Quasi-likelihood analysis for adaptive estimation of a degenerate diffusion process under relaxed balance conditions

Masayuki Uchida[1], Nakahiro Yoshida[2,3], and Shota Yano[1]

[1]Graduate School of Engineering Science, and Center for Mathematical Modeling and Date Science (MMDS) *

[2]Graduate School of Mathematical Sciences, University of Tokyo †

[3]Japan Science and Technology Agency CREST

[1]Graduate School of Mathematical Sciences, University of Tokyo ‡

Summary The adaptive quasi-likelihood analysis is developed for a degenerate diffusion process for relaxed balance condition $nh_n^p \to 0$. Asymptotic normality and moment convergence are uniformly established for hybrid estimators that combine the quasi-maximum likelihood estimator, quasi-Bayesian estimator, and multi-step estimator.

Keywords and phrases Degenerate diffusion, adaptive estimator, quasi-maximum likelihood estimator, quasi-Bayesian estimator, multi-step estimator, hybrid estimator, relaxed balance condition

1 Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$ be a standard stochastic basis. We will consider an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$-adapted process $Z_t = (X_t, Y_t)$ satisfying the following stochastic differential equation (SDE):

$$dZ_t = \begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} A(Z_t, \theta_2) \\ H(Z_t, \theta_3) \end{pmatrix} dt + \begin{pmatrix} B(Z_t, \theta_1) \\ O \end{pmatrix} dw_t, \tag{1}$$

where $A \in \mathrm{Map}(\mathbb{R}^{d_z} \times \overline{\Theta}_2; \mathbb{R}^{d_x})$, $B \in \mathrm{Map}(\mathbb{R}^{d_z} \times \overline{\Theta}_1; \mathbb{R}^{d_x} \otimes \mathbb{R}^r)$, $H \in \mathrm{Map}(\mathbb{R}^{d_z} \times \overline{\Theta}_3; \mathbb{R}^{d_y})$ and $w = (w_t)_{t\in\mathbb{R}_+}$ is an r-dimensional Brownian motion. The sets of unknown parameters $\Theta_l\,(l = 1, 2, 3)$ are open bounded subsets of $\mathbb{R}^{p_l}$, respectively. We denote the product set $\prod_{l=1}^3 \Theta_l$ by Θ. We denote by θ or $\bar{\theta}$ a running index in Θ and by $\boldsymbol{\theta} = (\theta, \bar{\theta})$ a running index in Θ^2. The set Θ^2 is assumed to admit the Sobolev inequality holds: for all $M > 2(\sum_{l=1}^3 p_l)$, there exists $C > 0$ such that

$$\sup_{\boldsymbol{\theta}\in\Theta^2} |f(\boldsymbol{\theta})|^M \le C \sum_{l=0,1} \int_{\Theta^2} |\partial_{\boldsymbol{\theta}}^l f(\boldsymbol{\theta})|^M d\boldsymbol{\theta} \text{ for all } f(\boldsymbol{\theta}) \in C^1(\Theta^2).$$

*Graduate School of Engineering Science, and Center for Mathematical Modeling and Date Science (MMDS), Osaka University, 1-3, Machikaneyama, Toyonaka, Osaka, 560-8531, Japan

†Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. e-mail: nakahiro@ms.u-tokyo.ac.jp

‡Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. e-mail: root1456849@g.ecc.u-tokyo.ac.jp

Let $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*) \in \Theta$ denote the true value of $\theta = (\theta_1, \theta_2, \theta_3) \in \Theta$ to be estimated. We aim to estimate $\theta = (\theta_1, \theta_2, \theta_3)$ from the data $(Z_{t_{n,j}})_{j=0,\cdots,n}$ with $t_{n,j} = jh_n$, where h_n satisfies $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

The statistical inference of stochastic differential equation models under discrete observation schemes has been actively studied since the 1980s. Prakasa Rao [1983, 1988] demonstrated the least square estimators are asymptotically normal under the condition $nh_n^2 \to 0$ as $n \to \infty$, known as the "rapidly increasing experimental design" condition.

Notably, Florens-Zmirou [1989] proposed an estimation method using the likelihood of a discretely approximated model created by the Euler-Maruyama method as a quasi-likelihood for stochastic differential equation models, which has become a standard approach for constructing quasi-likelihoods in subsequent research.

Research in this field has developed in three major directions. Florens-Zmirou [1989] and Yoshida [1992] showed that asymptotic normality of their estimators holds under the relaxed balance condition $nh_n^3 \to 0$, and Kessler [1997] further relaxed this to $nh_n^p \to 0$ $(p \geq 3)$.

Furthermore, the theory of quasi-likelihood analysis established by Yoshida [2011] made it possible to construct a wide range of estimators beyond the conventional quasi-maximum likelihood estimator (QMLE), including quasi-Bayesian estimator (QBE) and to prove their asymptotic normality with moment convergence in a unified framework. Under this theoretical framework, various estimation methods have been developed under the relaxed observation scheme $nh_n^p \to 0$, including the Bayesian-type estimator by Uchida and Yoshida [2012] and the multi-step estimator (MSE) by Kamatani and Uchida [2014]. Their method involves creating multiple types of quasi-likelihood functions and improving estimator accuracy by using them adaptively - a technique known as adaptive inference. According to numerical experiments in Kamatani and Uchida [2014], superior performance can be achieved in adaptive inference by constructing initial estimators through Bayesian estimation and then iteratively applying multi-step estimation methods.

However, all the literature mentioned above dealt with non-degenerate diffusion processes. Although there are many important degenerate equations, such as the harmonic oscillator and the FitzHugh-Nagumo model, their statistical inference involves fundamental difficulties compared to the non-degenerate case. For the degenerate diffusion process (1), the Euler-Maruyama approximation of the increment $Z_{t_{n,j}} - Z_{t_{n,j-1}}$ follows

$$Z_{t_{n,j}} - Z_{t_{n,j-1}} \approx \begin{pmatrix} A(Z_{t_{n,j-1}}, \theta_2) \\ H(Z_{t_{n,j-1}}, \theta_3) \end{pmatrix} h_n + \begin{pmatrix} B(Z_{t_{n,j-1}}, \theta_1) \\ O \end{pmatrix} (w_{t_{n,j}} - w_{t_{n,j-1}}).$$

Due to the degeneracy in the diffusion coefficient matrix, this standard Euler-Maruyama discretization fails to capture the underlying structure of the degenerate system: the distribution of this approximation is not absolutely continuous with respect to the Lebesgue measure, resulting in a singular measure. This makes it impossible to construct a quasi-likelihood function using the conventional local Gaussian approximation approach.

Recent studies Ditlevsen and Samson [2019], Gloter and Yoshida [2020, 2021, 2024a,b], Iguchi and Beskos [2023], Iguchi et al. [2024, 2025], Samson et al. [2025] have shown that this difficulty can be addressed through higher-order approximations based on the Itô-Taylor expansion, which introduces additional Gaussian terms with order $O(h_n^{3/2})$. Gloter and Yoshida [2020, 2021, 2024a] constructed adaptive and non-adaptive QMLE under the observation scheme $nh_n^2 \to 0$ by incorporating these higher-order Itô-Taylor expansion terms. Recently, Iguchi and Beskos [2023] extended the approach to construct adaptive QMLE for degenerate diffusion processes under the more general condition $nh_n^p \to 0$.

Furthermore, Iguchi et al. [2025] conducted Bayesian inference in the low-frequency context and Samson et al. [2025] discusses Bayesian inference for the FitzHugh-Nagumo model. However, there has not yet been any research demonstrating asymptotic normality for methods other than QMLE for degenerate diffusion processes in the high-frequency observation context.

In this paper, we further develop the adaptive inference for degenerate SDEs under the framework of the relaxed balance condition $nh_n^p \to 0$. We provide a unified framework that allows flexible combinations of estimation methods at each step: QMLE, QBE, and MSE. We establish asymptotic normality with moment convergence of all possible hybrid estimators under this general setting through quasi-likelihood analysis.

This paper is organized as follows. First, in Section 2, we explain all necessary notations and assumptions. In Section 3, the methods to construct our estimators and our main theorem (Theorem 3.2) are introduced. In Section 4, we present simulations for a linear model and the FitzHugh-Nagumo model. Section 5 and beyond are devoted to proving the main theorem. Section 5 introduce an appropriately modified version of quasi-likelihood analysis for this paper. In Section 6, we prove the main theorem using quasi-likelihood analysis assuming a central limit theorem (Proposition 6.1) and asymptotic analyses of our quasi-likelihoods (Proposition 6.2). In section 7, we prepare basic estimations utilized in the following sections. In Section 8, we conduct the Itô-Taylor expansion for the degenerate diffusion process, and in this Section, Proposition 6.1 is proved. In Section 9, under the preparation of Section 7,8, Proposition 6.2 is proved.

2 Notations and Assumptions

Let α_n, β_n be sequences of positive real numbers. We say $\alpha_n \Bumpeq \beta_n$ if there exist some $\epsilon, \mathcal{E}, c, C > 0$ such that for all n:

$$c\alpha_n^{\mathcal{E}} \leq \beta_n \leq C\alpha_n^{\epsilon}.$$

This forms an equivalence relation. From the definition, it's clear that $\alpha_n^\epsilon \Bumpeq \alpha_n$. Furthermore, when $\alpha_n \Bumpeq \beta_n$, we have:

$$(\alpha_n + \beta_n)^\epsilon \leq 2^\epsilon \alpha_n^\epsilon + 2^\epsilon \beta_n^\epsilon.$$

Therefore, $\alpha_n + \beta_n \Bumpeq \alpha_n$. Also, for any $c > 0$, it's clear that $c\alpha_n \Bumpeq \alpha_n$. In other words, the equivalence classes under $\sim$ form a convex cone. They are also closed under multiplication. Let p be some integer larger than or equal to 2 and we denote $\lfloor p/2 \rfloor$ by k_0. Hereafter, we assume [A1]-[A5] below.

[A1] $h_n \to 0, nh_n \to \infty, nh_n^p \to 0$ and there exists $\epsilon > 0$ such that $n^\epsilon \leq nh_n$

Assuming [A1], we note that for sufficiently large n: $n^\epsilon \leq nh_n \leq n$ and $h_n^p \leq \frac{1}{n} \leq h_n$. Therefore

$$h_n \Bumpeq \frac{1}{n} \Bumpeq \frac{1}{nh_n} \Bumpeq \frac{h_n}{n}$$

holds.

[A2] For any $M > 0$,

$$\sup_{t \in \mathbb{R}_+} E[|Z_t|^M] < \infty.$$

We denote the set of all polynomial growth continuous function on $\mathbb{R}^{dz}$ by $C_{\uparrow}(\mathbb{R}^{dz})$.

[A3] There exists a probability measure ν^* on $\mathbb{R}^{d_z}$ and $\epsilon > 0$ such that

$$\sup_n E\left[\left(\frac{\left|\frac{1}{nh_n}\int_0^{nh_n} f(Z_t)\,dt - \int_{\mathbb{R}^{d_z}} f(z)\,d\nu^*(x)\right|}{\frac{1}{(nh_n)^\epsilon}}\right)^M\right] < \infty$$

for all $M > 0$ and $f \in C_{\uparrow}(\mathbb{R}^{d_z})$.

Let E, F be finite-dimensional vector spaces and U a bounded open subset of some Euclidean space. We define two function spaces:

1. $C_{\uparrow}^m(E;F)$ is the set of all functions $f : E \ni z \mapsto f(z) \in F$ such that for all $m' \in \mathbb{Z}$ satisfying $0 \le m' \le m$, the derivative $\partial_z^{m'} f$ is defined as a continuous function on E and of at most polynomial growth in z.
2. $C_{\uparrow}^{m,l}(E \times U;F)$ is the set of all functions $f : E \times U \ni (z,u) \mapsto f(z,u) \in F$ such that for $m', l' \in \mathbb{Z}$ satisfying $0 \le m' \le m$ and $0 \le l' \le l$, the derivatives $\partial_z^{m'}\partial_u^{l'} f$ are defined as a continuous function on $E \times \bar{U}$ and of at most polynomial growth in z uniformly in u.

When the domain and codomain of a given function are clear from the context, we may abbreviate $f \in C_{\uparrow}^{m,l}(E \times U;F)$ as $f \in C_{\uparrow}^{m,l}$ and $f \in C_{\uparrow}^m(E;F)$ as $f \in C_{\uparrow}^m$.

[A4] $A \in C_{\uparrow}^{p-1,3}(\mathbb{R}^{d_z} \times \Theta_2; \mathbb{R}^{d_x}), B \in C_{\uparrow}^{p,3}(\mathbb{R}^{d_z} \times \Theta_1; \mathbb{R}^{d_x} \otimes \mathbb{R}^r)$ and $H \in C_{\uparrow}^{p+1,3}(\mathbb{R}^{d_z} \times \Theta_3; \mathbb{R}^{d_y})$.

Let $C := BB^\star$ and $V := (\partial_x H)C(\partial_x H)^\star$, where $\star$ denotes the transpose.

[A5] $\inf_{(z,\theta_1)\in\mathbb{R}^{d_z}\times\Theta_1} \det C(z,\theta_1) > 0$ and $\inf_{(z,\theta_1,\theta_3)\in\mathbb{R}^{d_z}\times\Theta_1\times\Theta_3} \det V(z,\theta_1,\theta_3) > 0$.

Let

$$\mathbb{Y}^{1'}(\theta_1) := -\frac{1}{2}\int \Big(\mathrm{Tr}\left(C^{-1}(z,\theta_1)C(z,\theta_1^*)\right) - d_x + \log\frac{\det C(z,\theta_1)}{\det C(z,\theta_1^*)}\Big) d\nu^*(z),$$

$$\mathbb{Y}^1(\theta_1) :=$$

$$-\frac{1}{2}\int\begin{pmatrix}\mathrm{Tr}\left(C^{-1}(z,\theta_1)C(z,\theta_1^*)\right) + \mathrm{Tr}\left(V^{-1}(z,\theta_1,\theta_3^*)V(z,\theta_1^*,\theta_3^*)\right) - d_z \\ + \log\dfrac{\det C(z,\theta_1)\det V(\theta_1,\theta_3^*)}{\det C(z,\theta_1^*)\det V(\theta_1^*,\theta_3^*)}\end{pmatrix} d\nu^*(z),$$

$$\mathbb{Y}^2(\theta_2) := -\frac{1}{2}\int C^{-1}(z,\theta_1^*)\left[\left(A(z,\theta_2) - A(z,\theta_2^*)\right)^{\otimes 2}\right]\nu^*(dz),$$

$$\mathbb{Y}^3(\theta_3|\bar{\theta}_3) := -\int 6V^{-1}(z,\theta_1^*,\bar{\theta}_3)\left[\left(H(z,\theta_3) - H(z,\theta_3^*)\right)^{\otimes 2}\right] d\nu^*(z),$$

where $\bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3)$ is a variable moving in Θ.

[A6] There exists $\epsilon > 0$ such that the following inequalities hold

(**i**) $\mathbb{Y}^{1'}(\theta_1) \le -\epsilon|\theta_1 - \theta_1^*|^2$ for all $\theta_1 \in \Theta_1$.

(**ii**) $\mathbb{Y}^1(\theta_1) \leq -\epsilon|\theta_1 - \theta_1^*|^2$ for all $\theta_1 \in \Theta_1$.

(**iii**) $\mathbb{Y}^2(\theta_2) \leq -\epsilon|\theta_2 - \theta_2^*|^2$ for all $\theta_2 \in \Theta_2$.

(**iv**) $\mathbb{Y}^3(\theta_3|\bar{\theta}_3) \leq -\epsilon|\theta_3 - \theta_3^*|^2$ for all $\theta_3, \bar{\theta}_3 \in \Theta_3$.

Let $f(u,v)$ be a function defined on some direct product set to $\mathbb{R}$. We denote $f(u,v) - f(u',v)$ by $f(u \backslash u', v)$. For example, $f(z, \theta_1, \theta_2 \backslash \theta_2^*, \theta_3) = f(z, \theta_1, \theta_2, \theta_3) - f(z, \theta_1, \theta_2^*, \theta_3)$.

