



HAL
open science

Strang splitting for parametric inference in second-order stochastic differential equations

Predrag Pilipovic, Adeline Samson, Susanne Ditlevsen

► **To cite this version:**

Predrag Pilipovic, Adeline Samson, Susanne Ditlevsen. Strang splitting for parametric inference in second-order stochastic differential equations. 2024. <hal-04600139>

HAL Id: hal-04600139

<https://hal.science/hal-04600139v1>

Preprint submitted on 4 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

STRANG SPLITTING FOR PARAMETRIC INFERENCE IN SECOND-ORDER STOCHASTIC DIFFERENTIAL EQUATIONS

A PREPRINT

✉ **Predrag Pilipovic**

Department of Mathematical Sciences
University of Copenhagen
2100 Copenhagen, Denmark
predrag@math.ku.dk
Bielefeld Graduate School of Economics and Management
University of Bielefeld
33501 Bielefeld, Germany
predrag.pilipovic@uni-bielefeld.de

Adeline Samson

Univ. Grenoble Alpes
CNRS, Grenoble INP, LJK
38000 Grenoble, France
adeline.leclercq-samson@univ-grenoble-alpes.fr

Susanne Ditlevsen

Department of Mathematical Sciences
University of Copenhagen
2100 Copenhagen, Denmark
susanne@math.ku.dk

ABSTRACT

We address parameter estimation in second-order stochastic differential equations (SDEs), prevalent in physics, biology, and ecology. Second-order SDE is converted to a first-order system by introducing an auxiliary velocity variable raising two main challenges. First, the system is hypoelliptic since the noise affects only the velocity, making the Euler-Maruyama estimator ill-conditioned. To overcome that, we propose an estimator based on the Strang splitting scheme. Second, since the velocity is rarely observed we adjust the estimator for partial observations. We present four estimators for complete and partial observations, using full likelihood or only velocity marginal likelihood. These estimators are intuitive, easy to implement, and computationally fast, and we prove their consistency and asymptotic normality. Our analysis demonstrates that using full likelihood with complete observations reduces the asymptotic variance of the diffusion estimator. With partial observations, the asymptotic variance increases due to information loss but remains unaffected by the likelihood choice. However, a numerical study on the Kramers oscillator reveals that using marginal likelihood for partial observations yields less biased estimators. We apply our approach to paleoclimate data from the Greenland ice core and fit it to the Kramers oscillator model, capturing transitions between metastable states reflecting observed climatic conditions during glacial eras.

Keywords Second-order stochastic differential equations, Hypoellipticity, Partial observations, Strang splitting estimator, Greenland ice core data, Kramers oscillator

1 Introduction

Second-order stochastic differential equations (SDEs) are an effective instrument for modeling complex systems showcasing both deterministic and stochastic dynamics, which incorporate the second derivative of a variable - the acceleration. These models are extensively applied in many fields, including physics (Rosenblum and Pikovsky, 2003), molecular dynamics (Leimkuhler and Matthews, 2015), ecology (Johnson et al., 2008; Michelot and Blackwell, 2019), paleoclimate research (Ditlevsen et al., 2002), and neuroscience (Ziv et al., 1994; Jansen and Rit, 1995).

The general form of a second-order SDE in Langevin form is given as follows:

$$\ddot{\mathbf{X}}_t = \mathbf{F}(\mathbf{X}_t, \dot{\mathbf{X}}_t; \boldsymbol{\beta}) + \boldsymbol{\Sigma} \boldsymbol{\xi}_t. \quad (1)$$

Here, $\mathbf{X}_t \in \mathbb{R}^d$ denotes the variable of interest, the dot indicates derivative with respect to time t , drift \mathbf{F} represents the deterministic force, and $\boldsymbol{\xi}_t$ is a white noise representing the system's random perturbations around the deterministic force. We assume that $\boldsymbol{\Sigma}$ is constant, that is the noise is additive.

The main goal of this study is to estimate parameters in second-order SDEs. We first reformulate the d -dimensional second-order SDE (1) into a $2d$ -dimensional SDE in Itô's form. We define an auxiliary velocity variable, and express the second-order SDE in terms of its position \mathbf{X}_t and velocity \mathbf{V}_t :

$$\begin{aligned} d\mathbf{X}_t &= \mathbf{V}_t dt, & \mathbf{X}_0 &= \mathbf{x}_0, \\ d\mathbf{V}_t &= \mathbf{F}(\mathbf{X}_t, \mathbf{V}_t; \boldsymbol{\beta}) dt + \boldsymbol{\Sigma} d\mathbf{W}_t, & \mathbf{V}_0 &= \mathbf{v}_0, \end{aligned} \quad (2)$$

where \mathbf{W}_t is a standard Wiener process. We refer to \mathbf{X}_t and \mathbf{V}_t as the smooth and rough coordinates, respectively.

A specific example of model (2) is $\mathbf{F}(\mathbf{x}, \mathbf{v}) = -\mathbf{c}(\mathbf{x}, \mathbf{v})\mathbf{v} - \nabla \mathbf{U}(\mathbf{x})$, for some function $\mathbf{c}(\cdot)$ and potential $\mathbf{U}(\cdot)$. Then, model (2) is called a stochastic damping Hamiltonian system. This system describes the motion of a particle subjected to potential, dissipative, and random forces (Wu, 2001). An example of a stochastic damping Hamiltonian system is the Kramers oscillator introduced in Section 2.1.

Let $\mathbf{Y}_t = (\mathbf{X}_t^\top, \mathbf{V}_t^\top)^\top$, $\tilde{\mathbf{F}}(\mathbf{x}, \mathbf{v}; \boldsymbol{\beta}) = (\mathbf{v}^\top, \mathbf{F}(\mathbf{x}, \mathbf{v}; \boldsymbol{\beta})^\top)^\top$ and $\tilde{\boldsymbol{\Sigma}} = (\mathbf{0}^\top, \boldsymbol{\Sigma}^\top)^\top$. Then (2) is formulated as

$$d\mathbf{Y}_t = \tilde{\mathbf{F}}(\mathbf{Y}_t; \boldsymbol{\beta}) dt + \tilde{\boldsymbol{\Sigma}} d\mathbf{W}_t, \quad \mathbf{Y}_0 = \mathbf{y}_0. \quad (3)$$

The notation $\tilde{\cdot}$ over an object indicates that it is associated with process \mathbf{Y}_t . Specifically, the object is of dimension $2d$ or $2d \times 2d$.

When it exists, the unique solution of (3) is called a diffusion or diffusion process. System (3) is usually not fully observed since the velocity \mathbf{V}_t is not observable. Thus, our primary objective is to estimate the underlying drift parameter $\boldsymbol{\beta}$ and the diffusion parameter $\boldsymbol{\Sigma}$, based on discrete observations of either \mathbf{Y}_t (referred to as complete observation case), or only \mathbf{X}_t (referred to as partial observation case). Diffusion \mathbf{Y}_t is said to be hypoelliptic since the matrix

$$\tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{\Sigma}}^\top = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top \end{bmatrix} \quad (4)$$

is not of full rank, while \mathbf{Y}_t admits a smooth density. Thus, (2) is a subclass of a larger class of hypoelliptic diffusions.

Parametric estimation for hypoelliptic diffusions is an active area of research. Ditlevsen and Sørensen (2004) studied discretely observed integrated diffusion processes. They proposed to use prediction-based estimating functions, which are suitable for non-Markovian processes and which do not require access to the unobserved component. They proved consistency and asymptotic normality of the estimators for $N \rightarrow \infty$, but without any requirements on the sampling interval h . Certain moment conditions are needed to obtain results for fixed h , which are often difficult to fulfill for nonlinear drift functions. The estimator was applied to paleoclimate data in Ditlevsen et al. (2002), similar to the data we analyze in Section 5.

Gloter (2006) also focused on parametric estimation for discretely observed integrated diffusion processes, introducing a contrast function using the Euler-Maruyama discretization. He studied the asymptotic properties as the sampling interval $h \rightarrow 0$ and the sample size $N \rightarrow \infty$, under the so-called rapidly increasing experimental design $Nh \rightarrow \infty$ and $Nh^2 \rightarrow 0$. To address the ill-conditioned contrast from the Euler-Maruyama discretization, he suggested using only the rough equations of the SDE. He proposed to recover the unobserved integrated component through the finite difference approximation $(\mathbf{X}_{t_{k+1}} - \mathbf{X}_{t_k})/h$. This approximation makes the estimator biased and requires a correction factor of $3/2$ in one of the terms of the contrast function for partial observations. Consequently, the correction increases the asymptotic variance of the estimator of the diffusion parameter. Samson and Thieullen (2012) expanded the ideas of (Gloter, 2006) and proved the results of (Gloter, 2006) in more general models. Similar to (Gloter, 2006), their focus was on contrasts using the Euler-Maruyama discretization limited to only the rough equations.

Pokern et al. (2009) proposed an Itô-Taylor expansion, adding a noise term of order $h^{3/2}$ to the smooth component in the numerical scheme. They argued against the use of finite differences for approximating unobserved components. Instead, he suggested using the Itô-Taylor expansion leading to non-degenerate conditionally Gaussian approximations of the transition density and using Markov Chain Monte Carlo (MCMC) Gibbs samplers for conditionally imputing missing components based on the observations. They found out that this approach resulted in a biased estimator of the drift parameter of the rough component.

Ditlevsen and Samson (2019) focused on both filtering and inference methods for complete and partial observations. They proposed a contrast estimator based on the strong order 1.5 scheme (Kloeden and Platen, 1992), which incorporates noise of order $h^{3/2}$ into the smooth component, similar to (Pokern et al., 2009). Moreover, they retained terms of order h^2 in the mean, which removed the bias in the drift parameters noted in (Pokern et al., 2009). They proved consistency and asymptotic normality under complete observations, with the standard rapidly increasing experimental design $Nh \rightarrow \infty$ and $Nh^2 \rightarrow 0$. They adopted an unconventional approach by using two separate contrast functions, resulting in marginal asymptotic results rather than a joint central limit theorem. The model was limited to a scalar smooth component and a diagonal diffusion coefficient matrix for the rough component.

Melnykova (2020) developed a contrast estimator using local linearization (LL) (Ozaki, 1985; Shoji and Ozaki, 1998; Ozaki et al., 2000) and compared it to the least-squares estimator. She employed local linearization of the drift function, providing a non-degenerate conditional Gaussian discretization scheme, enabling the construction of a contrast estimator that achieves asymptotic normality under the standard conditions $Nh \rightarrow \infty$ and $Nh^2 \rightarrow 0$. She proved a joint central limit theorem, bypassing the need for two separate contrasts as in Ditlevsen and Samson (2019). The models in Ditlevsen and Samson (2019) and Melnykova (2020) allow for parameters in the smooth component of the drift, in contrast to models based on second-order differential equations.

Recent work by Gloter and Yoshida (2020, 2021) introduced adaptive and non-adaptive methods in hypoelliptic diffusion models, proving asymptotic normality in the complete observation regime. In line with this work, we briefly review their non-adaptive estimator. It is based on a higher-order Itô-Taylor expansion that introduces additional Gaussian noise onto the smooth coordinates, accompanied by an appropriate higher-order mean approximation of the rough coordinates. The resulting estimator was later termed the local Gaussian (LG), which should be differentiated from LL. The LG estimator can be viewed as an extension of the estimator proposed in Ditlevsen and Samson (2019), with fewer restrictions on the class of models. Gloter and Yoshida (2020, 2021) found that using the full SDE to create a contrast reduces the asymptotic variance of the estimator of the diffusion parameter compared to methods using only rough coordinates in the case of complete observations.

The most recent contributions are Iguchi et al. (2023a,b); Iguchi and Beskos (2023), building on the foundation of the LG estimator and focusing on high-frequency regimes addressing limitations in earlier methods. Iguchi et al. (2023b) presented a new closed-form contrast estimator for hypoelliptic SDEs (denoted as Hypo-I) based on Edgeworth-type density expansion and Malliavin calculus that achieves asymptotic normality under the less restrictive condition of $Nh^3 \rightarrow 0$. Iguchi et al. (2023a) focused on a highly degenerate class of SDEs (denoted as Hypo-II) where smooth coordinates split into further sub-groups and proposed estimators for both complete and partial observation settings. Iguchi and Beskos (2023) further refined the conditions for estimators asymptotic normality for both Hypo-I and Hypo-II under a weak design $Nh^p \rightarrow 0$, for $p \geq 2$.

The existing methods are generally based on approximations with varying degrees of refinements to correct for possible nonlinearities. This implies that they quickly degrade for highly nonlinear models if the step size is increased. In particular, this is the case for Hamiltonian systems. Instead, we propose to use splitting schemes, more precisely the Strang splitting scheme.

Splitting schemes are established techniques initially developed for solving ordinary differential equations (ODEs) and have proven to be effective also for SDEs (Ableidinger et al., 2017; Buckwar et al., 2022; Pilipovic et al., 2024). These schemes yield accurate results in many practical applications since they incorporate nonlinearities in their construction. This makes them particularly suitable for second-order SDEs, where they have been widely used. Early work in dissipative particle dynamics (Shardlow, 2003; Serrano et al., 2006), applications to molecular dynamics (Vanden-Eijnden and Ciccotti, 2006; Melchionna, 2007; Leimkuhler and Matthews, 2015) and studies on internal particles (Pavliotis et al., 2009) all highlight the scheme's versatility. Burrage et al. (2007), Bou-Rabee and Owhadi (2010), and Abdulle et al. (2015) focused on the long-run statistical properties such as invariant measures. Bou-Rabee (2017); Bréhier and Goudenège (2019) and Adams et al. (2022) used splitting schemes for stochastic partial differential equations (SPDEs).

Despite the extensive use of splitting schemes in different areas, statistical applications have been lacking. We have recently proposed statistical estimators for elliptic SDEs (Pilipovic et al., 2024). The straightforward and intuitive schemes lead to robust, easy-to-implement estimators, offering an advantage over more numerically intensive and less user-friendly state-of-the-art methods. We use the Strang splitting scheme to approximate the transition density between two consecutive observations and derive the pseudo-likelihood function since the exact likelihood function is often unknown or intractable. Then, to estimate parameters, we employ maximum likelihood estimation (MLE). However, two specific statistical problems arise due to hypoellipticity and partial observations.

First, hypoellipticity leads to degenerate Euler-Maruyama transition schemes, which can be addressed by constructing the pseudo-likelihood solely from the rough equations of the SDE, referred to as the rough likelihood hereafter. The

Strang splitting technique enables the estimator to incorporate both smooth and rough components (referred to as the full likelihood). It is also possible to construct Strang splitting estimators using only the rough likelihood, raising the question of which estimator performs better. Our results are in line with Gloter and Yoshida (2020, 2021) in the complete observation setting, where we find that using the full likelihood reduces the asymptotic variance of the diffusion estimator. We found the same results in the simulation study for the LL estimator proposed by Melnykova (2020).

Second, we suggest to treat the unobserved velocity by approximating it using finite difference methods. While Gloter (2006) and Samson and Thieullen (2012) exclusively use forward differences, we investigate also central and backward differences. The forward difference approach leads to a biased estimator unless it is corrected. One of the main contributions of this work is finding suitable corrections of the pseudo-likelihoods for different finite difference approximations such that the Strang estimators are asymptotically unbiased. This also ensures consistency of the diffusion parameter estimator, at the cost of increasing its asymptotic variance.

When only partial observations are available, we explore the impact of using the full likelihood versus the rough likelihood and how different finite differentiation approximations influence the parametric inference. We find that the choice of likelihood does not affect the asymptotic variance of the estimator. However, our simulation study on the Kramers oscillator suggests that using the full likelihood in finite sample setups introduce more bias than using only the rough marginal likelihood, which is the opposite of the complete observation setting. Finally, we analyze a paleoclimate ice core dataset from Greenland using a second-order SDE.

The main contributions of this paper are:

1. We extend the Strang splitting estimator of (Pilipovic et al., 2024) to hypoelliptic models given by second-order SDEs, including appropriate correction factors to obtain consistency.
2. When complete observations are available, we show that the asymptotic variance of the estimator of the diffusion parameter is smaller when maximizing the full likelihood. In contrast, for partial observations, we show that the asymptotic variance remains unchanged regardless of using the full or marginal likelihood of the rough coordinates.
3. We discuss the influence on the statistical properties of using the forward difference approximation for imputing the unobserved velocity variables compared to using the backward or the central difference.
4. We evaluate the performance of the estimators through a simulation study of a second-order SDE, the Kramers oscillator. Additionally, we show numerically in a finite sample study that the marginal likelihood for partial observations is more favorable than the full likelihood.
5. We fit the Kramers oscillator to a paleoclimate ice core dataset from Greenland and estimate the average time needed to pass between two metastable states.

The structure of the paper is as follows. In Section 2, we introduce the class of SDE models, define hypoellipticity, introduce the Kramers oscillator, and explain the Strang splitting scheme and its associated estimators. The asymptotic properties of the estimator are established in Section 3. The theoretical results are illustrated in a simulation study on the Kramers Oscillator in Section 4. Section 5 illustrates our methodology on the Greenland ice core data, while the technical results and the proofs of the main theorems and properties are in Section 6 and Supplementary Material S1, respectively.

Notation. We use capital bold letters for random vectors, vector-valued functions, and matrices, while lowercase bold letters denote deterministic vectors. $\|\cdot\|$ denotes both the L^2 vector norm in \mathbb{R}^d . Superscript (i) on a vector denotes the i -th component, while on a matrix it denotes the i -th column. Double subscript ij on a matrix denotes the component in the i -th row and j -th column. The transpose is denoted by \top . Operator $\text{Tr}(\cdot)$ returns the trace of a matrix and $\det(\cdot)$ the determinant. \mathbf{I}_d denotes the d -dimensional identity matrix, while $\mathbf{0}_{d \times d}$ is a d -dimensional zero square matrix. We denote by $[a_i]_{i=1}^d$ a vector with coordinates a_i , and by $[b_{ij}]_{i,j=1}^d$ a matrix with coordinates b_{ij} , for $i, j = 1, \dots, d$. For a real-valued function $g : \mathbb{R}^d \rightarrow \mathbb{R}$, $\partial_{x^{(i)}} g(\mathbf{x})$ denotes the partial derivative with respect to $x^{(i)}$ and $\partial_{x^{(i)}x^{(j)}}^2 g(\mathbf{x})$ denotes the second partial derivative with respect to $x^{(i)}$ and $x^{(j)}$. The nabla operator $\nabla_{\mathbf{x}}$ denotes the gradient vector of g with respect of \mathbf{x} , that is, $\nabla_{\mathbf{x}} g(\mathbf{x}) = [\partial_{x^{(i)}} g(\mathbf{x})]_{i=1}^d$. \mathbb{H} denotes the Hessian matrix of function g , $\mathbb{H}_g(\mathbf{x}) = [\partial_{x^{(i)}x^{(j)}}^2 g(\mathbf{x})]_{i,j=1}^d$. For a vector-valued function $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, the differential operator $D_{\mathbf{x}}$ denotes the Jacobian matrix $D_{\mathbf{x}} \mathbf{F}(\mathbf{x}) = [\partial_{x^{(i)}} F^{(j)}(\mathbf{x})]_{i,j=1}^d$. Let \mathbf{R} represent a vector (or a matrix) valued function defined on $(0, 1) \times \mathbb{R}^d$ (or $(0, 1) \times \mathbb{R}^{d \times d}$), such that, for some constant C , $\|\mathbf{R}(a, \mathbf{x})\| < aC(1 + \|\mathbf{x}\|)^C$ for all a, \mathbf{x} . When denoted by R , it refers to a scalar function. For an open set A , the bar \bar{A} indicates closure. We write $\xrightarrow{\mathbb{P}}$ for convergence in probability \mathbb{P} .

2 Problem setup

Let $\mathbf{Y} = (\mathbf{Y}_t)_{t \geq 0}$ in (3) be defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$ with a complete right-continuous filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$, and let the d -dimensional Wiener process $\mathbf{W} = (\mathbf{W}_t)_{t \geq 0}$ be adapted to \mathcal{F}_t . The probability measure \mathbb{P}_θ is parameterized by the parameter $\theta = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Rewrite equation (3) as follows:

$$d\mathbf{Y}_t = \tilde{\mathbf{A}}(\boldsymbol{\beta})(\mathbf{Y}_t - \tilde{\mathbf{b}}(\boldsymbol{\beta})) dt + \tilde{\mathbf{N}}(\mathbf{Y}_t; \boldsymbol{\beta}) dt + \tilde{\boldsymbol{\Sigma}} d\mathbf{W}_t, \quad (5)$$

where

$$\tilde{\mathbf{A}}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{0}_{d \times d} & \mathbf{I}_d \\ \mathbf{A}_x(\boldsymbol{\beta}) & \mathbf{A}_v(\boldsymbol{\beta}) \end{bmatrix}, \quad \tilde{\mathbf{b}}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{b}(\boldsymbol{\beta}) \\ \mathbf{0}_d \end{bmatrix}, \quad \tilde{\mathbf{N}}(\mathbf{x}, \mathbf{v}; \boldsymbol{\beta}) = \begin{bmatrix} \mathbf{0}_d \\ \mathbf{N}(\mathbf{x}, \mathbf{v}; \boldsymbol{\beta}) \end{bmatrix}. \quad (6)$$

Function \mathbf{F} in (2) is thus split as $\mathbf{F}(\mathbf{x}, \mathbf{v}; \boldsymbol{\beta}) = \mathbf{A}_x(\boldsymbol{\beta})(\mathbf{x} - \mathbf{b}(\boldsymbol{\beta})) + \mathbf{A}_v(\boldsymbol{\beta})\mathbf{v} + \mathbf{N}(\mathbf{x}, \mathbf{v}; \boldsymbol{\beta})$.

Let $\bar{\Theta}_\beta \times \bar{\Theta}_\Sigma = \bar{\Theta}$ denote the closure of the parameter space with Θ_β and Θ_Σ being two convex open bounded subsets of \mathbb{R}^r and $\mathbb{R}^{d \times d}$, respectively. The function $\mathbf{N} : \mathbb{R}^{2d} \times \bar{\Theta}_\beta \rightarrow \mathbb{R}^d$ is assumed locally Lipschitz; functions \mathbf{A}_x and \mathbf{A}_v are defined on $\bar{\Theta}_\beta$ and take values in $\mathbb{R}^{d \times d}$; and the parameter matrix $\boldsymbol{\Sigma}$ takes values in $\mathbb{R}^{d \times d}$. The matrix $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ is assumed to be positive definite, shaping the variance of the rough coordinates. As any square root of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ induces the same distribution, $\boldsymbol{\Sigma}$ is identifiable only up to equivalence classes. Hence, estimation of the parameter $\boldsymbol{\Sigma}$ means estimation of $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$. The drift function $\tilde{\mathbf{F}}$ in (3) is divided into a linear part given by the matrix $\tilde{\mathbf{A}}$ and a nonlinear part given by $\tilde{\mathbf{N}}$.

The true value of the parameter is denoted by $\theta_0 = (\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$, and we assume that $\theta_0 \in \Theta$. When referring to the true parameters, we write $\mathbf{A}_{x,0}$, $\mathbf{A}_{v,0}$, \mathbf{b}_0 , $\mathbf{N}_0(\mathbf{x})$, $\mathbf{F}_0(\mathbf{x})$ and $\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^\top$ instead of $\mathbf{A}_x(\boldsymbol{\beta}_0)$, $\mathbf{A}_v(\boldsymbol{\beta}_0)$, $\mathbf{b}(\boldsymbol{\beta}_0)$, $\mathbf{N}(\mathbf{x}; \boldsymbol{\beta}_0)$, $\mathbf{F}(\mathbf{x}; \boldsymbol{\beta}_0)$ and $\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0^\top$, respectively. We write \mathbf{A}_x , \mathbf{A}_v , \mathbf{b} , $\mathbf{N}(\mathbf{x})$, $\mathbf{F}(\mathbf{x})$, and $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ for any parameter θ .

2.1 Example: The Kramers oscillator

The abrupt temperature changes during the ice ages, known as the Dansgaard–Oeschger (DO) events, are essential elements for understanding the climate (Dansgaard et al., 1993). These events occurred during the last glacial era spanning approximately the period from 115,000 to 12,000 years before present and are characterized by rapid warming phases followed by gradual cooling periods, revealing colder (stadial) and warmer (interstadial) climate states (Rasmussen et al., 2014).

To analyze the DO events in Section 5, we propose a stochastic model of the escape dynamics in metastable systems, the Kramers oscillator (Kramers, 1940), originally formulated to model the escape rate of Brownian particles from potential wells. The escape rate is related to the mean first passage time — the time needed for a particle to exceed the potential's local maximum for the first time, starting at a neighboring local minimum. This rate depends on variables such as the damping coefficient, noise intensity, temperature, and specific potential features, including the barrier's height and curvature at the minima and maxima. We apply this framework to quantify the rate of climate transitions between stadial and interstadial periods. This provides an estimate on the probability distribution of the occurrence of DO events, contributing to our understanding of the global climate system.

Following Arnold and Imkeller (2000), we introduce the Kramers oscillator as the stochastic Duffing oscillator - an example of a second-order SDE and a stochastic damping Hamiltonian system. The Duffing oscillator (Duffing, 1918) is a forced nonlinear oscillator, featuring a cubic stiffness term. The governing equation is given by:

$$\ddot{x}_t + \eta \dot{x}_t + \frac{d}{dx}U(x_t) = f(t), \quad \text{where } U(x) = -a\frac{x^2}{2} + b\frac{x^4}{4}, \quad \text{with } a, b > 0, \quad \eta \geq 0. \quad (7)$$

The parameter η in (7) indicates the damping level, a regulates the linear stiffness, and b determines the nonlinear component of the restoring force. In the special case where $b = 0$, the equation simplifies to a damped harmonic oscillator. Function f represents the driving force and is usually set to $f(t) = \eta \cos(\omega t)$, which introduces deterministic chaos (Korsch and Jodl, 1999).

