t_1)$ $(0 \leq i, j \leq q)$ is given by

$$\begin{aligned}
\mathrm{Cov}(\Delta W^{(i)}, \Delta W^{(j)}) &= \mathbf{E}\left[\int_{t_1}^{t_2}\int_{t_1}^{t_2} \frac{(t_2 - s)^i (t_2 - r)^j}{i!j!} \mathrm{d}W(s)\mathrm{d}W(r)\right] \\
&= \int_{t_1}^{t_2} \frac{(t_2 - \tau)^i (t_2 - \tau)^j}{i!j!} \mathrm{d}\tau \quad \text{(by Itô's isometry)} \\
&= \frac{(t_2 - t_1)^{i+j+1}}{(i + j + 1) i! j!}
\end{aligned}$$

where we implicitly used the fact that

$$\mathbf{E}\left[\int f(\omega, t)\mathrm{d}W(t)\right] = 0$$

for all f such that $\mathbf{E}\left[\int f^2(\omega, t)\mathrm{d}t\right] < \infty$ since $W(t)$ is a martingale.

Also note that for the integrated Wiener increment process, using Itô's lemma we can easily derive

$$\int_{t_1}^{t_2} \left(W(s) - W(t_1)\right) \mathrm{d}s = \int_{t_1}^{t_2} (t_2 - s)\mathrm{d}W(s)$$

which contributes the same stochastic integral as the integrated Wiener process. Therefore, we can construct the covariance structure for $q = 2$, i.e. we obtain a multivariate normal random vector $\boldsymbol{W}_n = (\Delta W_n^{(2)}, \Delta W_n^{(1)}, \Delta W_n)$

$$\boldsymbol{W}_n \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \frac{(\Delta t)^5}{20} & \frac{(\Delta t)^4}{8} & \frac{(\Delta t)^3}{6} \\ \frac{(\Delta t)^4}{8} & \frac{(\Delta t)^3}{3} & \frac{(\Delta t)^2}{2} \\ \frac{(\Delta t)^3}{6} & \frac{(\Delta t)^2}{2} & \Delta t \end{pmatrix}$$

as desired.

C.2 Proof of Lemma 4.2.1

Proof of (4.8). It is clear to see that $p(W, X, Y)$ is a multivariate normal so it suffices to find its mean and variance. To that end, let $\boldsymbol{Z}_W, \boldsymbol{Z}_X, \boldsymbol{Z}_Y$ be independent vectors of i.i.d.

normal random variables of size corresponding to the dimensions of $\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{Y}$ then we have

$$\begin{aligned}
\boldsymbol{W} &= \boldsymbol{\mu}_W + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W \\
\boldsymbol{X} &= \boldsymbol{W} + \boldsymbol{\mu}_{X|W} + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X \\
&= \boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W} + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X \\
\boldsymbol{Y} &= \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\Sigma}_Y^{1/2}\boldsymbol{Z}_Y \\
&= \boldsymbol{A}[\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W} + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y \\
&= \boldsymbol{A}[\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W}] + \boldsymbol{A}[\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y.
\end{aligned}$$

The following steps will be computing the covariance between each of the variables using the equations derived above

$$\begin{aligned}
\operatorname{cov}(\boldsymbol{W}, \boldsymbol{X}) &= \operatorname{cov}(\boldsymbol{\mu}_W + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W, \boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W} + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X) \\
&= \operatorname{cov}(\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W, \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X) \\
&= \boldsymbol{\Sigma}_W \\
\operatorname{cov}(\boldsymbol{W}, \boldsymbol{Y}) &= \operatorname{cov}\left(\boldsymbol{\mu}_W + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W, \boldsymbol{A}[\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W}] + \boldsymbol{A}[\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y\right) \\
&= \operatorname{cov}\left(\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W, \boldsymbol{A}[\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y\right) \\
&= \boldsymbol{\Sigma}_W\boldsymbol{A}' \\
\operatorname{cov}(\boldsymbol{X}, \boldsymbol{Y}) &= \operatorname{cov}\Big(\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W} + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X, \boldsymbol{A}[\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W}] + \\
&\quad \boldsymbol{A}[\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y\Big) \\
&= \operatorname{cov}\left(\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X, \boldsymbol{A}[\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y\right) \\
&= (\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W})\boldsymbol{A}'.
\end{aligned}$$