Given a function $f(z, \boldsymbol{\theta})$ on $\mathbb{R}^{d_z} \times \Theta^2$, we write $f_{n,j}(\boldsymbol{\theta}) = f(Z_{t_{n,j}}, \boldsymbol{\theta})$ and we use simplified notations: $Z_{n,j} = Z_{t_{n,j}}$, $X_{n,j} = X_{t_{n,j}}$, $Y_{n,j} = Y_{t_{n,j}}$ and $E_{n,j}[\cdot] = E[\cdot|\mathcal{F}_{t_{n,j}}]$.

We denote the tensor inner product by $\mathcal{A}[\mathcal{B}]$ or $\mathcal{A} \cdot \mathcal{B}$. For instance,

$$C[(X_{n,j} - X_{n,j-1})^{\otimes 2}] = \sum_{l,m} C_{l,m}(X_{n,j} - X_{n,j-1})_l (X_{n,j} - X_{n,j-1})_m,$$

where $C_{l,m}$ is the (l,m)-component of C and $(X_{n,j} - X_{n,j-1})_l$ is the l-component of $(X_{n,j} - X_{n,j-1})$, respectively. The symmetrized tensor product is defined as

$$\mathcal{A} \odot \mathcal{B} = \frac{1}{2}(\mathcal{A} \otimes \mathcal{B} + \mathcal{B} \otimes \mathcal{A}).$$

Let $f(z,\theta) = f(x,y,\theta)$ be a C^2 function on $\mathbb{R}^{d_z} \times \Theta$. We define the operator L by

$$Lf(z,\theta) = A(z,\theta_2) \cdot \partial_x f(z,\theta) + H(z,\theta_3) \cdot \partial_y f(z,\theta) + \frac{1}{2} C(z,\theta_1) \cdot \partial_x^2 f(z,\theta)$$

and set

$$L_0^l f = \begin{cases} L^l f / l! & (l \geq 0) \\ 0 & (l < 0) \end{cases}.$$

for sufficiently smooth functions f and integers l.

3 Construction of estimators and the main theorem

For $k = 0, \cdots, k_0$, we define the following random fields under [A4]:

$$\mathcal{D}_{n,j}^k(\theta_2, \theta_3, \bar{\theta}) := \begin{pmatrix} \mathcal{D}_{n,j}^{k,x}(\theta_2, \bar{\theta}) \\ \mathcal{D}_{n,j}^{k,y}(\theta_3, \bar{\theta}) \end{pmatrix}$$

$$:= \begin{cases} \begin{pmatrix} h_n^{-1/2}(X_{n,j} - X_{n,j-1}) \\ h_n^{-3/2}(Y_{n,j} - Y_{n,j-1}) \end{pmatrix} & (k=0) \\ \begin{pmatrix} h_n^{-1/2}(X_{n,j} - X_{n,j-1} - hA_{n,j-1}(\theta_2)) \\ h_n^{-3/2}(Y_{n,j} - Y_{n,j-1} - hH_{n,j-1}(\theta_3)) \end{pmatrix} & (k=1) \\ \begin{pmatrix} h_n^{-1/2}(X_{n,j} - X_{n,j-1} - hA_{n,j-1}(\theta_2) - \sum_{m=2}^k h_n^m (L_0^m \pi_x)_{n,j-1}(\bar{\theta})) \\ h_n^{-3/2}(Y_{n,j} - Y_{n,j-1} - hH_{n,j-1}(\theta_3) - \sum_{m=2}^{k+1} h_n^m (L_0^m \pi_y)_{n,j-1}(\bar{\theta})) \end{pmatrix} & (k \geq 2) \end{cases},$$

$$\mathcal{D}_{n,j}^k(\bar{\theta}) := \mathcal{D}_{n,j}^k(\bar{\theta}_2, \bar{\theta}_3, \bar{\theta}).$$

For any integer l, we define the following formal matrix-valued polynomials constructed from formal variables representing partial derivatives $\{\partial_z^{l_A} A, \partial_z^{l_B} B, \partial_z^{l_H} H\}_{l_A, l_B, l_H = 0}^{\infty}$ and z:

$$\mathscr{T}^l := \begin{pmatrix} \mathscr{T}^{l,xx} & \mathscr{T}^{l,xy} \\ \mathscr{T}^{l,yx} & \mathscr{T}^{l,yy} \end{pmatrix}$$

$$
\begin{aligned}
&:= \begin{pmatrix} L_0^{l+1}\pi_x^{\otimes 2} & L_0^{l+2}(\pi_x \otimes \pi_y) \\ L_0^{l+2}(\pi_y \otimes \pi_x) & L_0^{l+3}\pi_y^{\otimes 2} \end{pmatrix}, \\
\mathscr{U}^l &:= \begin{pmatrix} \mathscr{U}^{l,xx} & \mathscr{U}^{l,xy} \\ \mathscr{U}^{l,yx} & \mathscr{U}^{l,yy} \end{pmatrix} \\
&:= \begin{pmatrix} \sum_{m_1+m_2=l+1} L_0^{m_1}\pi_x \otimes L_0^{m_2}\pi_x & \sum_{m_1+m_2=l+2} L_0^{m_1}\pi_x \otimes L_0^{m_2}\pi_y \\ \sum_{m_1+m_2=l+2} L_0^{m_1}\pi_y \otimes L_0^{m_2}\pi_x & \sum_{m_1+m_2=l+3} L_0^{m_1}\pi_y \otimes L_0^{m_2}\pi_y \end{pmatrix}, \\
\mathscr{S}^l &:= \begin{pmatrix} \mathscr{S}^{l,xx} & \mathscr{S}^{l,xy} \\ \mathscr{S}^{l,yx} & \mathscr{S}^{l,yy} \end{pmatrix} \\
&:= \mathscr{T}^l - \mathscr{U}^l,
\end{aligned}
$$

where the tensor products are taken in the range space. For instance, $\pi_x^{\otimes 2}$ means $z = (x, y) \mapsto x^{\otimes 2}$. Then, $T^l(z,\theta), U^l(z,\theta)$ and $S^l(z,\theta)$ are defined as the evaluation of the formal polynomials $\mathscr{T}^l, \mathscr{U}^l$ and $\mathscr{S}^l$ at $\{\partial_z^{l_A} A(z,\theta), \partial_z^{l_B} B(z,\theta), \partial_z^{l_H} H(z,\theta)\}_{l_A,l_B,l_H=0}^{\infty}$ and z, respectively, where we set any non-existent partial derivatives to 0. Similarly, we define $T^{l,xx}(z,\theta), T^{l,xy}(z,\theta), T^{l,yx}(z,\theta)$ and $T^{l,yy}(z,\theta)$, as well as the corresponding components of $U^l(z,\theta)$ and $S^l(z,\theta)$, in the same manner, that is,

$$
\begin{aligned}
T^l(z,\theta) &= \begin{pmatrix} T^{l,xx}(z,\theta) & T^{l,xy}(z,\theta) \\ T^{l,yx}(z,\theta) & T^{l,yy}(z,\theta) \end{pmatrix}, \\
U^l(z,\theta) &= \begin{pmatrix} U^{l,xx}(z,\theta) & U^{l,xy}(z,\theta) \\ U^{l,yx}(z,\theta) & U^{l,yy}(z,\theta) \end{pmatrix}, \\
S^l(z,\theta) &= \begin{pmatrix} S^{l,xx}(z,\theta) & S^{l,xy}(z,\theta) \\ S^{l,yx}(z,\theta) & S^{l,yy}(z,\theta) \end{pmatrix}.
\end{aligned}
$$

For instance, when $d_x = d_y = r = 1, p = 2$,

$$
\begin{aligned}
\mathscr{T}^{0,xy} = \mathscr{T}^{0,yx} &= AH + \frac{1}{2}B^2\partial_x H \\
&\quad + x(\frac{1}{2}A\partial_x H + \frac{1}{4}B^2\partial_x^2 A + \frac{1}{2}H\partial_y H) \\
&\quad + y(\frac{1}{2}A\partial_x A + \frac{1}{4}B^2\partial_x^2 A + \frac{1}{2}H\partial_y A), \qquad (2) \\
\mathscr{U}^{0,xy} = \mathscr{U}^{0,yx} &= AH \\
&\quad + x(\frac{1}{2}A\partial_x H + \frac{1}{4}B^2\partial_x^2 A + \frac{1}{2}H\partial_y H) \\
&\quad + y(\frac{1}{2}A\partial_x A + \frac{1}{4}B^2\partial_x^2 A + \frac{1}{2}H\partial_y A), \qquad (3) \\
\mathscr{S}^{0,xy} = \mathscr{S}^{0,yx} &= \frac{1}{2}B^2\partial_x H. \qquad (4)
\end{aligned}
$$

If $p=2$, we only assume $A\in C^{1,3}_{\uparrow}(\mathbb{R}^{d_z}\times\Theta_2;\mathbb{R}^{d_x}), B\in C^{2,3}_{\uparrow}(\mathbb{R}^{d_z}\times\Theta_1;\mathbb{R}^{d_x}\otimes\mathbb{R}^r)$ and $H\in C^{3,2}_{\uparrow}(\mathbb{R}^{d_z}\times\Theta_3;\mathbb{R}^{d_y})$ in [A4]. Thus, if $p=2$, we obtain

$$
\begin{aligned}
T^{0,xy}=T^{0,yx}&=\begin{cases}
AH+\frac{1}{2}B^2\partial_x H\\
\quad+x(\frac{1}{2}A\partial_x H+\frac{1}{4}B^2\partial_x^2 A+\frac{1}{2}H\partial_y H) & \text{if } \partial_x^2 A \text{ exists,}\\
\quad+y(\frac{1}{2}A\partial_x A+\frac{1}{4}B^2\partial_x^2 A+\frac{1}{2}H\partial_y A),\\
AH+\frac{1}{2}B^2\partial_x H\\
\quad+x(\frac{1}{2}A\partial_x H+\frac{1}{2}H\partial_y H) & \text{if } \partial_x^2 A \text{ does not exist,}\\
\quad+y(\frac{1}{2}A\partial_x A+\frac{1}{2}H\partial_y A),
\end{cases}\\
U^{0,xy}=U^{0,yx}&=\begin{cases}
AH\\
\quad+x(\frac{1}{2}A\partial_x H+\frac{1}{4}B^2\partial_x^2 A+\frac{1}{2}H\partial_y H) & \text{if } \partial_x^2 A \text{ exists,}\\
\quad+y(\frac{1}{2}A\partial_x A+\frac{1}{4}B^2\partial_x^2 A+\frac{1}{2}H\partial_y A),\\
AH\\
\quad+x(\frac{1}{2}A\partial_x H+\frac{1}{2}H\partial_y H) & \text{if } \partial_x^2 A \text{ does not exist.}\\
\quad+y(\frac{1}{2}A\partial_x A+\frac{1}{2}H\partial_y A),
\end{cases}\\
S^{0,xy}=S^{0,yx}&=\frac{1}{2}B^2\partial_x^2 H.
\end{aligned}
$$

For $l\le k_0$, $\mathscr{T}^l$ and $\mathscr{U}^l$ are formally defined using derivatives of order higher than that assumed in [A4] as (2) and (3). However, due to the cancellation of higher-order terms, their difference $\mathscr{S}^l$ is actually a polynomial that involves only derivatives of A, B, and H up to the order assumed in [A4] as (4). Therefore, when the coefficient functions satisfy [A4], we can explicitly compute S^l by directly differentiating the expressions given above. From simple calculations, we have $\mathscr{S}^l=0$ for $l<0$ and

$$
\begin{aligned}
\mathscr{S}^0&=\begin{pmatrix} C & 2^{-1}C(\partial_x H)^\star\\ 2^{-1}(\partial_x H)C & 3^{-1}(\partial_x H)C(\partial_x H)^\star\end{pmatrix},\\
S^0&=\begin{pmatrix} C & 2^{-1}C(\partial_x H)^\star\\ 2^{-1}(\partial_x H)C & 3^{-1}(\partial_x H)C(\partial_x H)^\star\end{pmatrix}
\end{aligned}
$$

so from [A5], we have $\inf_{z,\theta_1,\theta_3}\det S^0(z,\theta_1,\theta_3)>0$. Also,

$$
(S^0)^{-1}=\begin{pmatrix} C^{-1}+3(\partial_x H)^\star V^{-1}\partial_x H & -6(\partial_x H)^\star V^{-1}\\ -6V^{-1}\partial_x H & 12V^{-1}\end{pmatrix}
$$

where $V=(\partial_x H)C(\partial_x H)^\star$. For $k\ge 1$, we define the quasi-likelihood functions corresponding to $\theta_1,\theta_2,\theta_3$ as follows:

$$
\begin{aligned}
V_n^{1',k}(\theta_1|\bar\theta):=&-\frac{1}{2n}\sum_{j=1}^{n}C_{n,j-1}^{-1}(\theta_1)\Big[(\mathcal{D}_{n,j}^{k-1,x})^{\otimes 2}(\bar\theta)-\sum_{l=1}^{k-1}h_n^l S_{n,j-1}^{l,xx}(\bar\theta)\Big]\\
&-\frac{1}{2n}\sum_{j=1}^{n}\log\det C_{n,j-1}(\theta_1),
\end{aligned}
$$

$$
\begin{aligned}
V_n^{1,k}(\theta_1|\bar{\theta}) :=& -\frac{1}{2n}\sum_{j=1}^{n}(S^0)^{-1}_{n,j-1}(\theta_1,\bar{\theta}_3)\Big[(\mathcal{D}^{k-1}_{n,j})^{\otimes 2}(\bar{\theta}) - \sum_{l=1}^{k-1} h_n^l S^l_{n,j-1}(\bar{\theta})\Big] \\
& -\frac{1}{2n}\sum_{j=1}^{n}\log\det S^0_{n,j-1}(\theta_1,\bar{\theta}_3), \\
V_n^{2,k}(\theta_2|\bar{\theta}) :=& -\frac{1}{2nh}\sum_{j=1}^{n} C^{-1}_{n,j-1}(\bar{\theta}_1)[(\mathcal{D}^{k,x})^{\otimes 2}_{n,j}(\theta_2,\bar{\theta})], \\
V_n^{3,k}(\theta_3|\bar{\theta}) :=& -\frac{h_n}{2n}\sum_{j} 12V^{-1}_{n,j-1}(\bar{\theta}_1,\bar{\theta}_3)[(\mathcal{D}^{k,y})^{\otimes 2}_{j}(\theta_3,\bar{\theta})] \\
& +\frac{h_n}{2n}\sum_{j=1}^{n} 12((\partial_x H)^\star V^{-1})_{n,j-1}(\bar{\theta}_1,\bar{\theta}_3)[\mathcal{D}^{k,x}_{n,j}(\bar{\theta}_2,\bar{\theta})\otimes\mathcal{D}^{k,y}_{n,j}(\theta_3,\bar{\theta})].
\end{aligned}
$$

We denote by (i',i) an element of the set $\{(1',1),(1,1),(2,2),(3,3)\}$. Let

$$
\begin{aligned}
a_n :=& \operatorname{diag}[a_n^1 I_{p_1}, a_n^2 I_{p_2}, a_n^3 I_{p_3}] \\
:=& \operatorname{diag}\Big[\frac{1}{n^{1/2}}I_{p_1}, \frac{1}{(nh_n)^{1/2}}I_{p_2}, \frac{h_n^{1/2}}{n^{1/2}}I_{p_3}\Big],
\end{aligned}
$$

where I_{p_1}, I_{p_2} and I_{p_3} are identity matrices of dimensions p_1, p_2 and p_3, respectively. Fix some continuous prior distribution π_i for θ_i satisfying $0<\inf_{\theta_i\in\Theta_i}\pi_i(\theta_i)\le\sup_{\theta_i\in\Theta_i}\pi_i(\theta_i)<\infty$. For $k\ge 1$ and a Θ-valued measurable function $\hat{\theta}_0=(\hat{\theta}_{0,1},\hat{\theta}_{0,2},\hat{\theta}_{0,3})$, let $\hat{\theta}^{\mathrm{M},k}_{i',n}(\hat{\theta}_0)$ be a Θ_i-valued measurable function such that if $\operatorname{argmax}_{\theta_i\in\Theta_i} V_n^{i',k}(\theta_i|\hat{\theta}_0)$ is not empty, $\hat{\theta}^{\mathrm{M},k}_{i',n}(\hat{\theta}_0)\in\operatorname{argmax}_{\theta_i\in\Theta_i} V_n^{i',k}(\theta_i|\hat{\theta}_0)$ holds. Furthermore, we define two Θ_i-valued measurable functions as

$$
\begin{aligned}
\hat{\theta}^{\mathrm{B},k}_{i',n}(\hat{\theta}_0) :=& \frac{\int_{\Theta_i}\theta_i\exp((a_n^i+h_n^{k+\delta_{i,3}})^{-2}V_n^{i',k}(\theta_i|\hat{\theta}_0))\pi_i(\theta_i)d\theta_i}{\int_{\Theta_i}\exp((a_n^i+h_n^{k+\delta_{i,3}})^{-2}V_n^{i',k}(\theta_i|\hat{\theta}_0))\pi_i(\theta_i)d\theta_i}, \\
\hat{\theta}^{\mathrm{S},k}_{i',n}(\hat{\theta}_0) :=& \begin{cases}\hat{\theta}_{0,i}-(\partial^2_{\theta_i}V_n^{i',k}(\hat{\theta}_{0,i}|\hat{\theta}_0))^{-1}[\partial_{\theta_i}V_n^{i',k}(\hat{\theta}_{0,i}|\hat{\theta}_0)] & (\in\Theta_i) \\ \hat{\theta}_{0,i} & (\text{otherwise})\end{cases},
\end{aligned}
$$

where $\delta_{i,3}$ is the Kronecker delta.