When the driving force is $f(t) = \sqrt{2\eta\Gamma}\xi(t)$, where $\xi(t)$ is white noise, equation (7) characterizes the stochastic movement of a particle within a bistable potential well, interpreting $\Gamma > 0$ as the temperature of a heat bath. Setting $\sigma = \sqrt{2\eta\Gamma}$, equation (7) can be reformulated as an Itô SDE for variables X_t and $V_t = \dot{X}_t$, expressed as:

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= \left(-\eta V_t - \frac{d}{dx}U(X_t) \right) dt + \sigma dW_t, \end{aligned} \quad (8)$$

where W_t denotes a standard Wiener process. The parameter set of SDE (8) is $\theta = \{\eta, a, b, \sigma^2\}$.

The existence and uniqueness of the invariant measure $\nu_0(dx, dy)$ of (8) is proved in Theorem 3 in (Arnold and Imkeller, 2000). The invariant measure ν_0 is linked to the invariant density π_0 through $\nu_0(dx, dy) = \pi_0(x, v) dx dy$. Here we write $\pi_0(x, v)$ instead of $\pi(x, v; \theta_0)$, and $\pi(x, v)$ instead of $\pi(x, v; \theta)$. The Fokker-Plank equation for π is given by

$$-v \frac{\partial}{\partial x} \pi(x, v) + \eta \pi(x, v) + \eta v \frac{\partial}{\partial v} \pi(x, v) + \frac{d}{dx} U(x) \frac{\partial}{\partial v} \pi(x, v) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial v^2} \pi(x, v) = 0. \quad (9)$$

The invariant density that solves the Fokker-Plank equation is:

$$\pi(x, v) = C \exp\left(-\frac{2\eta}{\sigma^2} U(x)\right) \exp\left(-\frac{\eta}{\sigma^2} v^2\right), \quad (10)$$

where C is the normalizing constant.

The marginal invariant probability of V_t is thus Gaussian with zero mean and variance $\sigma^2/(2\eta)$. The marginal invariant probability of X_t is bimodal driven by the potential $U(x)$:

$$\pi(x) = C \exp\left(-\frac{2\eta}{\sigma^2} U(x)\right). \quad (11)$$

At steady state, for a particle moving in any potential $U(x)$ and driven by random Gaussian noise, the position x and velocity v are independent of each other. This is reflected by the decomposition of the joint density $\pi(x, v)$ into $\pi(x)\pi(v)$.

Fokker-Plank equation (9) can also be used to derive the mean first passage time τ which is inversely related to Kramers' escape rate κ (Kramers, 1940):

$$\tau = \frac{1}{\kappa} \approx \frac{2\pi}{\left(\sqrt{1 + \frac{\eta^2}{4\omega^2}} - \frac{\eta}{2\omega}\right) \Omega} \exp\left(\frac{\Delta U}{T}\right),$$

where $x_{\text{barrier}} = 0$ is the local maximum of $U(x)$ and $x_{\text{well}} = \pm\sqrt{a/b}$ are the local minima, $\omega = \sqrt{|U''(x_{\text{barrier}})|} = \sqrt{a}$, $\Omega = \sqrt{U''(x_{\text{well}})} = \sqrt{2a}$, and $\Delta U = U(x_{\text{barrier}}) - U(x_{\text{well}}) = a^2/4b$. The formula is derived assuming strong friction, or an over-damped system ($\eta \gg \omega$), and a small parameter $T/\Delta U \ll 1$, indicating sufficiently deep potential wells. For the potential defined in (7), the mean waiting time τ is then approximated by

$$\tau \approx \frac{\sqrt{2}\pi}{\sqrt{a + \frac{\eta^2}{4} - \frac{\eta}{2}}} \exp\left(\frac{a^2\eta}{2b\sigma^2}\right). \quad (12)$$

2.2 Hypocoellipticity

The SDE (5) is said to be hypoelliptic if its quadratic diffusion matrix $\tilde{\Sigma}\tilde{\Sigma}^\top$ is not of full rank, while its solutions admit a smooth transition density with respect to the Lebesgue measure. According to Hörmander's theorem (Nualart, 2006), this is fulfilled if the SDE in its Stratonovich form satisfies the weak Hörmander condition. Since Σ does not depend on \mathbf{y} , the Itô and Stratonovich forms coincide.

We begin by recalling the concept of Lie brackets: for smooth vector fields $\mathbf{f}, \mathbf{g} : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d}$, the i -th component of the Lie bracket, $[\mathbf{f}, \mathbf{g}]^{(i)}$, is defined as

$$[\mathbf{f}, \mathbf{g}]^{(i)} := D_{\mathbf{y}}^\top g^{(i)}(\mathbf{y}) \mathbf{f}(\mathbf{y}) - D_{\mathbf{y}}^\top f^{(i)}(\mathbf{y}) \mathbf{g}(\mathbf{y}).$$

We define the set \mathcal{H} of vector fields by initially including $\tilde{\Sigma}^{(i)}$, $i = 1, 2, \dots, 2d$, and then recursively adding Lie brackets

$$H \in \mathcal{H} \Rightarrow [\tilde{\mathbf{F}}, H], [\tilde{\Sigma}^{(1)}, H], \dots, [\tilde{\Sigma}^{(2d)}, H] \in \mathcal{H}.$$

The weak Hörmander condition is met if the vectors in \mathcal{H} span \mathbb{R}^{2d} at every point $\mathbf{y} \in \mathbb{R}^{2d}$. The initial vectors span $\{(\mathbf{0}, \mathbf{v}) \in \mathbb{R}^{2d} \mid \mathbf{v} \in \mathbb{R}^d\}$, a d -dimensional subspace. We therefore need to verify the existence of some $H \in \mathcal{H}$ with a non-zero first element. The first iteration of the system yields

$$\begin{aligned} [\tilde{\mathbf{F}}, \tilde{\Sigma}^{(i)}]^{(1)} &= -\Sigma^{(i)}, \\ [\tilde{\Sigma}^{(i)}, \tilde{\Sigma}^{(j)}]^{(1)} &= \mathbf{0}, \end{aligned}$$

for $i, j = 1, 2, \dots, 2d$. The first equation is non-zero, as are all subsequent iterations. Thus, the second-order SDE defined in (5) is always hypoelliptic.

2.3 Assumptions

The following assumptions are a generalization of those presented in (Pilipovic et al., 2024).

Let $T > 0$ be the length of the observed time interval. We assume that (5) has a unique strong solution $\mathbf{Y} = \{\mathbf{Y}_t \mid t \in [0, T]\}$, adapted to $\mathcal{F} = \{\mathcal{F}_t \mid t \in [0, T]\}$, which follows from the following first two assumptions (Theorem 2 in Alyushina (1988), Theorem 1 in Krylov (1991), Theorem 3.5 in Mao (2007)). We need the last three assumptions to prove the properties of the estimators.

- (A1) Function \mathbf{N} is twice continuously differentiable with respect to both \mathbf{y} and $\boldsymbol{\theta}$, i.e., $\mathbf{N} \in C^2$. Moreover, it is globally one-sided Lipschitz continuous with respect to \mathbf{y} on $\mathbb{R}^{2d} \times \bar{\Theta}_\beta$. That is, there exists a constant $C > 0$ such that for all $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^{2d}$,

$$(\mathbf{y}_1 - \mathbf{y}_2)^\top (\mathbf{N}(\mathbf{y}_1; \boldsymbol{\beta}) - \mathbf{N}(\mathbf{y}_2; \boldsymbol{\beta})) \leq C \|\mathbf{y}_1 - \mathbf{y}_2\|^2.$$

- (A2) Function \mathbf{N} exhibits at most polynomial growth in \mathbf{y} , uniformly in $\boldsymbol{\theta}$. Specifically, there exist constants $C > 0$ and $\chi \geq 1$ such that for all $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^{2d}$,

$$\|\mathbf{N}(\mathbf{y}_1; \boldsymbol{\beta}) - \mathbf{N}(\mathbf{y}_2; \boldsymbol{\beta})\|^2 \leq C (1 + \|\mathbf{y}_1\|^{2\chi-2} + \|\mathbf{y}_2\|^{2\chi-2}) \|\mathbf{y}_1 - \mathbf{y}_2\|^2.$$

Additionally, its derivatives exhibit polynomial growth in \mathbf{y} , uniformly in $\boldsymbol{\theta}$.

- (A3) The solution \mathbf{Y} to SDE (5) has invariant probability $\nu_0(d\mathbf{y})$.

- (A4) $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ is invertible on $\bar{\Theta}_\Sigma$.

- (A5) $\boldsymbol{\beta}$ is identifiable, that is, if $\mathbf{F}(\mathbf{y}, \boldsymbol{\beta}_1) = \mathbf{F}(\mathbf{y}, \boldsymbol{\beta}_2)$ for all $\mathbf{y} \in \mathbb{R}^{2d}$, then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

Assumption (A1) ensures finiteness of the moments of the solution \mathbf{X} (Tretyakov and Zhang, 2013), i.e.,

$$\mathbb{E}[\sup_{t \in [0, T]} \|\mathbf{Y}_t\|^{2p}] < C(1 + \|\mathbf{y}_0\|^{2p}), \quad \forall p \geq 1. \quad (13)$$

Assumption (A3) is necessary for the ergodic theorem to ensure convergence in distribution. Assumption (A4) ensures that the model (5) is hypoelliptic. Assumption (A5) ensures the identifiability of the drift parameter.

2.4 Strang splitting scheme

Consider the following splitting of (5):

$$d\mathbf{Y}_t^{[1]} = \tilde{\mathbf{A}}(\mathbf{Y}_t^{[1]} - \tilde{\mathbf{b}}) dt + \tilde{\boldsymbol{\Sigma}} d\mathbf{W}_t, \quad \mathbf{Y}_0^{[1]} = \mathbf{y}_0, \quad (14)$$

$$d\mathbf{Y}_t^{[2]} = \tilde{\mathbf{N}}(\mathbf{Y}_t^{[2]}) dt, \quad \mathbf{Y}_0^{[2]} = \mathbf{y}_0. \quad (15)$$

There are no assumptions on the choice of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$, and thus the nonlinear function $\tilde{\mathbf{N}}$. Indeed, we show that the asymptotic results hold for any choice of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ in both the complete and the partial observation settings. This extends the results in Pilipovic et al. (2024), where it is shown to hold in the elliptic complete observation case, as well. While asymptotic results are invariant to the choice of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$, finite sample properties of the scheme and the corresponding estimators are very different, and it is important to choose the splitting wisely. Intuitively, when the process is close to a fixed point of the drift, the linear dynamics are dominating, whereas far from the fixed points, the nonlinearities might be dominating. If the drift has a fixed point \mathbf{y}^* , we therefore suggest setting $\tilde{\mathbf{A}} = D_{\mathbf{y}}\tilde{\mathbf{F}}(\mathbf{y}^*)$ and $\tilde{\mathbf{b}} = \mathbf{y}^*$. This choice is confirmed in simulations (for more details see Pilipovic et al. (2024)).

Solution of SDE (14) is an Ornstein–Uhlenbeck (OU) process given by the following h -flow:

$$\mathbf{Y}_{t_k}^{[1]} = \Phi_h^{[1]}(\mathbf{Y}_{t_{k-1}}^{[1]}) = \tilde{\boldsymbol{\mu}}_h(\mathbf{Y}_{t_{k-1}}^{[1]}; \boldsymbol{\beta}) + \tilde{\boldsymbol{\varepsilon}}_{h,k}, \quad (16)$$

$$\tilde{\boldsymbol{\mu}}_h(\mathbf{y}; \boldsymbol{\beta}) := e^{\tilde{\mathbf{A}}h}(\mathbf{y} - \tilde{\mathbf{b}}) + \tilde{\mathbf{b}}, \quad (17)$$

$$\tilde{\boldsymbol{\Omega}}_h = \int_0^h e^{\tilde{\mathbf{A}}(h-u)} \tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{\Sigma}}^\top e^{\tilde{\mathbf{A}}^\top(h-u)} du, \quad (18)$$

where $\tilde{\boldsymbol{\varepsilon}}_{h,k} \stackrel{i.i.d.}{\sim} \mathcal{N}_{2d}(\mathbf{0}, \tilde{\boldsymbol{\Omega}}_h)$ for $k = 1, \dots, N$. It is useful to rewrite $\tilde{\boldsymbol{\Omega}}_h$ in the following block matrix form,

$$\tilde{\boldsymbol{\Omega}}_h = \begin{bmatrix} \boldsymbol{\Omega}_h^{[SS]} & \boldsymbol{\Omega}_h^{[SR]} \\ \boldsymbol{\Omega}_h^{[RS]} & \boldsymbol{\Omega}_h^{[RR]} \end{bmatrix}, \quad (19)$$

where S in the superscript stands for smooth and R stands for rough. The Schur complement of $\tilde{\Omega}_h$ with respect to $\Omega_h^{[RR]}$ and the determinant of $\tilde{\Omega}_h$ are given by:

$$\Omega_h^{[S|R]} := \Omega_h^{[SS]} - \Omega_h^{[SR]}(\Omega_h^{[RR]})^{-1}\Omega_h^{[RS]}, \quad \det \tilde{\Omega}_h = \det \Omega_h^{[RR]} \det \Omega_h^{[S|R]}.$$

Assumptions (A1)-(A2) ensure the existence and uniqueness of the solution of (15) (Theorem 1.2.17 in Humphries and Stuart (2002)). Thus, there exists a unique function $\tilde{f}_h : \mathbb{R}^{2d} \times \Theta_\beta \rightarrow \mathbb{R}^{2d}$, for $h \geq 0$, such that

$$\mathbf{Y}_{t_k}^{[2]} = \Phi_h^{[2]}(\mathbf{Y}_{t_{k-1}}^{[2]}) = \tilde{f}_h(\mathbf{Y}_{t_{k-1}}^{[2]}; \beta). \quad (20)$$

For all $\beta \in \Theta_\beta$, the h -flow \tilde{f}_h fulfills the following semi-group properties:

$$\tilde{f}_0(\mathbf{y}; \beta) = \mathbf{y}, \quad \tilde{f}_{t+s}(\mathbf{y}; \beta) = \tilde{f}_t(\tilde{f}_s(\mathbf{y}; \beta); \beta), \quad t, s \geq 0.$$

For $\mathbf{y} = (\mathbf{x}^\top, \mathbf{v}^\top)^\top$, we have:

$$\tilde{f}_h(\mathbf{x}, \mathbf{v}; \beta) = \begin{bmatrix} \mathbf{x} \\ \mathbf{f}_h(\mathbf{x}, \mathbf{v}; \beta) \end{bmatrix}, \quad (21)$$

where $\mathbf{f}_h(\mathbf{x}, \mathbf{v}; \beta)$ is the solution of the ODE with vector field $\mathbf{N}(\mathbf{x}, \mathbf{v}; \beta)$.

We introduce another assumption needed to define the pseudo-likelihood based on the splitting scheme.

(A6) Inverse function $\tilde{f}_h^{-1}(\mathbf{y}; \beta)$ is defined asymptotically for all $\mathbf{y} \in \mathbb{R}^{2d}$ and all $\beta \in \Theta_\beta$, when $h \rightarrow 0$.

Then, the inverse of \tilde{f}_h can be decomposed as:

$$\tilde{f}_h^{-1}(\mathbf{x}, \mathbf{v}; \beta) = \begin{bmatrix} \mathbf{x} \\ \mathbf{f}_h^{*-1}(\mathbf{x}, \mathbf{v}; \beta) \end{bmatrix}, \quad (22)$$

where $\mathbf{f}_h^{*-1}(\mathbf{x}, \mathbf{v}; \beta)$ is the rough part of the inverse of \tilde{f}_h^{-1} . It does not equal \mathbf{f}_h^{-1} since the inverse does not propagate through coordinates when \mathbf{f}_h depends on \mathbf{x} .

We are now ready to define the Strang splitting scheme for model (5).

Definition 2.1 (Strang splitting) *Let Assumptions (A1)-(A2) hold. The Strang approximation of the solution of (5) is given by:*

$$\Phi_h^{[\text{str}]}(\mathbf{Y}_{t_{k-1}}^{[\text{str}]}) = (\Phi_{h/2}^{[2]} \circ \Phi_h^{[1]} \circ \Phi_{h/2}^{[2]})(\mathbf{Y}_{t_{k-1}}^{[\text{str}]}) = \tilde{f}_{h/2}(\tilde{\mu}_h(\tilde{f}_{h/2}(\mathbf{Y}_{t_{k-1}}^{[\text{str}]}) + \tilde{\varepsilon}_{h,k})). \quad (23)$$

Remark 1 *The order of composition in the splitting schemes is not unique. Changing the order in the Strang splitting leads to a sum of 2 independent random variables, one Gaussian and one non-Gaussian, whose likelihood is not trivial. Thus, we only use the splitting (23).*

2.5 Strang splitting estimators

In this section, we introduce four estimators, all based on the Strang splitting scheme. We distinguish between estimators based on complete observations (denoted by C when both \mathbf{X} and \mathbf{V} are observed) and partial observations (denoted by P when only \mathbf{X} is observed). In applications, we typically only have access to partial observations, however, the full observation estimator is used as a building block for the partial observation case. Additionally, we distinguish the estimators based on the type of likelihood function employed. These are the full likelihood (denoted by F) and the marginal likelihood of the rough component (denoted by R). We furthermore use the conditional likelihood based on the smooth component given the rough part (denoted by S | R) to decompose the full likelihood.

2.5.1 Complete observations

Assume we observe the complete sample $\mathbf{Y}_{0:t_N} := (\mathbf{Y}_{t_k})_{k=1}^N$ from (5) at time steps $0 = t_0 < t_1 < \dots < t_N = T$. For notational simplicity, we assume equidistant step size $h = t_k - t_{k-1}$. Strang splitting scheme (23) is a nonlinear transformation of a Gaussian random variable $\tilde{\mu}_h(\tilde{f}_{h/2}(\mathbf{Y}_{t_{k-1}}^{[\text{str}]}) + \tilde{\varepsilon}_{h,k})$. We define:

$$\tilde{\mathbf{Z}}_{k,k-1}(\beta) := \tilde{f}_{h/2}^{-1}(\mathbf{Y}_{t_k}; \beta) - \tilde{\mu}_h(\tilde{f}_{h/2}(\mathbf{Y}_{t_{k-1}}; \beta); \beta), \quad (24)$$

and apply change of variables to get:

$$p(\mathbf{y}_{t_k} | \mathbf{y}_{t_{k-1}}) = p_{\mathcal{N}(\mathbf{0}, \tilde{\Omega}_h)}(\tilde{\mathbf{z}}_{k,k-1} | \mathbf{y}_{t_{k-1}}) |\det D_{\mathbf{y}} \tilde{\mathbf{f}}_{h/2}^{-1}(\mathbf{y}_{t_k})|.$$

Using $-\log |\det D_{\mathbf{y}} \tilde{\mathbf{f}}_{h/2}^{-1}(\mathbf{y}; \boldsymbol{\beta})| = \log |\det D_{\mathbf{y}} \tilde{\mathbf{f}}_{h/2}(\mathbf{y}; \boldsymbol{\beta})|$ and $\det D_{\mathbf{y}} \tilde{\mathbf{f}}_{h/2}(\mathbf{y}; \boldsymbol{\beta}) = \det D_{\mathbf{v}} \mathbf{f}_{h/2}(\mathbf{y}; \boldsymbol{\beta})$, together with the Markov property of $\mathbf{Y}_{0:t_N}$, we get the following objective function based on the full log-likelihood:

$$\mathcal{L}^{[\text{CF}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) := \sum_{k=1}^N \left(\log \det \tilde{\Omega}_h(\boldsymbol{\theta}) + \tilde{\mathbf{z}}_{k,k-1}(\boldsymbol{\beta})^\top \tilde{\Omega}_h(\boldsymbol{\theta})^{-1} \tilde{\mathbf{z}}_{k,k-1}(\boldsymbol{\beta}) + 2 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}(\mathbf{Y}_{t_k}; \boldsymbol{\beta})| \right). \quad (25)$$

Now, split $\tilde{\mathbf{z}}_{k,k-1}$ from (24) into the smooth and rough parts $\tilde{\mathbf{z}}_{k,k-1} = ((\mathbf{z}_{k,k-1}^{[\text{S}]})^\top, (\mathbf{z}_{k,k-1}^{[\text{R}]})^\top)^\top$ defined as:

$$\mathbf{z}_{k,k-1}^{[\text{S}]}(\boldsymbol{\beta}) := [\tilde{Z}_{k,k-1}^{(i)}(\boldsymbol{\beta})]_{i=1}^d = \mathbf{X}_{t_k} - \boldsymbol{\mu}_h^{[\text{S}]}(\tilde{\mathbf{f}}_{h/2}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}); \boldsymbol{\beta}), \quad (26)$$

$$\mathbf{z}_{k,k-1}^{[\text{R}]}(\boldsymbol{\beta}) := [\tilde{Z}_{k,k-1}^{(i)}(\boldsymbol{\beta})]_{i=d+1}^{2d} = \mathbf{f}_{h/2}^{*-1}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}) - \boldsymbol{\mu}_h^{[\text{R}]}(\tilde{\mathbf{f}}_{h/2}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}); \boldsymbol{\beta}), \quad (27)$$

where

$$\boldsymbol{\mu}_h^{[\text{S}]}(\mathbf{y}; \boldsymbol{\beta}) := [\tilde{\mu}_h^{(i)}(\mathbf{y}; \boldsymbol{\beta})]_{i=1}^d, \quad \boldsymbol{\mu}_h^{[\text{R}]}(\mathbf{y}; \boldsymbol{\beta}) := [\tilde{\mu}_h^{(i)}(\mathbf{y}; \boldsymbol{\beta})]_{i=d+1}^{2d}. \quad (28)$$

We also define the following sequence of vectors

$$\mathbf{z}_{k,k-1}^{[\text{S|R}]}(\boldsymbol{\beta}) := \mathbf{z}_{k,k-1}^{[\text{S}]}(\boldsymbol{\beta}) - \Omega_h^{[\text{SR}]}(\Omega_h^{[\text{RR}]})^{-1} \mathbf{z}_{k,k-1}^{[\text{R}]}(\boldsymbol{\beta}). \quad (29)$$

The formula for jointly normal distributions yields:

$$p_{\mathcal{N}(\mathbf{0}, \tilde{\Omega}_h)}(\tilde{\mathbf{z}}_{k,k-1} | \mathbf{y}_{t_{k-1}}) = p_{\mathcal{N}(\mathbf{0}, \Omega_h^{[\text{RR}]})}(\mathbf{z}_{k,k-1}^{[\text{R}]} | \mathbf{y}_{t_{k-1}}) \cdot p_{\mathcal{N}(\Omega_h^{[\text{SR}]}(\Omega_h^{[\text{RR}]})^{-1} \mathbf{z}_{k,k-1}^{[\text{R}]}, \Omega_h^{[\text{S|R}]})}(\mathbf{z}_{k,k-1}^{[\text{S}]} | \mathbf{z}_{k,k-1}^{[\text{R}]}, \mathbf{y}_{t_{k-1}}).$$

This leads to dividing the full log-likelihood $\mathcal{L}^{[\text{CF}]}$ into a sum of the marginal log-likelihood $\mathcal{L}^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta})$ and the smooth-given-rough log-likelihood $\mathcal{L}^{[\text{CS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta})$:

$$\mathcal{L}^{[\text{CF}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \mathcal{L}^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) + \mathcal{L}^{[\text{CS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}),$$

where

$$\mathcal{L}^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) := \sum_{k=1}^N \left(\log \det \Omega_h^{[\text{RR}]}(\boldsymbol{\theta}) + \mathbf{z}_{k,k-1}^{[\text{R}]}(\boldsymbol{\beta})^\top \Omega_h^{[\text{RR}]}(\boldsymbol{\theta})^{-1} \mathbf{z}_{k,k-1}^{[\text{R}]}(\boldsymbol{\beta}) + 2 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}(\mathbf{Y}_{t_k}; \boldsymbol{\beta})| \right), \quad (30)$$

$$\mathcal{L}^{[\text{CS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) := \sum_{k=1}^N \left(\log \det \Omega_h^{[\text{S|R}]}(\boldsymbol{\theta}) + \mathbf{z}_{k,k-1}^{[\text{S|R}]}(\boldsymbol{\beta})^\top \Omega_h^{[\text{S|R}]}(\boldsymbol{\theta})^{-1} \mathbf{z}_{k,k-1}^{[\text{S|R}]}(\boldsymbol{\beta}) \right). \quad (31)$$

The terms containing the drift parameter in $\mathcal{L}^{[\text{CR}]}$ in (30) are of order $h^{1/2}$, as in the elliptic case, whereas the terms containing the drift parameter in $\mathcal{L}^{[\text{CS|R}]}$ in (31) are of order $h^{3/2}$. Consequently, under a rapidly increasing experimental design where $Nh \rightarrow \infty$ and $Nh^2 \rightarrow 0$, the objective function (31) is degenerate for estimating the drift parameter. However, it contributes to the estimation of the diffusion parameter when the full objective function (25) is used. We show in later sections that employing (25) results in a lower asymptotic variance for the diffusion parameter making it more efficient in complete observation scenarios.

The estimators based on complete observations are then defined as:

$$\hat{\boldsymbol{\theta}}_N^{[\text{obj}]} := \arg \min_{\boldsymbol{\theta}} \mathcal{L}^{[\text{obj}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}), \quad \text{obj} \in \{[\text{CF}], [\text{CR}]\}. \quad (32)$$

Although the full objective function is based on twice as many equations as the marginal likelihood, its implementation complexity, speed, and memory requirements are similar to the marginal objective function. Therefore, if the complete observations are available, we recommend using the objective function (25) based on the full likelihood.