Similarly, we have the variance of each variable

$$\begin{aligned}
\text{var}(\boldsymbol{W}) &= \text{var}(\boldsymbol{\mu}_W + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W) \\
&= \boldsymbol{\Sigma}_W \\
\text{var}(\boldsymbol{X}) &= \text{var}(\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W} + \boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X) \\
&= \boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W} \\
\text{var}(\boldsymbol{Y}) &= \text{var}(\boldsymbol{A}[\boldsymbol{\mu}_W + \boldsymbol{\mu}_{X|W}] + \boldsymbol{A}[\boldsymbol{\Sigma}_W^{1/2}\boldsymbol{Z}_W + \boldsymbol{\Sigma}_{X|W}^{1/2}\boldsymbol{Z}_X] + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_Y) \\
&= \boldsymbol{A}(\boldsymbol{\Sigma}_W + \boldsymbol{\Sigma}_{X|W})\boldsymbol{A}' + \boldsymbol{\Omega}.
\end{aligned}$$

□

Proof of (4.9). Given the joint distribution from (2), we can directly apply the formula for the conditional distribution of a multivariate normal to obtain $p(W|Y) \sim \mathcal{N}(\boldsymbol{\mu}_{W|Y}, \boldsymbol{\Sigma}_{W|Y})$. In this context we have

$$\begin{aligned}
\boldsymbol{\mu}_{W|Y} &= \boldsymbol{\mu}_W + \boldsymbol{\Sigma}_W\boldsymbol{A}'\boldsymbol{\Sigma}_Y^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}_Y) \\
\boldsymbol{\Sigma}_{W|Y} &= \boldsymbol{\Sigma}_W - \boldsymbol{\Sigma}_W\boldsymbol{A}'\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{A}\boldsymbol{\Sigma}_W
\end{aligned}$$

which completes the final formulation. □

C.3 A General Derivation of the Result in Section 4.3.5

Proposition C.3.1. *Let $\boldsymbol{W}(t)$ be a d-dimensional continuous-time Gaussian process with $E[\boldsymbol{W}(t)] = \boldsymbol{\mu}(t)$ such that $\boldsymbol{W}(t) - \boldsymbol{\mu}(t)$ is a homogeneous Markov process, i.e., we have matrix valued functions $\tilde{\boldsymbol{\Sigma}}(t)$ and $\tilde{\boldsymbol{R}}(t)$ such that*

$$\boldsymbol{W}(t+h) \mid \{\boldsymbol{W}(s) : s \leq t\} \sim \mathcal{N}\big(\boldsymbol{\mu}(t+h) + \tilde{\boldsymbol{R}}(h)(\boldsymbol{W}(t) - \boldsymbol{\mu}(t)), \tilde{\boldsymbol{\Sigma}}(h)\big).$$

Consider time points $t_0 < t_1 < t_2$, and suppose that

$$\boldsymbol{Y} \mid \boldsymbol{W}(t_0), \boldsymbol{W}(t_1), \boldsymbol{W}(t_2) \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{W}(t_2), \boldsymbol{\Omega}).$$

Then we can find the joint distribution $p(\boldsymbol{W}(t_2), \boldsymbol{W}(t_1), \boldsymbol{Y} \mid \boldsymbol{W}_0)$ and the conditional distribution $p(\boldsymbol{W}(t_1) \mid \boldsymbol{W}(t_0), \boldsymbol{Y})$ both of which are multivariate normal.

Remark C.3.1. *In Section 4.3.5, we actually have a special case where $\boldsymbol{W}(t_2) = \boldsymbol{W}_m^{(m)}$, $\boldsymbol{W}(t_1) = \boldsymbol{W}_{k+1}^{(m)}$ and $\boldsymbol{W}(t_0) = \boldsymbol{W}_k^{(m)}$.*

Proof. Let $\tilde{\boldsymbol{W}}(t) = \boldsymbol{W}(t) - \boldsymbol{\mu}(t)$, and let $\boldsymbol{W}_i = \boldsymbol{W}(t_i)$, $\tilde{\boldsymbol{W}}_i = \tilde{\boldsymbol{W}}(t_i)$, etc. First, we derive the joint distribution $p(\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_0)$ from the following factorization

$$\begin{aligned}
\tilde{\boldsymbol{W}}_1 \mid \tilde{\boldsymbol{W}}_0 &\sim \mathcal{N}(\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0, \tilde{\boldsymbol{\Sigma}}_1) \\
\tilde{\boldsymbol{W}}_2 \mid \tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_0 &\sim \mathcal{N}(\tilde{\boldsymbol{R}}_2 \tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{\Sigma}}_2) \\
\boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_2, \tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_0 &\sim \mathcal{N}(\boldsymbol{A}\tilde{\boldsymbol{W}}_2 + \boldsymbol{\beta}, \boldsymbol{\Omega}),
\end{aligned}$$

where $\tilde{\boldsymbol{\Sigma}}_i = \tilde{\boldsymbol{\Sigma}}(t_i - t_{i-1})$, $\tilde{\boldsymbol{R}}_i = \tilde{\boldsymbol{R}}(t_i - t_{i-1})$, and $\boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\mu}(t_2)$.