Remark 3.1. If $h_n^{k+\delta_{i,3}}/a_n^i\to 0$ as $n\to\infty$ (which holds for $k\ge k_0$), then $a_n^i+h_n^{k+\delta_{i,3}}$ in the definition of $\hat{\theta}^{\mathrm{B},k}_{i',n}(\hat{\theta}_0)$ can be replaced by a_n^i without affecting the validity of Theorem 3.2. In this case, the estimator coincides with the standard form of the Bayes estimator.

Let $\hat{\theta}^{0,0}_n:=(\hat{\theta}^{0,0}_{1,n},\hat{\theta}^{0,0}_{2,n},\hat{\theta}^{0,0}_{3,n})\in\Theta$ be any measurable functions (can be constant) and $\mathrm{A}=(\mathrm{A}^1_1,\mathrm{A}^1_2,\mathrm{A}^1_3,\mathrm{A}^2_1,\mathrm{A}^2_2,\mathrm{A}^2_3,\cdots)\in\{\mathrm{M},\mathrm{B}\}^3\times\{\mathrm{M},\mathrm{B},\mathrm{S}\}^{\mathbb{N}}$. For any integer $k\ge 1$, we construct an estimators $\hat{\theta}^k_n=(\hat{\theta}^k_{1,n},\hat{\theta}^k_{2,n},\hat{\theta}^k_{3,n})$ using the following algorithm:

1. For $k\ge 1$, define the following estimators inductively:

$$
\begin{aligned}
\hat{\theta}^{k,0}_{1,n} :=& \hat{\theta}^{\mathrm{A}^k_1,k}_{1',n}(\hat{\theta}^{k-1,0}_{1,n},\hat{\theta}^{k-1,0}_{2,n},\hat{\theta}^{k-1,0}_{3,n}), \\
\hat{\theta}^{k,0}_{2,n} :=& \hat{\theta}^{\mathrm{A}^k_2,k}_{2,n}(\hat{\theta}^{k,0}_{1,n},\hat{\theta}^{k-1,0}_{2,n},\hat{\theta}^{k-1,0}_{3,n}), \\
\hat{\theta}^{k,0}_{3,n} :=& \hat{\theta}^{\mathrm{A}^k_3,k}_{3,n}(\hat{\theta}^{k,0}_{1,n},\hat{\theta}^{k,0}_{2,n},\hat{\theta}^{k-1,0}_{3,n}), \\
\hat{\theta}^{k,0}_{n} :=& (\hat{\theta}^{k,0}_{1,n},\hat{\theta}^{k,0}_{2,n},\hat{\theta}^{k,0}_{3,n}).
\end{aligned}
$$

2. Let

$$\begin{aligned}
\hat{\theta}_{1,n}^{k} &:= \hat{\theta}_{1,n}^{\mathrm{A}_1^{k+1},k+1}(\hat{\theta}_{1,n}^{k,0},\hat{\theta}_{2,n}^{k,0},\hat{\theta}_{3,n}^{k,0}),\\
\hat{\theta}_{2,n}^{k} &:= \hat{\theta}_{2,n}^{k,0},\\
\hat{\theta}_{3,n}^{k} &:= \begin{cases}\hat{\theta}_{3,n}^{k,0} & k\geq 2\\ \hat{\theta}_{3,n}^{\mathrm{A}_3^2,1}(\hat{\theta}_{1,n}^{1,0},\hat{\theta}_{2,n}^{1,0},\hat{\theta}_{3,n}^{1,0}) & k=1\end{cases},\\
\hat{\theta}_{n}^{k} &:= (\hat{\theta}_{1,n}^{k},\hat{\theta}_{2,n}^{k},\hat{\theta}_{3,n}^{k}).
\end{aligned}$$

Let

$$\begin{aligned}
\Gamma^{1'}(\theta_1) &:= \frac{1}{2}\int \mathrm{Tr}\Big(C^{-1}(\partial_{\theta_1}C)C^{-1}(\partial_{\theta_1}C)(z,\theta_1)\Big)\nu^*(dz),\\
\Gamma^{1}(\theta_1|\bar{\theta}_3) &:= \Gamma^{1'}(\theta_1) + \frac{1}{2}\int \mathrm{Tr}\Big(V^{-1}(\partial_{\theta_1}V)V^{-1}(\partial_{\theta_1}V)(z,\theta_1,\bar{\theta}_3)\Big)\nu^*(dz),\\
\Gamma^{2}(\theta_2|\bar{\theta}_1) &:= \int C^{-1}(z,\bar{\theta}_1)[(\partial_{\theta_2}A)^{\otimes 2}(z,\theta_2)]\nu^*(dz),\\
\Gamma^{3}(\theta_3|\bar{\theta}_1,\bar{\theta}_3) &:= \int 12V^{-1}(z,\bar{\theta}_1,\bar{\theta}_3)[(\partial_{\theta_3}H)^{\otimes 2}(z,\theta_3)]\nu^*(dz)
\end{aligned}$$

and $\Gamma := \mathrm{diag}[\Gamma^1(\theta_1^*|\theta_3^*),\Gamma^2(\theta_2^*|\theta_1^*),\Gamma^3(\theta_3^*|\theta_1^*,\theta_3^*)]$. We denote the expectation with respect to $N(0,\Gamma^{-1})$ by $\mathbb{E}$.

Theorem 3.2. The convergence

$$E[f(a_n^{-1}(\hat{\theta}_n^{k_0}-\theta^*))] \to \mathbb{E}[f]$$

holds as n tends to ∞ for all $f\in C_{\uparrow}$.

The proof of Theorem 3.2 will be given in Section 6.

Remark 3.3. If $\mathrm{A}_3^k \in \{\mathrm{M},\mathrm{B}\}$ for any $k\geq 1$, the regularity assumption on H in [A4] can be relaxed to $H\in C_{\uparrow}^{p+1,2}(\mathbb{R}^{d_z}\times\Theta_3;\mathbb{R}^{d_y})$.

4 Simulation

We investigate the performance of adaptive quasi-maximum likelihood estimators through simulation studies of two models: a linear model and the FitzHugh-Nagumo model. For each model, we constructed estimators $\hat{\theta}_n^{k_0}$ by setting $\mathrm{A}=\mathrm{M}$ at each estimation step for various values of k_0. We computed these estimators in closed form and conducted 400 iterations to obtain sample means and standard deviations for each parameter.

4.1 Linear model

First, we consider the following linear model:

$$\begin{aligned}
dX_t &= (-\theta_{2,1}X_t-\theta_{2,2}Y_t)dt+\theta_1 dw_t,\\
dY_t &= \theta_3 X_t dt.
\end{aligned}$$

We fix the true values as $(\theta_1,\theta_{2,1},\theta_{2,2},\theta_3)=(1.0,1.0,1.0,1.0)$ and simulate the model with different time step sizes $h_n=0.1,0.2$ while keeping the total time interval nh_n fixed at

100. The simulation results are summarized in the following tables, where the entries show the sample means of the estimators with their sample standard deviations in parentheses.

	$\hat{\theta}_1$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$	$\hat{\theta}_3$
True value	1.0	1.0	1.0	1.0
$k_0 = 1$	0.975 (1.58e-2)	1.014 (1.39e-1)	0.979 (1.35e-1)	0.999 (4.50e-3)
$k_0 = 2$	0.999 (1.66e-2)	1.017 (1.53e-1)	1.029 (1.42e-1)	1.001 (4.51e-3)
$k_0 = 3$	0.999 (1.66e-2)	1.017 (1.54e-1)	1.032 (1.42e-1)	1.001 (4.52e-3)
$k_0 = 4$	0.999 (1.66e-2)	1.017 (1.55e-1)	1.032 (1.42e-1)	1.001 (4.52e-3)
$k_0 = 5$	0.999 (1.66e-2)	1.017 (1.55e-1)	1.032 (1.42e-1)	1.001 (4.52e-3)

Table 1: $nh_n = 100, h_n = 0.1$, Linear model.

	$\hat{\theta}_1$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$	$\hat{\theta}_3$
True value	1.0	1.0	1.0	1.0
$k_0 = 1$	0.954 (2.21e-2)	1.009 (1.31e-1)	0.928 (1.29e-1)	0.995 (9.04e-3)
$k_0 = 2$	0.998 (2.41e-2)	1.020 (1.59e-1)	1.021 (1.43e-1)	1.000 (9.08e-3)
$k_0 = 3$	0.998 (2.41e-2)	1.020 (1.63e-1)	1.032 (1.44e-1)	1.001 (9.12e-3)
$k_0 = 4$	0.998 (2.41e-2)	1.018 (1.64e-1)	1.032 (1.44e-1)	1.001 (9.12e-3)
$k_0 = 5$	0.998 (2.41e-2)	1.018 (1.64e-1)	1.032 (1.44e-1)	1.001 (9.12e-3)

Table 2: $nh_n = 100, h_n = 0.2$, Linear model.

	$\hat{\theta}_1$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$	$\hat{\theta}_3$
True value	1.0	1.0	1.0	1.0
$k_0 = 1$	0.915 (3.81e-2)	0.973 (1.09e-1)	0.774 (1.08e-1)	0.964 (2.64e-2)
$k_0 = 2$	0.995 (4.05e-2)	1.026 (1.65e-1)	0.963 (1.37e-1)	0.989 (2.72e-2)
$k_0 = 3$	0.998 (4.10e-2)	1.038 (1.83e-1)	1.017 (1.48e-1)	0.999 (2.83e-2)
$k_0 = 4$	0.998 (4.10e-2)	1.029 (1.90e-1)	1.031 (1.50e-1)	1.001 (2.87e-2)
$k_0 = 5$	0.997 (4.09e-2)	1.023 (1.92e-1)	1.034 (1.51e-1)	1.001 (2.88e-2)
$k_0 = 6$	0.996 (4.09e-2)	1.019 (1.92e-1)	1.033 (1.50e-1)	1.001 (2.88e-2)
$k_0 = 7$	0.996 (4.09e-2)	1.018 (1.92e-1)	1.033 (1.50e-1)	1.001 (2.88e-2)
$k_0 = 8$	0.996 (4.09e-2)	1.017 (1.92e-1)	1.032 (1.50e-1)	1.001 (2.88e-2)
$k_0 = 9$	0.996 (4.09e-2)	1.017 (1.91e-1)	1.032 (1.50e-1)	1.001 (2.88e-2)
$k_0 = 10$	0.996 (4.09e-2)	1.017 (1.91e-1)	1.032 (1.50e-1)	1.001 (2.88e-2)

Table 3: $nh_n = 100, h_n = 0.5$, Linear model.

	$\hat{\theta}_1$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$	$\hat{\theta}_3$
True value	1.0	1.0	1.0	1.0
$k_0 = 1$	0.935 (7.47e-2)	0.899 (8.67e-2)	0.589 (8.44e-2)	0.885 (5.76e-2)
$k_0 = 2$	1.011 (6.71e-2)	1.032 (1.62e-1)	0.828 (1.22e-1)	0.950 (6.20e-2)
$k_0 = 3$	1.019 (7.03e-2)	1.096 (1.94e-1)	0.941 (1.50e-1)	0.994 (7.02e-2)
$k_0 = 4$	1.017 (6.54e-2)	1.092 (2.21e-1)	0.999 (1.61e-1)	1.006 (7.35e-2)
$k_0 = 5$	1.010 (6.36e-2)	1.079 (2.40e-1)	1.034 (1.72e-1)	1.010 (7.61e-2)
$k_0 = 6$	1.006 (6.38e-2)	1.062 (2.50e-1)	1.050 (1.78e-1)	1.011 (7.77e-2)
$k_0 = 7$	1.004 (6.38e-2)	1.048 (2.54e-1)	1.055 (1.80e-1)	1.010 (7.85e-2)
$k_0 = 8$	1.003 (6.38e-2)	1.036 (2.55e-1)	1.054 (1.81e-1)	1.008 (7.88e-2)
$k_0 = 9$	1.002 (6.39e-2)	1.029 (2.54e-1)	1.052 (1.80e-1)	1.007 (7.88e-2)
$k_0 = 10$	1.002 (6.39e-2)	1.024 (2.51e-1)	1.048 (1.78e-1)	1.006 (7.87e-2)

Table 4: $nh_n = 100, h_n = 0.9$, Linear model.

The simulation results for the linear model demonstrate several key findings. For small time steps ($h_n = 0.1, 0.2$), the estimators show rapid convergence when $k_0 \geq 2$, with parameter estimates closely approaching their true values. As the time step increases ($h_n = 0.5, 0.9$), larger values of k_0 are required to achieve accurate estimates. Recall $k_0 = \lfloor p/2 \rfloor$, which shows these results are consistent with our theoretical findings. For the linear model, we observe that even under challenging conditions with $h_n = 0.9$, increasing k_0 leads to highly accurate estimates. The diffusion coefficient θ_1 shows remarkable stability across different time steps, while the drift parameters θ_{21} and θ_{22} exhibit higher variances. For small time steps, the estimates converge after $k_0 = 3$ or 4, whereas larger time steps require k_0 values up to 7 or 8 for the estimates to reach steady values.

4.2 FitzHugh-Nagumo model

Next, we consider the FitzHugh-Nagumo model, a well-known non-linear system used to model the behaviour of neurons. The model is defined by the following stochastic differential equations:

$$
\begin{aligned}
dX_t &= (\theta_{2,1} Y_t - X_t + \theta_{2,2})dt + \theta_1 dw_t, \\
dY_t &= \frac{1}{\theta_{3,1}}(Y_t - Y_t^3 - X_t + \theta_{3,2})dt.
\end{aligned}
$$

We set the true parameter values as $(\theta_1, \theta_{2,1}, \theta_{2,2}, \theta_{3,1}, \theta_{3,2}) = (0.3, 1.5, 0.8, 0.1, 0)$ and simulate the model with different total time interval nh_n and time step sizes h_n. The simulation results are summarized in the following tables, where the entries show the sample means of the estimators with their sample standard deviations in parentheses.

		$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\epsilon}$	$\hat{s}$
	True value	0.3	1.5	0.8	0.1	0
$k_0 = 1$	Mean (std)	0.331 (4.83e-3)	1.462 (5.51e-2)	0.777 (4.45e-2)	0.104 (7.52e-4)	1.29e-3 (4.35e-4)
$k_0 = 2$	Mean (std)	0.308 (4.07e-3)	1.496 (5.65e-2)	0.796 (4.58e-2)	0.101 (8.00e-4)	9.55e-4 (4.33e-4)
$k_0 = 3$	Mean (std)	0.300 (3.74e-3)	1.504 (5.71e-2)	0.803 (4.63e-2)	0.100 (8.99e-4)	2.45e-4 (4.31e-4)
$k_0 = 4$	Mean (std)	0.302 (3.80e-3)	1.504 (5.72e-2)	0.803 (4.64e-2)	0.100 (9.04e-4)	1.02e-4 (4.31e-4)
$k_0 = 5$	Mean (std)	0.301 (3.78e-3)	1.504 (5.72e-2)	0.803 (4.64e-2)	0.100 (9.06e-4)	8.08e-5 (4.31e-4)

Table 5: $nh_n = 100, h_n = 0.05$, FHN model.