2.5.2 Partial observations

Assume we only observe the smooth coordinates $\mathbf{X}_{0:t_N} := (\mathbf{X}_{t_k})_{k=0}^N$. The observed process \mathbf{X}_t alone is not a Markov process, although the complete process \mathbf{Y}_t is. To approximate \mathbf{V}_{t_k} , we define the backward difference process:

$$\Delta_h \mathbf{X}_{t_k} := \frac{\mathbf{X}_{t_k} - \mathbf{X}_{t_{k-1}}}{h}. \quad (33)$$

From SDE (2) it follows that

$$\Delta_h \mathbf{X}_{t_k} = \frac{1}{h} \int_{t_{k-1}}^{t_k} \mathbf{V}_t dt. \quad (34)$$

We propose to approximate \mathbf{V}_{t_k} using $\Delta_h \mathbf{X}_{t_k}$ by any of the three approaches:

1. Backward difference approximation: $\mathbf{V}_{t_k} \approx \Delta_h \mathbf{X}_{t_k}$;
2. Forward difference approximation: $\mathbf{V}_{t_k} \approx \Delta_h \mathbf{X}_{t_{k+1}}$;
3. Central difference approximation: $\mathbf{V}_{t_k} \approx \frac{\Delta_h \mathbf{X}_{t_k} + \Delta_h \mathbf{X}_{t_{k+1}}}{2}$.

The forward difference approximation performs best in our simulation study, which is also the approximation method employed in Gloter (2006) and Samson and Thieullen (2012).

In the field of numerical approximations of ODEs, backward and forward finite differences have the same order of convergence, whereas the central difference has a higher convergence rate. However, the diffusion parameter estimator based on the central difference $(\mathbf{X}_{t_{k+1}} - \mathbf{X}_{t_{k-1}})/2h$ is less suitable because this approximation skips a data point and thus increases the estimator's variance. For further discussion, see Remark 6.

Thus, we focus exclusively on forward differences, following Gloter (2006); Samson and Thieullen (2012), and all proofs are done for this approximation. Similar results also hold for the backward difference, with some adjustments needed in the conditional moments due to filtration issues.

We start by approximating $\tilde{\mathbf{Z}}$ for the case of partial observations denoted by $\tilde{\mathbf{Z}}$:

$$\tilde{\mathbf{Z}}_{k+1,k,k-1}(\boldsymbol{\beta}) := \tilde{\mathbf{f}}_{h/2}^{-1}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}; \boldsymbol{\beta}) - \tilde{\boldsymbol{\mu}}_h(\tilde{\mathbf{f}}_{h/2}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}; \boldsymbol{\beta}); \boldsymbol{\beta}). \quad (35)$$

The smooth and rough parts of $\tilde{\mathbf{Z}}$ are thus equal to:

$$\bar{\mathbf{Z}}_{k,k-1}^{[S]}(\boldsymbol{\beta}) := \mathbf{X}_{t_k} - \boldsymbol{\mu}_h^{[S]}(\tilde{\mathbf{f}}_{h/2}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}; \boldsymbol{\beta}); \boldsymbol{\beta}), \quad (36)$$

$$\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\boldsymbol{\beta}) := \mathbf{f}_{h/2}^{*-1}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}; \boldsymbol{\beta}) - \boldsymbol{\mu}_h^{[R]}(\tilde{\mathbf{f}}_{h/2}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}; \boldsymbol{\beta}); \boldsymbol{\beta}), \quad (37)$$

and

$$\bar{\mathbf{Z}}_{k+1,k,k-1}^{[S|R]}(\boldsymbol{\beta}) := \bar{\mathbf{Z}}_{k,k-1}^{[S]}(\boldsymbol{\beta}) - \boldsymbol{\Omega}_h^{[SR]}(\boldsymbol{\Omega}_h^{[RR]})^{-1} \bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\boldsymbol{\beta}). \quad (38)$$

Compared to $\mathbf{Z}_{k,k-1}^{[R]}$ in (27), $\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}$ in (37) depends on three consecutive data points, with the additional point $\mathbf{X}_{t_{k+1}}$ entering through $\Delta_h \mathbf{X}_{t_{k+1}}$. Furthermore, \mathbf{X}_{t_k} enters both $\mathbf{f}_{h/2}^{*-1}$ and $\tilde{\boldsymbol{\mu}}_h^{[R]}$, rendering them coupled. This coupling has a significant influence on later derivations of the estimator's asymptotic properties, in contrast to the elliptic case where the derivations simplify.

While it might seem straightforward to incorporate $\tilde{\mathbf{Z}}$, $\bar{\mathbf{Z}}_{k,k-1}^{[S]}$ and $\bar{\mathbf{Z}}_{k,k-1}^{[R]}$ into the objective functions (25), (30) and (31), it introduces bias in the estimators of the diffusion parameters, as also discussed in (Gloter, 2006; Samson and Thieullen, 2012). The bias arises because \mathbf{X}_{t_k} enters in both $\mathbf{f}_{h/2}^{*-1}$ and $\tilde{\boldsymbol{\mu}}_h^{[R]}$, and the covariances of $\tilde{\mathbf{Z}}$, $\bar{\mathbf{Z}}_{k,k-1}^{[S]}$, and $\bar{\mathbf{Z}}_{k,k-1}^{[R]}$ differ from their complete observation counterparts. To eliminate this bias, Gloter (2006); Samson and Thieullen (2012) applied a correction of $2/3$ multiplied to $\log \det$ of the covariance term in the objective functions, which is $\log \det \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top$ in the Euler-Maruyama discretization. We also need appropriate corrections to our objective functions (25), (30) and (31), however, caution is necessary because $\log \det \tilde{\boldsymbol{\Omega}}_h(\boldsymbol{\theta})$ depends on both drift and diffusion parameters. To counterbalance this, we also incorporate an adjustment to h in $\boldsymbol{\Omega}_h$. Moreover, we add the term $4 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}|$ to objective function (31) to obtain consistency of the drift estimator under partial observations. The detailed derivation of these correction factors will be elaborated in the following sections.

We thus propose the following objective functions:

$$\begin{aligned} \mathcal{L}^{[\text{PF}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}) &:= \frac{4}{3}(N-2) \log \det \tilde{\boldsymbol{\Omega}}_{3h/4}(\boldsymbol{\theta}) \\ &+ \sum_{k=1}^{N-1} \left(\tilde{\mathbf{z}}_{k+1,k,k-1}(\boldsymbol{\beta})^\top \tilde{\boldsymbol{\Omega}}_h(\boldsymbol{\theta})^{-1} \tilde{\mathbf{z}}_{k+1,k,k-1}(\boldsymbol{\beta}) + 6 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}; \boldsymbol{\beta})| \right), \end{aligned} \quad (39)$$

$$\begin{aligned} \mathcal{L}^{[\text{PR}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}) &:= \frac{2}{3}(N-2) \log \det \boldsymbol{\Omega}_{3h/2}^{[\text{RR}]}(\boldsymbol{\theta}) \\ &+ \sum_{k=1}^{N-1} \left(\bar{\mathbf{z}}_{k+1,k,k-1}^{[\text{R}]}(\boldsymbol{\beta})^\top \boldsymbol{\Omega}_h^{[\text{RR}]}(\boldsymbol{\theta})^{-1} \bar{\mathbf{z}}_{k+1,k,k-1}^{[\text{R}]}(\boldsymbol{\beta}) + 2 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}; \boldsymbol{\beta})| \right), \end{aligned} \quad (40)$$

$$\mathcal{L}^{[\text{PS|R}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}) := 2(N-2) \log \det \boldsymbol{\Omega}_h^{[\text{S|R}]}(\boldsymbol{\theta}) \quad (41)$$

$$+ \sum_{k=1}^{N-1} \left(\bar{\mathbf{z}}_{k+1,k,k-1}^{[\text{S|R}]}(\boldsymbol{\beta})^\top \boldsymbol{\Omega}_h^{[\text{S|R}]}(\boldsymbol{\theta})^{-1} \bar{\mathbf{z}}_{k+1,k,k-1}^{[\text{S|R}]}(\boldsymbol{\beta}) + 4 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}; \boldsymbol{\beta})| \right). \quad (42)$$

Remark 2 Due to the correction factors in the objective functions, we now have that

$$\mathcal{L}^{[\text{PF}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}) \neq \mathcal{L}^{[\text{PR}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}) + \mathcal{L}^{[\text{PS|R}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}). \quad (43)$$

However, when expanding the objective functions (39)-(41) using Taylor series to the lowest necessary order in h , their approximations will satisfy equality in (43), as shown in Section 6.

Remark 3 Adding the extra term $4 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}|$ in (41) is necessary to keep the consistency of the drift parameter. However, this term is not initially present in objective function (31), making this correction somehow artificial. This can potentially make the objective function further from the true log-likelihood.

The estimators based on the partial sample are then defined as:

$$\hat{\boldsymbol{\theta}}_N^{[\text{obj}]} := \arg \min_{\boldsymbol{\theta}} \mathcal{L}^{[\text{obj}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta}), \quad \text{obj} \in \{[\text{PF}], [\text{PR}]\}. \quad (44)$$

In the partial observation case, the asymptotic variances of the diffusion estimators are identical whether using (39) or (40), in contrast to the complete observation scenario. This variance is shown to be $9/4$ times higher than the variance of the estimator $\hat{\boldsymbol{\theta}}_N^{[\text{CF}]}$, and $9/8$ times higher than that of the estimator based on the marginal likelihood $\hat{\boldsymbol{\theta}}_N^{[\text{CR}]}$.

The numerical study in Section 4 shows that the estimator based on the marginal objective function (40) is less biased than the one based on the full objective function (39) in finite sample scenarios with partial observations. A potential reason for this is discussed in Remark 3. Therefore, we recommend using the objective function (40) for partial observations.

3 Main results

This section states the two main results – consistency and asymptotic normality of all four proposed estimators. The key ideas for proofs are presented in Supplementary Materials S1.

First, we state the consistency of the estimators in both complete and partial observation cases. Let $\mathcal{L}_N^{[\text{obj}]}$ be one of the objective functions (25), (30), (39) or (40) and $\hat{\boldsymbol{\theta}}_N^{[\text{obj}]}$ the corresponding estimator. Thus,

$$\text{obj} \in \{[\text{CF}], [\text{CR}], [\text{PF}], [\text{PR}]\}.$$

We use superscript $[\text{C}\cdot]$ to refer to any objective function in the complete observation case. Likewise, $[\cdot\text{R}]$ stands for an objective function based on the rough marginal likelihood either in the complete or the partial observation case.

Theorem 3.1 (Consistency of the estimators) Assume (A1)-(A6), $h \rightarrow 0$, and $Nh \rightarrow \infty$. Then under the complete observation or partial observation case, it holds:

$$\hat{\boldsymbol{\beta}}_N^{[\text{obj}]} \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \boldsymbol{\beta}_0, \quad \widehat{\boldsymbol{\Sigma}}_N^{[\text{obj}]} \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top.$$

Remark 4 We split the full objective function (25) into the sum of the rough marginal likelihood (30) and the conditional smooth-given-rough likelihood (31). Even if (31) cannot identify the drift parameter β , it is an important intermediate step in understanding the full objective function (25). This can be seen in the proof of Theorem 3.1, where we first establish consistency of the diffusion estimator with a convergence rate of \sqrt{N} , which is faster than \sqrt{Nh} , the convergence rate of the drift estimators. Then, under complete observations, we show that

$$\frac{1}{Nh}(\mathcal{L}_N^{[\text{CR}]}(\beta, \sigma_0) - \mathcal{L}_N^{[\text{CR}]}(\beta_0, \sigma_0)) \xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}} \int (\mathbf{F}_0(\mathbf{y}) - \mathbf{F}(\mathbf{y}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}_0(\mathbf{y}) - \mathbf{F}(\mathbf{y})) d\nu_0(\mathbf{y}). \quad (45)$$

The right-hand side of (45) is non-negative, with a unique zero for $\mathbf{F} = \mathbf{F}_0$. Conversely, for objective function (31), it holds:

$$\frac{1}{Nh}(\mathcal{L}_N^{[\text{CS|R}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{CS|R}]}(\beta_0, \sigma)) \xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}} 0. \quad (46)$$

Hence, (46) does not have a unique minimum, making the drift parameter unidentifiable. Similar conclusions are drawn in the partial observation case.

Now, we state the asymptotic normality of the estimator. First, we need some preliminaries. Let $\rho > 0$ and $\mathcal{B}_\rho(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \Theta \mid \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \rho\}$ be a ball around $\boldsymbol{\theta}_0$. Since $\boldsymbol{\theta}_0 \in \Theta$, for sufficiently small $\rho > 0$, $\mathcal{B}_\rho(\boldsymbol{\theta}_0) \in \Theta$. For $\hat{\boldsymbol{\theta}}_N^{[\text{obj}]} \in \mathcal{B}_\rho(\boldsymbol{\theta}_0)$, the mean value theorem yields:

$$\left(\int_0^1 \mathbb{H}_{\mathcal{L}_N^{[\text{obj}]}}(\boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}}_N^{[\text{obj}]} - \boldsymbol{\theta}_0)) dt \right) (\hat{\boldsymbol{\theta}}_N^{[\text{obj}]} - \boldsymbol{\theta}_0) = -\nabla_{\boldsymbol{\theta}} \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}_0). \quad (47)$$

Define:

$$\mathbf{C}_N^{[\text{obj}]}(\boldsymbol{\theta}) := \begin{bmatrix} \left[\frac{1}{Nh} \partial_{\beta^{(i_1)} \beta^{(i_2)}}^2 \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}) \right]_{i_1, i_2=1}^r & \left[\frac{1}{N\sqrt{h}} \partial_{\beta^{(i)} \sigma^{(j)}}^2 \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}) \right]_{i=1, j=1}^{r, s} \\ \left[\frac{1}{N\sqrt{h}} \partial_{\sigma^{(j)} \beta^{(i)}}^2 \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}) \right]_{i=1, j=1}^{r, s} & \left[\frac{1}{N} \partial_{\sigma^{(j_1)} \sigma^{(j_2)}}^2 \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}) \right]_{j_1, j_2=1}^s \end{bmatrix}, \quad (48)$$

$$\mathbf{s}_N^{[\text{obj}]} := \begin{bmatrix} \sqrt{Nh}(\hat{\boldsymbol{\beta}}_N^{[\text{obj}]} - \boldsymbol{\beta}_0) \\ \sqrt{N}(\hat{\boldsymbol{\sigma}}_N^{[\text{obj}]} - \boldsymbol{\sigma}_0) \end{bmatrix}, \quad \boldsymbol{\lambda}_N^{[\text{obj}]} := \begin{bmatrix} -\frac{1}{\sqrt{Nh}} \nabla_{\boldsymbol{\beta}} \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}_0) \\ -\frac{1}{\sqrt{N}} \nabla_{\boldsymbol{\sigma}} \mathcal{L}_N^{[\text{obj}]}(\boldsymbol{\theta}_0) \end{bmatrix}, \quad (49)$$

and $\mathbf{D}_N^{[\text{obj}]} := \int_0^1 \mathbf{C}_N^{[\text{obj}]}(\boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}}_N^{[\text{obj}]} - \boldsymbol{\theta}_0)) dt$. Then, (47) is equivalent to $\mathbf{D}_N^{[\text{obj}]} \mathbf{s}_N^{[\text{obj}]} = \boldsymbol{\lambda}_N^{[\text{obj}]}$. Let:

$$[\mathbf{C}_\beta(\boldsymbol{\theta}_0)]_{i_1, i_2} := \int (\partial_{\beta^{(i_1)}} \mathbf{F}_0(\mathbf{y}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} (\partial_{\beta^{(i_2)}} \mathbf{F}_0(\mathbf{y})) d\nu_0(\mathbf{y}), \quad 1 \leq i_1, i_2 \leq r, \quad (50)$$

$$[\mathbf{C}_\sigma(\boldsymbol{\theta}_0)]_{j_1, j_2} := \text{Tr}((\partial_{\sigma^{(j_1)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} (\partial_{\sigma^{(j_2)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1}), \quad 1 \leq j_1, j_2 \leq s. \quad (51)$$

Theorem 3.2 Let assumptions (A1)-(A6) hold, and let $h \rightarrow 0$, $Nh \rightarrow \infty$, and $Nh^2 \rightarrow 0$. Then under complete observations, it holds:

$$\begin{bmatrix} \sqrt{Nh}(\hat{\boldsymbol{\beta}}_N^{[\text{CR}]} - \boldsymbol{\beta}_0) \\ \sqrt{N}(\hat{\boldsymbol{\sigma}}_N^{[\text{CR}]} - \boldsymbol{\sigma}_0) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{C}_\beta(\boldsymbol{\theta}_0)^{-1} & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & 2\mathbf{C}_\sigma(\boldsymbol{\theta}_0)^{-1} \end{bmatrix} \right),$$

$$\begin{bmatrix} \sqrt{Nh}(\hat{\boldsymbol{\beta}}_N^{[\text{CF}]} - \boldsymbol{\beta}_0) \\ \sqrt{N}(\hat{\boldsymbol{\sigma}}_N^{[\text{CF}]} - \boldsymbol{\sigma}_0) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{C}_\beta(\boldsymbol{\theta}_0)^{-1} & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \mathbf{C}_\sigma(\boldsymbol{\theta}_0)^{-1} \end{bmatrix} \right),$$

under \mathbb{P}_{θ_0} . If only partial observations are available and the unobserved coordinates are approximated using the forward or backward differences, then

$$\begin{bmatrix} \sqrt{Nh}(\hat{\boldsymbol{\beta}}_N^{[\text{PR}]} - \boldsymbol{\beta}_0) \\ \sqrt{N}(\hat{\boldsymbol{\sigma}}_N^{[\text{PR}]} - \boldsymbol{\sigma}_0) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{C}_\beta(\boldsymbol{\theta}_0)^{-1} & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \frac{9}{4} \mathbf{C}_\sigma(\boldsymbol{\theta}_0)^{-1} \end{bmatrix} \right),$$

$$\begin{bmatrix} \sqrt{Nh}(\hat{\boldsymbol{\beta}}_N^{[\text{PF}]} - \boldsymbol{\beta}_0) \\ \sqrt{N}(\hat{\boldsymbol{\sigma}}_N^{[\text{PF}]} - \boldsymbol{\sigma}_0) \end{bmatrix} \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{C}_\beta(\boldsymbol{\theta}_0)^{-1} & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \frac{9}{4} \mathbf{C}_\sigma(\boldsymbol{\theta}_0)^{-1} \end{bmatrix} \right),$$

under \mathbb{P}_{θ_0} .

Here, we only outline the proof. According to Theorem 1 in Kessler (1997) or Theorem 1 in Sørensen and Uchida (2003), Lemmas 3.3 and 3.4 below are enough for establishing asymptotic normality of $\hat{\theta}_N$. For more details, see proof of Theorem 1 in Sørensen and Uchida (2003).

Lemma 3.3 *Let $\mathbf{C}_N(\theta_0)$ be defined in (48). For $h \rightarrow 0$ and $Nh \rightarrow \infty$, it holds:*

$$\begin{aligned} \mathbf{C}_N^{[\text{CR}]}(\theta_0) &\xrightarrow{\mathbb{P}_{\theta_0}} \begin{bmatrix} 2\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \mathbf{C}_\sigma(\theta_0) \end{bmatrix}, & \mathbf{C}_N^{[\text{PR}]}(\theta_0) &\xrightarrow{\mathbb{P}_{\theta_0}} \begin{bmatrix} 2\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \frac{2}{3}\mathbf{C}_\sigma(\theta_0) \end{bmatrix}, \\ \mathbf{C}_N^{[\text{CF}]}(\theta_0) &\xrightarrow{\mathbb{P}_{\theta_0}} \begin{bmatrix} 2\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & 2\mathbf{C}_\sigma(\theta_0) \end{bmatrix}, & \mathbf{C}_N^{[\text{PF}]}(\theta_0) &\xrightarrow{\mathbb{P}_{\theta_0}} \begin{bmatrix} 2\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \frac{8}{3}\mathbf{C}_\sigma(\theta_0) \end{bmatrix}. \end{aligned}$$

Moreover, let ρ_N be a sequence such that $\rho_N \rightarrow 0$, then in all cases it holds:

$$\sup_{\|\theta\| \leq \rho_N} \|\mathbf{C}_N^{[\text{obj}]}(\theta_0 + \theta) - \mathbf{C}_N^{[\text{obj}]}(\theta_0)\| \xrightarrow{\mathbb{P}_{\theta_0}} 0.$$

Lemma 3.4 *Let λ_N be defined (49). For $h \rightarrow 0$, $Nh \rightarrow \infty$ and $Nh^2 \rightarrow 0$, it holds:*

$$\begin{aligned} \lambda_N^{[\text{CR}]} &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 4\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & 2\mathbf{C}_\sigma(\theta_0) \end{bmatrix}\right), & \lambda_N^{[\text{PR}]} &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 4\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & \mathbf{C}_\sigma(\theta_0) \end{bmatrix}\right), \\ \lambda_N^{[\text{CF}]} &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 4\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & 4\mathbf{C}_\sigma(\theta_0) \end{bmatrix}\right), & \lambda_N^{[\text{PF}]} &\xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 4\mathbf{C}_\beta(\theta_0) & \mathbf{0}_{r \times s} \\ \mathbf{0}_{s \times r} & 16\mathbf{C}_\sigma(\theta_0) \end{bmatrix}\right), \end{aligned}$$

under \mathbb{P}_{θ_0} .

Now, the two previous lemmas suggest

$$\mathbf{s}_N^{[\text{obj}]} = (\mathbf{D}_n^{[\text{obj}]})^{-1} \lambda_N^{[\text{obj}]} \xrightarrow{d} \mathbf{C}_N^{[\text{obj}]}(\theta_0)^{-1} \lambda_N^{[\text{obj}]}.$$

The previous line is not completely formal, but it gives the intuition. For more details on formally deriving the result, see Section 7.4 in Pilipovic et al. (2024) or proof of Theorem 1 in Sørensen and Uchida (2003).

4 Simulation study

This Section illustrates the simulation study of the Kramers oscillator (8), demonstrating the theoretical aspects and comparing our proposed estimators against estimators based on the EM and LL approximations. We chose to compare our proposed estimators to these two, because the EM estimator is routinely used in applications, and the LL estimator has shown to be one of the best state-of-the-art methods, see Pilipovic et al. (2024) for the elliptic case. The true parameters are set to $\eta_0 = 6.5$, $a_0 = 1$, $b_0 = 0.6$ and $\sigma_0^2 = 0.1$. We outline the estimators specifically designed for the Kramers oscillator, explain the simulation procedure, describe the optimization implemented in the R programming language R Core Team (2022), and then present and interpret the results.

4.1 Estimators used in the study

For the Kramers oscillator (8), the EM transition distribution is:

$$\begin{bmatrix} X_{t_k} \\ V_{t_k} \end{bmatrix} \mid \begin{bmatrix} X_{t_{k-1}} \\ V_{t_{k-1}} \end{bmatrix} = \begin{bmatrix} x \\ v \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} x + hv \\ v + h(-\eta v + ax - bx^3) \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & h\sigma^2 \end{bmatrix}\right).$$

The ill-conditioned variance of this discretization restricts us to an estimator that only uses the marginal likelihood of the rough coordinate. The estimator for complete observations directly follows from the Gaussian distribution. The estimator for partial observations is defined as (Samson and Thieullen, 2012):

$$\hat{\theta}_{\text{EM}}^{[\text{PR}]} = \arg \min_{\theta} \left\{ \frac{2}{3}(N-3) \log \sigma^2 + \frac{1}{h\sigma^2} \sum_{k=1}^{N-2} (\Delta_h X_{t_{k+1}} - \Delta_h X_{t_k} - h(-\eta \Delta_h X_{t_{k-1}} + a X_{t_{k-1}} - b X_{t_{k-1}}^3))^2 \right\}.$$

To our knowledge, the LL estimator has not previously been applied to partial observations. Given the similar theoretical and computational performance of the Strang and LL discretizations, we suggest (without formal proof) to adjust the LL objective functions with the same correction factors as used in the Strang approach. The numerical evidence indicates

that the LL estimator has the same asymptotic properties as those proved for the Strang estimator. We omit the definition of the LL estimator due to its complexity (see Melnykova (2020); Pilipovic et al. (2024) and accompanying code).

To define S estimators based on the Strang splitting scheme, we first split SDE (8) as follows:

$$d \begin{bmatrix} X_t \\ V_t \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -2a & -\eta \end{bmatrix}}_{\mathbf{A}} \left(\begin{bmatrix} X_t \\ V_t \end{bmatrix} - \underbrace{\begin{bmatrix} x_{\pm}^* \\ 0 \end{bmatrix}}_{\mathbf{b}} \right) dt + \underbrace{\begin{bmatrix} 0 \\ aX_t - bX_t^3 + 2a(X_t - x_{\pm}^*) \end{bmatrix}}_{\mathbf{N}(X_t, V_t)} dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dW_t,$$

where $x_{\pm}^* = \pm\sqrt{a/b}$ are the two stable points of the dynamics. Since there are two stable points, we suggest splitting with x_+^* , when $X_t > 0$, and x_-^* , when $X_t < 0$. This splitting follows the guidelines from (Pilipovic et al., 2024). Note that the nonlinear ODE driven by $\mathbf{N}(x, v)$ has a trivial solution where x is a constant. To obtain Strang estimators, we plug in the corresponding components in the objective functions (25), (30), (39) and (40).

4.2 Trajectory simulation

We simulate a sample path using the EM discretization with a step size of $h^{\text{sim}} = 0.0001$ to ensure good performance. To reduce discretization errors, we sub-sample from the path at wider intervals to get time step $h = 0.1$. The path has $N = 5000$ data points. We repeat the simulations to obtain 250 data sets.

4.3 Optimization in R

For optimizing the objective functions, we proceed as in Pilipovic et al. (2024) using the R package `torch` (Falbel and Luraschi, 2022), which allows automatic differentiation. The optimization employs the resilient backpropagation algorithm, `optim_rprop`. We use the default hyperparameters and limit the number of optimization iterations to 2000. The convergence criterion is set to a precision of 10^{-5} for the difference between estimators in consecutive iterations. The initial parameter values are set to $(-0.1, -0.1, 0.1, 0.1)$.