We know that $p(\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_0)$ is multivariate normal, so we just need to find its mean and variance. To do this, note that we can sample from $p(\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_0)$ by first generating $\boldsymbol{Z}_0, \boldsymbol{Z}_1, \boldsymbol{Z}_2$ as independent vectors of *i.i.d.* normal random variables of compatible dimension, then setting

$$\begin{aligned}
\tilde{\boldsymbol{W}}_1 &= \tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0 \\
\tilde{\boldsymbol{W}}_2 &= \tilde{\boldsymbol{R}}_2 \tilde{\boldsymbol{W}}_1 + \boldsymbol{L}_2 \boldsymbol{Z}_1 \\
&= \tilde{\boldsymbol{R}}_2 (\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0) + \boldsymbol{L}_2 \boldsymbol{Z}_1 \\
\boldsymbol{Y} &= \boldsymbol{A}\tilde{\boldsymbol{W}}_2 + \boldsymbol{\beta} + \boldsymbol{\Omega}^{1/2} \boldsymbol{Z}_2 \\
&= \boldsymbol{A}(\tilde{\boldsymbol{R}}_2 (\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0) + \boldsymbol{L}_2 \boldsymbol{Z}_1) + \boldsymbol{\beta} + \boldsymbol{\Omega}^{1/2} \boldsymbol{Z}_2,
\end{aligned}$$

where $\boldsymbol{L}_i$ is the lower Cholesky factor of $\tilde{\boldsymbol{\Sigma}}_i = \boldsymbol{L}_i \boldsymbol{L}_i'$. Thus, we may calculate

$$\begin{aligned}
\operatorname{cov}\left(\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2 \mid \tilde{\boldsymbol{W}}_0\right) &= \operatorname{cov}\left(\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0,\ \tilde{\boldsymbol{R}}_2 (\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0) + \boldsymbol{L}_2 \boldsymbol{Z}_1\right) \\
&= \operatorname{cov}\left(\boldsymbol{L}_1 \boldsymbol{Z}_0,\ \tilde{\boldsymbol{R}}_2 \boldsymbol{L}_1 \boldsymbol{Z}_0 + \boldsymbol{L}_2 \boldsymbol{Z}_1\right) \\
&= \boldsymbol{L}_1 \operatorname{cov}\left(\boldsymbol{Z}_0, \boldsymbol{Z}_0\right) \boldsymbol{L}_1' \tilde{\boldsymbol{R}}_2' + \boldsymbol{L}_1 \operatorname{cov}\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1\right) \boldsymbol{L}_2' \\
&= \tilde{\boldsymbol{\Sigma}}_1 \tilde{\boldsymbol{R}}_2'.
\end{aligned}$$

The rest can be calculated similarly

$$\begin{aligned}
\operatorname{cov}(\tilde{\boldsymbol{W}}_1, \boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_0) &= \operatorname{cov}\left(\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0,\ \boldsymbol{A}(\tilde{\boldsymbol{R}}_2 (\tilde{\boldsymbol{R}}_1 \tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1 \boldsymbol{Z}_0) + \boldsymbol{L}_2 \boldsymbol{Z}_1) + \boldsymbol{\beta} + \boldsymbol{\Omega}^{1/2} \boldsymbol{Z}_2\right) \\
&= \operatorname{cov}\left(\boldsymbol{L}_1 \boldsymbol{Z}_0,\ \boldsymbol{A}\tilde{\boldsymbol{R}}_2 \boldsymbol{L}_1 \boldsymbol{Z}_0 + \boldsymbol{A}\boldsymbol{L}_2 \boldsymbol{Z}_1 + \boldsymbol{\Omega}^{1/2} \boldsymbol{Z}_2\right) \\
&= \boldsymbol{L}_1 \operatorname{cov}\left(\boldsymbol{Z}_0, \boldsymbol{Z}_0\right) \boldsymbol{L}_1' \tilde{\boldsymbol{R}}_2' \boldsymbol{A}' + \boldsymbol{0} \\
&= \tilde{\boldsymbol{\Sigma}}_1 \tilde{\boldsymbol{R}}_2' \boldsymbol{A}'.
\end{aligned}$$