		$\hat{\sigma}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\epsilon}$	$\hat{s}$
	True value	0.3	1.5	0.8	0.1	0
$k_0 = 1$	Mean (std)	0.334 (1.74e-2)	1.556 (0.348)	0.850 (0.300)	0.105 (3.42e-3)	1.33e-3 (3.82e-3)
$k_0 = 2$	Mean (std)	0.312 (1.40e-2)	1.595 (0.362)	0.874 (0.312)	0.102 (4.08e-3)	1.08e-3 (1.83e-3)
$k_0 = 3$	Mean (std)	0.304 (1.38e-2)	1.606 (0.369)	0.883 (0.318)	0.100 (4.05e-3)	3.53e-4 (1.44e-3)
$k_0 = 4$	Mean (std)	0.306 (1.37e-2)	1.606 (0.369)	0.883 (0.318)	0.101 (3.97e-3)	1.99e-4 (1.49e-3)
$k_0 = 5$	Mean (std)	0.305 (1.36e-2)	1.606 (0.369)	0.883 (0.319)	0.101 (3.96e-3)	1.71e-4 (1.45e-3)

Table 6: $nh_n = 10, h_n = 0.05$, FHN model.

The FitzHugh-Nagumo model simulations reveal interesting patterns in the estimation performance. In almost all cases, the estimates for $nh_n = 100$ have means closer to the true values and lower standard deviations compared to those for $nh_n = 10$. Firstly, the estimates of the degenerate drift coefficients ϵ and s have almost no bias and very small standard deviations, which is consistent with the theoretically fastest convergence rate. Furthermore, the diffusion parameter σ can also be estimated with nearly no bias and monotonically decreasing standard deviations with respect to k_0. However, for the non-degenerate drift parameters γ and β, there is considerable bias when $nh_n = 10$, suggesting that asymptotic expansion corrections could potentially make the performance better.

5 Modified version of quasi-likelihood analysis

Definition 5.1. We promise $0/0 = 0$ in this definition. For a family of random variables $\Xi_{\lambda_1,\cdots,\lambda_m}$ and non-random $\alpha_{\lambda_1,\cdots,\lambda_m} \geq 0$ dependent on indices $\lambda_1, \cdots, \lambda_{m_0}, \cdots, \lambda_m$, we write $\Xi_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{\lambda_1,\cdots,\lambda_{m_0}}(\alpha_{\lambda_1,\cdots,\lambda_m})$ if $\sup_{\lambda_{m_0+1},\cdots,\lambda_m} \frac{|\Xi_{\lambda_1,\cdots,\lambda_m}|}{\alpha_{\lambda_1,\cdots,\lambda_m}}$ is measurable and for any $M > 0$,

$$\sup_{\lambda_1,\cdots,\lambda_{m_0}} E\left[\sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left(\frac{|\Xi_{\lambda_1,\cdots,\lambda_m}|}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right)^M\right] < \infty. \tag{5}$$

In particular we write $\Xi_{\lambda_1,\cdots,\lambda_m} = O_{\lambda_1,\cdots,\lambda_m}(\alpha_{\lambda_1,\cdots,\lambda_m})$ if for any $M > 0$,

$$\sup_{\lambda_1,\cdots,\lambda_m} E\left[\left(\frac{|\Xi_{\lambda_1,\cdots,\lambda_m}|}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right)^M\right] < \infty$$

and $\Xi_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_1,\cdots,\lambda_m}(\alpha_{\lambda_1,\cdots,\lambda_m})$ if for any $M > 0$,

$$E\left[\sup_{\lambda_1,\cdots,\lambda_m}\left(\frac{|\Xi_{\lambda_1,\cdots,\lambda_m}|}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right)^M\right] < \infty.$$

The proof of Theorem 3.2 relies on quasi-likelihood analysis. We present a modified version of the framework. Fix $d \in \mathbb{N}$ and consider an open set $\Theta_\circ \subset \mathbb{R}^d$ with a fixed point $\theta^*_\circ \in \Theta_\circ$ and a C^2-class random function $\mathbb{Y}_n(\theta_\circ)$ satisfying $\mathbb{Y}_n(\theta^*_\circ) = 0$. Let $\hat{\theta}^{\mathrm{M}}_{\circ,n}(\hat{\theta}_0)$ be a $\Theta_\circ$-valued measurable function such that if $\mathrm{argmax}_{\theta_\circ\in\Theta_\circ} \mathbb{Y}_n(\theta_\circ)$ is non-empty, then $\hat{\theta}^{\mathrm{M}}_{\circ,n} \in \mathrm{argmax}_{\theta_\circ\in\Theta_\circ} \mathbb{Y}_n(\theta_\circ)$ holds. Assume π is a continuous prior distribution on $\Theta_\circ$ satisfying $0 < \inf_{\theta_\circ\in\Theta_\circ} \pi(\theta_\circ) \leq \sup_{\theta_\circ\in\Theta_\circ} \pi(\theta_\circ) < \infty$. Define

$$\hat{\theta}^{\mathrm{B}}_{\circ,n} := \frac{\int_{\Theta_\circ} \theta_\circ \exp((a^\circ_n)^{-2}\mathbb{Y}_n(\theta_\circ))\pi(\theta_\circ)d\theta_\circ}{\int_{\Theta_\circ} \exp((a^\circ_n)^{-2}\mathbb{Y}_n(\theta_\circ))\pi(\theta_\circ)d\theta_\circ}.$$

Given a $\Theta_\circ$-valued measurable function $\hat{\theta}^0_{\circ,n}$, set

$$\hat{\theta}^{\mathrm{S}}_{\circ,n} := \begin{cases} \hat{\theta}^0_{\circ,n} - (\partial^2_{\theta_\circ}\mathbb{Y}_n(\hat{\theta}^0_{\circ,n}))^{-1}[\partial_{\theta_\circ}\mathbb{Y}_n(\hat{\theta}^0_{\circ,n})] & (\in \Theta_\circ), \\ \hat{\theta}^0_{\circ,n} & (\text{otherwise}). \end{cases}$$

Assume a°_n and b°_n are sequences of positive real numbers satisfying $a^\circ_n \to 0$ as $n \to \infty$, Γ_n is a random $d \times d$ symmetric matrix, and $\mathbb{Y}(\theta_\circ)$ is a deterministic continuous function on $\Theta_\circ$. Since the following theorem essentially follows from Yoshida [2011], we omit the proof.

Theorem 5.2. Suppose there exists $\epsilon > 0$ such that the following conditions hold:

1. $\mathbb{Y}(\theta_\circ) \leq -\epsilon|\theta_\circ - \theta^*_\circ|^2$
2. $\mathbb{Y}_n(\theta_\circ) = \mathbb{Y}(\theta_\circ) + O^{\theta_\circ}_n((a^\circ_n)^{3\epsilon})$
3. $\partial_{\theta_\circ}\mathbb{Y}_n(\theta^*_\circ) = O_n(a^\circ_n)$
4. $\lambda_{\min}(\Gamma_n) \geq \epsilon$ a.s.
5. $\partial^2_{\theta_\circ}\mathbb{Y}_n(\theta_\circ) = \partial^2_{\theta_\circ}\mathbb{Y}_n(\theta^*_\circ) + O^\circ_n(b^\circ_n + |\theta_\circ - \theta^*_\circ|)$
   $= -\Gamma_n + O^{\theta_\circ}_n((a^\circ_n)^\epsilon + |\theta_\circ - \theta^*_\circ|)$

6. $|\hat{\theta}^0_{\circ,n} - \theta^*_\circ|^2 + b^\circ_n|\hat{\theta} - \theta^*_\circ| = O_n(a^\circ_n)$

Then for each $\mathrm{A} \in \{\mathrm{M}, \mathrm{B}, \mathrm{S}\}$, we have:

$$\hat{\theta}^{\mathrm{A}}_{\circ,n} - \theta^*_\circ = O_n(a^\circ_n),$$
$$(a^\circ_n)^{-1}(\hat{\theta}^{\mathrm{A}}_{\circ,n} - \theta^*_\circ) - \Gamma_n^{-1}(a^\circ_n)^{-1}\partial_{\theta_\circ}\mathbb{Y}_n(\theta^*_\circ) \xrightarrow{P} 0.$$

6 The proof of Theorem 3.2

Let

$$\begin{aligned} M_n^{1'} :=\ & -\frac{1}{2n^{1/2}}\sum_{j=1}^n \partial_{\theta_1}(C^{-1})_{n,j-1}(\theta^*_1) \\ & \cdot [(h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]))^{\otimes 2} - E_{n,j-1}[(h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]))^{\otimes 2}]], \end{aligned}$$

$$\begin{aligned} M_n^1 :=\ & -\frac{1}{2n^{1/2}}\sum_{j=1}^n (\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta^*_1, \theta^*_3) \\ & \cdot \left[\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right]\right], \end{aligned}$$

$$\begin{aligned} M_n^2 :=\ & \frac{1}{n^{1/2}}\sum_{j=1}^n C^{-1}_{n,j-1}(\theta^*_1)[\partial_{\theta_2}A_{n,j-1}(\theta^*_2) \odot h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}])], \\ M_n^3 :=\ & \frac{1}{n^{1/2}}\sum_{j=1}^n 12V^{-1}_{n,j-1}(\theta^*_1, \theta^*_3)[\partial_{\theta_3}H_{n,j-1}(\theta^*_3) \odot h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}])] \\ & - \frac{1}{n^{1/2}}\sum_{j=1}^n 6(\partial_x H)^\star V^{-1}_{n,j-1}(\theta^*_1, \theta^*_3)[h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \otimes \partial_{\theta_3}H_{n,j}(\theta^*_3)] \end{aligned}$$

and let $M_n := (M_n^1, M_n^2, M_n^3)$. The following proposition is a central limit theorem and its proof is given in Section 8.

Propositon 6.1. The convergence

$$M_n \xrightarrow{d} N(0, \Gamma)$$

holds as n tends to ∞.

For $(i', i) \in \{(1', 1), (1, 1), (2, 2), (3, 3)\}$ and $k \geq 1$, let

$$\mathbb{Y}_n^{i',k}(\theta_i|\bar{\theta}) := V_n^{i',k}(\theta_i \backslash \theta^*_i|\bar{\theta}).$$

and we write $\delta_i := |\theta_i - \theta^*_i|, \bar{\delta}_i = |\bar{\theta}_i - \theta^*_i|$ and $\bar{\delta} = |\bar{\theta} - \theta^*|$. The proof of the following proposition is given in Section 9.

Propositon 6.2. For $k = 1, \cdots, k_0$, there exists an $\epsilon > 0$ such that the following hold:

1. (a) $\mathbb{Y}_n^{1',k}(\theta_1|\bar{\theta}) = \mathbb{Y}^{1'}(\theta_1) + O_n^{\theta_1,\bar{\theta}}(h_n^{\epsilon})$

   (b) $$\partial_{\theta_1}\mathbb{Y}_n^{1',k}(\theta_1^*|\bar{\theta}) = n^{-1/2}M_n^{1'} + O_n^{\bar{\theta}}\Big(\Big(h_n + \frac{1}{n^{1/2}}\Big)\bar{\delta} + h_n^{(p/2)\wedge k}\Big) = O_n^{\bar{\theta}}\Big(h_n\bar{\delta} + h_n^{(p/2)\wedge k} + \frac{1}{n^{1/2}}\Big)$$

   (c) $$\partial_{\theta_1}^2\mathbb{Y}_n^{1',k}(\theta_1|\bar{\theta}) = \partial_{\theta_1}^2\mathbb{Y}_n^{1',k}(\theta_1^*|\bar{\theta}) + O_n^{\theta_1,\bar{\theta}}\Big(h_n^{(p/2)\wedge k} + h_n\bar{\delta} + \delta_1 + \frac{1}{n^{1/2}}\Big) = -\Gamma^{1'}(\theta_1^*) + O_n^{\theta_1,\bar{\theta}}(h_n^{\epsilon} + \bar{\delta}_1)$$

2. (a) $\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta}) = \mathbb{Y}^1(\theta_1) + O_n^{\theta_1,\bar{\theta}}(h_n^{-1}\bar{\delta}_3^2 + \bar{\delta}_3 + h_n^{\epsilon})$

   (b) $$\partial_{\theta_1}\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta}) = n^{-1/2}M_n^1 + O_n^{\bar{\theta}}\Big(h_n^{-1}\bar{\delta}_3^2 + \bar{\delta}_3 + h_n\bar{\delta} + h_n^{(p/2)\wedge k} + \frac{h^{1/2}}{n^{1/2}}\bar{\delta}\Big) = O_n^{\bar{\theta}}\Big(h_n^{-1}\bar{\delta}_3^2 + \bar{\delta}_3 + h_n\bar{\delta} + h_n^{(p/2)\wedge k} + \frac{1}{n^{1/2}}\Big)$$

   (c) $$\partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta}) = \partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta}) + O_{n,j}^{\theta_1,\bar{\theta}}\Big(h_n^{(p/2)\wedge k} + h_n^{-1}\bar{\delta}_3^2 + \delta_1 + \bar{\delta}_3 + h_n\bar{\delta} + \frac{1}{n^{1/2}}\Big) = \Gamma^1(\theta_1^*|\bar{\theta}_3) + O_n^{\theta_1,\bar{\theta}}\Big(h_n^{\epsilon} + \delta_1 + \bar{\delta}_3 + h_n^{-1}\bar{\delta}_3\Big)$$

3. (a) $\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}) = \mathbb{Y}^2(\theta_2) + O_n^{\theta_2,\bar{\theta}}(\bar{\delta}_1 + h_n^{\epsilon})$

   (b) $$\partial_{\theta_2}\mathbb{Y}_n^{2,k}(\theta_2^*|\bar{\theta}) = \frac{1}{(nh_n)^{1/2}}M_n^2 + O_n^{\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{\bar{\delta}_1}{(nh_n)^{1/2}}\Big) = O_n^{\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{1}{(nh_n)^{1/2}}\Big)$$

   (c) $$\partial_{\theta_2}^2\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}) = \partial_{\theta_2}^2\mathbb{Y}_n^{2,k}(\theta_2^*|\bar{\theta}) + O_n^{\theta_2,\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{1}{(nh_n)^{1/2}} + \delta_2\Big) = \Gamma^2(\theta_2^*|\bar{\theta}) + O_n^{\theta_2,\bar{\theta}}(h_n^{\epsilon} + \delta_2)$$

4. (a) $\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta}) = \mathbb{Y}^3(\theta_3|\bar{\theta}_3) + O_n^{\theta_3,\bar{\theta}}(\bar{\delta}_1 + h_n^{\epsilon})$

   (b) $$\partial_{\theta_3}\mathbb{Y}_n^{3,k}(\theta_3^*|\bar{\theta}) = \frac{h_n^{1/2}}{n^{1/2}}M_n^3 + O_n^{\bar{\theta}}\Big(h_n(\bar{\delta}_2 + \bar{\delta}1_{\{k\geq 2\}}) + h_n^{((p+1)/2)\wedge(k+3/2)} + \frac{h_n^{1/2}}{n^{1/2}}\bar{\delta}\Big) = O_n^{\bar{\theta}}\Big(h_n(\bar{\delta}_2 + \bar{\delta}1_{\{k\geq 2\}}) + h_n^{((p+1)/2)\wedge(k+3/2)} + \frac{h_n^{1/2}}{n^{1/2}}\Big)$$

   (c) $$\partial_{\theta_3}^2\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta}) = \partial_{\theta_3}^2\mathbb{Y}_n^{3,k}(\theta_3^*|\bar{\theta}) + O_n^{\theta_3,\bar{\theta}}\Big(\delta_3 + h_n\bar{\delta} + h_n^{(p+1)/2\wedge(k+1)} + \frac{h_n^{1/2}}{n^{1/2}}\Big) = -\Gamma^3(\theta_3^*|\bar{\theta}_1,\bar{\theta}_3) + O_n^{\theta_3,\bar{\theta}}(h_n^{\epsilon} + \delta_3)$$

Proof of Theorem 3.2. For integers $k_1, k_2, k_3 \geq 0$, only in this proof, we denote$(\hat{\theta}_{1,n}^{k_1,0}, \hat{\theta}_{2,n}^{k_2,0}, \hat{\theta}_{3,n}^{k_3,0})$ by $\hat{\theta}_n^{(k_1,k_2,k_3),0}$. For each $k = 1, 2, \cdots, k_0$, we can apply Theorem 5.2 for