4.4 Results

The results of the simulation study are presented in Figure 1. Figure 1A) presents the distributions of the normalized estimators in the complete and partial observation cases. The S and LL estimators exhibit nearly identical performance, particularly in the complete observation scenario. In contrast, the EM method displays significant underperformance and notable bias. The variances of the S and LL rough-likelihood estimators of σ^2 are higher compared to those derived from the full likelihood, aligning with theoretical expectations. Interestingly, in the partial observation scenario, Figure 1A) reveals that estimators employing the full likelihood display greater finite sample bias compared to those based on the rough likelihood. Possible reasons for this bias are discussed in Remark 3. However, it is noteworthy that this bias is eliminated for smaller time steps, e.g. $h = 0.0001$ (not shown), thus confirming the theoretical asymptotic results. This observation suggests that the rough likelihood is preferable under partial observations due to its lower bias. Backward finite difference approximations of the velocity variables perform similarly to the forward differences and are therefore excluded from the figure for clarity.

We closely examine the variances of the S estimators of σ^2 in Figure 1B). The LL estimators are omitted due to their similarity to the S estimators, and because the computation times for the LL estimators are prohibitive. To align more closely with the asymptotic predictions, we opt for $h = 0.02$ and conduct 1000 simulations. Additionally, we set $\sigma_0^2 = 100$ to test different noise levels. Atop each empirical distribution, we overlay theoretical normal densities that match the variances as per Theorem 3.2. The theoretical variance is derived from $C_{\sigma^2}(\boldsymbol{\theta}_0)$ in (51), which for the Kramers oscillator in (8) is:

$$C_{\sigma^2}(\boldsymbol{\theta}_0) = \frac{1}{\sigma_0^4}. \quad (52)$$

Figure 1 illustrates that the lowest variance of the diffusion estimator is observed when using the full likelihood with complete observations. The second lowest variance is achieved using the rough likelihood with complete observations. The largest variance is observed in the partial observation case; however, it remains independent of whether the full or rough likelihood is used. Once again, we observe that using the full likelihood introduces additional finite sample bias.

In Figure 1C), we compare running times calculated using the `tictoc` package in R. Running times are measured from the start of the optimization step until convergence. The figure depicts the median over 250 repetitions to mitigate the influence of outliers. The EM method is notably the fastest; however, the S estimators exhibit only slightly slower performance. The LL estimators are 10-100 times slower than the S estimators, depending on whether complete or partial observations are used and whether the full or rough likelihood is employed.

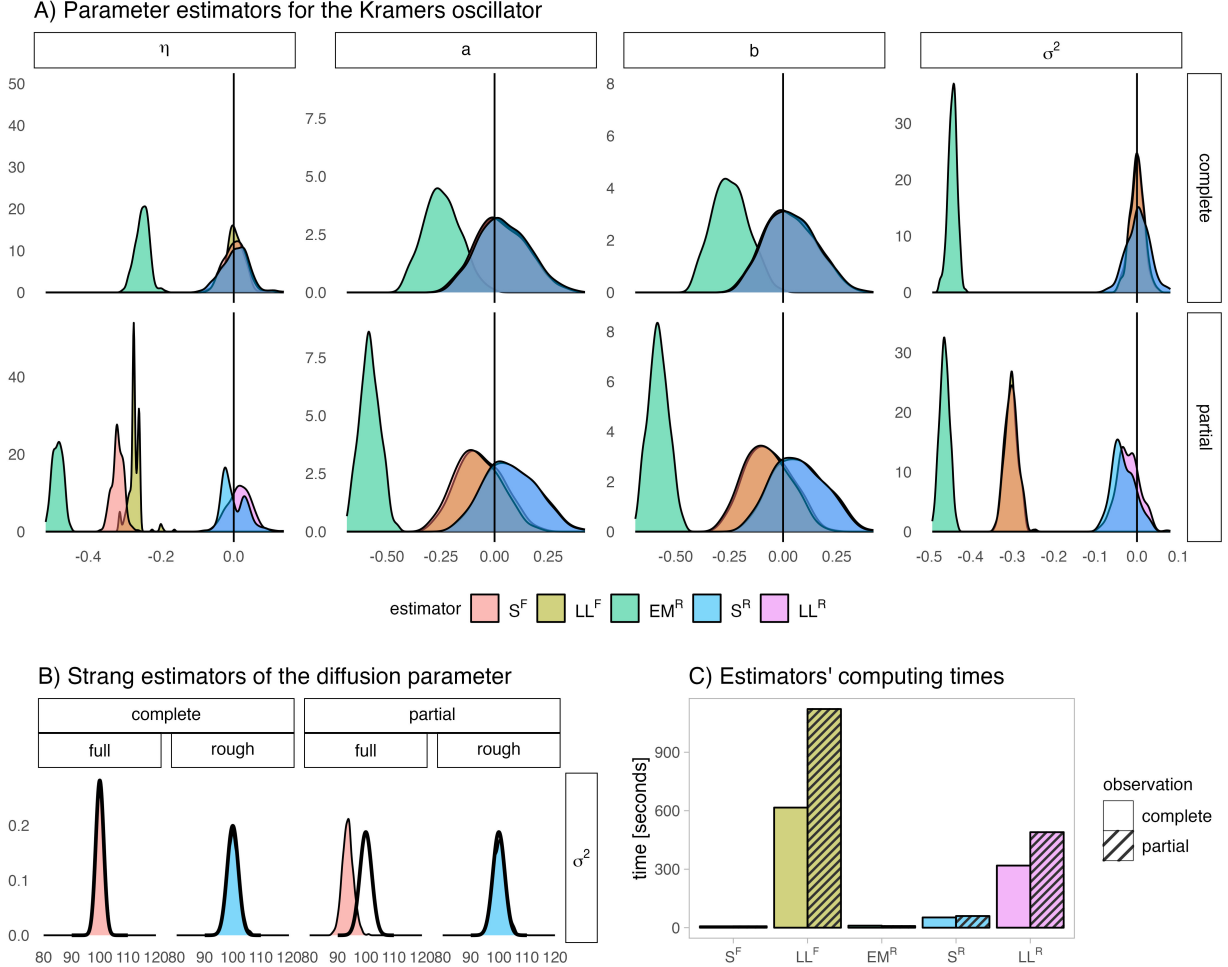


Figure 1: Parameter estimates in a simulation study for the Kramers oscillator, eq. (8). The color code remains consistent across all three figures. **A)** Normalized distributions of parameter estimation errors $(\hat{\theta}_N - \theta_0) \oslash \theta_0$ in both complete and partial observation cases, based on 250 simulated data sets with $h = 0.1$ and $N = 5000$. Each column corresponds to a different parameter, while the color indicates the type of estimator. Estimators are distinguished by superscripted objective functions (F for full and R for rough). **B)** Distribution of $\hat{\sigma}_N^2$ estimators based on 1000 simulations with $h = 0.02$ and $N = 5000$ across different observation settings (complete or partial) and likelihood choices (full or rough) using the Strang splitting scheme. The true value of σ^2 is set to $\sigma_0^2 = 100$. Theoretical normal densities are overlaid for comparison. Theoretical variances are calculated based on $C_{\sigma^2}(\theta_0)$, eq. (52). **C)** Median computing time in seconds for one estimation of various estimators based on 250 simulations with $h = 0.1$ and $N = 5000$. Shaded color patterns represent times in the partial observation case, while no color pattern indicates times in the complete observation case.

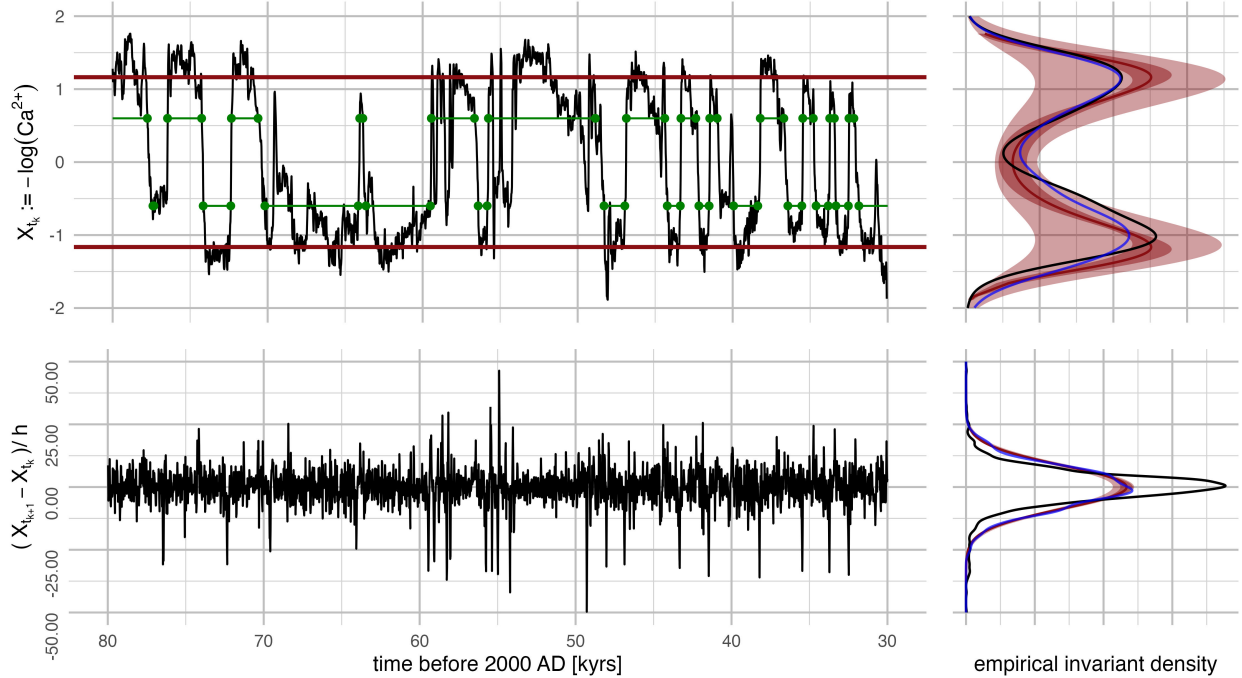


Figure 2: Ice core data from Greenland. **Left:** Trajectories over time (in kilo years) of the centered negative logarithm of the Ca^{2+} measurements (top) and forward difference approximations of its rate of change (bottom). The two vertical dark red lines represent the estimated stable equilibria of the double-well potential function. Green points denote up- and down-crossings of level ± 0.6 , conditioned on having crossed the other level. Green vertical lines indicate empirical estimates of occupancy in either of the two metastable states. **Right:** Empirical densities (black) alongside estimated invariant densities with confidence intervals (dark red), prediction intervals (light red), and the empirical density of a simulated sample from the estimated model (blue).

5 Application to Greenland Ice Core Data

During the last glacial period, significant climatic shifts known as Dansgaard-Oeschger (DO) events have been documented in paleoclimatic records (Dansgaard et al., 1993). Proxy data from Greenland ice cores, particularly stable water isotope composition ($\delta^{18}\text{O}$) and calcium ion concentrations (Ca^{2+}), offer valuable insights into these past climate variations (Boers et al., 2017, 2018; Boers, 2018; Ditlevsen et al., 2002; Lohmann and Ditlevsen, 2019; Hassanibesheli et al., 2020).

The $\delta^{18}\text{O}$ ratio, reflecting the relative abundance of ^{18}O and ^{16}O isotopes in ice, serves as a proxy for paleotemperatures during snow deposition. Conversely, calcium ions, originating from dust deposition, exhibit a strong negative correlation with $\delta^{18}\text{O}$, with higher calcium ion levels indicating colder conditions. Here, we prioritize Ca^{2+} time series due to its finer temporal resolution.

In Greenland ice core records, the DO events manifest as abrupt transitions from colder climates (stadials) to approximately 10 degrees warmer climates (interstadials) within a few decades. Although the waiting times between state switches last a couple of thousand years, their spacing exhibits significant variability. The underlying mechanisms driving these changes remain largely elusive, prompting discussions on whether they follow cyclic patterns, result from external forcing, or emerge from noise-induced processes (Boers, 2018; Ditlevsen et al., 2007). We aim to determine if the observed data can be explained by noise-induced transitions of the Kramers oscillator.

The measurements were conducted at the summit of the Greenland ice sheet as part of the Greenland Icecore Project (GRIP) (Anklin et al., 1993; Andersen et al., 2004). Originally, the data were sampled at 5 cm intervals, resulting in a non-equidistant time series due to ice compression at greater depths, where 5 cm of ice core spans longer time periods. For our analysis, we use a version of the data transformed into a uniformly spaced series through 20-year binning and averaging. This transformation simplifies the analysis and highlights significant climatic trends. The dataset is available in the supplementary material of (Rasmussen et al., 2014; Seierstad et al., 2014).

To address the large amplitudes and negative correlation with temperature, we transform the data to minus the logarithm of Ca^{2+} , where higher values of the transformed variable indicate warmer climates at the time of snow deposition. Additionally, we center the transformed measurements around zero. With the 20-year binning, to obtain one point per 20 years, we average across the bins, resulting in a time step of $h = 0.02\text{kyr}$ ($1\text{kyr} = 1000$ years). Additionally, we addressed a few missing values using the `na.approx` function from the `zoo` package. Following the approach of Hassanibesheli et al. (2020), we analyze a subset of the data with a sufficiently good signal-to-noise ratio. Hassanibesheli et al. (2020) examined the data from 30 to 60kyr before present. Here, we extend the analysis to cover 30kyr to 80kyr, resulting in a time interval of $T = 50\text{kyr}$ and a sample size of $N = 2500$. We approximate the velocity of the transformed Ca^{2+} by the forward difference method. The trajectories and empirical invariant distributions are illustrated in Figure 2.

We fit the Kramers oscillator to the $-\log \text{Ca}^{2+}$ time series and estimate parameters using the Strang estimator. Following Theorem 3.2, we compute $\mathbf{C}_\beta(\theta_0)$ from (50). Applying the invariant density $\pi_0(x, v)$ from (10), which decouples into $\pi_0(x)$ (11) and a Gaussian zero-mean and $\sigma_0^2/(2\eta_0)$ variance, leads us to:

$$\mathbf{C}_\beta(\theta_0) = \begin{bmatrix} \frac{1}{2\eta_0} & 0 & 0 \\ 0 & \frac{1}{\sigma_0^2} \int_{-\infty}^{\infty} x^2 \pi_0(x) dx & -\frac{1}{\sigma_0^2} \int_{-\infty}^{\infty} x^4 \pi_0(x) dx \\ 0 & -\frac{1}{\sigma_0^2} \int_{-\infty}^{\infty} x^4 \pi_0(x) dx & \frac{1}{\sigma_0^2} \int_{-\infty}^{\infty} x^6 \pi_0(x) dx \end{bmatrix}. \quad (53)$$

Thus, to obtain 95% confidence intervals (CI) for the estimated parameters, we plug $\hat{\theta}_N$ into (52) and (53). The estimators and confidence intervals are shown in Table 1. We also calculate the expected waiting time τ , eq. (12), of crossing from one state to another, and its confidence interval using the Delta Method.

Parameter	Estimate	95% CI
η	62.5	59.4 – 65.6
a	296.7	293.6 – 299.8
b	219.1	156.4 – 281.7
σ^2	9125	8589 – 9662
τ	3.97	3.00 – 4.94

Table 1: Estimated parameters of the Kramers oscillator from Greenland ice core data.

The model fit is assessed in the right panels of Figure 2. Here, we present the empirical distributions of the two coordinates along with the fitted theoretical invariant distribution and a 95% confidence interval. Additionally, a prediction interval for the distribution is provided by simulating 1000 datasets from the fitted model, matching the size of the empirical data. We estimate the empirical distributions for each simulated dataset and construct a 95% prediction interval using the pointwise 2.5th and 97.5th percentiles of these estimates. A single example trace is included in blue. While the fitted distribution for $-\log \text{Ca}^{2+}$ appears to fit well, even with this symmetric model, the velocity variables are not adequately captured. This discrepancy is likely due to the presence of extreme values in the data that are not effectively accounted for by additive Gaussian noise. Consequently, the model compensates by estimating a large variance.

We estimate the waiting time between metastable states to be approximately 4000 years. However, this approximation relies on certain assumptions, namely $62.5 \approx \eta \gg \sqrt{a} \approx 17.2$ and $73 \approx \sigma^2/2\eta \ll a^2/4b \approx 100$. Thus, the accuracy of the approximation may not be highly accurate.

Defining the current state of the process is not straightforward. One method involves identifying successive up- and down-crossings of predefined thresholds within the smoothed data. However, the estimated occupancy time in each state depends on the level of smoothing applied and the distance of crossing thresholds from zero. Using a smoothing technique involving running averages within windows of 11 data points (equivalent to 220 years) and detecting down- and up-crossings of levels ± 0.6 , we find an average occupancy time of 4058 years in stadial states and 3550 years in interstadial states. Nevertheless, the actual occupancy times exhibit significant variability, ranging from 60 to 6900 years, with the central 50% of values falling between 665 and 2115 years. This classification of states is depicted in green in Figure 2. Overall, the estimated mean occupancy time inferred from the Kramers oscillator appears reasonable.

6 Technical results

In this Section, we present all the necessary technical properties that are used to derive the main results of the paper.

We start by expanding $\tilde{\Omega}_h$ and its block components $\Omega_h^{[RR]}(\theta)^{-1}$, $\Omega_h^{[SR]}(\theta)^{-1}$, $\log \det \Omega_h^{[RR]}(\theta)$, $\log \det \Omega_h^{[SR]}(\theta)$ and $\log |\det Df_{h/2}(\mathbf{y}; \beta)|$ when h goes to zero. Then, we expand $\tilde{\mathbf{Z}}_{k,k-1}(\beta)$ and $\tilde{\mathbf{Z}}_{k+1,k,k-1}(\beta)$ around $\mathbf{Y}_{t_{k-1}}$ when h goes to zero. The main tools used are Itô's lemma, Taylor expansions, and Fubini's theorem. The final result is stated in Propositions 6.6 and 6.7. The approximations depend on the drift function \mathbf{F} , the nonlinear part \mathbf{N} , and some correlated sequences of Gaussian random variables. Finally, we obtain approximations of the objective functions (25), (30), (31) and (39) - (41). Proofs of all the stated propositions and lemmas in this section are in Supplementary Material S1.

6.1 Covariance matrix $\tilde{\Omega}_h$

The covariance matrix $\tilde{\Omega}_h$ is approximated by:

$$\begin{aligned} \tilde{\Omega}_h &= \int_0^h e^{\tilde{\mathbf{A}}(h-u)} \tilde{\Sigma} \tilde{\Sigma}^\top e^{\tilde{\mathbf{A}}^\top(h-u)} du \\ &= h \tilde{\Sigma} \tilde{\Sigma}^\top + \frac{h^2}{2} (\tilde{\mathbf{A}} \tilde{\Sigma} \tilde{\Sigma}^\top + \tilde{\Sigma} \tilde{\Sigma}^\top \tilde{\mathbf{A}}^\top) + \frac{h^3}{6} (\tilde{\mathbf{A}}^2 \tilde{\Sigma} \tilde{\Sigma}^\top + 2 \tilde{\mathbf{A}} \tilde{\Sigma} \tilde{\Sigma}^\top \tilde{\mathbf{A}}^\top + \tilde{\Sigma} \tilde{\Sigma}^\top (\tilde{\mathbf{A}}^2)^\top) \\ &\quad + \frac{h^4}{24} (\tilde{\mathbf{A}}^3 \tilde{\Sigma} \tilde{\Sigma}^\top + 3 \tilde{\mathbf{A}}^2 \tilde{\Sigma} \tilde{\Sigma}^\top \tilde{\mathbf{A}}^\top + 3 \tilde{\mathbf{A}} \tilde{\Sigma} \tilde{\Sigma}^\top (\tilde{\mathbf{A}}^2)^\top + \tilde{\Sigma} \tilde{\Sigma}^\top (\tilde{\mathbf{A}}^3)^\top) + \mathbf{R}(h^5, \mathbf{y}_0). \end{aligned} \quad (54)$$

The following lemma approximates each block of $\tilde{\Omega}_h$ up to the first two leading orders of h . The result follows directly from equations (4), (6), and (54).

Lemma 6.1 *The covariance matrix $\tilde{\Omega}_h$ defined in (54)-(19) approximates block-wise as:*

$$\begin{aligned} \Omega_h^{[SS]}(\theta) &= \frac{h^3}{3} \Sigma \Sigma^\top + \frac{h^4}{8} (\mathbf{A}_v(\beta) \Sigma \Sigma^\top + \Sigma \Sigma^\top \mathbf{A}_v(\beta)^\top) + \mathbf{R}(h^5, \mathbf{y}_0), \\ \Omega_h^{[SR]}(\theta) &= \frac{h^2}{2} \Sigma \Sigma^\top + \frac{h^3}{6} (\mathbf{A}_v(\beta) \Sigma \Sigma^\top + 2 \Sigma \Sigma^\top \mathbf{A}_v(\beta)^\top) + \mathbf{R}(h^4, \mathbf{y}_0), \\ \Omega_h^{[RS]}(\theta) &= \frac{h^2}{2} \Sigma \Sigma^\top + \frac{h^3}{6} (2 \mathbf{A}_v(\beta) \Sigma \Sigma^\top + \Sigma \Sigma^\top \mathbf{A}_v(\beta)^\top) + \mathbf{R}(h^4, \mathbf{y}_0), \\ \Omega_h^{[RR]}(\theta) &= h \Sigma \Sigma^\top + \frac{h^2}{2} (\mathbf{A}_v(\beta) \Sigma \Sigma^\top + \Sigma \Sigma^\top \mathbf{A}_v(\beta)^\top) + \mathbf{R}(h^3, \mathbf{y}_0). \end{aligned}$$

Building on Lemma 6.1, we calculate products, inverses, and logarithms of the components of $\tilde{\Omega}_h$ in the following lemma.

Lemma 6.2 *For the covariance matrix $\tilde{\Omega}_h$ defined in (54) it holds:*

- (i) $\Omega_h^{[RR]}(\theta)^{-1} = \frac{1}{h} (\Sigma \Sigma^\top)^{-1} - \frac{1}{2} ((\Sigma \Sigma^\top)^{-1} \mathbf{A}_v(\beta) + \mathbf{A}_v(\beta)^\top (\Sigma \Sigma^\top)^{-1}) + \mathbf{R}(h, \mathbf{y}_0);$
- (ii) $\Omega_h^{[SR]}(\theta) \Omega_h^{[RR]}(\theta)^{-1} = \frac{h}{2} \mathbf{I} - \frac{h^2}{12} (\mathbf{A}_v - \Sigma \Sigma^\top \mathbf{A}_v(\beta)^\top (\Sigma \Sigma^\top)^{-1}) + \mathbf{R}(h^3, \mathbf{y}_0);$
- (iii) $\Omega_h^{[SR]}(\theta) \Omega_h^{[RR]}(\theta)^{-1} \Omega_h^{[RS]}(\theta) = \frac{h^3}{4} \Sigma \Sigma^\top + \frac{h^4}{8} (\mathbf{A}_v(\beta) \Sigma \Sigma^\top + \Sigma \Sigma^\top \mathbf{A}_v(\beta)^\top) + \mathbf{R}(h^5, \mathbf{y}_0);$
- (iv) $\Omega_h^{[SR]}(\theta) = \frac{h^3}{12} \Sigma \Sigma^\top + \mathbf{R}(h^5, \mathbf{y}_0);$
- (v) $\log \det \Omega_h^{[RR]}(\theta) = d \log h + \log \det \Sigma \Sigma^\top + h \operatorname{Tr} \mathbf{A}_v(\beta) + R(h^2, \mathbf{y}_0);$
- (vi) $\log \det \Omega_h^{[SR]}(\theta) = 3d \log h + \log \det \Sigma \Sigma^\top + R(h^2, \mathbf{y}_0);$
- (vii) $\log \det \tilde{\Omega}_h(\theta) = 4d \log h + 2 \log \det \Sigma \Sigma^\top + h \operatorname{Tr} \mathbf{A}_v(\beta) + R(h^2, \mathbf{y}_0).$

Remark 5 *We adjusted the objective functions for partial observations using the term $c \log \det \Omega_{h/c}^{[\cdot]}$, where c is a correction constant. This adjustment keeps the term $h \operatorname{Tr} \mathbf{A}_v(\beta)$ in (v)-(vii) constant, not affecting the asymptotic distribution of the drift parameter. There is no h^4 -term in $\Omega_h^{[SR]}(\theta)$ which simplifies the approximation of $\Omega_h^{[SR]}(\theta)^{-1}$ and $\log \det \Omega_h^{[SR]}(\theta)$. Consequently, this makes (41) a bad choice for estimating the drift parameter.*

6.2 Nonlinear solution $\tilde{\mathbf{f}}_h$

We now state a useful proposition for the nonlinear solution $\tilde{\mathbf{f}}_h$ (Section 1.8 in (Hairer et al., 1993)).

Proposition 6.3 *Let Assumptions (A1), (A2) and (A6) hold. When $h \rightarrow 0$, the h -flow of (15) approximates as:*

$$\tilde{\mathbf{f}}_h(\mathbf{y}) = \mathbf{y} + h\tilde{\mathbf{N}}(\mathbf{y}) + \frac{h^2}{2}(D_{\mathbf{y}}\tilde{\mathbf{N}}(\mathbf{y}))\tilde{\mathbf{N}}(\mathbf{y}) + \mathbf{R}(h^3, \mathbf{y}), \quad (55)$$

$$\tilde{\mathbf{f}}_h^{-1}(\mathbf{y}) = \mathbf{y} - h\tilde{\mathbf{N}}(\mathbf{y}) + \frac{h^2}{2}(D_{\mathbf{y}}\tilde{\mathbf{N}}(\mathbf{y}))\tilde{\mathbf{N}}(\mathbf{y}) + \mathbf{R}(h^3, \mathbf{y}). \quad (56)$$

Applying the previous proposition on (21) and (22), we get:

$$\mathbf{f}_h(\mathbf{y}) = \mathbf{v} + h\mathbf{N}(\mathbf{y}) + \frac{h^2}{2}(D_{\mathbf{v}}\mathbf{N}(\mathbf{y}))\mathbf{N}(\mathbf{y}) + \mathbf{R}(h^3, \mathbf{y}), \quad (57)$$

$$\mathbf{f}_h^{*-1}(\mathbf{y}) = \mathbf{v} - h\mathbf{N}(\mathbf{y}) + \frac{h^2}{2}(D_{\mathbf{v}}\mathbf{N}(\mathbf{y}))\mathbf{N}(\mathbf{y}) + \mathbf{R}(h^3, \mathbf{y}). \quad (58)$$

The following lemma approximates $\log |\det D\mathbf{f}_{h/2}(\mathbf{y}; \boldsymbol{\beta})|$ in the objective functions and connects it with Lemma 6.2.