$$
\begin{aligned}
\operatorname{cov}(\tilde{\boldsymbol{W}}_2, \boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_0) &= \operatorname{cov}\left(\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{W}}_1 + \boldsymbol{L}_2\boldsymbol{Z}_1,\ \boldsymbol{A}(\tilde{\boldsymbol{R}}_2(\tilde{\boldsymbol{R}}_1\tilde{\boldsymbol{W}}_0 + \boldsymbol{L}_1\boldsymbol{Z}_0) + \boldsymbol{L}_2\boldsymbol{Z}_1) + \boldsymbol{\beta} + \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{Z}_2\right) \\
&= \operatorname{cov}\left(\tilde{\boldsymbol{R}}_2\boldsymbol{L}_1\boldsymbol{Z}_0 + \boldsymbol{L}_2\boldsymbol{Z}_1,\ \boldsymbol{A}\tilde{\boldsymbol{R}}_2\boldsymbol{L}_1\boldsymbol{Z}_0 + \boldsymbol{A}\boldsymbol{L}_2\boldsymbol{Z}_1 + \boldsymbol{\Omega}^{1/2}\boldsymbol{Z}_2\right) \\
&= \tilde{\boldsymbol{R}}_2\boldsymbol{L}_1 \operatorname{cov}(\boldsymbol{Z}_0, \boldsymbol{Z}_0)\boldsymbol{L}_1'\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' + \boldsymbol{L}_2 \operatorname{cov}(\boldsymbol{Z}_1, \boldsymbol{Z}_1)\boldsymbol{L}_2'\boldsymbol{A}' \\
&= \tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' + \tilde{\boldsymbol{\Sigma}}_2\boldsymbol{A}'.
\end{aligned}
$$

Once we have the joint distribution $p(\tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \boldsymbol{Y} \mid \tilde{\boldsymbol{W}}_0)$, we can use the formula for the conditional distribution of a multivariate normal to obtain $p(\tilde{\boldsymbol{W}}_1 \mid \tilde{\boldsymbol{W}}_0, \boldsymbol{Y})$, and then $p(\boldsymbol{W}_1 \mid \boldsymbol{W}_0, \boldsymbol{Y})$. The joint distribution of $p(\tilde{\boldsymbol{W}}_1 \mid \tilde{\boldsymbol{W}}_0, \boldsymbol{Y})$ is given by

$$
\begin{bmatrix} \tilde{\boldsymbol{W}}_1 \\ \tilde{\boldsymbol{W}}_2 \\ \boldsymbol{Y} \end{bmatrix} |\tilde{\boldsymbol{W}}_0 \sim \mathcal{N}\left(\begin{bmatrix} \tilde{\boldsymbol{R}}_1\tilde{\boldsymbol{W}}_0 \\ \tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{R}}_1\tilde{\boldsymbol{W}}_0 \\ \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{R}}_1\tilde{\boldsymbol{W}}_0 + \boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2' & \tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' \\ \tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1 & \tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2' + \tilde{\boldsymbol{\Sigma}}_2 & \tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' + \tilde{\boldsymbol{\Sigma}}_2\boldsymbol{A}' \\ \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1 & \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2' + \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_2 & \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' + \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_2\boldsymbol{A}' + \boldsymbol{\Omega} \end{bmatrix}\right),
$$

and

$$
\begin{bmatrix} \boldsymbol{W}_1 \\ \boldsymbol{Y} \end{bmatrix} \mid \boldsymbol{W}_0 \sim \mathcal{N}\left(\begin{bmatrix} \tilde{\boldsymbol{R}}_1(\boldsymbol{W}_0 - \boldsymbol{\mu}(t_0)) + \boldsymbol{\mu}(t_1) \\ \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{R}}_1\boldsymbol{W}_0 \end{bmatrix}, \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' \\ \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1' & \boldsymbol{A}\tilde{\boldsymbol{R}}_2\tilde{\boldsymbol{\Sigma}}_1\tilde{\boldsymbol{R}}_2'\boldsymbol{A}' + \boldsymbol{A}\tilde{\boldsymbol{\Sigma}}_2\boldsymbol{A}' + \boldsymbol{\Omega} \end{bmatrix}\right).
$$

□