1. $\Theta_\circ = \Theta_1, \theta_\circ^* = \theta_1^*, \mathbb{Y}_n(\theta_1) = \mathbb{Y}_n^{1',k}(\theta_1|\hat{\theta}_n^{(k-1,k-1,k-1),0}), \mathbb{Y}(\theta_1) = \mathbb{Y}^1(\theta_1), \Gamma_n = \Gamma^{1'}(\theta_1^*), a_n^\circ = a_n^1 + h_n^{(p/2)\wedge k}, b_n^\circ = h_n + n^{-1/2}$,

2. $\Theta_\circ = \Theta_2, \theta_\circ^* = \theta_2^*, \mathbb{Y}_n(\theta_2) = \mathbb{Y}_n^{2,k}(\theta_2|\hat{\theta}_n^{(k,k-1,k-1),0}), \mathbb{Y}(\theta_2) = \mathbb{Y}^2(\theta_2), \Gamma'_n(\theta_2) = \Gamma^2(\theta_2|\hat{\theta}_{1,n}^{k,0}), a_n^\circ = a_n^2 + h_n^{((p-1)/2)\wedge k}, b_n^\circ = h_n^{(p-1)/2\wedge 1} + (nh_n)^{-1/2}$,

3. $\Theta_\circ = \Theta_3, \theta_\circ^* = \theta_3^*, \mathbb{Y}_n(\theta_3) = \mathbb{Y}_n^{3,k}(\theta_3|\hat{\theta}_n^{(k,k,k-1),0}), \mathbb{Y}(\theta_3) = \mathbb{Y}^3(\theta_3), \Gamma'_n(\theta_3) = \Gamma^3(\theta_3|\hat{\theta}_{1,n}^{k,0}, \hat{\theta}_{3,n}^{k-1,0}), a_n^\circ = a_n^3 + h_n^{((p+1)/2)\wedge(k+1)}, b_n^\circ = h_n\bar{\delta} + h_n^{1/2}n^{-1/2}$,

and as a result, we have

[C1] $\hat{\theta}_{1,n}^{k,0} - \theta_1^* = O_n(a_n^1 + h_n^{(p/2)\wedge k})$,

[C2] $\hat{\theta}_{2,n}^{k,0} - \theta_2^* = O_n(a_n^2 + h_n^{((p-1)/2)\wedge k})$,

[C3] $\hat{\theta}_{3,n}^{k,0} - \theta_3^* = O_n(a_n^3 + h_n^{((p+1)/2)\wedge(k+1)})$.

The above is easily checked by induction with Proposition 6.2. Finally, we can use Theorem 5.2 for

1. $\Theta_\circ = \Theta_1, \theta_\circ^* = \theta_1^*, \mathbb{Y}_n(\theta_1) = \mathbb{Y}_n^{1,k_0+1}(\theta_1|\hat{\theta}_n^{(k_0,k_0,k_0),0}), \mathbb{Y}(\theta_1) = \mathbb{Y}^1(\theta_1), \Gamma'_n(\theta_1) = \Gamma^1(\theta_1|\hat{\theta}_{3,n}^{k_0,0}), a_n^\circ = a_n^1 + h_n^{(p/2)\wedge(k_0+1)}, b_n^\circ = h_n + n^{-1/2}$,

2. If $k_0 = 1$, $\Theta_\circ = \Theta_3, \theta_\circ^* = \theta_3^*, \mathbb{Y}_n(\theta_3) = \mathbb{Y}_n^3(\theta_3|\hat{\theta}_n^{(1,1,1),0}), \mathbb{Y}(\theta_3) = \mathbb{Y}^3(\theta_3), \Gamma'_n(\theta_3) = \Gamma^3(\theta_3|\hat{\theta}_{1,n}^{1,0}, \hat{\theta}_{3,n}^{1,0}), a_n^\circ = a_n^3 + h_n^2$

and we obtain

[D1] $\hat{\theta}_{1,n}^{k_0} - \theta_1^* = O_n(a_n^1 + h_n^{p/2})$,

[D2] $\hat{\theta}_{2,n}^{k_0} - \theta_2^* = O_n(a_n^2 + h_n^{(p-1)/2})$,

[D3] $\hat{\theta}_{3,n}^{k_0} - \theta_3^* = O_n(a_n^3 + h_n^{(p+1)/2})$.

From $nh_n^p \to 0$, we see

$$\frac{h_n^{p/2}}{a_n^1}, \frac{h_n^{(p-1)/2}}{a_n^2}, \frac{h_n^{(p+1)/2}}{a_n^3} \to 0.$$

Thus, it follows from [D1]-[D3] that

$$a_n^{-1}(\hat{\theta}_n^{k_0} - \theta^*) = O_n(1).$$

Moreover, by Theorem 5.2, the following asymptotic equivalence holds:

$$\begin{aligned} a_n^{-1}(\hat{\theta}_n^{k_0} - \theta^*) &\sim \begin{pmatrix} \Gamma^1(\theta_1^*|\hat{\theta}_{3,n}^{k_0,0}) & O & O \\ O & \Gamma^2(\theta_2^*|\hat{\theta}_{1,n}^{k_0,0}) & O \\ O & O & \Gamma^3(\theta_3^*|\hat{\theta}_{1,n}^{k_0,0}, \hat{\theta}_{3,n}^{(k_0-1)\vee 1,0}) \end{pmatrix}^{-1} \\ &\qquad \bullet \begin{pmatrix} (a_n^1)^{-1}\partial_{\theta_1}\mathbb{Y}_n^{1,k_0}(\theta_1^*|\hat{\theta}_n^{(k_0,k_0,k_0),0}) \\ (a_n^2)^{-1}\partial_{\theta_2}\mathbb{Y}_n^{2,k_0}(\theta_2^*|\hat{\theta}_n^{(k_0,k_0-1,k_0-1),0}) \\ (a_n^3)^{-1}\partial_{\theta_3}\mathbb{Y}_n^{3,k_0}(\theta_3^*|\hat{\theta}_n^{(k_0,k_0,(k_0-1)\vee 1),0}) \end{pmatrix} \\ &\sim \Gamma^{-1}M_n \\ &\qquad \text{(by Proposition 6.2 and [C1]-[C3]).} \end{aligned}$$

□

7 Basic estimations with O-notation

In this section, we introduce a notation and basic techniques used in later sections.

Remark 7.1. If (5) holds for any M larger than some $M_0 > 0$, then, by the monotonicity of the L^p-norm, it follows that

$$\Xi_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{\lambda_1,\cdots,\lambda_{m_0}}(\alpha_{\lambda_1,\cdots,\lambda_m}).$$

We can paraphrase [A2] and [A3] as

[A2] $Z_t = O_t(1)$.

[A3] There exists a probability measure ν^* on $\mathbb{R}^{d_z}$ and $\epsilon > 0$ such that

$$\frac{1}{nh_n}\int_0^{nh_n} f(Z_t)dt - \int_{\mathbb{R}^{d_z}} f(z)d\nu^*(x) = O_n\Big(\frac{1}{(nh)^\epsilon}\Big)$$

for all $f \in C_\uparrow(\mathbb{R}^{d_z})$.

respectively. Remark that by [A1] and [A3], we can derive there exists $\epsilon > 0$ such that

$$\frac{1}{nh_n}\int_0^{nh_n} f(Z_t)dt - \int_{\mathbb{R}^{d_z}} f(z)d\nu^*(x) = O^{\boldsymbol{\theta}}_n(h_n^\epsilon). \tag{6}$$

for all $f \in C_\uparrow(\mathbb{R}^{dz})$. For $f \in C^{0,0}_\uparrow(\mathbb{R}^{d_z}, \Theta^2)$, the family of random variables $f(Z_t, \boldsymbol{\theta})$ dependent on $t \in [0, \infty)$ and $\boldsymbol{\theta} \in \Theta^2$ satisfies

$$\sup_{t\in\mathbb{R}_+} E[\sup_{\boldsymbol{\theta}} |f(Z_t, \boldsymbol{\theta})|^M] < \infty$$

for all $M > 0$ by [A2] so we can write $f(Z_t, \boldsymbol{\theta}) = O^{\boldsymbol{\theta}}_t(1)$.

Lemma 7.2. 1. Let $(\Lambda_1, \mathcal{B}, \mu)$ be a measure space, and let $\mathcal{B}_0$ be a collection of finite-measure sets. Consider a family of random variables $\Xi_{\lambda_1,\ldots,\lambda_m}$ depending on $\lambda_1 \in \Lambda_1$ and some indices $\lambda_2, \ldots, \lambda_m$, satisfying

$$\Xi_{\lambda_1,\ldots,\lambda_m} = O^{\lambda_{m_0+1},\ldots,\lambda_m}_{\lambda_1,\ldots,\lambda_{m_0}}(\alpha_{\lambda_1,\ldots,\lambda_m}),$$

where the random function $\lambda_1 \mapsto \Xi_{\lambda_1,\ldots,\lambda_m}$ is always measurable and

$$\Lambda_1 \times \Omega \ni (\lambda_1, \omega) \mapsto \sup_{\lambda_{m_0+1},\cdots,\lambda_m} \left|\frac{\Xi_{\lambda_1,\cdots,\lambda_m}}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right| \in [0, \infty]$$

is $\mathcal{B} \otimes \mathcal{F}$-measurable for each fixed $\lambda_2, \ldots, \lambda_m$. Moreover, $\sup_{\lambda_1\in A} \alpha_{\lambda_1,\cdots,\lambda_m} < \infty$ and

$$\sup_{\lambda_{m_0+1},\cdots,\lambda_m} \frac{\int_A |\Xi_{\lambda_1,\cdots,\lambda_m}|\, d\mu(\lambda_1)}{\sup_{\lambda_1\in A} a_{\lambda_1,\cdots,\lambda_m}}$$

is measurable for each fixed $\lambda_2, \ldots, \lambda_m, A \in \mathcal{B}_0$. Then $\int_A |\Xi_{\lambda_1,\cdots,\lambda_m}|d\mu(\lambda_1) < \infty$ a.s. for all $\lambda_2, \cdots, \lambda_m$ and

$$\int_A \Xi_{\lambda_1,\cdots,\lambda_m} d\mu(\lambda_1) = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{A,\lambda_2,\cdots,\lambda_{m_0}}\Big(\sup_{\lambda_1\in A} \alpha_{\lambda_1,\cdots,\lambda_m}\mu(A)\Big)$$

with A varying in $\mathcal{B}_0$.

2. For a family of random variables $\Xi^i_{\lambda_1,\cdots,\lambda_m}$ depending on i in a finite set I and indices $\lambda_1,\cdots,\lambda_{m_0},\cdots,\lambda_m$, if $\Xi^i_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{\lambda_1,\cdots,\lambda_{m_0}}(\alpha^i_{\lambda_1,\cdots,\lambda_m})$ for each $i \in I$, then

$$\sum_{i\in I} \Xi^i_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{\lambda_1,\cdots,\lambda_{m_0}}\Big(\sum_{i\in I}\alpha^i_{\lambda_1,\cdots,\lambda_m}\Big), \tag{7}$$

$$\prod_{i\in I} \Xi^i_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{\lambda_1,\cdots,\lambda_{m_0}}\Big(\prod_{i\in I}\alpha^i_{\lambda_1,\cdots,\lambda_m}\Big). \tag{8}$$

Proof. 1: For $M > 1$ and each $\lambda_2,\cdots,\lambda_{m_0}, A \in \mathcal{B}_0$, we have

$$\begin{aligned}
\sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{\int_A |\Xi_{\lambda_1,\cdots,\lambda_m}| d\mu(\lambda_1)}{\sup_{\lambda_1\in A}\alpha_{\lambda_1,\cdots,\lambda_m}\mu(A)}\right|^M &\le \sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{1}{\mu(A)}\int_A \frac{|\Xi_{\lambda_1,\cdots,\lambda_m}|}{\alpha_{\lambda_1,\cdots,\lambda_m}} d\mu(\lambda_1)\right|^M \\
&\le \sup_{\lambda_{m_0+1},\cdots,\lambda_m}\frac{1}{\mu(A)}\int_A \left|\frac{\Xi_{\lambda_1,\cdots,\lambda_m}}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right|^M d\mu(\lambda_1) \\
&\quad \text{(by the Jensen inequality)} \\
&\le \frac{1}{\mu(A)}\int_A \sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{\Xi_{\lambda_1,\cdots,\lambda_m}}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right|^M d\mu(\lambda_1).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sup_{A,\lambda_2,\cdots,\lambda_{m_0}} E\left[\sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{\int_A |\Xi_{\lambda_1,\cdots,\lambda_m}| d\mu(\lambda_1)}{\sup_{\lambda_1\in A}\alpha_{\lambda_1,\cdots,\lambda_m}\mu(A)}\right|^M\right] \\
&\le \sup_{A,\lambda_2,\cdots,\lambda_{m_0}}\frac{1}{\mu(A)}\int_A E\left[\sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{\Xi_{\lambda_1,\cdots,\lambda_m}}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right|^M\right] d\mu(\lambda_1) \\
&\le \sup_{A,\lambda_2,\cdots,\lambda_{m_0}}\frac{1}{\mu(A)}\int_A \sup_{\lambda_1} E\left[\sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{\Xi_{\lambda_1,\cdots,\lambda_m}}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right|^M\right] d\mu(\lambda_1) \\
&= \sup_{\lambda_1,\cdots,\lambda_{m_0}} E\left[\sup_{\lambda_{m_0+1},\cdots,\lambda_m}\left|\frac{\Xi_{\lambda_1,\cdots,\lambda_m}}{\alpha_{\lambda_1,\cdots,\lambda_m}}\right|^M\right] < \infty.
\end{aligned}$$

2: Note that $\Xi^i_{\lambda_1,\cdots,\lambda_m} = O^{\lambda_{m_0+1},\cdots,\lambda_m}_{i,\lambda_1,\cdots,\lambda_{m_0}}\Big(\sum_{i\in I}\alpha^i_{\lambda_1,\cdots,\lambda_m}\Big)$. By applying the statement 1 with the counting measure on I, we obtain (7). The equation (8) follows from the Hölder inequality.

□

Lemma 7.3. 1. For a family of $\mathcal{F}$-adapted continuous stochastic processes $\Phi^{\lambda_2,\cdots,\lambda_m} = \{\Phi^{\lambda_2,\cdots,\lambda_m}_t\}_{t\ge 0}$ and non-random $\alpha_{t,\lambda_2,\cdots,\lambda_m} \ge 0$ depending on $t \in [0,\infty)$ and some indices $\lambda_2,\cdots,\lambda_m$ satisfying

$$\Phi^{\lambda_2,\cdots,\lambda_m}_t = O_{t,\lambda_2,\cdots,\lambda_{m_0}}(\alpha_{t,\lambda_2,\cdots,\lambda_m}).$$

Then,

$$\int_s^t \Phi^{\lambda_2,\cdots,\lambda_m}_u du = O_{s,t,\lambda_2,\cdots,\lambda_m}\Big(\sup_{s\le u\le t}\alpha_{t,\lambda_2,\cdots,\lambda_m}|t-s|\Big), \tag{9}$$

$$\int_s^t \Phi^{\lambda_2,\cdots,\lambda_m}_u dw_u = O_{s,t,\lambda_2,\cdots,\lambda_m}\Big(\sup_{s\le u\le t}\alpha_{t,\lambda_2,\cdots,\lambda_m}|t-s|^{1/2}\Big) \tag{10}$$

with $0 \le s \le t < \infty$.

2. For $f \in C^{1,0}_{\uparrow}(\mathbb{R}^{d_Z}, \Theta^2)$,

$$f(Z_t, \boldsymbol{\theta}) - f(Z_s, \boldsymbol{\theta}) = O^{\boldsymbol{\theta}}_{s,t}(|t-s|^{1/2})$$

with $0 \le s \le t < \infty$.