Lemma 6.4 *Let $\tilde{\mathbf{f}}_h$ be the function defined in (21). It holds:*

$$\begin{aligned} 2 \log |\det D\mathbf{f}_{h/2}(\mathbf{Y}_{t_k}; \boldsymbol{\beta})| &= h \operatorname{Tr} D_{\mathbf{v}}\mathbf{N}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) + R(h^{3/2}, \mathbf{Y}_{t_{k-1}}), \\ 2 \log |\det D\mathbf{f}_{h/2}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}; \boldsymbol{\beta})| &= h \operatorname{Tr} D_{\mathbf{v}}\mathbf{N}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) + R(h^{3/2}, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

An immediate consequence of the previous lemma and that $D_{\mathbf{v}}\mathbf{F}(\mathbf{y}; \boldsymbol{\beta}) = \mathbf{A}_{\mathbf{v}}(\boldsymbol{\beta}) + D_{\mathbf{v}}\mathbf{N}(\mathbf{y}; \boldsymbol{\beta})$ is

$$\log \det \boldsymbol{\Omega}_h^{\text{[RR]}}(\boldsymbol{\theta}) + 2 \log |\det D\mathbf{f}_{h/2}(\mathbf{Y}_{t_k}; \boldsymbol{\beta})| = \log \det h\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\top} + h \operatorname{Tr} D_{\mathbf{v}}\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) + R(h^{3/2}, \mathbf{Y}_{t_{k-1}}).$$

The same equality holds when \mathbf{Y}_{t_k} is approximated by $(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}})$. The following lemma expands function $\mu_h(\tilde{\mathbf{f}}_{h/2}(\mathbf{y}))$ up to the highest necessary order of h .

Lemma 6.5 *For the functions $\tilde{\mathbf{f}}_h$ in (21) and $\tilde{\boldsymbol{\mu}}_h$ in (28), it holds*

$$\boldsymbol{\mu}_h^{\text{[S]}}(\tilde{\mathbf{f}}_{h/2}(\mathbf{y})) = \mathbf{x} + h\mathbf{v} + \frac{h^2}{2}\mathbf{F}(\mathbf{y}) + \mathbf{R}(h^3, \mathbf{y}), \quad (59)$$

$$\boldsymbol{\mu}_h^{\text{[R]}}(\tilde{\mathbf{f}}_{h/2}(\mathbf{y})) = \mathbf{v} + h(\mathbf{F}(\mathbf{y}) - \frac{1}{2}\mathbf{N}(\mathbf{y})) + \mathbf{R}(h^2, \mathbf{y}). \quad (60)$$

6.3 Random variables $\tilde{\mathbf{Z}}_{k,k-1}$ and $\tilde{\mathbf{Z}}_{k+1,k,k-1}$

To approximate the random variables $\mathbf{Z}_{k,k-1}^{\text{[S]}}(\boldsymbol{\beta})$, $\mathbf{Z}_{k,k-1}^{\text{[R]}}(\boldsymbol{\beta})$, $\bar{\mathbf{Z}}_{k,k-1}^{\text{[S]}}(\boldsymbol{\beta})$, and $\bar{\mathbf{Z}}_{k+1,k,k-1}^{\text{[R]}}(\boldsymbol{\beta})$ around $\mathbf{Y}_{t_{k-1}}$, we start by defining the following random sequences:

$$\boldsymbol{\eta}_{k-1} := \frac{1}{h^{1/2}} \int_{t_{k-1}}^{t_k} d\mathbf{W}_t, \quad (61)$$

$$\boldsymbol{\xi}_{k-1} := \frac{1}{h^{3/2}} \int_{t_{k-1}}^{t_k} (t - t_{k-1}) d\mathbf{W}_t, \quad \boldsymbol{\xi}'_k := \frac{1}{h^{3/2}} \int_{t_k}^{t_{k+1}} (t_{k+1} - t) d\mathbf{W}_t, \quad (62)$$

$$\boldsymbol{\zeta}_{k-1} := \frac{1}{h^{5/2}} \int_{t_{k-1}}^{t_k} (t - t_{k-1})^2 d\mathbf{W}_t, \quad \boldsymbol{\zeta}'_k := \frac{1}{h^{5/2}} \int_{t_k}^{t_{k+1}} (t_{k+1} - t)^2 d\mathbf{W}_t. \quad (63)$$

The random variables (61)-(63) are Gaussian with mean zero. Moreover, at time t_k they are $\mathcal{F}_{t_{k+1}}$ measurable and independent of \mathcal{F}_{t_k} . The following linear combinations of (61)-(63) appear in the expansions in the partial observation case:

$$\mathbf{U}_{k,k-1} := \boldsymbol{\xi}'_k + \boldsymbol{\xi}_{k-1}, \quad (64)$$

$$\mathbf{Q}_{k,k-1} := \boldsymbol{\zeta}'_k + 2\boldsymbol{\eta}_{k-1} - \boldsymbol{\zeta}_{k-1}. \quad (65)$$

It is not hard to check that $\xi'_k + \eta_{k-1} - \xi'_{k-1} = \mathbf{U}_{k,k-1}$. This alternative representation of $\mathbf{U}_{k,k-1}$ will be used later in proofs.

The Itô isometry yields:

$$\mathbb{E}_{\theta_0}[\eta_{k-1}\eta_{k-1}^\top | \mathcal{F}_{t_{k-1}}] = \mathbf{I}, \quad \mathbb{E}_{\theta_0}[\eta_{k-1}\xi_{k-1}^\top | \mathcal{F}_{t_{k-1}}] = \mathbb{E}_{\theta_0}[\eta_{k-1}\xi'_{k-1}{}^\top | \mathcal{F}_{t_{k-1}}] = \frac{1}{2}\mathbf{I}, \quad (66)$$

$$\mathbb{E}_{\theta_0}[\xi_{k-1}\xi'_{k-1}{}^\top | \mathcal{F}_{t_{k-1}}] = \frac{1}{6}\mathbf{I}, \quad \mathbb{E}_{\theta_0}[\xi_{k-1}\xi_k^\top | \mathcal{F}_{t_{k-1}}] = \mathbb{E}_{\theta_0}[\xi'_k\xi_k{}^\top | \mathcal{F}_{t_{k-1}}] = \frac{1}{3}\mathbf{I}, \quad (67)$$

$$\mathbb{E}_{\theta_0}[\mathbf{U}_{k,k-1}\mathbf{U}_{k,k-1}^\top | \mathcal{F}_{t_{k-1}}] = \frac{2}{3}\mathbf{I}, \quad \mathbb{E}_{\theta_0}[\mathbf{U}_{k,k-1}(\mathbf{U}_{k,k-1} + 2\xi'_{k-1})^\top | \mathcal{F}_{t_{k-1}}] = \mathbf{I}. \quad (68)$$

The covariances of other combinations of the random variables (61)-(63) are not needed for the proofs. However, to derive asymptotic properties, we need some fourth moments calculated in Supplementary Materials S1.

The following two propositions are the last building blocks for approximating the objective functions (30)-(31) and (40)-(41).

Proposition 6.6 *The random variables $\tilde{\mathbf{Z}}_{k,k-1}(\beta)$ in (24) and $\tilde{\mathbf{Z}}_{k+1,k,k-1}(\beta)$ in (35) are approximated by:*

$$\begin{aligned} \mathbf{Z}_{k,k-1}^{[S]}(\beta) &= h^{3/2}\Sigma_0\xi'_{k-1} + \frac{h^2}{2}(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) + \frac{h^{5/2}}{2}D_{\mathbf{v}}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\Sigma_0\xi'_{k-1} + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}), \\ \mathbf{Z}_{k,k-1}^{[R]}(\beta) &= h^{1/2}\Sigma_0\eta_{k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h^{3/2}}{2}D_{\mathbf{v}}\mathbf{N}(\mathbf{Y}_{t_{k-1}})\Sigma_0\eta_{k-1} \\ &\quad + h^{3/2}D_{\mathbf{v}}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\Sigma_0\xi'_{k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}), \\ \bar{\mathbf{Z}}_{k,k-1}^{[S]}(\beta) &= -\frac{h^2}{2}\mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \frac{h^{5/2}}{2}D_{\mathbf{v}}\mathbf{F}(\mathbf{Y}_{t_{k-1}})\Sigma_0\xi'_{k-1} + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}), \\ \bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\beta) &= h^{1/2}\Sigma_0\mathbf{U}_{k,k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h^{3/2}}{2}D_{\mathbf{v}}\mathbf{N}(\mathbf{Y}_{t_{k-1}})\Sigma_0\mathbf{U}_{k,k-1} \\ &\quad - h^{3/2}D_{\mathbf{v}}\mathbf{F}(\mathbf{Y}_{t_{k-1}})\Sigma_0\xi'_{k-1} + \frac{h^{3/2}}{2}D_{\mathbf{v}}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\Sigma_0\mathbf{Q}_{k,k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

Remark 6 *Proposition 6.6 yield*

$$\begin{aligned} \mathbb{E}_{\theta_0}[\mathbf{Z}_{k,k-1}^{[R]}(\beta)\mathbf{Z}_{k,k-1}^{[R]}(\beta)^\top | \mathbf{Y}_{t_{k-1}}] &= h\Sigma\Sigma_0^\top + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}) = \Omega_h^{[RR]} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}), \\ \mathbb{E}_{\theta_0}[\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\beta)\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\beta)^\top | \mathbf{Y}_{t_{k-1}}] &= \frac{2}{3}h\Sigma\Sigma_0^\top + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}) = \frac{2}{3}\Omega_h^{[RR]} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

Thus, the correction factor 2/3 in (40) compensates for the underestimation of the covariance of $\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\beta)$. Similarly, it can be shown that the same underestimation happens when using the backward difference. On the other hand, when using the central difference, it can be shown that

$$\mathbb{E}_{\theta_0}[\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R],\text{central}}(\beta)\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R],\text{central}}(\beta)^\top | \mathbf{Y}_{t_{k-1}}] = \frac{5}{12}h\Sigma\Sigma_0^\top + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}),$$

which is a larger deviation from $\Omega_h^{[RR]}$, yielding a larger correcting factor and larger asymptotic variance of the diffusion parameter estimator.

Proposition 6.7 *Let $\tilde{\mathbf{Z}}_{k,k-1}(\beta)$ and $\tilde{\mathbf{Z}}_{k+1,k,k-1}(\beta)$ be defined in (24) and (35), respectively. Then,*

$$\begin{aligned} \mathbf{Z}_{k,k-1}^{[S|R]}(\beta) &= -\frac{h^{3/2}}{2}\Sigma_0(\eta_{k-1} - 2\xi'_{k-1}) + \frac{h^{5/2}}{12}(\mathbf{A}_{\mathbf{v}} - \Sigma\Sigma^\top\mathbf{A}_{\mathbf{v}}(\beta)^\top(\Sigma\Sigma^\top)^{-1})\Sigma_0\eta_{k-1} \\ &\quad + \frac{h^{5/2}}{4}D_{\mathbf{v}}\mathbf{N}(\mathbf{Y}_{t_{k-1}})\Sigma_0\eta_{k-1} - \frac{h^{5/2}}{2}D_{\mathbf{v}}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\Sigma_0(\xi'_{k-1} - \zeta'_{k-1}) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}), \\ \bar{\mathbf{Z}}_{k+1,k,k-1}^{[S|R]}(\beta) &= -\frac{h^{3/2}}{2}\Sigma_0\mathbf{U}_{k,k-1} - \frac{h^2}{2}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \frac{h^{5/2}}{12}(\mathbf{A}_{\mathbf{v}} - \Sigma\Sigma^\top\mathbf{A}_{\mathbf{v}}(\beta)^\top(\Sigma\Sigma^\top)^{-1})\Sigma_0\mathbf{U}_{k,k-1} \\ &\quad + \frac{h^{5/2}}{4}D_{\mathbf{v}}\mathbf{N}(\mathbf{Y}_{t_{k-1}})\Sigma_0\mathbf{U}_{k,k-1} - \frac{h^{5/2}}{4}D_{\mathbf{v}}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\Sigma_0\mathbf{Q}_{k,k-1} + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

6.4 Objective functions

Starting with the complete observation case, we approximate objective functions (30) and (31) up to order $R(h^{3/2}, \mathbf{Y}_{t_{k-1}})$ to prove the asymptotic properties of the estimators $\hat{\boldsymbol{\theta}}_N^{[\text{CR}]}$ and $\hat{\boldsymbol{\theta}}_N^{[\text{CS|R}]}$. After omitting the terms of order $R(h, \mathbf{Y}_{t_{k-1}})$ that do not depend on $\boldsymbol{\beta}$, we obtain the following approximations:

$$\mathcal{L}_N^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = (N-1) \log \det \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top + \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \quad (69)$$

$$\begin{aligned} &+ 2\sqrt{h} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\ &+ h \sum_{k=1}^N (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\ &- h \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} + h \sum_{k=1}^N \text{Tr} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}), \end{aligned}$$

$$\mathcal{L}_N^{[\text{CS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = (N-1) \log \det \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top + 3 \sum_{k=1}^N (\boldsymbol{\eta}_{k-1} - 2\xi'_{k-1})^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 (\boldsymbol{\eta}_{k-1} - 2\xi'_{k-1}) \quad (70)$$

$$\begin{aligned} &- 3h \sum_{k=1}^N (\boldsymbol{\eta}_{k-1} - 2\xi'_{k-1})^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \\ &- h \sum_{k=1}^N (\boldsymbol{\eta}_{k-1} - 2\xi'_{k-1})^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{A}_{\mathbf{v}}(\boldsymbol{\beta}) - \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top \mathbf{A}_{\mathbf{v}}(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \end{aligned}$$

$$\mathcal{L}_N^{[\text{CF}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \mathcal{L}_N^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) + \mathcal{L}_N^{[\text{CS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}). \quad (71)$$

The two last sums in (70) converge to zero because $\mathbb{E}_{\boldsymbol{\theta}_0}[(\boldsymbol{\eta}_{k-1} - 2\xi'_{k-1}) \boldsymbol{\eta}_{k-1}^\top | \mathcal{F}_{t_{k-1}}] = \mathbf{0}$. Moreover, (70) lacks the quadratic form of $\mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})$, that is crucial for the asymptotic variance of the drift estimator. This implies that the objective function $\mathcal{L}_N^{[\text{CS|R}]}$ is not suitable for estimating the drift parameter. Conversely, (70) provides a correct and consistent estimator of the diffusion parameter, indicating that the full objective function (the sum of $\mathcal{L}_N^{[\text{CR}]}$ and $\mathcal{L}_N^{[\text{CS|R}]}$) consistently estimates $\boldsymbol{\theta}$.

Similarly, the approximated objective functions in the partial observation case are:

$$\mathcal{L}_N^{[\text{PR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \frac{2}{3} (N-2) \log \det \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top + \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \quad (72)$$

$$\begin{aligned} &+ 2\sqrt{h} \sum_{k=1}^N \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\ &+ h \sum_{k=1}^{N-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\ &- h \sum_{k=1}^{N-1} (\mathbf{U}_{k,k-1} + 2\xi'_{k-1})^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h \sum_{k=1}^{N-1} \text{Tr} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}), \end{aligned}$$

$$\mathcal{L}_N^{[\text{PS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = 2(N-2) \log \det \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top + 3 \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \quad (73)$$

$$\begin{aligned} &+ 6\sqrt{h} \sum_{k=1}^N \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) \\ &- 3h \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + 2h \sum_{k=1}^{N-1} \text{Tr} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}), \end{aligned}$$

$$\mathcal{L}_N^{[\text{PF}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \mathcal{L}_N^{[\text{PR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) + \mathcal{L}_N^{[\text{PS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}). \quad (74)$$

This time, the term with $\mathbf{A}_v(\boldsymbol{\beta}) - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}$ vanishes because

$$\text{Tr}(\boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} (\mathbf{A}_v(\boldsymbol{\beta}) - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1})) = 0$$

due to the symmetry of the matrices and the trace cyclic property.

Even though the partial observation objective function $\mathcal{L}^{[\text{PR}]}(\mathbf{X}_{0:t_N}; \boldsymbol{\theta})$ (40) depends only on $\mathbf{X}_{0:t_N}$, we could approximate it with $\mathcal{L}_N^{[\text{PR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta})$ (72). This is useful for proving the asymptotic normality of the estimator since its asymptotic distribution will depend on the invariant probability ν_0 defined for the solution \mathbf{Y} .

The absence of the quadratic form $\mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})$ in (73) indicates that $\mathcal{L}_N^{[\text{PS|R}]}$ is not suitable for estimating the drift parameter. Additionally, the penultimate term in (73) does not vanish, needing an additional correction term of $2h \sum_{k=1}^{N-1} \text{Tr} D_v \mathbf{N}(\mathbf{Y}_{t_k}; \boldsymbol{\beta})$ for consistency. This correction is represented as $4 \log |\det D_v \mathbf{f}_{h/2}|$ in (41). Notably, this term is absent in the complete objective function (31), making this adjustment somewhat artificial and could potentially deviate further from the true log-likelihood. Consequently, the objective function based on the full likelihood (39) inherits this characteristic from (73), suggesting that in the partial observation scenario, using only the rough likelihood (72) may be more appropriate.

7 Conclusion

Many fundamental laws of physics and chemistry are formulated as second-order differential equations, a model class important for understanding complex dynamical systems in various fields such as biology and economics. The extension of these deterministic models to stochastic second-order differential equations represents a natural generalization, allowing for the incorporation of uncertainties and variability inherent in real-world systems. However, robust statistical methods for analyzing data generated from such stochastic models have been lacking, presenting a significant challenge due to the inherent degeneracy of the noise and partial observation.

In this study, we propose estimating model parameters using a recently developed methodology of Strang splitting estimator for SDEs. This estimator has demonstrated finite sample efficiency with relatively large sample time steps, particularly in handling highly nonlinear models. We adjust the estimator to the partial observation setting and employ either the full likelihood or only the marginal likelihood based on the rough coordinates. For all four obtained estimators, we establish the consistency and asymptotic normality.

The application of the Strang estimator to a historical paleoclimate dataset obtained from ice cores in Greenland has yielded valuable insights and analytical tools for comprehending abrupt climate shifts throughout history. Specifically, we employed the stochastic Duffing oscillator, also known as the Kramers oscillator, to analyze the data.

While our focus in this paper has been primarily confined to second-order SDEs with no parameters in the smooth components, we are confident that our findings can be extended to encompass models featuring parameters in the drift of the smooth coordinates. This opens up directions for further exploration and application of our methodology to a broader range of complex dynamical systems, promising deeper insights into their behavior and underlying mechanisms.

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956107, "Economic Policy in Complex Environments (EPOC)"; and Novo Nordisk Foundation NNF20OC0062958.

References

- A. Abdule, G. Vilmart, and K. C. Zygalakis. Long Time Accuracy of Lie–Trotter Splitting Methods for Langevin Dynamics. *SIAM Journal on Numerical Analysis*, 53(1):1–16, 2015.
- M. Ableidinger, E. Buckwar, and H. Hinterleitner. A stochastic version of the jansen and rit neural mass model: Analysis and numerics. *Journal of Mathematical Neuroscience*, 7, 2017. ISSN 2190-8567.
- D. Adams, M. H. Duong, and G. dos Reis. Operator-splitting schemes for degenerate, non-local, conservative-dissipative systems. *Discrete and Continuous Dynamical Systems*, 42(11):5453–5486, 2022.
- L. A. Alyushina. Euler Polygonal Lines for Itô Equations with Monotone Coefficients. *Theory of Probability & Its Applications*, 32(2):340–345, 1988.

- K. Andersen, N. Azuma, J. Barnola, M. Bigler, P. Biscaye, N. Caillon, J. Chappellaz, H. Clausen, D. Dahl-Jensen, H. Fischer, J. Flückiger, D. Fritzsche, Y. Fujii, K. Goto-Azuma, K. Grønvold, N. Gundestrup, M. Hansson, C. Huber, C. Hvidberg, and J. White. High-resolution record of northern hemisphere climate extending into the last interglacial period. *Nature*, 431:147–51, 10 2004.
- M. Anklin, J. M. Barnola, J. Beer, T. Blunier, J. Chappellaz, H. B. Clausen, D. Dahljensen, W. Dansgaard, M. Deangelis, R. Delmas, P. Duval, M. Fratta, A. Fuchs, K. Fuhrer, N. Gundestrup, C. Hammer, P. Iversen, S. Johnsen, J. Jouzel, and E. W. Wolff. Climate instability during the last interglacial period recorded in the grip ice core. *Nature*, 364: 203–207, 07 1993.
- L. Arnold and P. Imkeller. The Kramers Oscillator Revisited. In J. A. Freund and T. Pöschel, editors, *Stochastic Processes in Physics, Chemistry, and Biology*, pages 280–291. Springer Berlin Heidelberg, 2000.
- N. Boers. Early-warning signals for Dansgaard-Oeschger events in a high-resolution ice core record. *Nature Communications*, 9, 07 2018.
- N. Boers, M. D. Chekroun, H. Liu, D. Kondrashov, D.-D. Rousseau, A. Svensson, M. Bigler, and M. Ghil. Inverse stochastic–dynamic models for high-resolution Greenland ice core records. *Earth System Dynamics*, 8(4):1171–1190, 2017.
- N. Boers, M. Ghil, and D.-D. Rousseau. Ocean circulation, ice shelf, and sea ice interactions explain Dansgaard–Oeschger cycles. *Proceedings of the National Academy of Sciences*, 115:E11005–E11014, 11 2018.
- G. W. Bohrnstedt and A. S. Goldberger. On the Exact Covariance of Products of Random Variables. *Journal of the American Statistical Association*, 64(328):1439–1442, 1969.
- N. Bou-Rabee. Cayley splitting for second-order langevin stochastic partial differential equations. *arXiv: Probability*, 2017. URL <https://api.semanticscholar.org/CorpusID:119132768>.
- N. Bou-Rabee and H. Owhadi. Long-Run Accuracy of Variational Integrators in the Stochastic Context. *SIAM Journal on Numerical Analysis*, 48(1):278–297, Jan. 2010.
- C.-E. Bréhier and L. Goudenège. Analysis of some splitting schemes for the stochastic Allen-Cahn equation. *Discrete and Continuous Dynamical Systems - B*, 24(8):4169–4190, 2019.
- E. Buckwar, A. Samson, M. Tamborrino, and I. Tubikanec. A splitting method for SDEs with locally Lipschitz drift: Illustration on the FitzHugh-Nagumo model. *Applied Numerical Mathematics*, 179:191–220, 2022.
- K. Burrage, I. Lenane, and G. Lythe. Numerical Methods for Second-Order Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 29(1):245–264, 2007.
- I. Crimaldi and L. Pratelli. Convergence results for multivariate martingales. *Stochastic Processes and their Applications*, 115(4):571–577, 2005.
- W. Dansgaard, S. Johnsen, and H. e. a. Clausen. Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, 364:218–220, 1993.
- P. D. Ditlevsen, S. Ditlevsen, and K. K. Andersen. The fast climate fluctuations during the stadial and interstadial climate states. *Annals of Glaciology*, 35:457–462, 2002.
- P. D. Ditlevsen, K. K. Andersen, and A. Svensson. The DO-climate events are probably noise induced: statistical investigation of the claimed 1470 years cycle. *Climate of the Past*, 3(1):129–134, 2007.
- S. Ditlevsen and A. Samson. Hypoelliptic diffusions: filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 81(2):361–384, 2019.
- S. Ditlevsen and M. Sørensen. Inference for observations of integrated diffusion processes. *Scandinavian Journal of Statistics*, 31(3):417–429, 2004.
- G. Duffing. *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*. Vieweg, 1918.
- D. Falbel and J. Luraschi. *torch: Tensors and Neural Networks with 'GPU' Acceleration*, 2022.
- V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l’I.H.P. Probabilités et statistiques*, 29(1):119–151, 1993.
- A. Gloter. Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: Probability and Statistics*, 4:205–227, 2000.
- A. Gloter. Parameter Estimation for a Discretely Observed Integrated Diffusion Process. *Scandinavian Journal of Statistics*, 33(1):83–104, 2006.
- A. Gloter and N. Yoshida. Adaptive and non-adaptive estimation for degenerate diffusion processes, 2020.