Proof. 1: The equation (9) follows directly from Lemma 7.2. For (10), we have for $M > 1$,

$$\begin{aligned} E\left[\left|\int_s^t \Phi_u^{\lambda_2,\cdots,\lambda_m} dw_u\right|^M\right] &\lesssim E\left[\left|\int_s^t |\Phi_u^{\lambda_2,\cdots,\lambda_m}|^2 du\right|^{M/2}\right] \\ &\quad \text{(by the Burkholder-Davis-Gundy inequality)} \\ &\lesssim \sup_{s\le u\le t} \alpha^M_{t,\lambda_2,\cdots,\lambda_m}(t-s)^{M/2} \\ &\quad \left(\text{by } \int_s^t |\Phi_u^{\lambda_2,\cdots,\lambda_m}|^2 du = O_{s,t}\Big(\sup_{s\le u\le t} \alpha^2_{t,\lambda_2,\cdots,\lambda_m}|t-s|\Big)\right), \end{aligned}$$

which proves (10).

2: Note that

$$f(Z_t, \boldsymbol{\theta}) - f(Z_s, \boldsymbol{\theta}) = \int_0^1 \partial_z f(Z_s + u(Z_t - Z_s), \boldsymbol{\theta}) du [Z_t - Z_s].$$

By $\partial_{\boldsymbol{\theta}} f(Z_s + u(Z_t - Z_s), \boldsymbol{\theta}) = O^{\boldsymbol{\theta}}_{s,t,u}(1)$ and (9), $\int_0^1 \partial_{\boldsymbol{\theta}} f(Z_s + u(Z_t - Z_s), \boldsymbol{\theta}) du = O^{\boldsymbol{\theta}}_{s,t}(1)$. Now, we only need to prove that $Z_t - Z_s = O_{t,s}(|t-s|^{1/2})$. Let

$$\boldsymbol{A}(z) := \begin{pmatrix} A(z, \theta_2^*) \\ H(z, \theta_3^*) \end{pmatrix}, \boldsymbol{B}(z) := \begin{pmatrix} B(z, \theta_1^*) \\ O \end{pmatrix}$$

Then, with $\boldsymbol{A}, \boldsymbol{B} \in C_{\uparrow}$, we have

$$\begin{aligned} Z_t - Z_s &= \int_s^t \boldsymbol{A}(Z_u) du + \int_s^t \boldsymbol{B}(Z_u) dw_u \\ &= O_{s,t}(|t-s|) + O_{s,t}(|t-s|^{1/2}) \\ &\quad \text{(by the statement 1).} \end{aligned}$$

Since $Z_t = O_t(1)$,

$$\begin{aligned} Z_t - Z_s &= O_{s,t}\big((|t-s| + |t-s|^{1/2}) \wedge 1\big) \\ &= O_{s,t}(|t-s|^{1/2}). \end{aligned}$$

□

The following propositions are utilized only in Section 9. Let $\Xi_\lambda(\boldsymbol{\theta})$ be a family of random variables depending on an index λ and parameters $\boldsymbol{\theta} \in \Theta^2$. Suppose that for each λ, the random function $\Xi_\lambda(\boldsymbol{\theta})$ belongs to the class C^2. In the following lemma, We denote (θ^*, θ^*) by $\boldsymbol{\theta}^*$.

Lemma 7.4. 1. Suppose that $\partial^l_{\boldsymbol{\theta}} \Xi_\lambda(\boldsymbol{\theta}) = O_{\lambda,\boldsymbol{\theta}}(1) (l = 0, 1)$, then $\Xi_\lambda(\boldsymbol{\theta}) = O^{\boldsymbol{\theta}}_\lambda(1)$.

2. Suppose that $\partial^l_{\boldsymbol{\theta}} \Xi_\lambda(\boldsymbol{\theta}) = O_{\lambda,\boldsymbol{\theta}}(1) (l = 0, 1, 2)$, then, $\Xi_\lambda(\boldsymbol{\theta}) - \Xi_\lambda(\boldsymbol{\theta}^*) = O^{\boldsymbol{\theta}}_\lambda(|\boldsymbol{\theta} - \boldsymbol{\theta}^*|)$.

Remark 7.5. Without using the O-notation, the claim can be expressed as follows:

1. $\sup_{\lambda,\boldsymbol{\theta}} E[|\partial_{\boldsymbol{\theta}}^l \Xi_\lambda(\boldsymbol{\theta})|^M] < \infty$ for $l = 0, 1 \Rightarrow \sup_\lambda E\Big[\sup_{\boldsymbol{\theta}} |\Xi_\lambda(\boldsymbol{\theta})|^M\Big] < \infty$.

2. $\sup_{\lambda,\boldsymbol{\theta}} E[|\partial_{\boldsymbol{\theta}}^l \Xi_\lambda(\boldsymbol{\theta})|^M] < \infty$ for $l = 0, 1, 2 \Rightarrow \sup_\lambda E\left[\sup_{\boldsymbol{\theta}\neq\boldsymbol{\theta}^*} \frac{|\Xi_\lambda(\boldsymbol{\theta}) - \Xi_\lambda(\boldsymbol{\theta}^*)|}{|\boldsymbol{\theta} - \boldsymbol{\theta}^*|}\right] < \infty$.

Proof of Lemma 7.4. 1: For all $M > M_0$, by the Sobolev inequality, we have

$$\begin{aligned}
\sup_\lambda E\Big[\sup_{\boldsymbol{\theta}} |\Xi_\lambda(\boldsymbol{\theta})|^M\Big] &\leq C \sup_\lambda \sum_{l=0,1} E\left[\int_{\Theta^2} |\partial_{\boldsymbol{\theta}}^l \Xi_\lambda(\boldsymbol{\theta})|^M d\boldsymbol{\theta}\right] \\
&\leq C \sum_{l=0,1} \sup_{\lambda,\boldsymbol{\theta}} E[|\partial_{\boldsymbol{\theta}}^l \Xi_\lambda(\boldsymbol{\theta})|^M] \int_{\Theta^2} d\boldsymbol{\theta} < \infty.
\end{aligned}$$

2: Take $\epsilon > 0$ such that the open ball of radius ϵ centred at $\boldsymbol{\theta}^*$ is contained in Θ^2. Then,

$$\sup_\lambda E\left[\sup_{\boldsymbol{\theta}\neq\boldsymbol{\theta}^*} \frac{|\Xi_\lambda(\boldsymbol{\theta}) - \Xi_\lambda(\boldsymbol{\theta}^*)|}{|\boldsymbol{\theta} - \boldsymbol{\theta}^*|}\right] \leq \sup_\lambda E\left[\sup_{\boldsymbol{\theta}} |\partial_{\boldsymbol{\theta}} \Xi_\lambda(\boldsymbol{\theta})| + \frac{2}{\epsilon} \sup_{\boldsymbol{\theta}} |\Xi_\lambda(\boldsymbol{\theta})|\right].$$

By the statement 1, we have $\partial_{\boldsymbol{\theta}}^l \Xi_\lambda(\boldsymbol{\theta}) = O_\lambda^{\boldsymbol{\theta}}(1)$ for $l = 0, 1$, which implies that the right-hand side is finite. □

Let $\xi_{n,j}(\boldsymbol{\theta})$ be a family of random variables depending on an index $n, j \geq 0$ and parameters $\boldsymbol{\theta} \in \Theta^2$. Suppose that for each $n, j \geq 0$, the random function $\xi_\lambda(\boldsymbol{\theta})$ belongs to the class C^2 and $\mathcal{F}_{t_{n,j}}$-adapted.

Lemma 7.6. 1. Suppose that

$$\begin{aligned}
&E_{n,j-1}[\partial_{\boldsymbol{\theta}}^l \xi_{n,j}(\boldsymbol{\theta})] = O_{n,j,\boldsymbol{\theta}}(\alpha_n), \\
&\partial_{\boldsymbol{\theta}}^l \xi_{n,j}(\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}}^l \xi_{n,j}(\boldsymbol{\theta})] = O_{n,j,\boldsymbol{\theta}}(\beta_n)
\end{aligned} \tag{11}$$

holds for $l = 0, 1$, then $\sum_{j=1}^n \xi_{n,j}(\boldsymbol{\theta}) = O_n^{\boldsymbol{\theta}}(n\alpha_n + n^{1/2}\beta_n)$.

2. Suppose that (11) holds for $l = 0, 1, 2$, then $\sum_{j=1}^n \xi_{n,j}(\boldsymbol{\theta}) - \sum_{j=1}^n \xi_{n,j}(\boldsymbol{\theta}^*) = O_n^{\boldsymbol{\theta}}(|\boldsymbol{\theta} - \boldsymbol{\theta}^*|(n\alpha_n + n^{1/2}\beta_n))$.

Proof. Assume (11) holds for fixed l. Then, we have

$$\begin{aligned}
\sum_{j=1}^n \partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta}) &= \sum_{j=1}^n E_{n,j-1}[\partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta})] + \sum_{j=1}^n \partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta})] \\
&= O_{n,\boldsymbol{\theta}}(n\alpha_n) + \sum_{j=1}^n \partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta})].
\end{aligned}$$

For $M \geq 2$,

$$E\left[\left|\sum_{j=1}^n \partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta})]\right|^M\right] \lesssim E\left[\left|\sum_{j=1}^n (\partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}} \xi_{n,j}(\boldsymbol{\theta})])^2\right|^{M/2}\right]$$

(by the Burkholder-Davis-Gundy inequality)

$$\lesssim (n^{1/2}\beta_n)^M$$
$$\Big(\text{by } \sum_{j=1}^n (\partial_{\boldsymbol{\theta}}\xi_{n,j}(\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}}\xi_{n,j}(\boldsymbol{\theta})])^2 = O_{n,\boldsymbol{\theta}}(n\beta_n^2)\Big),$$

which implies

$$\sum_{j=1}^n \partial_{\boldsymbol{\theta}}\xi_{n,j}(\boldsymbol{\theta}) = O_{n,\boldsymbol{\theta}}(n\alpha_n + n^{1/2}\beta_n).$$

Thus we can apply Lemma (7.4). □

Lemma 7.7. Let $f \in C_{\uparrow}^{1,1}(\mathbb{R}^{d_z} \times \Theta^2)$. There exits $\epsilon > 0$ such that

$$\frac{1}{n}\sum_{j=1}^n f_{n,j-1}(\boldsymbol{\theta}) = \int_{\mathbb{R}^{d_Z}} f(z,\boldsymbol{\theta})d\nu^*(z) + O_n^{\boldsymbol{\theta}}(h_n^{\epsilon}).$$

Proof. First, from Lemma 7.4 and (6), there exists $\epsilon > 0$ such that

$$\frac{1}{nh_n}\int_0^{nh_n} f(Z_t,\boldsymbol{\theta})dt - \int_{\mathbb{R}^{d_z}} f(z,\boldsymbol{\theta})d\nu^*(z) = O_n^{\boldsymbol{\theta}}(h_n^{\epsilon})$$

for all $f \in C_{\uparrow}^{0,1}(\mathbb{R}^{d_Z}, \Theta^2)$. Then, it suffices to show that

$$\frac{1}{nh}\int_0^{nh} f(Z_s,\boldsymbol{\theta})ds - \frac{1}{n}\sum_{j=1}^n f_{j-1}(\boldsymbol{\theta})ds = O_n^{\boldsymbol{\theta}}(h_n^{\epsilon}).$$

Note that

$$\frac{1}{nh}\int_0^{nh} f(Z_s,\boldsymbol{\theta})ds - \frac{1}{n}\sum_{j=1}^n f_{j-1}(\boldsymbol{\theta})ds = \sum_{j=1}^n \frac{1}{nh}\int_{t_{j-1}}^{t_j} f(Z_s,\boldsymbol{\theta}) - f(Z_{t_{j-1}},\boldsymbol{\theta})ds.$$

For $l = 0, 1$,

$$\frac{1}{nh}\int_{t_{j-1}}^{t_j} E_{n,j-1}[\partial_{\boldsymbol{\theta}}^l f(Z_s,\boldsymbol{\theta}) - \partial_{\boldsymbol{\theta}}^l f(Z_{t_{j-1}},\boldsymbol{\theta})]ds = O_{n,j}^{\boldsymbol{\theta}}\Big(\frac{h^{1/2}}{n}\Big), \tag{12}$$
$$\frac{1}{nh}\int_{t_{j-1}}^{t_j} \partial_{\boldsymbol{\theta}}^l f(Z_s,\boldsymbol{\theta}) - E_{n,j-1}[\partial_{\boldsymbol{\theta}}^l f(Z_s,\boldsymbol{\theta})]ds = O_{n,j}^{\boldsymbol{\theta}}\Big(\frac{1}{n}\Big). \tag{13}$$

By Lemma 7.3,

$$\Big(\partial_{\boldsymbol{\theta}}^l f(Z_s,\boldsymbol{\theta}) - \partial_{\boldsymbol{\theta}}^l f(Z_{t_{j-1}},\boldsymbol{\theta})\Big)1_{t_{j-1}\le s\le t_j} = O_{n,s,j}^{\theta}(h^{1/2}),$$

which implies (12) by Lemma 7.2. Similarly, (13) follows from $\partial_{\boldsymbol{\theta}}^l f(Z_s,\boldsymbol{\theta}) = O_s^{\boldsymbol{\theta}}(1)$. The conclusion now follows from Lemma 7.6. □

8 Ito-Taylor expansion

In addition to L, define an operator M for a C^2 function $f(z,\theta) = f(x,y,\theta)$, as:

$$Mf(z,\theta) := \partial_x f(z,\theta)\cdot B(z,\theta_1).$$

Note that if f takes values in $\mathbb{R}^m$, then Mf takes values in $\mathbb{R}^m \otimes \mathbb{R}^r$. We treat L and M as formal symbols and call a finite sequence W of terms, each being either L or M, a word. For a word $W = (W_1, W_2, \cdots)$, we use the same symbol W to denote the composition operator $W_1 \circ W_2 \circ \cdots$. Following word convention, we write $LLLMMLLM = L^3M^2L^2M$, etc. Let $\mathrm{len}(W)$ denote the length of the finite sequence W, and $\mathrm{len}_L(W)$, $\mathrm{len}_M(W)$ denote the number of L's and M's in W respectively. Further, define $\mathrm{ord}(W) := \mathrm{len}_L(W) + \frac{1}{2}\mathrm{len}_M(W)$. For example, for $W = LM$, $\mathrm{len}(W) = 2$, $\mathrm{len}_L(W) = 1$, $\mathrm{len}_M(W) = 1$, $\mathrm{ord}(W) = \frac{3}{2}$ and for a sufficiently smooth f, $Wf = LMf$. For tensors $\mathcal{B}, \mathcal{A}_1, \mathcal{A}_2, \cdots$, we say $\mathcal{B} = \mathrm{pol}(\mathcal{A}_1, \mathcal{A}_2, \cdots)$ if each component of $\mathcal{B}$ can be expressed as a polynomial of the components of $\mathcal{A}_1, \mathcal{A}_2, \cdots$. The following lemma follows from straightforward calculations, and thus we omit the proof.

Lemma 8.1. Let $k = 0, 1, \cdots, k_0$. Then the following statements hold:

1. $$L^k\pi_x = \mathrm{pol}(\{\partial_z^l A\}_{l=0}^{2k-2}, \{\partial_z^l B\}_{l=0}^{2k-4}, \{\partial_z^l H\}_{l=0}^{2k-4}).$$

   In particular $L^k\pi_x \in C_\uparrow^{p-2k+1,2}$.

2. For a word W' with $\mathrm{len}(W') = k, \mathrm{len}_M(W') \geq 1$,

   $$W'\pi_x = \mathrm{pol}(\{\partial_z^l A\}_{l=0}^{2k-3}, \{\partial_z^l B\}_{l=0}^{2k-2}, \{\partial_z^l H\}_{l=0}^{2k-4}).$$

   In particular $W'\pi_x \in C_\uparrow^{p-2k+2,2}$.

3. $$L^{k+1}\pi_y = \mathrm{pol}(\{\partial_z^l A\}_{l=0}^{2k-2}, \{\partial_z^l B\}_{l=0}^{2k-2}, \{\partial_z^l H\}_{l=0}^{2k}).$$

   In particular $L^{k+1}\pi_x \in C_\uparrow^{p-2k+1,2}$.

4. For a word W' with $\mathrm{len}(W') = k+1, \mathrm{len}_M(W') \geq 1$,

   $$W'\pi_y = \mathrm{pol}(\{\partial_z^l A\}_{l=0}^{2k-3}, \{\partial_z^l B\}_{l=0}^{2k-2}, \{\partial_z^l H\}_{l=0}^{2k-1}).$$

   In particular $W'\pi_y \in C_\uparrow^{p-2k+2,2}$.