- A. Gloter and N. Yoshida. Adaptive estimation for degenerate diffusion processes. *Electronic Journal of Statistics*, 15(1):1424 – 1472, 2021.
- E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I (2nd Revised. Ed.): Nonstiff Problems*. Springer-Verlag, Berlin, Heidelberg, 1993.
- F. Hassanibesheli, N. Boers, and J. Kurths. Reconstructing complex system dynamics from time series: a method comparison. *New journal of physics*, 2020.
- A. R. Humphries and A. M. Stuart. *Deterministic and random dynamical systems: theory and numerics*, pages 211–254. Springer Netherlands, Dordrecht, 2002.
- Y. Iguchi and A. Beskos. Parameter inference for hypo-elliptic diffusions under a weak design condition, 2023.
- Y. Iguchi, A. Beskos, and M. Graham. Parameter inference for degenerate diffusion processes, 2023a.
- Y. Iguchi, A. Beskos, and M. M. Graham. Parameter Estimation with Increased Precision for Elliptic and Hypo-elliptic Diffusions, 2023b.
- B. H. Jansen and V. G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological cybernetics*, 73(4):357–366, 1995.
- D. S. Johnson, J. M. London, M.-A. Lea, and J. W. Durban. Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215, 2008.
- M. Kessler. Estimation of an Ergodic Diffusion from Discrete Observations. *Scandinavian Journal of Statistics*, 24(2): 211–229, 1997.
- P. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 1992.
- H. Korsch and H. Jodl. *Chaos: A Program Collection for the PC*. Springer, 1999.
- H. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284 – 304, 1940.
- N. V. Krylov. A Simple Proof of the Existence of a Solution of Itô’s Equation with Monotone Coefficients. *Theory of Probability & Its Applications*, 35(3):583–587, 1991.
- B. Leimkuhler and C. Matthews. Molecular dynamics. *Interdisciplinary applied mathematics*, 36, 2015.
- J. Lohmann and P. Ditlevsen. A consistent statistical model selection for abrupt glacial climate changes. *Climate Dynamics*, 52, 06 2019.
- X. Mao. *Stochastic differential equations and applications*. Elsevier, 2007.
- S. Melchionna. Design of quasisymplectic propagators for Langevin dynamics. *The Journal of Chemical Physics*, 127(4):044108, 07 2007.
- A. Melnykova. Parametric inference for hypoelliptic ergodic diffusions with full observations, 2020.
- T. Michelot and P. G. Blackwell. State-switching continuous-time correlated random walks. *Methods in Ecology and Evolution*, 10(5):637–649, 2019.
- D. Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. Springer Berlin Heidelberg, 2006. ISBN 9783540283294.
- T. Ozaki. Statistical Identification of Storage Models with Application to Stochastic Hydrology. *Journal of The American Water Resources Association*, 21:663–675, 1985.
- T. Ozaki, J. C. Jimenez, and V. Haggan-Ozaki. The Role of the Likelihood Function in the Estimation of Chaos Models. *Journal of Time Series Analysis*, 21(4):363–387, 2000.
- G. Pavliotis, A. Stuart, and K. Zygalakis. Calculating effective diffusivities in the limit of vanishing molecular diffusion. *Journal of Computational Physics*, 228(4):1030–1055, 2009.
- P. Pilipovic, A. Samson, and S. Ditlevsen. Parameter estimation in nonlinear multivariate stochastic differential equations based on splitting schemes. *arXiv preprint arXiv:2211.11884*, 2024. To appear in The Annals of Statistics.
- Y. Pokern, A. M. Stuart, and P. Wiberg. Parameter Estimation for Partially Observed Hypoelliptic Diffusions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(1):49–73, 01 2009.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.

- S. O. Rasmussen, M. Bigler, S. P. Blockley, T. Blunier, S. L. Buchardt, H. B. Clausen, I. Cvijanovic, D. Dahl-Jensen, S. J. Johnsen, H. Fischer, V. Gkinis, M. Guillevic, W. Z. Hoek, J. J. Lowe, J. B. Pedro, T. Popp, I. K. Seierstad, J. P. Steffensen, A. M. Svensson, P. Vallelonga, B. M. Vinther, M. J. Walker, J. J. Wheatley, and M. Winstrup. A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. *Quaternary Science Reviews*, 106:14–28, 2014.
- M. Rosenblum and A. Pikovsky. Synchronization: from pendulum clocks to chaotic lasers and chemical oscillators. *Contemporary Physics*, 44(5):401–416, 2003.
- A. Samson and M. Thieullen. Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 122(7):2521–2552, 2012.
- I. Seierstad, P. Abbott, M. Bigler, T. Blunier, A. Bourne, E. Brook, S. L. Buchardt, C. Buizert, H. Clausen, E. Cook, D. Dahl-Jensen, S. Davies, M. Guillevic, S. Johnsen, D. Pedersen, T. Popp, S. Rasmussen, J. Severinghaus, A. Svensson, and B. Vinther. Consistently dated records from the Greenland GRIP, GISP2 and NGRIP ice cores for the past 104 ka reveal regional millennial-scale $\delta^{18}\text{O}$ gradients with possible Heinrich event imprint. *Quaternary Science Reviews*, 106, 11 2014.
- M. Serrano, G. De Fabritiis, P. Español, and P. Coveney. A stochastic Trotter integration scheme for dissipative particle dynamics. *Mathematics and Computers in Simulation*, 72(2):190–194, 2006.
- T. Shardlow. Splitting for dissipative particle dynamics. *SIAM Journal on Scientific Computing*, 24(4):1267–1282, Dec. 2003.
- I. Shoji and T. Ozaki. Estimation for nonlinear stochastic differential equations by a local linearization method. *Stochastic Analysis and Applications*, 16(4):733–752, 1998.
- M. Sørensen and M. Uchida. Small-diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli*, 9 (6):1051 – 1069, 2003. ISSN 1350-7265.
- M. V. Tretyakov and Z. Zhang. A Fundamental Mean-Square Convergence Theorem for SDEs with Locally Lipschitz Coefficients and Its Applications. *SIAM Journal on Numerical Analysis*, 51(6):3135–3162, 2013.
- E. Vanden-Eijnden and G. Ciccotti. Second-order integrators for Langevin equations with holonomic constraints. *Chemical Physics Letters - CHEM PHYS LETT*, 429, 09 2006.
- L. Wu. Large and moderate deviations and exponential convergence for stochastic damping hamiltonian systems. *Stochastic Processes and their Applications*, 91(2):205–238, 2001.
- N. Yoshida. Asymptotic behavior of M-estimator and related random field for diffusion process. *Annals of the Institute of Statistical Mathematics*, 42(2):221–251, June 1990.
- I. Ziv, D. Baxter, and J. Byrne. Simulator for neural networks and action-potentials - description and application. *Journal of Neurophysiology*, 71(1):294–308, 1994.

S1 Supplementary Material

In this section, we provide proofs for all the nontrivial lemmas, propositions, and theorems presented in Sections 3 and 6. The majority of these proofs, especially those in Section 6, heavily rely on Itô or Taylor expansions in h around $\mathbf{Y}_{t_{k-1}}$. Additionally, we frequently employ Fubini's theorem as a useful tool. Our initial focus is on the results from Section 6, as they constitute technical auxiliary properties essential for understanding the main results outlined in Section 3.

S1.1 Proofs of results from Section 6

Proof of Lemma 6.2 To prove (i), calculate

$$\begin{aligned}\Omega_h^{[\text{RR}]}(\boldsymbol{\theta})^{-1} &= \frac{1}{h}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}(\mathbf{I} + \frac{h}{2}(\mathbf{A}_v(\boldsymbol{\beta}) + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1})^{-1}) + \mathbf{R}(h, \mathbf{y}_0) \\ &= \frac{1}{h}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}(\mathbf{I} - \frac{h}{2}(\mathbf{A}_v(\boldsymbol{\beta}) + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}) + \mathbf{R}(h, \mathbf{y}_0) \\ &= \frac{1}{h}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} - \frac{1}{2}((\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \mathbf{A}_v(\boldsymbol{\beta}) + \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}) + \mathbf{R}(h, \mathbf{y}_0).\end{aligned}$$

Proof of (ii):

$$\begin{aligned}\Omega_h^{[\text{SR}]}(\boldsymbol{\theta})\Omega_h^{[\text{RR}]}(\boldsymbol{\theta})^{-1} &= (\frac{h^2}{2}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top) + \frac{h^3}{6}(\mathbf{A}_v(\boldsymbol{\beta})(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top) + 2\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top))\frac{1}{h}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \\ &\quad - \frac{h^2}{4}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)((\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \mathbf{A}_v(\boldsymbol{\beta}) + \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}) + \mathbf{R}(h^3, \mathbf{y}_0) \\ &= \frac{h}{2}\mathbf{I} - \frac{h^2}{12}(\mathbf{A}_v - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}) + \mathbf{R}(h^3, \mathbf{y}_0).\end{aligned}$$

To prove (iii), use the previous result to obtain:

$$\begin{aligned}\Omega_h^{[\text{SR}]}(\boldsymbol{\theta})\Omega_h^{[\text{RR}]}(\boldsymbol{\theta})^{-1}\Omega_h^{[\text{RS}]}(\boldsymbol{\theta}) \\ &= (\frac{h}{2}\mathbf{I} - \frac{h^2}{12}(\mathbf{A}_v - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}))(\frac{h^2}{2}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top) + \frac{h^3}{6}(2\mathbf{A}_v(\boldsymbol{\beta})(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top) + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top)) + \mathbf{R}(h^5, \mathbf{y}_0) \\ &= \frac{h^3}{4}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + \frac{h^4}{8}(\mathbf{A}_v(\boldsymbol{\beta})\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top) + \mathbf{R}(h^5, \mathbf{y}_0).\end{aligned}$$

Proof of (iv) follows from (iii) and Lemma 6.1. To prove (v), approximate the log-determinant as:

$$\begin{aligned}\log \det \Omega_h^{[\text{RR}]}(\boldsymbol{\theta}) &= \log \det \left(h\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + \frac{h^2}{2}(\mathbf{A}_v(\boldsymbol{\beta})\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top) \right) + R(h^2, \mathbf{y}_0) \\ &= d \log h + \log \det \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + \log \det \left(\mathbf{I} + \frac{h}{2}(\mathbf{A}_v(\boldsymbol{\beta}) + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}) \right) + R(h^2, \mathbf{y}_0) \\ &= d \log h + \log \det \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + \frac{h}{2} \text{Tr}(\mathbf{A}_v(\boldsymbol{\beta}) + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \mathbf{A}_v(\boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1}) + R(h^2, \mathbf{y}_0) \\ &= d \log h + \log \det \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top + h \text{Tr} \mathbf{A}_v(\boldsymbol{\beta}) + R(h^2, \mathbf{y}_0).\end{aligned}\tag{S1}$$

To prove (vi), repeat the previous reasoning on (iv). The proof of (vii) follows from $\det \tilde{\boldsymbol{\Omega}}_h = \det \Omega_h^{[\text{RR}]} \det \Omega_h^{[\text{S|R}]}$ and properties (v) and (vi).

Proof of Lemma 6.4 Using the same approximation as in the previous proof of (v), we obtain:

$$\begin{aligned}2 \log |\det D\mathbf{f}_{h/2}(\mathbf{y}; \boldsymbol{\beta})| &= 2 \log |\det(\mathbf{I} + \frac{h}{2}D\mathbf{N}(\mathbf{y}; \boldsymbol{\beta}))| + R(h^2, \mathbf{y}) \\ &= 2 \log |1 + \frac{h}{2} \text{Tr} D\mathbf{N}(\mathbf{y}; \boldsymbol{\beta})| + R(h^2, \mathbf{y}) \\ &= h \text{Tr} D\mathbf{N}(\mathbf{y}; \boldsymbol{\beta}) + R(h^2, \mathbf{y}) = h \text{Tr} D_v \mathbf{N}(\mathbf{y}; \boldsymbol{\beta}) + R(h^2, \mathbf{y}).\end{aligned}\tag{S2}$$

In complete observation, put \mathbf{Y}_{t_k} instead of \mathbf{y} and use Itô's lemma on $\mathbf{N}(\mathbf{Y}_{t_k})$ as in (S6). In partial observation, put $(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}})$ instead of \mathbf{y} and approximate $\mathbf{N}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}})$ as in (S7).

Proof of Lemma 6.5 We use definition (6) and approximation (55), and plug them in (17) to obtain:

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_h(\tilde{\mathbf{f}}_{h/2}(\mathbf{y})) &= e^{\tilde{\mathbf{A}}h}(\tilde{\mathbf{f}}_{h/2}(\mathbf{y}) - \tilde{\mathbf{b}}) + \tilde{\mathbf{b}} \\
&= \left(\mathbf{I}_{2d} + h\tilde{\mathbf{A}} + \frac{h^2}{2}\tilde{\mathbf{A}}^2 + \mathbf{R}(h^3, \mathbf{y}) \right) (\tilde{\mathbf{f}}_{h/2}(\mathbf{y}) - \tilde{\mathbf{b}}) + \tilde{\mathbf{b}} \\
&= \begin{bmatrix} \mathbf{I}_d + \frac{h^2}{2}\mathbf{A}_x + \mathbf{R}(h^3, \mathbf{y}) & h\mathbf{I}_d + \frac{h^2}{2}\mathbf{A}_v + \mathbf{R}(h^3, \mathbf{y}) \\ h\mathbf{A}_x + \mathbf{R}(h^2, \mathbf{y}) & \mathbf{I}_d + h\mathbf{A}_v + \mathbf{R}(h^2, \mathbf{y}) \end{bmatrix} \begin{bmatrix} \mathbf{x} - \mathbf{b} \\ \mathbf{v} + \frac{h}{2}\mathbf{N}(\mathbf{y}) + \mathbf{R}(h^2, \mathbf{y}) \end{bmatrix} + \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{x} + h\mathbf{v} + \frac{h^2}{2}\mathbf{F}(\mathbf{y}) + \mathbf{R}(h^3, \mathbf{y}) \\ \mathbf{v} + h(\mathbf{F}(\mathbf{y}) - \frac{1}{2}\mathbf{N}(\mathbf{y})) + \mathbf{R}(h^2, \mathbf{y}) \end{bmatrix}.
\end{aligned}$$

This concludes the proof.

To prove Proposition 6.6, we need the following lemma that provides expansion of $\Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k}$.

Lemma S1.1 For process $\Delta_h \mathbf{X}_{t_{k+1}}$ (33) it holds:

$$\Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k} = \sqrt{h}\boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \frac{h^{3/2}}{2}D_v \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\boldsymbol{\Sigma}_0 \mathbf{Q}_{k,k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}), \quad (\text{S3})$$

$$\Delta_h \mathbf{X}_{t_k} - \mathbf{V}_{t_{k-1}} = \sqrt{h}\boldsymbol{\Sigma}_0 \boldsymbol{\zeta}'_{k-1} + \frac{h}{2}\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \frac{h^{3/2}}{2}D_v \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})\boldsymbol{\Sigma}_0 \boldsymbol{\zeta}'_{k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}). \quad (\text{S4})$$

Proof of Lemma S1.1 Proof of (S3). Equation (2) in integral form and (33) yield:

$$\begin{aligned}
\Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k} &= \frac{1}{h} \int_{t_k}^{t_{k+1}} \mathbf{V}_t dt - \frac{1}{h} \int_{t_{k-1}}^{t_k} \mathbf{V}_t dt = \frac{1}{h} \int_{t_k}^{t_{k+1}} (\mathbf{V}_t - \mathbf{V}_{t_k}) dt + \frac{1}{h} \int_{t_{k-1}}^{t_k} (\mathbf{V}_{t_k} - \mathbf{V}_t) dt \\
&= \frac{1}{h} \int_{t_k}^{t_{k+1}} \int_{t_k}^t \mathbf{F}_0(\mathbf{Y}_s) ds dt + \frac{1}{h} \boldsymbol{\Sigma}_0 \int_{t_k}^{t_{k+1}} \int_{t_k}^t d\mathbf{W}_s dt \\
&\quad + \frac{1}{h} \int_{t_{k-1}}^{t_k} \int_t^{t_k} \mathbf{F}_0(\mathbf{Y}_s) ds dt + \frac{1}{h} \boldsymbol{\Sigma}_0 \int_{t_{k-1}}^{t_k} \int_t^{t_k} d\mathbf{W}_s dt.
\end{aligned}$$

Apply Fubini's theorem on double integrals to obtain:

$$\Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k} = \frac{1}{h} \int_{t_k}^{t_{k+1}} (t_{k+1} - t) \mathbf{F}_0(\mathbf{Y}_t) dt + \frac{1}{h} \int_{t_{k-1}}^{t_k} (t - t_{k-1}) \mathbf{F}_0(\mathbf{Y}_t) dt + \sqrt{h}\boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1}. \quad (\text{S5})$$

Applying Itô's lemma on $\mathbf{F}_0(\mathbf{Y}_t)$ yields the following approximation:

$$\mathbf{F}_0(\mathbf{Y}_t) = \mathbf{F}_0(\mathbf{Y}_{t_k}) + D_v \mathbf{F}_0(\mathbf{Y}_{t_k}) \boldsymbol{\Sigma}_0 \int_{t_k}^t d\mathbf{W}_s + \mathbf{R}(h, \mathbf{Y}_{t_k}). \quad (\text{S6})$$

Plugging (S6) into (S5) gives:

$$\begin{aligned}
\Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k} &= h\mathbf{F}_0(\mathbf{Y}_{t_k}) + \frac{1}{h} D_v \mathbf{F}_0(\mathbf{Y}_{t_k}) \boldsymbol{\Sigma}_0 \int_{t_k}^{t_{k+1}} \int_{t_k}^t (t_{k+1} - t) d\mathbf{W}_s dt \\
&\quad - \frac{1}{h} D_v \mathbf{F}_0(\mathbf{Y}_{t_k}) \boldsymbol{\Sigma}_0 \int_{t_{k-1}}^{t_k} \int_t^{t_k} (t - t_{k-1}) d\mathbf{W}_s dt + \sqrt{h}\boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).
\end{aligned}$$

Once again, apply Fubini's theorem on the double integrals to get:

$$\int_{t_k}^{t_{k+1}} \int_{t_k}^t (t_{k+1} - t) d\mathbf{W}_s dt = \frac{1}{2} h^{5/2} \boldsymbol{\zeta}'_k, \quad \int_{t_{k-1}}^{t_k} \int_t^{t_k} (t - t_{k-1}) d\mathbf{W}_s dt = \frac{1}{2} h^{5/2} \boldsymbol{\zeta}'_{k-1}.$$

So far, we have

$$\Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k} = \sqrt{h}\boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h\mathbf{F}_0(\mathbf{Y}_{t_k}) + \frac{h^{3/2}}{2} D_v \mathbf{F}_0(\mathbf{Y}_{t_k}) \boldsymbol{\Sigma}_0 (\boldsymbol{\zeta}'_k - \boldsymbol{\zeta}'_{k-1}) + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).$$

To conclude the proof, use Itô's lemma to get $\mathbf{F}_0(\mathbf{Y}_{t_k}) = \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \sqrt{h} D_v \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} + \mathbf{R}(h, \mathbf{Y}_{t_{k-1}})$.

Proof of (S4). As before, start with (33) and use (S6) to get:

$$\begin{aligned}\Delta_h \mathbf{X}_{t_k} - \mathbf{V}_{t_{k-1}} &= \frac{1}{h} \int_{t_{k-1}}^{t_k} (t_k - t) \mathbf{F}_0(\mathbf{Y}_t) dt + \frac{1}{h} \Sigma_0 \int_{t_{k-1}}^{t_k} (t_k - t) d\mathbf{W}_t \\ &= \sqrt{h} \Sigma_0 \xi'_{k-1} + \frac{h}{2} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \frac{1}{h} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_k}) \Sigma_0 \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^t (t_k - t) d\mathbf{W}_s dt + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

This concludes the proof.

Proof of Proposition 6.6 The expansion of $\mathbf{Z}_{k,k-1}^{[S]}$ follows directly from Lemma 6.5 and S1.1. Indeed, it holds

$$\begin{aligned}\mathbf{Z}_{k,k-1}^{[S]}(\beta) &= \mathbf{X}_{t_k} - \mathbf{X}_{t_{k-1}} - h \mathbf{V}_{t_{k-1}} - \frac{h^2}{2} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}) \\ &= h(\Delta_h \mathbf{X}_{t_k} - \mathbf{V}_{t_{k-1}}) - \frac{h^2}{2} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

To expand $\mathbf{Z}_{k,k-1}^{[R]}$, we use definition (27) and approximations (58) and (60), as follows:

$$\begin{aligned}\mathbf{Z}_{k,k-1}^{[R]}(\beta) &= \mathbf{V}_{t_k} - \mathbf{V}_{t_{k-1}} - h \mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \frac{h}{2} (\mathbf{N}(\mathbf{Y}_{t_k}) - \mathbf{N}(\mathbf{Y}_{t_{k-1}})) + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}) \\ &= \Sigma_0 \int_{t_{k-1}}^{t_k} d\mathbf{W}_t + \int_{t_{k-1}}^{t_k} \mathbf{F}_0(\mathbf{Y}_t) dt - h \mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \frac{h}{2} (\mathbf{N}(\mathbf{Y}_{t_k}) - \mathbf{N}(\mathbf{Y}_{t_{k-1}})) + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}) \\ &= \sqrt{h} \Sigma_0 \eta_{k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h}{2} (\mathbf{N}(\mathbf{Y}_{t_k}) - \mathbf{N}(\mathbf{Y}_{t_{k-1}})) \\ &\quad + D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \Sigma_0 \int_{t_{k-1}}^{t_k} \int_{t_{k-1}}^t d\mathbf{W}_s dt + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

In the last line, we used Itô's lemma on $\mathbf{F}_0(\mathbf{Y}_t)$ as in (S6). Again, apply Itô's lemma on $\mathbf{N}(\mathbf{Y}_{t_k})$ to get:

$$\begin{aligned}\mathbf{Z}_{k,k-1}^{[R]}(\beta) &= \sqrt{h} \Sigma_0 \eta_{k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \Sigma_0 \eta_{k-1} \\ &\quad + h^{3/2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \Sigma_0 \xi'_{k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

The expansion of $\mathbf{Z}_{k,k-1}^{[S]}$ follows from definition (36) and plugging $(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})$ in approximation (60):

$$\begin{aligned}\bar{\mathbf{Z}}_{k+1,k,k-1}^{[S]}(\beta) &= \mathbf{X}_{t_k} - \mathbf{X}_{t_{k-1}} - h \Delta_h \mathbf{X}_{t_k} - \frac{h^2}{2} \mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}) \\ &= -\frac{h^2}{2} \mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

Use Taylor's formula on $\mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})$ to get

$$\mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}) = \mathbf{F}(\mathbf{Y}_{t_{k-1}}) + D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) (\Delta_h \mathbf{X}_{t_k} - \mathbf{V}_{t_{k-1}}) + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}). \quad (\text{S7})$$

Now, the rest follows from Lemma S1.1.

Finally, to expand $\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}$, start with definition (37) and approximations (58), and (60):

$$\begin{aligned}\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\beta) &= \Delta_h \mathbf{X}_{t_{k+1}} - \Delta_h \mathbf{X}_{t_k} - h \mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k}) \\ &\quad - \frac{h}{2} (\mathbf{N}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}) - \mathbf{N}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})) + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

Lemma S1.1 yields:

$$\begin{aligned}\bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\beta) &= \sqrt{h} \Sigma_0 \mathbf{U}_{k,k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})) \\ &\quad - \frac{h}{2} (\mathbf{N}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}}) - \mathbf{N}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})) + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \Sigma_0 \mathbf{Q}_{k,k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}).\end{aligned}$$

Apply Taylor's formula on $\mathbf{F}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})$, $\mathbf{N}(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_{k+1}})$, and $\mathbf{N}(\mathbf{X}_{t_{k-1}}, \Delta_h \mathbf{X}_{t_k})$, to get:

$$\begin{aligned} \bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\boldsymbol{\beta}) &= \sqrt{h} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h}{2}(\mathbf{N}(\mathbf{Y}_{t_k}) - \mathbf{N}(\mathbf{Y}_{t_{k-1}})) \\ &\quad + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{Q}_{k,k-1} - h^{3/2} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_k}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_k \\ &\quad + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

Finally, applying Itô's lemma on $\mathbf{N}(\mathbf{Y}_{t_k})$ yields

$$\begin{aligned} \bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\boldsymbol{\beta}) &= \sqrt{h} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \\ &\quad + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{Q}_{k,k-1} - h^{3/2} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_k \\ &\quad + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}) \\ &= \sqrt{h} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \\ &\quad + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{Q}_{k,k-1} - h^{3/2} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} + \mathbf{R}(h^2, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

This concludes the proof.

Proof of Lemma 6.7 Combining Proposition 6.6 and property (ii) of Lemma 6.2 yields:

$$\begin{aligned} \mathbf{Z}_{k,k-1}^{[S|R]}(\boldsymbol{\beta}) &= \mathbf{Z}_{k,k-1}^{[S]}(\boldsymbol{\beta}) - \boldsymbol{\Omega}_h^{[SR]}(\boldsymbol{\theta}) \boldsymbol{\Omega}_h^{[RR]}(\boldsymbol{\theta})^{-1} \mathbf{Z}_{k,k-1}^{[R]}(\boldsymbol{\beta}) \\ &= h^{3/2} \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} + \frac{h^2}{2} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) + \frac{h^{5/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} \\ &\quad - \left(\frac{h}{2} \mathbf{I} - \frac{h^2}{12} (\mathbf{A}_{\mathbf{v}} - \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{\top} \mathbf{A}_{\mathbf{v}}(\boldsymbol{\beta})^{\top} (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{\top})^{-1}) \right) \left(h^{1/2} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) \right. \\ &\quad \left. - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} + h^{3/2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} \right) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}) \\ &= h^{3/2} \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} + \frac{h^2}{2} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) + \frac{h^{5/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} \\ &\quad - h^{3/2} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} - \frac{h^2}{2} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) + \frac{h^{5/2}}{4} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \\ &\quad - \frac{h^{5/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} + \frac{h^{5/2}}{12} (\mathbf{A}_{\mathbf{v}} - \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{\top} \mathbf{A}_{\mathbf{v}}(\boldsymbol{\beta})^{\top} (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{\top})^{-1}) \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}). \end{aligned}$$

Additionally,

$$\begin{aligned} \bar{\mathbf{Z}}_{k+1,k,k-1}^{[S|R]}(\boldsymbol{\beta}) &= \bar{\mathbf{Z}}_{k,k-1}^{[S]}(\boldsymbol{\beta}) - \boldsymbol{\Omega}_h^{[SR]}(\boldsymbol{\theta}) \boldsymbol{\Omega}_h^{[RR]}(\boldsymbol{\theta})^{-1} \bar{\mathbf{Z}}_{k+1,k,k-1}^{[R]}(\boldsymbol{\beta}) \\ &= -\frac{h^2}{2} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \frac{h^{5/2}}{2} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} - \left(\frac{h}{2} \mathbf{I} - \frac{h^2}{12} (\mathbf{A}_{\mathbf{v}} - \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{\top} \mathbf{A}_{\mathbf{v}}(\boldsymbol{\beta})^{\top} (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^{\top})^{-1}) \right) \\ &\quad \cdot \left(h^{1/2} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} + h(\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) - \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \right. \\ &\quad \left. + \frac{h^{3/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{Q}_{k,k-1} - h^{3/2} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} \right) + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}) \\ &= -\frac{h^2}{2} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \frac{h^{5/2}}{2} D_{\mathbf{v}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} - h^{3/2} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} - \frac{h^2}{2} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) \\ &\quad + \frac{h^{5/2}}{4} D_{\mathbf{v}} \mathbf{N}(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} - \frac{h^{5/2}}{4} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \mathbf{Q}_{k,k-1} + \frac{h^{5/2}}{2} D_{\mathbf{v}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \boldsymbol{\Sigma}_0 \boldsymbol{\xi}'_{k-1} \end{aligned}$$

$$+ \frac{h^{5/2}}{12} (\mathbf{A}_v - \Sigma \Sigma^\top \mathbf{A}_v (\beta)^\top (\Sigma \Sigma^\top)^{-1}) \Sigma_0 \mathbf{U}_{k,k-1} + \mathbf{R}(h^3, \mathbf{Y}_{t_{k-1}}).$$

S1.2 Proofs from Section 3

Before we start the proofs, we state the following ergodic property, which is proved in Kessler (1997) in case of complete observations, and in Samson and Thieullen (2012) for both complete and partial observation.