Also, for $s \in \mathbb{R}_+$, let $s^L = s$, $s^M = w_s$. For a word $W = (W_1, \cdots, W_l)$ of length l, a $\mathbb{R}^m \otimes (\mathbb{R}^r)^{\otimes \mathrm{len}_M(W)}$-valued $(m \geq 1)$ $\mathcal{F}$-adapted continuous stochastic process $\Phi = \{\Phi_t\}_t$ and $n, j \in \mathbb{N}$, we define the iterated integral $I_{n,j}^W(\Phi)$ which takes values in $\mathbb{R}^m$ as follows:

$$I_{n,j}^W(\Phi) := \int_{t_{n,j-1}}^{t_{n,j}} (\int_{t_{n,j-1}}^{s_l} \cdots (\int_{t_{n,j-1}}^{s_3} (\int_{t_{n,j-1}}^{s_2} (\Phi_{s_1}\cdot ds_1^{W_1})\cdot ds_2^{W_2}) \cdots ds_{l-1}^{W_{l-1}}) \cdot ds_l^{W_l}.$$

We also define $I_{n,j}^W$ as a $(\mathbb{R}^r)^{\otimes \mathrm{len}_M(W)}$-valued random variable by setting

$$I_{n,j}^W := \int_{t_{n,j-1}}^{t_{n,j}} \int_{t_{n,j-1}}^{s_l} \cdots \int_{t_{n,j-1}}^{s_3} \int_{t_{n,j-1}}^{s_2} ds_1^{W_1} ds_2^{W_2} \cdots ds_{l-1}^{W_{l-1}} ds_l^{W_l}.$$

For instance, $I^{L^k}_{n,j} = \frac{h_n^k}{k!}$ and the (l_1, l_2)-th component of $I^{M^2}_{n,j}$ is

$$\int_{t_{n,j-1}}^{t_{n,j}} \int_{t_{n,j-1}}^{t} dw_s^{l_1} dw_t^{l_2},$$

where w_t^l is the l-th component of w_t.

Lemma 8.2. For a family of $\mathbb{R}^m \otimes (\mathbb{R}^r)^{\otimes \mathrm{len}_M(W)}$-valued ($m \geq 1$) $\mathcal{F}$-adapted continuous stochastic processes $\Phi^n = \{\Phi^n_t\}_{t \geq 0}$ and non-random $\alpha_n \geq 0$ depending on $n \in \mathbb{N}$ satisfying

$$\Phi^n_t = O_{t,n}(\alpha_n).$$

1. For a word W, $I^W_{n,j}(\Phi^n) = O_{n,j}(\alpha_n h_n^{\mathrm{ord}(W)})$.
2. For words W_1, W_2,
$$\begin{aligned} &E_{n,j-1}[I^{W_1}_{n,j} \otimes I^{W_2}_{n,j}(\Phi^n)] \\ &\quad = \begin{cases} O_{n,j}(\alpha_n h_n^{\mathrm{ord}(W_1)+\mathrm{ord}(W_2)}) & (\mathrm{len}_M(W_1) \geq \mathrm{len}_M(W_2)) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}$$
3. There exists $c(W_1, W_2)$ depending only on words W_1, W_2 such that $E_{n,j-1}[I^{W_1}_{n,j} \otimes I^{W_2}_{n,j}] = h_n^{\mathrm{ord}(W_1)+\mathrm{ord}(W_2)} c(W_1, W_2)$

Proof. 1: By Lemma 7.3,

$$\sum_{j=1}^{\infty} \int_{t_{n,j-1}}^{t} \Phi^n_s \, ds 1_{(t_{n,j-1}, t_{n,j}]}(t) = O_{n,t}(\alpha_n h_n),$$
$$\sum_{j=1}^{\infty} \int_{t_{n,j-1}}^{t} \Phi^n_s \, dw_s 1_{(t_{n,j-1}, t_{n,j}]}(t) = O_{n,t}(\alpha_n h_n^{1/2}).$$

The desired result then follows by mathematical induction.

2: For a word $W = W_1 W_2 \cdots W_{\mathrm{len}(W)}$, we denote the indices i where $W_i = L$ by the sequence $l^W_1, \cdots, l^W_{\mathrm{len}_L(W)}$ and the indices where $W_i = M$ by the sequence $m^W_1, \cdots, m^W_{\mathrm{len}_M(W)}$. For example, consider a word $W = LMLMMMLLLMM$. In this case: The sequence $l^W_1, \cdots, l^W_{\mathrm{len}_L(W)}$ represents the positions where L appears, which are $1, 3, 7, 8, 9$. Similarly, the sequence $m^W_1, \cdots, m^W_{\mathrm{len}_M(W)}$ represents the positions where M appears, which are $2, 4, 5, 6, 10$.

To simplify, we prove the case where $r = 1, m = 1$. Then, $I^W_{n,j} = I^W_{n,j}\mathbf{1}$, where $\mathbf{1}$ denotes the constant function identically equal to 1. For a $\mathcal{F}$-adapted continuous stochastic process Φ and a word W, by the Fubini theorem for stochastic integrals,

$$\begin{aligned} I^W_{n,j}(\Phi) &= \int_{t_{n,j-1}}^{t_{n,j}} \cdots \int_{t_{n,j-1}}^{s_2} \Phi_{s_1} ds_1^{W_1} \cdots ds_l^{W_l} \\ &= \int_{t_{n,j-1}}^{t_{n,j}} \cdots \int_{t_{n,j-1}}^{s_{m^W_3}} \int_{t_{n,j-1}}^{s_{m^W_2}} \Phi^W_{s_{m^W_1}, s_{m^W_2}, \cdots, s_{m^W_{\mathrm{len}_M(W)}}} dw_{s_{m^W_1}} dw_{s_{m^W_2}} \cdots dw_{m^W_{\mathrm{len}_M(W)}}, \end{aligned}$$

where

$$\Phi^W_{s_{m^W_1}, s_{m^W_2}, \cdots, s_{m^W_{\mathrm{len}_M(W)}}} = \int_{[t_{n,j-1}, t_{n,j}]^{\mathrm{len}_L(W)}} \Phi_{s_1} 1_{\{t_{n,j-1} \leq s_1 \leq s_2 \leq \cdots \leq s_l \leq t_{n,j}\}} ds_{l^W_1} \cdots ds_{l^W_{\mathrm{len}_L(W)}}.$$

In particular, $\mathbf{1}^W_{s_{m_1^W}, s_{m_2^W}, \cdots, s_{m^W_{\mathrm{len}_M(W)}}}$ is a deterministic function. We write $l_1 := \mathrm{len}_M(W_1)$ and $l_2 := \mathrm{len}_M(W_2)$ respectively. Then, when $l_1 < l_2$, by the Itô isometry,

$$
\begin{aligned}
&E_{n,j-1}[I^{l_1}_{n,j} \cdot I^{l_2}_{n,j}(\Phi^n)] \\
&= E_{n,j-1}\Bigg[\int_{t_{n,j-1}}^{t_{n,j}} \cdots \int_{t_{n,j-1}}^{s_{l_2-l_1+2}} \Bigg(\mathbf{1}^{l_1}_{s_{l_2-l_1+1},\cdots,s_{l_2}} \int_{t_{n,j-1}}^{s_{l_2-l_1+1}} \cdots \int_{t_{n,j-1}}^{s_2} \Phi^{W_2}_{s_1,\cdots,s_{l_2}} dw_1 \cdots dw_{l_2-l_1} \Bigg) \\
&\qquad ds_{l_2-l_1+1} \cdots ds_{l_2} \Bigg] \\
&= \int_{t_{n,j-1}}^{t_{n,j}} \cdots \int_{t_{n,j-1}}^{s_{l_2-l_1+2}} \Bigg(\mathbf{1}^{l_1}_{s_{l_2-l_1+1},\cdots,s_{l_2}} E_{n,j-1}\Bigg[\int_{t_{n,j-1}}^{s_{l_2-l_1+1}} \cdots \int_{t_{n,j-1}}^{s_2} \Phi^{W_2}_{s_1,\cdots,s_{l_2}} dw_1 \cdots dw_{l_2-l_1} \Bigg] \Bigg) \\
&\qquad ds_{l_2-l_1+1} \cdots ds_{l_2} \\
&= 0.
\end{aligned}
$$

When $l_1 \geq l_2$, the conclusion follows from the statement 1.

3: We prove the case where $r = 1$ using the same notation in the proof of the statement 2. When $l_1 \neq l_2$, $c(W_1, W_2) = 0$. Furthermore, when $l_1 = l_2$,

$$
c(W_1, W_2) = \int_{t_{n,j-1}}^{t_{n,j}} \cdots \int_{t_{n,j-1}}^{s_2} \mathbf{1}^{W_1}_{s_1,\cdots,s_{l_1}} \mathbf{1}^{W_2}_{s_1,\cdots,s_{l_2}} ds_1 \cdots ds_{l_1}
$$

by the Ito isometry. □

Given a stochastic process Φ, we define a new process $\Delta^n \Phi$ by

$$
\Delta^n_t \Phi := \Phi_t - \Phi_{t_{n,j-1}} \qquad (t_{n,j-1} \leq t < t_{n,j}).
$$

For $k = 0, \cdots, k_0$, let

$$
\begin{aligned}
\mathcal{M}^{0,k}_{n,j}(z,\theta) &:= \begin{pmatrix} \mathcal{M}^{0,k,x}_{n,j}(z,\theta) \\ \mathcal{M}^{0,k,y}_{n,j}(z,\theta) \end{pmatrix} \\
&:= \begin{pmatrix} h_n^{-1/2} \sum_{\substack{1 \leq \mathrm{len}(W) \leq k \\ \mathrm{len}_M(W) \geq 1}} W\pi_x(z,\theta) \cdot I^W_{n,j} \\ h_n^{-3/2} \sum_{\substack{1 \leq \mathrm{len}(W) \leq k+1 \\ \mathrm{len}_M(W), \mathrm{len}_L(W) \geq 1}} W\pi_y(z,\theta) \cdot I^W_{n,j} \end{pmatrix}, \\
\mathcal{M}^{1,k}_{n,j}(z,\theta) &:= \begin{pmatrix} \mathcal{M}^{1,k,x}_{n,j}(z,\theta) \\ \mathcal{M}^{1,k,y}_{n,j}(z,\theta) \end{pmatrix} \\
&:= \begin{pmatrix} h_n^{-1/2} \sum_{\substack{\mathrm{len}(W) = k \\ \mathrm{len}_M(W) \geq 1}} LW\pi_x(z,\theta) \cdot I^{LW}_{n,j} + MW\pi_x(z,\theta) \cdot I^{MW}_{n,j} \\ h_n^{-3/2} \sum_{\substack{\mathrm{len}(W) = k+1 \\ \mathrm{len}_M(W), \mathrm{len}_L(W) \geq 1}} LW\pi_y(z,\theta) \cdot I^{LW}_{n,j} + MW\pi_y(z,\theta) \cdot I^{MW}_{n,j} \end{pmatrix}, \\
\mathcal{M}^{k}_{n,j}(z,\theta) &:= \begin{pmatrix} \mathcal{M}^{k,x}_{n,j}(z,\theta) \\ \mathcal{M}^{k,y}_{n,j}(z,\theta) \end{pmatrix} \\
&:= \mathcal{M}^{0,k}_{n,j}(z,\theta) + \mathcal{M}^{1,k}_{n,j}(z,\theta), \\
\mathcal{M}^{0,k}_{n,j}(\theta) &:= \mathcal{M}^{0,k}_{n,j}(Z_{n,j-1},\theta), \\
\mathcal{M}^{1,k}_{n,j}(\theta) &:= \mathcal{M}^{1,k}_{n,j}(Z_{n,j-1},\theta),
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{M}^k_{n,j}(\theta) &:= \mathcal{M}^k_{n,j}(Z_{n,j-1},\theta),\\
\mathcal{R}^{M,k}_{n,j} &:= \begin{pmatrix}\mathcal{R}^{M,k,x}_{n,j}\\ \mathcal{R}^{M,k,y}_{n,j}\end{pmatrix}\\
&:= \begin{pmatrix} h_n^{-1/2}\displaystyle\sum_{\substack{\mathrm{len}(W)=k\\ \mathrm{len}_M(W)\ge 1}} I^{LW}_{n,j}\left(\Delta^n LW\pi_x(Z,\theta^*)\right)+I^{MW}_{n,j}\left(\Delta^n MW\pi_x(Z,\theta^*)\right)\\ h_n^{-3/2}\displaystyle\sum_{\substack{\mathrm{len}(W)=k+1\\ \mathrm{len}_M(W),\mathrm{len}_L(W)\ge 1}} I^{LW}_{n,j}\left(\Delta^n LW\pi_y(Z,\theta^*)\right)+I^{MW}_{n,j}\left(\Delta^n MW\pi_y(Z,\theta^*)\right)\end{pmatrix},\\
\mathcal{R}^{L,k}_{n,j} &:= \begin{pmatrix}\mathcal{R}^{L,k,x}_{n,j}\\ \mathcal{R}^{L,k,y}_{n,j}\end{pmatrix}\\
&:= \begin{pmatrix} h_n^{-1/2} I^{L^k}_{n,j}\left(\Delta^n L^k\pi_x(Z,\theta^*)\right)\\ h_n^{-3/2} I^{L^{k+1}}_{n,j}\left(\Delta^n L^{k+1}\pi_y(Z,\theta^*)\right)\end{pmatrix},\\
\mathcal{R}^k_{n,j} &:= \mathcal{R}^{M,k}_{n,j}+\mathcal{R}^{L,k}_{n,j}.
\end{aligned}
$$

The above definitions are well defined by Lemma 8.1. Note that by applying the Itô-Taylor expansion to π_x and π_y, we obtain the following equation:

$$
\mathcal{D}^k_{n,j}(\theta^*)=\mathcal{M}^k_{n,j}(\theta^*)+\mathcal{R}^k_{n,j}.
$$

Furthermore, let $\tilde{\mathscr{S}}^l$ denote formal matrix-valued formal polynomials constructed from formal variables representing partial derivatives $\{\partial_z^{l_A}A,\partial_z^{l_B}B,\partial_z^{l_H}h_n\}_{l_A,l_B,l_H=0}^{\infty}$, defined through the following expansion:

$$
E_{n,j-1}[(\mathcal{M}^{k_0}_{n,j})^{\otimes 2}(z,\theta)]=\sum_{l=0}^{\infty}h_n^l\tilde{\mathscr{S}}^l.
$$

By Lemma 8.2, we can analytically compute $\tilde{\mathscr{S}}^l$. And we define $S^l(z,\theta)$ as the evaluation of $\tilde{\mathscr{S}}^l$ at $\{\partial_z^{l_A}A(z,\theta),\partial_z^{l_B}B(z,\theta),\partial_z^{l_H}h_n(z,\theta)\}_{l_A,l_B,l_H=0}^{\infty}$. Moreover, we can easily check that $\tilde{S}^l\in C^{0,2}_{\uparrow}$ and $\tilde{S}^l=0$ for sufficiently large l. In the proof of the following lemma, we will show that $\mathscr{S}^l=\tilde{\mathscr{S}}^l$ for $k=0,\cdots,k_0$.