Lemma S1.2 (Proposition 4 in Samson and Thieullen (2012)) *Let Assumptions (A1), (A2) and (A3) hold, and let \mathbf{Y} be the solution to (3). Let $g : \mathbb{R}^{2d} \times \Theta \rightarrow \mathbb{R}$ be a differentiable function with respect to \mathbf{y} and θ with derivatives of polynomial growth in \mathbf{y} , uniformly in θ . If $h \rightarrow 0$ and $Nh \rightarrow \infty$, then,*

$$\frac{1}{N-1} \sum_{k=1}^N g(\mathbf{Y}_{t_k}; \theta) \xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}} \int g(\mathbf{y}; \theta) d\nu_0(\mathbf{y}), \quad (\text{S8})$$

$$\frac{1}{N-2} \sum_{k=1}^{N-1} g(\mathbf{X}_{t_k}, \Delta_h \mathbf{X}_{t_k}; \theta) \xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}} \int g(\mathbf{y}; \theta) d\nu_0(\mathbf{y}), \quad (\text{S9})$$

uniformly in θ .

S1.3 Proof of consistency

Proof of Theorem 3.1 The proof of the consistency of the estimators follows a similar path as in Theorem 5.1 of (Pilipovic et al., 2024). With $\sigma := \text{vech}(\Sigma \Sigma^\top) = ([\Sigma \Sigma^\top]_{11}, [\Sigma \Sigma^\top]_{12}, [\Sigma \Sigma^\top]_{22}, \dots, [\Sigma \Sigma^\top]_{1d}, \dots, [\Sigma \Sigma^\top]_{dd})$, we half-vectorize $\Sigma \Sigma^\top$ to avoid working with tensors when computing derivatives with respect to $\Sigma \Sigma^\top$. Since $\Sigma \Sigma^\top$ is a symmetric $d \times d$ matrix, σ is of dimension $s = d(d+1)/2$. For a diagonal matrix, instead of a half-vectorization, we use $\sigma := \text{diag}(\Sigma \Sigma^\top)$ and $s = d$ in that case.

We start by finding the limits in \mathbb{P}_{θ_0} of

$$\frac{1}{N-1} \mathcal{L}_N^{[\text{C}\cdot]}(\beta, \sigma) \quad \text{and} \quad \frac{1}{N-2} \mathcal{L}_N^{[\text{P}\cdot]}(\beta, \sigma), \quad (\text{S10})$$

for $Nh \rightarrow \infty$, $h \rightarrow 0$, uniformly in θ . We apply Lemma 9 in Genon-Catalot and Jacod (1993) to prove the convergence and use Proposition A1 in Gloter (2006) to prove the uniform convergence. For more detailed derivations, see proofs in Pilipovic et al. (2024). Taking the expectations of (69)-(74), we conclude that:

$$\begin{aligned} \frac{1}{N-1} \mathcal{L}_N^{[\text{CR}]}(\beta, \sigma) &\rightarrow \log \det(\Sigma \Sigma^\top) + \text{Tr}((\Sigma \Sigma^\top)^{-1} \Sigma \Sigma_0^\top), \\ \frac{1}{N-1} \mathcal{L}_N^{[\text{CS|R}]}(\beta, \sigma) &\rightarrow \log \det(\Sigma \Sigma^\top) + \text{Tr}((\Sigma \Sigma^\top)^{-1} \Sigma \Sigma_0^\top), \\ \frac{1}{N-1} \mathcal{L}_N^{[\text{CF}]}(\beta, \sigma) &\rightarrow 2 \log \det(\Sigma \Sigma^\top) + 2 \text{Tr}((\Sigma \Sigma^\top)^{-1} \Sigma \Sigma_0^\top), \\ \frac{1}{N-2} \mathcal{L}_N^{[\text{PR}]}(\beta, \sigma) &\rightarrow \frac{2}{3} \log \det(\Sigma \Sigma^\top) + \frac{2}{3} \text{Tr}((\Sigma \Sigma^\top)^{-1} \Sigma \Sigma_0^\top), \\ \frac{1}{N-2} \mathcal{L}_N^{[\text{PS|R}]}(\beta, \sigma) &\rightarrow 2 \log \det(\Sigma \Sigma^\top) + 2 \text{Tr}((\Sigma \Sigma^\top)^{-1} \Sigma \Sigma_0^\top), \\ \frac{1}{N-2} \mathcal{L}_N^{[\text{PF}]}(\beta, \sigma) &\rightarrow \frac{8}{3} \log \det(\Sigma \Sigma^\top) + \frac{8}{3} \text{Tr}((\Sigma \Sigma^\top)^{-1} \Sigma \Sigma_0^\top), \end{aligned}$$

in \mathbb{P}_{θ_0} , for $Nh \rightarrow \infty$, $h \rightarrow 0$, uniformly in θ . From here, the rest of the proof of consistency for $\widehat{\Sigma}_N^{\top[\text{C}\cdot]}$ and $\widehat{\Sigma}_N^{\top[\text{P}\cdot]}$ is the same as in (Pilipovic et al., 2024). The coefficients in front of $\log \det$ terms in the partial observation setup correspond to the correcting factors in definitions of objective functions (39)-(41). They are needed to match coefficients in front of Tr terms that come from the forward difference's under- or over-estimation of the noise effects.

To prove the consistency of the drift estimators $\widehat{\beta}_N^{[\text{CR}]}$ and $\widehat{\beta}_N^{[\text{PR}]}$, we start by finding the limits in \mathbb{P}_{θ_0} of

$$\frac{1}{(N-1)h} (\mathcal{L}_N^{[\text{CR}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{CR}]}(\beta_0, \sigma)) \quad \text{and} \quad \frac{1}{(N-2)h} (\mathcal{L}_N^{[\text{PR}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{PR}]}(\beta_0, \sigma)), \quad (\text{S11})$$

for $Nh \rightarrow \infty$, $h \rightarrow 0$, uniformly in θ . Starting with expressions (69) and (72) we get

$$\begin{aligned}
\frac{1}{(N-1)h} (\mathcal{L}_N^{[\text{CR}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{CR}]}(\beta_0, \sigma)) &= \frac{2}{(N-1)\sqrt{h}} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) \\
&+ \frac{1}{N-1} \sum_{k=1}^N (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) \\
&- \frac{1}{N-1} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}}(\mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \\
&+ \frac{1}{N-1} \sum_{k=1}^N \text{Tr} D_{\mathbf{v}}(\mathbf{F}(\mathbf{Y}_{t_k}) - \mathbf{F}_0(\mathbf{Y}_{t_k})), \\
\frac{1}{(N-2)h} (\mathcal{L}_N^{[\text{PR}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{PR}]}(\beta_0, \sigma)) &= \frac{2}{(N-2)\sqrt{h}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) \\
&+ \frac{1}{N-2} \sum_{k=1}^{N-1} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}(\mathbf{Y}_{t_{k-1}})) \\
&- \frac{1}{N-2} \sum_{k=1}^{N-1} (\mathbf{U}_{k,k-1} + 2\xi'_{k-1})^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}}(\mathbf{F}(\mathbf{Y}_{t_{k-1}}) - \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \\
&+ \frac{1}{N-2} \sum_{k=1}^{N-1} \text{Tr} D_{\mathbf{v}}(\mathbf{F}(\mathbf{Y}_{t_k}) - \mathbf{F}_0(\mathbf{Y}_{t_k})).
\end{aligned}$$

To prove the convergence in probability of the previous two sequences, we use Lemma S1.2 and Lemma 9 in Genon-Catalot and Jacod (1993). To apply Lemma 9 from Genon-Catalot and Jacod (1993), we need to show that the sum of expectations converges to a certain value, while the sum of covariances converges to zero. Here, we only show the former. Moreover, standard tools like Proposition A1 in Gloter (2006) or Lemma 3.1 in Yoshida (1990) can be used to prove uniform convergence. Thus, we just look at the expectation to find the limits of these sequences. We use the known covariances (66) and (68) to get:

$$\begin{aligned}
\frac{1}{Nh} (\mathcal{L}_N^{[\cdot\text{R}]}(\beta, \sigma) - \mathcal{L}_N^{[\cdot\text{R}]}(\beta_0, \sigma)) &\xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}} \int (\mathbf{F}_0(\mathbf{y}) - \mathbf{F}(\mathbf{y}))^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}_0(\mathbf{y}) - \mathbf{F}(\mathbf{y})) d\nu_0(\mathbf{y}) \\
&+ \int \text{Tr}(D_{\mathbf{v}}(\mathbf{F}_0(\mathbf{y}) - \mathbf{F}(\mathbf{y})) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} - \mathbf{I})) d\nu_0(\mathbf{y}). \quad (\text{S12})
\end{aligned}$$

Thus, the consistency of the drift estimator in the partial case coincides with the complete case when using rough objective functions. This is because the right-hand side of (S12) is non-negative when $\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top = \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top$, and the left-hand side is non-positive, following the definition of the likelihood. The remainder of the proof is analogous to that in Pilipovic et al. (2024), and is therefore not repeated here.

Here, we also illustrate why the objective functions based on the conditional likelihood of smooth given rough coordinates do not provide identifiable drift estimators. Starting with the complete observations objective function (70) and using that $\mathbb{E}_{\theta_0}[(\boldsymbol{\eta}_{k-1} - 2\xi'_{k-1}) \boldsymbol{\eta}_{k-1}^\top | \mathcal{F}_{t_{k-1}}] = \mathbf{0}$, we conclude::

$$\frac{1}{Nh} (\mathcal{L}_N^{[\text{CS|R}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{CS|R}]}(\beta_0, \sigma)) \xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}} 0. \quad (\text{S13})$$

In the partial observation case, we need to add the term $4 \log |\det D_{\mathbf{v}} \mathbf{f}_{h/2}|$ in (31). Due to this correction, we obtain the consistency of the drift estimator from the following derivations and the fact that the diffusion estimator converges faster:

$$\begin{aligned}
\frac{1}{(N-2)h} (\mathcal{L}_N^{[\text{PS|R}]}(\beta, \sigma) - \mathcal{L}_N^{[\text{PS|R}]}(\beta_0, \sigma)) &= \frac{2}{N-2} \sum_{k=1}^{N-1} \text{Tr} D_{\mathbf{v}}(\mathbf{N}(\mathbf{Y}_{t_k}) - \mathbf{N}_0(\mathbf{Y}_{t_k})) \\
&+ \frac{3}{N-2} \sum_{k=1}^{N-1} \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}}(\mathbf{N}_0(\mathbf{Y}_{t_{k-1}}) - \mathbf{N}(\mathbf{Y}_{t_{k-1}}))^\top)
\end{aligned}$$

$$\xrightarrow[Nh \rightarrow \infty]{\mathbb{P}_{\theta_0}, h \rightarrow 0} 2 \int \text{Tr}(D_{\mathbf{v}}(\mathbf{N}_0(\mathbf{y}) - \mathbf{N}(\mathbf{y}))(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^\top(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} - \mathbf{I})) d\nu_0(\mathbf{y}).$$

Finally, the consistency of the estimators based on the full objective functions follows from the previous proofs, (71), and (74). That concludes the proof of consistency.

Proof of Lemma 3.3 We start by proving the first part of the lemma, for both complete and partial cases using the rough objective functions (69) and (72). First, we find their second derivatives with respect to $\boldsymbol{\beta}$:

$$\begin{aligned} \frac{1}{(N-1)h} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathcal{L}_N^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) &= \frac{2}{(N-1)\sqrt{h}} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\ &+ \frac{1}{N-1} \sum_{k=1}^N \text{Tr} D_{\mathbf{v}} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}) + \frac{2}{N-1} \sum_{k=1}^N \partial_{\beta^{(i_1)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_2)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\ &- \frac{2}{N-1} \sum_{k=1}^N \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\ &- \frac{1}{N-1} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1}, \\ \frac{1}{(N-2)h} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathcal{L}_N^{[\text{PR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) &= \frac{2}{(N-2)\sqrt{h}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\ &+ \frac{1}{N-2} \sum_{k=1}^{N-1} \text{Tr} D_{\mathbf{v}} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}) + \frac{2}{N-2} \sum_{k=1}^{N-1} \partial_{\beta^{(i_1)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_2)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\ &- \frac{2}{N-2} \sum_{k=1}^{N-1} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\ &- \frac{1}{N-2} \sum_{k=1}^{N-1} (\mathbf{U}_{k,k-1} + 2\xi'_{k-1})^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1}. \end{aligned}$$

As in the proof of consistency, it holds:

$$\frac{1}{Nh} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathcal{L}_N^{[\text{R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{\mathbb{P}_{\theta_0}} 2 \int \partial_{\beta^{(i_1)}} \mathbf{F}_0(\mathbf{y})^\top (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_2)}} \mathbf{F}_0(\mathbf{y}) d\nu_0(\mathbf{y}).$$

Now, we investigate the limit of $\frac{1}{(N-1)\sqrt{h}} \partial_{\beta^{(i_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{CR}]}(\boldsymbol{\theta})$ and $\partial_{\beta^{(i_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{PR}]}(\boldsymbol{\theta})$:

$$\begin{aligned} \frac{1}{(N-1)\sqrt{h}} \partial_{\beta^{(i_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) &= -\frac{2}{N-1} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top \partial_{\sigma^{(j_2)}} (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_1)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\ &+ \frac{1}{N-1} \sum_{k=1}^N R(\sqrt{h}, \mathbf{Y}_{t_{k-1}}) \\ \frac{1}{(N-2)\sqrt{h}} \partial_{\beta^{(i_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{PR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) &= -\frac{2}{N-2} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top \partial_{\sigma^{(j_2)}} (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i_1)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\ &+ \frac{1}{N-2} \sum_{k=1}^{N-1} R(\sqrt{h}, \mathbf{Y}_{t_{k-1}}) \end{aligned}$$

Both previous sequences converge to zero due to Lemma 9 in Genon-Catalot and Jacod (1993). Next, we look at the limits of $\frac{1}{N-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{CR}]}(\boldsymbol{\theta})$ and $\frac{1}{N-2} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{PR}]}(\boldsymbol{\theta})$:

$$\frac{1}{N-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \log \det(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top)$$

$$\begin{aligned}
& + \frac{1}{N-1} \sum_{k=1}^N \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \text{Tr}(\Sigma_0 \boldsymbol{\eta}_{k-1} \boldsymbol{\eta}_{k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1}) + \frac{1}{N-1} \sum_{k=1}^N R(\sqrt{h}, \mathbf{Y}_{t_{k-1}}) \\
& = \text{Tr}((\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \Sigma \Sigma^\top) - \text{Tr}((\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top) \\
& \quad - \frac{1}{N-1} \sum_{k=1}^N \text{Tr}(\Sigma_0 \boldsymbol{\eta}_{k-1} \boldsymbol{\eta}_{k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1}) \\
& \quad + \frac{1}{N-1} \sum_{k=1}^N \text{Tr}(\Sigma_0 \boldsymbol{\eta}_{k-1} \boldsymbol{\eta}_{k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1}) \\
& \quad + \frac{1}{N-1} \sum_{k=1}^N \text{Tr}(\Sigma_0 \boldsymbol{\eta}_{k-1} \boldsymbol{\eta}_{k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1}) \\
& \quad + \frac{1}{N-1} \sum_{k=1}^N R(\sqrt{h}, \mathbf{Y}_{t_{k-1}}), \\
\frac{1}{N-2} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{\text{[PR]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) & = \frac{2}{3} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \log \det(\Sigma \Sigma^\top) \\
& \quad + \frac{1}{N-2} \sum_{k=1}^{N-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \text{Tr}(\Sigma_0 \mathbf{U}_{k,k-1} \mathbf{U}_{k,k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1}) + \frac{1}{N-2} \sum_{k=1}^{N-1} R(\sqrt{h}, \mathbf{Y}_{t_{k-1}}) \\
& = \frac{2}{3} \text{Tr}((\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \Sigma \Sigma^\top) - \frac{2}{3} \text{Tr}((\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top) \\
& \quad - \frac{1}{N-2} \sum_{k=1}^{N-1} \text{Tr}(\Sigma_0 \mathbf{U}_{k,k-1} \mathbf{U}_{k,k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1}) \\
& \quad + \frac{1}{N-2} \sum_{k=1}^{N-1} \text{Tr}(\Sigma_0 \mathbf{U}_{k,k-1} \mathbf{U}_{k,k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1}) \\
& \quad + \frac{1}{N-2} \sum_{k=1}^{N-1} \text{Tr}(\Sigma_0 \mathbf{U}_{k,k-1} \mathbf{U}_{k,k-1}^\top \Sigma_0^\top (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1}) \\
& \quad + \frac{1}{N-2} \sum_{k=1}^{N-1} R(\sqrt{h}, \mathbf{Y}_{t_{k-1}}).
\end{aligned}$$

Using the second moments of $\boldsymbol{\eta}_{k-1}$ and $\mathbf{U}_{k,k-1}$ with additional calculations, we can conclude that:

$$\begin{aligned}
\frac{1}{N-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{\text{[CR]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} & \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \text{Tr}((\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top), \\
\frac{1}{N-2} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{\text{[PR]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} & \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \frac{2}{3} \text{Tr}((\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top).
\end{aligned}$$

To extend the previous results on the objective functions (70) and (73), we start by acknowledging that

$$\frac{1}{Nh} \partial_{\beta^{(i_1)\beta^{(i_2)}}}^2 \mathcal{L}_N^{\text{[S|R]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} 0.$$

The reasons behind this are the same as in the proof of consistency. The same can be said for the limit of $\frac{1}{N\sqrt{h}} \partial_{\beta\sigma} \mathcal{L}_N^{\text{[S|R]}}(\boldsymbol{\theta})$. Finally, repeating the same derivations as before, we get:

$$\begin{aligned}
\frac{1}{N-1} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{\text{[CS|R]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} & \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \text{Tr}((\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top), \\
\frac{1}{N-2} \partial_{\sigma^{(j_1)\sigma^{(j_2)}}}^2 \mathcal{L}_N^{\text{[PS|R]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} & \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} 2 \text{Tr}((\Sigma \Sigma^\top)^{-1} (\partial_{\sigma^{(j_1)}} \Sigma \Sigma^\top) (\Sigma \Sigma^\top)^{-1} \partial_{\sigma^{(j_2)}} \Sigma \Sigma^\top).
\end{aligned}$$

This concludes the first part of the lemma. The second part follows from the fact that all limits are continuous in $\boldsymbol{\theta}$.

To prove Lemma 3.4, we state another useful property that provides a general formula to calculate moments of a product of two quadratic forms with Gaussian vectors.

Lemma S1.3 Let $(\alpha_k)_{k=1}^N, (\beta_k)_{k=1}^N$, be two sequences of independent $\mathcal{F}_{t_{k+1}}$ -measurable Gaussian random variables with mean zero. If $\mathbb{E}_{\theta_0}[\alpha_{k-1}\beta_{k-1}^\top | \mathcal{F}_{t_{k-1}}]$ is diagonal, and \mathbf{A} and \mathbf{B} are two symmetric positive definite matrices, then:

$$\begin{aligned} \mathbb{E}_{\theta_0}[\alpha_{k-1}^\top \mathbf{A} \alpha_{k-1} \beta_{k-1}^\top \mathbf{B} \beta_{k-1} | \mathcal{F}_{t_{k-1}}] &= 2 \operatorname{Tr}(\mathbb{E}_{\theta_0}[\alpha_{k-1} \beta_{k-1}^\top | \mathcal{F}_{t_{k-1}}] \mathbf{A} \mathbb{E}_{\theta_0}[\alpha_{k-1} \beta_{k-1}^\top | \mathcal{F}_{t_{k-1}}] \mathbf{B}) \\ &\quad + \operatorname{Tr}(\mathbf{A} \mathbb{E}_{\theta_0}[\alpha_{k-1} \alpha_{k-1}^\top | \mathcal{F}_{t_{k-1}}]) \operatorname{Tr}(\mathbf{B} \mathbb{E}_{\theta_0}[\beta_{k-1} \beta_{k-1}^\top | \mathcal{F}_{t_{k-1}}]). \end{aligned}$$

Proof We start with

$$\begin{aligned} \mathbb{E}_{\theta_0}[\alpha_{k-1}^\top \mathbf{A} \alpha_{k-1} \beta_{k-1}^\top \mathbf{B} \beta_{k-1} | \mathcal{F}_{t_{k-1}}] &= \sum_{i,j,l,m=1}^d A_{ij} B_{lm} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \alpha_{k-1}^{(j)} \beta_{k-1}^{(l)} \beta_{k-1}^{(m)} | \mathcal{F}_{t_{k-1}}] \\ &= \sum_{i,j,l,m=1}^d A_{ij} B_{lm} \operatorname{cov}(\alpha_{k-1}^{(i)} \alpha_{k-1}^{(j)}, \beta_{k-1}^{(l)} \beta_{k-1}^{(m)}) \tag{S14} \\ &\quad + \sum_{i,j,l,m=1}^d A_{ij} B_{lm} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \alpha_{k-1}^{(j)} | \mathcal{F}_{t_{k-1}}] \mathbb{E}_{\theta_0}[\beta_{k-1}^{(l)} \beta_{k-1}^{(m)} | \mathcal{F}_{t_{k-1}}]. \end{aligned}$$

We compute (S14) using the formula for the covariance of products of centered Gaussian random variables (Bohrstedt and Goldberger, 1969). Then we get

$$\begin{aligned} \mathbb{E}_{\theta_0}[\alpha_{k-1}^\top \mathbf{A} \alpha_{k-1} \beta_{k-1}^\top \mathbf{B} \beta_{k-1} | \mathcal{F}_{t_{k-1}}] &= \sum_{i,j,l,m=1}^d A_{ij} B_{lm} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \beta_{k-1}^{(l)} | \mathcal{F}_{t_{k-1}}] \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(j)} \beta_{k-1}^{(m)} | \mathcal{F}_{t_{k-1}}] \\ &\quad + \sum_{i,j,l,m=1}^d A_{ij} B_{lm} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \beta_{k-1}^{(m)} | \mathcal{F}_{t_{k-1}}] \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(j)} \beta_{k-1}^{(l)} | \mathcal{F}_{t_{k-1}}] \\ &\quad + \sum_{i,j,l,m=1}^d A_{ij} B_{lm} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \alpha_{k-1}^{(j)} | \mathcal{F}_{t_{k-1}}] \mathbb{E}_{\theta_0}[\beta_{k-1}^{(l)} \beta_{k-1}^{(m)} | \mathcal{F}_{t_{k-1}}] \\ &= 2 \sum_{i,j=1}^d A_{ij} B_{ij} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \beta_{k-1}^{(i)} | \mathcal{F}_{t_{k-1}}] \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(j)} \beta_{k-1}^{(j)} | \mathcal{F}_{t_{k-1}}] \\ &\quad + \sum_{i,j=1}^d A_{ii} B_{jj} \mathbb{E}_{\theta_0}[\alpha_{k-1}^{(i)} \alpha_{k-1}^{(i)} | \mathcal{F}_{t_{k-1}}] \mathbb{E}_{\theta_0}[\beta_{k-1}^{(j)} \beta_{k-1}^{(j)} | \mathcal{F}_{t_{k-1}}]. \end{aligned}$$

That concludes the proof.

Applying the previous Lemma to our setup, corollary S1.4 follows immediately.

Corollary S1.4 Let $(\eta_k)_{k=1}^N, (\xi_k)_{k=1}^N, (\xi'_k)_{k=1}^N$ be random sequences as defined in (61) and (62). Let \mathbf{B}_{j_1} and \mathbf{B}_{j_2} be two symmetric positive definite matrices. Then, it holds:

$$\mathbb{E}_{\theta_0}[\eta_{k-1}^\top \mathbf{B}_{j_1} \eta_{k-1} \eta_{k-1}^\top \mathbf{B}_{j_2} \eta_{k-1} | \mathcal{F}_{t_{k-1}}] = 2 \operatorname{Tr}(\mathbf{B}_{j_1} \mathbf{B}_{j_2}) + \operatorname{Tr} \mathbf{B}_{j_1} \operatorname{Tr} \mathbf{B}_{j_2}, \tag{S15}$$

$$\mathbb{E}_{\theta_0}[\xi_{k-1}^\top \mathbf{B}_{j_1} \xi_{k-1} \xi_{k-1}^\top \mathbf{B}_{j_2} \xi_{k-1} | \mathcal{F}_{t_{k-1}}] = \frac{2}{9} \operatorname{Tr}(\mathbf{B}_{j_1} \mathbf{B}_{j_2}) + \frac{1}{9} \operatorname{Tr} \mathbf{B}_{j_1} \operatorname{Tr} \mathbf{B}_{j_2}, \tag{S16}$$

$$\mathbb{E}_{\theta_0}[\xi'_{k-1} \mathbf{B}_{j_1} \xi'_{k-1} \xi'_{k-1} \mathbf{B}_{j_2} \xi'_{k-1} | \mathcal{F}_{t_{k-1}}] = \frac{2}{9} \operatorname{Tr}(\mathbf{B}_{j_1} \mathbf{B}_{j_2}) + \frac{1}{9} \operatorname{Tr} \mathbf{B}_{j_1} \operatorname{Tr} \mathbf{B}_{j_2}, \tag{S17}$$

$$\mathbb{E}_{\theta_0}[\xi_{k-1}^\top \mathbf{B}_{j_1} \xi_{k-1} \xi'_{k-1} \mathbf{B}_{j_2} \xi'_{k-1} | \mathcal{F}_{t_{k-1}}] = \frac{1}{18} \operatorname{Tr}(\mathbf{B}_{j_1} \mathbf{B}_{j_2}) + \frac{1}{9} \operatorname{Tr} \mathbf{B}_{j_1} \operatorname{Tr} \mathbf{B}_{j_2}. \tag{S18}$$

Proof of Lemma 3.4 To prove the lemma, we need to compute $\lambda_N^{[\text{obj}]}$. The main part of the proof focuses only on the rough estimators, while at the end of the proof we discuss how the same ideas can be adapted for other estimators. Thus, we start with $-\frac{1}{\sqrt{N}h} \partial_{\beta^{(i)}} \mathcal{L}_N^{[\text{R}]}$:

$$-\frac{1}{\sqrt{(N-1)h}} \partial_{\beta^{(i)}} \mathcal{L}_N^{[\text{CR}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \frac{2}{\sqrt{N-1}} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})$$

$$\begin{aligned}
& + 2\sqrt{\frac{h}{N-1}} \sum_{k=1}^N \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\
& + \sqrt{\frac{h}{N-1}} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} - \sqrt{\frac{h}{N-1}} \sum_{k=1}^N \text{Tr} D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}), \\
& - \frac{1}{\sqrt{(N-2)h}} \partial_{\beta^{(i)}} \mathcal{L}_N^{\text{[PR]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = \frac{2}{\sqrt{N-2}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}) \\
& + 2\sqrt{\frac{h}{N-2}} \sum_{k=1}^{N-1} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\
& + \sqrt{\frac{h}{N-2}} \sum_{k=1}^{N-1} (\mathbf{U}_{k,k-1} + 2\xi_{k-1})^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \\
& - \sqrt{\frac{h}{N-2}} \sum_{k=1}^{N-1} \text{Tr} D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}(\mathbf{Y}_{t_k}; \boldsymbol{\beta}).
\end{aligned}$$

Similarly, for $-\frac{1}{\sqrt{N}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{\text{[R]}}$, we get:

$$\begin{aligned}
& - \frac{1}{\sqrt{(N-1)}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{\text{[CR]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = - \frac{1}{\sqrt{N-1}} \sum_{k=1}^N \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) \\
& + \frac{1}{\sqrt{N-1}} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \\
& + 2\sqrt{\frac{h}{N-1}} \sum_{k=1}^N \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\
& + \sum_{k=1}^N R\left(\frac{h}{\sqrt{N}}, \mathbf{Y}_{t_{k-1}}\right), \\
& - \frac{1}{\sqrt{(N-2)}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{\text{[PR]}}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = - \frac{2}{3\sqrt{N-2}} \sum_{k=1}^{N-1} \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) \\
& + \frac{1}{\sqrt{N-2}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \\
& + 2\sqrt{\frac{h}{N-2}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) - \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta})) \\
& + \sum_{k=1}^{N-1} R\left(\frac{h}{\sqrt{N}}, \mathbf{Y}_{t_{k-1}}\right).
\end{aligned}$$

To prove the convergence in distribution of $\boldsymbol{\lambda}_N^{\text{[R]}}$, we introduce the following triangular arrays that arise from the previous calculations:

$$\begin{aligned}
\phi_{N,k-1}^{\text{[CR]}(i)}(\boldsymbol{\theta}_0) & := \frac{2}{\sqrt{N-1}} \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \\
& + \sqrt{\frac{h}{N-1}} (\text{Tr}(\boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} \boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1}) - \text{Tr} D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})),
\end{aligned} \tag{S19}$$

$$\begin{aligned}
\phi_{N,k-1}^{\text{[PR]}(i)}(\boldsymbol{\theta}_0) & := \frac{2}{\sqrt{N-2}} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) \\
& + \sqrt{\frac{h}{N-2}} (\text{Tr}(\boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} (\mathbf{U}_{k,k-1} + 2\xi_{k-1})^\top \boldsymbol{\Sigma}_0^\top D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}})^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1}) - \text{Tr} D_{\mathbf{v}} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_k})),
\end{aligned} \tag{S20}$$

$$\rho_{N,k-1}^{[\text{CR}](j)}(\boldsymbol{\theta}_0) := \frac{1}{\sqrt{N-1}} (\boldsymbol{\eta}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\eta}_{k-1} - \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)), \quad (\text{S21})$$

$$\rho_{N,k-1}^{[\text{PR}](j)}(\boldsymbol{\theta}_0) := \frac{1}{\sqrt{N-2}} (\mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} - \frac{2}{3} \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)). \quad (\text{S22})$$

Then, $\boldsymbol{\lambda}_N^{[\text{R}]}$ rewrites as:

$$\boldsymbol{\lambda}_N^{[\text{R}]} = \sum_{k=1}^N \begin{bmatrix} \phi_{N,k-1}^{[\text{R}](1)}(\boldsymbol{\theta}_0) \\ \vdots \\ \phi_{N,k-1}^{[\text{R}](r)}(\boldsymbol{\theta}_0) \\ \rho_{N,k-1}^{[\text{R}](1)}(\boldsymbol{\theta}_0) \\ \vdots \\ \rho_{N,k-1}^{[\text{R}](s)}(\boldsymbol{\theta}_0) \end{bmatrix} + \frac{1}{N} \sum_{k=1}^N R(\sqrt{Nh^2}, \mathbf{Y}_{t_{k-1}}). \quad (\text{S23})$$

Thus, to establish estimators' asymptotic normality, we need an extra convergence condition $Nh^2 \rightarrow 0$. This is common in literature, and it is necessary for most estimators.

To finish the proof, we apply the central limit theorem for martingale difference arrays (Proposition 3.1 in Crimaldi and Pratelli (2005)). However, we can not apply the same reasoning in complete and partial observation cases.

First, we notice that in both complete and partial cases, $\phi_{N,k-1}^{[\text{R}](i)}(\boldsymbol{\theta}_0)$ and $\rho_{N,k-1}^{[\text{R}](j)}(\boldsymbol{\theta}_0)$ are centered conditionally to $\mathcal{F}_{t_{k-1}}$. Moreover, in the complete case, $\phi_{N,k-1}^{[\text{CR}](i)}(\boldsymbol{\theta}_0)$ and $\rho_{N,k-1}^{[\text{CR}](j)}(\boldsymbol{\theta}_0)$ are adapted to the filtration \mathcal{F}_{t_k} . Thus, the proof follows directly by applying Proposition 3.1 in Crimaldi and Pratelli (2005). This proposition assumes a martingale difference array centered conditionally to $\mathcal{F}_{t_{k-1}}$ and \mathcal{F}_{t_k} -measurable.

In the partial observation case, $\mathbf{U}_{k,k-1}$ is $\mathcal{F}_{t_{k+1}}$ -measurable as it depends on random variables $\boldsymbol{\xi}_{k-1}$ and $\boldsymbol{\xi}'_k$. Consequently, to apply Proposition 3.1 in Crimaldi and Pratelli (2005), it is not enough for $\phi_{N,k-1}^{[\text{PR}](i)}(\boldsymbol{\theta}_0)$ and $\rho_{N,k-1}^{[\text{PR}](j)}(\boldsymbol{\theta}_0)$ to be centered conditionally to $\mathcal{F}_{t_{k-1}}$, they also need to be centered conditionally to \mathcal{F}_{t_k} . The previous condition, however, does not hold. Thus, we use the idea of reordering the sum in $\boldsymbol{\lambda}_N^{[\text{PR}]}$ (S23) to obtain the \mathcal{F}_{t_k} -measurable and centered conditionally on $\mathcal{F}_{t_{k-1}}$, as proposed by Gloter (2000, 2006) and later used by Samson and Thieullen (2012).

First, use Lemma 9 from Genon-Catalot and Jacod (1993) to notice that:

$$\sum_{k=1}^{N-1} \phi_{N,k-1}^{[\text{PR}](i)}(\boldsymbol{\theta}_0) = \frac{2}{\sqrt{N-2}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1).$$

Then, reorder the sum of $\phi_{N,k-1}^{[\text{PR}](i)}(\boldsymbol{\theta}_0)$ as follows:

$$\begin{aligned} \sum_{k=1}^{N-1} \phi_{N,k-1}^{[\text{PR}](i)}(\boldsymbol{\theta}_0) &= \frac{2}{\sqrt{N-2}} (\boldsymbol{\xi}_0^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_0}) + \boldsymbol{\xi}'_{N-1}{}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{N-2}})) \\ &+ \frac{2}{\sqrt{N-2}} \sum_{k=2}^{N-1} (\boldsymbol{\xi}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \boldsymbol{\xi}'_{k-1}{}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-2}})) + o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1) \\ &= \frac{2}{\sqrt{N-2}} \sum_{k=2}^{N-1} (\boldsymbol{\xi}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \boldsymbol{\xi}'_{k-1}{}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-2}})) + o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1). \end{aligned}$$

Now, the triangular arrays under the sum are centered conditionally on $\mathcal{F}_{t_{k-1}}$ and \mathcal{F}_{t_k} measurable. Thus, define:

$$\phi_{N,k-1}^{*[\text{PR}](i)}(\boldsymbol{\theta}_0) := \frac{2}{\sqrt{N-2}} (\boldsymbol{\xi}_{k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}) + \boldsymbol{\xi}'_{k-1}{}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\beta^{(i)}} \mathbf{F}_0(\mathbf{Y}_{t_{k-2}})).$$

To apply Proposition 3.1 from Crimaldi and Pratelli (2005), we need the following limits in probability:

$$\sum_{k=2}^{N-1} \mathbb{E}_{\boldsymbol{\theta}_0} [\phi_{N,k-1}^{*[\text{PR}](i_1)}(\boldsymbol{\theta}_0) \phi_{N,k-1}^{*[\text{PR}](i_2)}(\boldsymbol{\theta}_0) \mid \mathcal{F}_{t_{k-1}}] \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} 4[\mathbf{C}_\beta(\boldsymbol{\theta}_0)]_{i_1 i_2},$$

$$\sum_{k=2}^{N-1} \mathbb{E}_{\theta_0} [(\phi_{N,k-1}^{*\text{[PR]}(i_1)}(\theta_0) \phi_{N,k-1}^{*\text{[PR]}(i_2)}(\theta_0))^2(\theta_0) \mid \mathcal{F}_{t_{k-1}}] \xrightarrow{\mathbb{P}_{\theta_0}} 0.$$

The first limit follows from properties (66) and (67). The second limit follows due to an additional order of $1/N$.

When looking at $\rho_{N,k-1}^{[\text{R}](j)}$, we repeat the same reasoning. For notational simplicity, start with defining:

$$\mathbf{B}_j(\theta_0) := \Sigma_0(\Sigma \Sigma_0^\top)^{-1}(\partial_{\sigma^{(j)}} \Sigma \Sigma_0^\top)(\Sigma \Sigma_0^\top)^{-1} \Sigma_0.$$

It follows immediately that $\text{Tr}(\mathbf{B}_j(\theta_0)) = \text{Tr}((\Sigma \Sigma_0^\top)^{-1} \partial_{\sigma^{(j)}} \Sigma \Sigma_0^\top)$. Again, reorder the sum of $\rho_{N,k-1}^{[\text{PR}](j)}(\theta_0)$ as follows:

$$\begin{aligned} \sum_{k=1}^{N-1} \rho_{N,k-1}^{[\text{PR}](j)}(\theta_0) &= \frac{1}{\sqrt{N-2}} \sum_{k=1}^{N-1} (\mathbf{U}_{k,k-1}^\top \mathbf{B}_j(\theta_0) \mathbf{U}_{k,k-1} - \frac{2}{3} \text{Tr}(\mathbf{B}_j(\theta_0))) \\ &= \frac{1}{\sqrt{N-2}} \sum_{k=2}^{N-1} \left(\xi_{k-1}^\top \mathbf{B}_j(\theta_0) \xi_{k-1} + 2 \xi_{k-2}^\top \mathbf{B}_j(\theta_0) \xi'_{k-1} + \xi_{k-1}^\top \mathbf{B}_j(\theta_0) \xi'_{k-1} - \frac{2}{3} \text{Tr}(\mathbf{B}_j(\theta_0)) \right) \\ &\quad + \frac{1}{\sqrt{N-2}} \left(\xi_0^\top \mathbf{B}_j(\theta_0) \xi_0 + 2 \xi_{N-2}^\top \mathbf{B}_j(\theta_0) \xi'_{N-1} + \xi_{N-1}^\top \mathbf{B}_j(\theta_0) \xi'_{N-1} - \frac{2}{3} \text{Tr}(\mathbf{B}_j(\theta_0)) \right). \end{aligned}$$

Since the last term in the previous equation is $o_{\mathbb{P}_{\theta_0}}(1)$, we focus only on:

$$\rho_{N,k-1}^{*\text{[PR]}(j)}(\theta_0) := \frac{1}{\sqrt{N-2}} \left(\xi_{k-1}^\top \mathbf{B}_j(\theta_0) \xi_{k-1} + 2 \xi_{k-2}^\top \mathbf{B}_j(\theta_0) \xi'_{k-1} + \xi_{k-1}^\top \mathbf{B}_j(\theta_0) \xi'_{k-1} - \frac{2}{3} \text{Tr}(\mathbf{B}_j(\theta_0)) \right).$$

Notice that $\rho_{N,k-1}^{*\text{[PR]}(j)}(\theta_0)$ is \mathcal{F}_{t_k} measurable and centered conditionally on $\mathcal{F}_{t_{k-1}}$. Again, to apply Proposition 3.1 from Crimaldi and Pratelli (2005), we need the following limits in probability:

$$\begin{aligned} &\sum_{k=2}^{N-1} \mathbb{E}_{\theta_0} [\rho_{N,k-1}^{*\text{[PR]}(j_1)}(\theta_0) \rho_{N,k-1}^{*\text{[PR]}(j_2)}(\theta_0) \mid \mathcal{F}_{t_{k-1}}] \xrightarrow{\mathbb{P}_{\theta_0}} [\mathbf{C}_\sigma(\theta_0)]_{j_1 j_2}, \\ &\sum_{k=2}^{N-1} \mathbb{E}_{\theta_0} [(\rho_{N,k-1}^{*\text{[PR]}(j_1)}(\theta_0) \rho_{N,k-1}^{*\text{[PR]}(j_2)}(\theta_0))^2(\theta_0) \mid \mathcal{F}_{t_{k-1}}] \xrightarrow{\mathbb{P}_{\theta_0}} 0. \end{aligned}$$

Once again, the second limit follows trivially. To prove the first limit, start by noticing that:

$$\mathbb{E}_{\theta_0} [\xi_{k-1}^\top \mathbf{B}_j(\theta_0) \xi_{k-1} \mid \mathcal{F}_{t_{k-1}}] = \mathbb{E}_{\theta_0} [\xi_{k-1}'^\top \mathbf{B}_j(\theta_0) \xi'_{k-1} \mid \mathcal{F}_{t_{k-1}}] = \frac{1}{3} \text{Tr}(\mathbf{B}_j(\theta_0)).$$

Then, we multiply the expectation with $N-2$ for notational simplicity and compute:

$$\begin{aligned} &(N-2) \mathbb{E}_{\theta_0} [\rho_{N,k-1}^{*\text{[PR]}(j_1)}(\theta_0) \rho_{N,k-1}^{*\text{[PR]}(j_2)}(\theta_0) \mid \mathcal{F}_{t_{k-1}}] \\ &= \mathbb{E}_{\theta_0} [\xi_{k-1}^\top \mathbf{B}_{j_1}(\theta_0) \xi_{k-1} \xi_{k-1}^\top \mathbf{B}_{j_2}(\theta_0) \xi_{k-1} \mid \mathcal{F}_{t_{k-1}}] + 4 \mathbb{E}_{\theta_0} [\xi_{k-2}^\top \mathbf{B}_{j_1}(\theta_0) \xi'_{k-1} \xi_{k-2}^\top \mathbf{B}_{j_2}(\theta_0) \xi'_{k-1} \mid \mathcal{F}_{t_{k-1}}] \\ &\quad + \mathbb{E}_{\theta_0} [\xi_{k-1}'^\top \mathbf{B}_{j_1}(\theta_0) \xi'_{k-1} \xi_{k-1}'^\top \mathbf{B}_{j_2}(\theta_0) \xi'_{k-1} \mid \mathcal{F}_{t_{k-1}}] + \mathbb{E}_{\theta_0} [\xi_{k-1}^\top \mathbf{B}_{j_1}(\theta_0) \xi_{k-1} \xi_{k-1}'^\top \mathbf{B}_{j_2}(\theta_0) \xi'_{k-1} \mid \mathcal{F}_{t_{k-1}}] \\ &\quad + \mathbb{E}_{\theta_0} [\xi_{k-1}'^\top \mathbf{B}_{j_1}(\theta_0) \xi'_{k-1} \xi_{k-1}^\top \mathbf{B}_{j_2}(\theta_0) \xi_{k-1} \mid \mathcal{F}_{t_{k-1}}] - \frac{4}{9} \text{Tr}(\mathbf{B}_{j_1}(\theta_0)) \text{Tr}(\mathbf{B}_{j_2}(\theta_0)). \end{aligned} \tag{S24}$$

Applying Corollary S1.4 on (S24) yields:

$$\begin{aligned} \sum_{k=2}^{N-1} \mathbb{E}_{\theta_0} [\rho_{N,k-1}^{*\text{[PR]}(j_1)}(\theta_0) \rho_{N,k-1}^{*\text{[PR]}(j_2)}(\theta_0) \mid \mathcal{F}_{t_{k-1}}] &= \frac{5}{9} \text{Tr}(\mathbf{B}_{j_1}(\theta_0) \mathbf{B}_{j_2}(\theta_0)) \\ &\quad + \frac{4}{3} \frac{1}{N-2} \sum_{k=2}^{N-1} \xi_{k-2}^\top \mathbf{B}_{j_1}(\theta_0) \mathbf{B}_{j_2}(\theta_0) \xi_{k-2}. \end{aligned}$$

Once again, applying Proposition 3.1 from Crimaldi and Pratelli (2005) yields:

$$\frac{4}{3} \frac{1}{N-2} \sum_{k=2}^{N-1} \xi_{k-2}^\top \mathbf{B}_{j_1}(\theta_0) \mathbf{B}_{j_2}(\theta_0) \xi_{k-2} \xrightarrow{\mathbb{P}_{\theta_0}} \frac{4}{9} \text{Tr}(\mathbf{B}_{j_1}(\theta_0) \mathbf{B}_{j_2}(\theta_0)),$$

since

$$\begin{aligned} & \frac{1}{N-2} \sum_{k=2}^{N-1} \mathbb{E}_{\theta_0} [\boldsymbol{\xi}_{k-2}^\top \mathbf{B}_{j_1}(\boldsymbol{\theta}_0) \mathbf{B}_{j_2}(\boldsymbol{\theta}_0) \boldsymbol{\xi}_{k-2} \mid \mathcal{F}_{t_{k-2}}] \xrightarrow{\mathbb{P}_{\theta_0}} \frac{1}{3} \text{Tr}(\mathbf{B}_{j_1}(\boldsymbol{\theta}_0) \mathbf{B}_{j_2}(\boldsymbol{\theta}_0)), \\ & \frac{1}{(N-2)^2} \sum_{k=2}^{N-1} \mathbb{E}_{\theta_0} [(\boldsymbol{\xi}_{k-2}^\top \mathbf{B}_{j_1}(\boldsymbol{\theta}_0) \mathbf{B}_{j_2}(\boldsymbol{\theta}_0) \boldsymbol{\xi}_{k-2})^2 \mid \mathcal{F}_{t_{k-2}}] \xrightarrow{\mathbb{P}_{\theta_0}} 0. \end{aligned}$$

This concludes the convergence in distribution of $\boldsymbol{\lambda}_N^{[\text{PR}]}$.

To find the asymptotic distributions of $\boldsymbol{\lambda}_N^{[\text{S|R}]}$, the main issue is the fact that $-\frac{1}{\sqrt{Nh}} \partial_{\beta^{(i)}} \mathcal{L}_N^{[\text{S|R}]} \rightarrow 0$ in probability. The proof of this follows the same ideas as in the proof of consistency. Thus, we focus only on $-\frac{1}{\sqrt{N}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{[\text{S|R}]}$. This is then used together with equations (71) and (74) to obtain the asymptotic distributions of $\boldsymbol{\lambda}_N^{[\text{F}]}$. Thus, we start with $-\frac{1}{\sqrt{N}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{[\text{S|R}]}$:

$$\begin{aligned} & -\frac{1}{\sqrt{(N-1)}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{[\text{CS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = -\frac{1}{\sqrt{N-1}} \sum_{k=1}^N \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) \\ & + \frac{3}{\sqrt{N-1}} \sum_{k=1}^N (\boldsymbol{\eta}_{k-1} - 2\boldsymbol{\xi}'_{k-1})^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 (\boldsymbol{\eta}_{k-1} - 2\boldsymbol{\xi}'_{k-1}) + \sum_{k=1}^N R\left(\frac{h}{\sqrt{N}}, \mathbf{Y}_{t_{k-1}}\right), \\ & -\frac{1}{\sqrt{(N-2)}} \partial_{\sigma^{(j)}} \mathcal{L}_N^{[\text{PS|R}]}(\mathbf{Y}_{0:t_N}; \boldsymbol{\theta}) = -\frac{2}{\sqrt{N-2}} \sum_{k=1}^{N-1} \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) \\ & + \frac{3}{\sqrt{N-2}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} \\ & - 6\sqrt{\frac{h}{N-2}} \sum_{k=1}^{N-1} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \mathbf{F}(\mathbf{Y}_{t_{k-1}}; \boldsymbol{\beta}_0) + \sum_{k=1}^{N-1} R\left(\frac{h}{\sqrt{N}}, \mathbf{Y}_{t_{k-1}}\right). \end{aligned}$$

Once again, we define:

$$\begin{aligned} \boldsymbol{\rho}_{N,k-1}^{[\text{CS|R}](j)}(\boldsymbol{\theta}_0) & := \frac{1}{\sqrt{N-1}} \left(3(\boldsymbol{\eta}_{k-1} - 2\boldsymbol{\xi}'_{k-1})^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \boldsymbol{\Sigma}_0 (\boldsymbol{\eta}_{k-1} - 2\boldsymbol{\xi}'_{k-1}) \right. \\ & \quad \left. - \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) \right), \end{aligned} \quad (\text{S25})$$

$$\begin{aligned} \boldsymbol{\rho}_{N,k-1}^{[\text{PS|R}](j)}(\boldsymbol{\theta}_0) & := \frac{3}{\sqrt{N-2}} (\mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \boldsymbol{\Sigma}_0 \mathbf{U}_{k,k-1} - \frac{2}{3} \text{Tr}((\boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)^{-1} \partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_0^\top)) \\ & \quad - 6\sqrt{\frac{h}{N-2}} \mathbf{U}_{k,k-1}^\top \boldsymbol{\Sigma}_0^\top (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} (\partial_{\sigma^{(j)}} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top) (\boldsymbol{\Sigma} \boldsymbol{\Sigma}^\top)^{-1} \mathbf{F}_0(\mathbf{Y}_{t_{k-1}}). \end{aligned} \quad (\text{S26})$$

We skip the proof of the complete case, but it can be shown analogously that:

$$\begin{aligned} & \sum_{k=1}^N \mathbb{E}_{\theta_0} [\boldsymbol{\rho}_{N,k-1}^{[\text{PS|R}](j_1)}(\boldsymbol{\theta}_0) \boldsymbol{\rho}_{N,k-1}^{[\text{PS|R}](j_2)}(\boldsymbol{\theta}_0) \mid \mathcal{F}_{t_{k-1}}] \xrightarrow{\mathbb{P}_{\theta_0}} [\mathbf{C}_\sigma(\boldsymbol{\theta}_0)]_{j_1 j_2}, \\ & \sum_{k=1}^N \mathbb{E}_{\theta_0} [(\boldsymbol{\rho}_{N,k-1}^{[\text{PS|R}](j_1)}(\boldsymbol{\theta}_0) \boldsymbol{\rho}_{N,k-1}^{[\text{PS|R}](j_2)}(\boldsymbol{\theta}_0))^2 \mid \mathcal{F}_{t_{k-1}}] \xrightarrow{\mathbb{P}_{\theta_0}} 0. \end{aligned}$$

Focusing on the partial case, we first notice:

$$\boldsymbol{\rho}_{N,k-1}^{[\text{PS|R}](j)}(\boldsymbol{\theta}_0) = 3\boldsymbol{\rho}_{N,k-1}^{[\text{PR}](j)}(\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\theta_0}}(1).$$

Thus, the same derivations from before hold. Moreover,

$$\boldsymbol{\rho}_{N,k-1}^{[\text{PF}](j)}(\boldsymbol{\theta}_0) = 4\boldsymbol{\rho}_{N,k-1}^{[\text{PR}](j)}(\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\theta_0}}(1),$$

which concludes the proof.