Lemma 8.3. For $k=0,\cdots,k_0$, the following hold:

1.

$$
\mathcal{D}^k_{n,j}(\theta^*)=\begin{pmatrix}h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}+O_{n,j}(h_n^{(p/2)\wedge(k+1/2)}),
$$

2.

$$
\begin{aligned}
&\mathcal{D}^k_{n,j}(\theta^*)^{\otimes 2}-\sum_{l=0}^{k}h_n^l S^l_{n,j-1}(\theta^*)\\
&\quad=\begin{pmatrix}h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2}-E_{n,j-1}\left[\begin{pmatrix}h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2}\right]\\
&\qquad+2E_{n,j-1}[\mathcal{D}^k_{n,j}(\theta^*)]\odot\begin{pmatrix}h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\\
&\qquad+O_{n,j}(h_n^{(p/2)\wedge(k+1)}).
\end{aligned}
$$

Remark 8.4. From Lemma 8.3, it immediately follows that

$$E_{n,j-1}[\mathcal{D}^k_{n,j}(\theta^*)] = O_{n,j}(h_n^{(p/2)\wedge(k+1/2)}).$$

Proof of Lemma 8.3. 1: Considering $E_{n,j-1}[\mathcal{M}_{n,j}(\theta^*)] = E_{n,j-1}[\mathcal{R}^M_{n,j}] = 0$, we have

$$\begin{aligned}
\mathcal{D}^k_{n,j}(\theta^*) - \begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix} &= \begin{pmatrix} h_n^{-1/2}\Big(E_{n,j-1}[X_{n,j}] - \sum_{l=0}^k h_n^l (L_0^l \pi_x)_{n,j-1}(\bar\theta)\Big) \\ h_n^{-3/2}\Big(E_{n,j-1}[Y_{n,j}] - \sum_{l=0}^{k+1} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar\theta)\Big) \end{pmatrix} \\
&= E_{n,j-1}[\mathcal{D}^k_{n,j}(\theta^*)] \\
&= E_{n,j-1}[\mathcal{R}^{L,k}_{n,j}].
\end{aligned}$$

For $f \in C^1_\uparrow$, it follows from Lemma 7.3 that

$$\Delta^n_t f(Z) = O_{n,t}(h_n^{1/2}).$$

From Lemma8.1, $L^k\pi_x$ and $L^{k+1}\pi_y$ belong to the class $C^{p-2k+1,2}_\uparrow$. Therefore, we obtain

$$\mathcal{R}^{L,k_0}_{n,j} = O_{n,j}(h_n^{k_0+1/2}) \tag{14}$$

by Lemma 8.2. When $p = 2k_0 + 1, k = k_0$, it leads to our conclusion. Otherwise, $L^k\pi_x$ and $L^{k+1}\pi_y$ belong to the class $C^{2,2}_\uparrow$. Hence

$$\mathcal{R}^{L,k}_{n,j} = \begin{pmatrix} h_n^{-1/2} I^{L^{k+1}}_{n,j}\big(L^{k+1}\pi_x(Z,\theta^*)\big) \\ h_n^{-3/2} I^{L^{k+2}}_{n,j}\big(L^{k+2}\pi_y(Z,\theta^*)\big) \end{pmatrix} + \begin{pmatrix} h_n^{-1/2} I^{ML^k}_{n,j}\big(ML^k\pi_x(Z,\theta^*)\big) \\ h_n^{-3/2} I^{ML^{k+1}}_{n,j}\big(ML^{k+1}\pi_y(Z,\theta^*)\big) \end{pmatrix}$$

holds. Thus we have

$$\begin{aligned}
E_{n,j-1}[\mathcal{R}^{L,k}_{n,j}] &= E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2} I^{L^{k+1}}_{n,j}\big(L^{k+1}\pi_x(Z,\theta^*)\big) \\ h_n^{-3/2} I^{L^{k+2}}_{n,j}\big(L^{k+2}\pi_y(Z,\theta^*)\big) \end{pmatrix}\right] \\
&= O_{n,j}(h_n^{k+1/2}),
\end{aligned}$$

by Lemma 8.2, which complete our conclusion.

2: From Equation 1, it suffices to show that

$$\sum_{l=0}^k h_n^l S^l_{n,j-1}(\theta^*) = E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right] + O_{n,j}(h_n^{(p/2)\wedge(k+1)}).$$

We prove this statement in three steps.

(i) First, We show that

$$\sum_{l=0}^k h_n^l \tilde S^l_{n,j-1}(\theta^*) = E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right] + O_{n,j}(h_n^{(p/2)\wedge(k+1)}).$$

Observe that

$$\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix} = \mathcal{D}^{k_0}_{n,j}(\theta^*) - E_{n,j-1}[\mathcal{D}^{k_0}_{n,j}(\theta^*)]$$

$$= \mathcal{M}^{k_0}_{n,j}(\theta^*) + \mathcal{R}^{M,k_0}_{n,j}.$$

Thus, our task reduces to showing that

$$2E_{n,j-1}[\mathcal{M}^{k_0}_{n,j}(\theta^*) \odot \mathcal{R}^{M,k_0}_{n,j}] + E_{n,j-1}[(\mathcal{R}^{M,k_0}_{n,j})^{\otimes 2}] = O_{n,j}(h_n^{p/2}).$$

Moreover, by (14), it suffices to prove that

$$E_{n,j-1}[\mathcal{M}^{k_0,y}_{n,j}(\theta^*) \odot \mathcal{R}^{M,k_0,y}_{n,j}] = O_{n,j}(h_n^{p/2}).$$

For notational convenience, we rewrite

$$\begin{aligned}
\mathcal{M}^{k_0,y}_{n,j}(\theta^*) &= h_n^{-3/2} \sum_{\alpha} (W_\alpha \pi_y)_{n,j-1}(\theta^*) \cdot I^{W_\alpha}_{n,j}, \\
\mathcal{R}^{M,k_0,y}_{n,j} &= h_n^{-3/2} \sum_{\beta} I^{W_\beta}_{n,j}\left(\Delta^n W_\beta \pi_y(Z, \theta^*)\right).
\end{aligned}$$

Consider the typical term in $E_{n,j-1}[\mathcal{M}^{k_0,y}_{n,j}(\theta^*) \odot \mathcal{R}^{M,k_0,y}_{n,j}]$:

$$\begin{aligned}
&h_n^{-3} E_{n,j-1}\Big[(W_\alpha \pi_y)_{n,j-1}(\theta^*) \cdot I^{W_\alpha}_{n,j} \odot I^{W_\beta}_{n,j}\left(\Delta^n W_\beta \pi_y(Z, \theta^*)\right)\Big] \\
&\quad = h_n^{-3} (W_\alpha \pi_y)_{n,j-1}(\theta^*) \cdot E_{n,j-1}\Big[I^{W_\alpha}_{n,j} \odot I^{W_\beta}_{n,j}\left(\Delta^n W_\beta \pi_y(Z, \theta^*)\right)\Big].
\end{aligned}$$

Note that this term vanishes unless $\mathrm{len}_M(W_\alpha) \geq \mathrm{len}_M(W_\beta)$. Moreover, since $\mathrm{len}_L(W_\alpha) \geq 1$ and $\mathrm{len}(W_\beta) = k_0 + 2$, we can assume $\mathrm{ord}(W_\alpha) + \mathrm{ord}(W_\beta) = k_0 + 3$. Furthermore,

$$W_\beta \pi_y \in \begin{cases} C^0_\uparrow & (p = 2k_0), \\ C^1_\uparrow & (p = 2k_0 + 1), \end{cases}$$

thus

$$\Delta^n_t W_\beta \pi_y(Z, \theta^*) = \begin{cases} O_{n,t}(1) & (p = 2k_0), \\ O_{n,t}(h_n^{1/2}) & (p = 2k_0 + 1). \end{cases}$$

Therefore, we have

$$\begin{aligned}
&h_n^{-3} (W_\alpha \pi_y)_{n,j-1}(\theta^*) \cdot E_{n,j-1}\Big[I^{W_\alpha}_{n,j} \odot I^{W_\beta}_{n,j}\left(\Delta^n W_\beta \pi_y(Z, \theta^*)\right)\Big] \\
&= \begin{cases} h_n^{-3} O_{n,j}\Big(1 \times h_n^{k_0+2+1}\Big) & (p = 2k_0), \\ h_n^{-3} O_{n,j}\Big(h_n^{1/2} \times h_n^{k_0+2+1}\Big) & (p = 2k_0 + 1), \end{cases} \\
&\quad \text{(by Lemma 8.2)} \\
&= O_{n,j}(h_n^{p/2}).
\end{aligned}$$

(ii) We then establish that $\mathscr{S}^l = \tilde{\mathscr{S}}^l$ for $l = 0, \cdots, k_0$. To see this, we consider a general stochastic differential equation of the same type:

$$dZ^\circ_t = \begin{pmatrix} dX^\circ_t \\ dY^\circ_t \end{pmatrix} = \begin{pmatrix} A^\circ(Z^\circ_t) \\ {h_n}^\circ(Z^\circ_t) \end{pmatrix} dt + \begin{pmatrix} B^\circ(Z^\circ_t) \\ O \end{pmatrix} dw_t$$

where $A^\circ \in C^\infty_c(\mathbb{R}^{d_z}; \mathbb{R}^{d_X})$, $B^\circ \in C^\infty_c(\mathbb{R}^{d_Z}; \mathbb{R}^{d_X} \otimes \mathbb{R}^r)$, $h^\circ_n \in C^\infty_c(\mathbb{R}^{d_Z}; \mathbb{R}^{d_Y})$ and $Z^\circ_0 = z \in \mathbb{R}^{d_Z}$. Under these conditions, there exists a unique strong solution. For this general case, we define $T^l_\circ(z), U^l_\circ(z), S^l_\circ(z)$ and $\tilde{S}^l_\circ(z)$ analogously to the evaluation of $\mathscr{T}^l, \mathscr{U}^l, \mathscr{S}^l$ and $\tilde{\mathscr{S}}^l$ at $\{\partial_z^{l_A} A^\circ(z), \partial_z^{l_B} B^\circ(z), \partial_z^{l_H} h_n^\circ(z)\}_{l_A,l_B,l_H=0}^\infty$ and z. Then, by following the same argument as in (i), we can show that

$$\sum_{l=0}^k h^l \tilde{S}^l_\circ(z) = E\left[\begin{pmatrix} h^{-1/2}(X_h^\circ - E[X_h^\circ]) \\ h^{-3/2}(Y_h^\circ - E[Y_h^\circ]) \end{pmatrix}^{\otimes 2}\right] + O_P(h^{k+1}), \tag{15}$$

Here, $O_P(h^{k+1})$ denotes tightness in the probabilistic sense, meaning that the sequence of random variables associated with this term is bounded in probability. Note that this usage of O-notation should not be confused with other instances of O-notation in this context. Furthermore, the standard asymptotic expansion yields

$$E\left[\begin{pmatrix} h^{-1/2}X_h^\circ \\ h^{-3/2}Y_h^\circ \end{pmatrix}^{\otimes 2}\right] = \sum_{l=-3}^k h^l T^l_\circ(z) + O_P(h^{k+1}),$$
$$\left(E\left[\begin{pmatrix} h^{-1/2}X_h^\circ \\ h^{-3/2}Y_h^\circ \end{pmatrix}\right]\right)^{\otimes 2} = \sum_{l=-3}^k h^l U^l_\circ(z) + O_P(h^{k+1}).$$

Note that $S^l_\circ = O$ for $l < 0$. Thus, it follows immediately from these expansions together with the variance decomposition formula that

$$\sum_{l=0}^k h^l S^l_\circ(z) = E\left[\begin{pmatrix} h^{-1/2}(X_h^\circ - E[X_h^\circ]) \\ h^{-3/2}(Y_h^\circ - E[Y_h^\circ]) \end{pmatrix}^{\otimes 2}\right] + O_P(h^{k+1}). \tag{16}$$

From (15) and (16), we conclude that

$$S^l_\circ = \tilde{S}^l_\circ$$

for $l = 0, \cdots, k_0$. Since $A^\circ \in C_c^\infty(\mathbb{R}^{d_z}; \mathbb{R}^{d_X})$, $B^\circ \in C_c^\infty(\mathbb{R}^{d_Z}; \mathbb{R}^{d_X} \otimes \mathbb{R}^r)$, $h_n^\circ \in C_c^\infty(\mathbb{R}^{d_Z}; \mathbb{R}^{d_Y})$ and $Z_0^\circ = z \in \mathbb{R}^{d_Z}$ are arbitrary, we can derive

$$\mathscr{S}^l = \tilde{\mathscr{S}}^l.$$

Combining the results from (i) and (ii), we obtain

$$\sum_{l=0}^k h_n^l S^l_{n,j-1}(\theta^*) = E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right] + O_{n,j}(h_n^{(p/2)\wedge(k+1)}).$$

□

Proof of Proposition 6.1. We denote $\mathcal{M}^{1,1}_{n,j}(\theta^*) + \mathcal{R}^1_{n,j} - E_{n,j-1}[\mathcal{R}^1_{n,j}]$ by $\mathscr{M}_{n,j}$. Then it is hold that $E_{n,j-1}[\mathscr{M}_{n,j}] = 0$ and

$$\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix} = \mathcal{M}^{0,1}_{n,j}(\theta^*) + \mathscr{M}_{n,j}.$$

Moreover, by Lemma 8.2, we have $\mathcal{M}^{0,1}_{n,j}(\theta^*) = O_{n,j}(1)$ and $\mathscr{M}_{n,j} = O_{n,j}(h_n^{1/2})$. Thus,

$$\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right]$$

$$
\begin{aligned}
&= \mathcal{M}^{0,1}_{n,j}(\theta^*)^{\otimes 2} - E_{n,j-1}[\mathcal{M}^{0,1}_{n,j}(\theta^*)^{\otimes 2}] \\
&\quad + 2\Big(\mathcal{M}^{0,1}_{n,j}(\theta^*) \odot \mathscr{M}_{n,j} - E_{n,j-1}[\mathcal{M}^{0,1}_{n,j}(\theta^*) \odot \mathscr{M}_{n,j}]\Big) \\
&\quad + \mathscr{M}^{\otimes 2}_{n,j} - E_{n,j-1}[\mathscr{M}^{\otimes 2}_{n,j}] \\
&= \mathcal{M}^{0,1}_{n,j}(\theta^*)^{\otimes 2} - E_{n,j-1}[\mathcal{M}^{0,1}_{n,j}(\theta^*)^{\otimes 2}] + O_{n,j}(h_n^{1/2}).
\end{aligned}
$$

Note that

$$
\mathcal{M}^{0,1}_{n,j}(\theta^*) = \begin{pmatrix} h_n^{-1/2} B_{n,j-1}(\theta_1^*) \cdot I^M_{n,j} \\ h_n^{-3/2}\big(B(\partial_x H)^\star\big)_{n,j-1}(\theta_1^*, \theta_3^*) \cdot I^{ML}_{n,j} \end{pmatrix}.
$$

By Lemma 8.2, we have

$$
\mathcal{M}^{0,1}_{n,j}(\theta^*)^{\otimes 2} - E_{n,j-1}[\mathcal{M}^{0,1}_{n,j}(\theta^*)^{\otimes 2}] = \begin{pmatrix} M_{11} & O \\ O & M_{22} \end{pmatrix},
$$

where

$$
\begin{aligned}
M_{11} &= (h_n^{-1/2} B_{n,j-1}(\theta_1^*) \cdot I^M_{n,j})^{\otimes 2} - E_{n,j-1}[(h_n^{-1/2} B_{n,j-1}(\theta_1^*) \cdot I^M_{n,j})^{\otimes 2}], \\
M_{22} &= (h_n^{-3/2}\big(B(\partial_x H)^\star\big)_{n,j-1}(\theta_1^*, \theta_3^*) \cdot I^{ML}_{n,j})^{\otimes 2} \\
&\qquad - E_{n,j-1}[(h_n^{-3/2}\big(B(\partial_x H)^\star\big)_{n,j-1}(\theta_1^*, \theta_3^*) \cdot I^{ML}_{n,j})^{\otimes 2}].
\end{aligned}
$$

Therefore, we have the following by the orthogonality:

$$
\begin{aligned}
M_n^{1'} &:= -\frac{1}{2n^{1/2}} \sum_{j=1}^n \partial_{\theta_1}(C^{-1})_{n,j-1}(\theta_1^*)[M_{11}] \\
&\quad + O_n(h_n^{1/2}), \\
M_n^{1} &:= -\frac{1}{2n^{1/2}} \sum_{j=1}^n (\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*, \theta_3^*) \left[\begin{pmatrix} M_{11} & O \\ O & M_{22} \end{pmatrix}\right] \\
&\quad + O_n(h_n^{1/2}), \\
M_n^{2} &:= \frac{1}{n^{1/2}} \sum_{j=1}^n C^{-1}_{n,j-1}(\theta_1^*)[\partial_{\theta_2} A_{n,j-1}(\theta_2^*) \odot h_n^{-1/2} B_{n,j-1}(\theta_1^*) \cdot I^M_{n,j}] \\
&\quad + O_n(h_n^{1/2}), \\
M_n^{3} &:= \frac{1}{n^{1/2}} \sum_{j=1}^n 12 V^{-1}_{n,j-1}(\theta_1^*, \theta_3^*)\Big[\partial_{\theta_3} H_{n,j-1}(\theta_3^*) \odot h_n^{-3/2}\Big(\big(B(\partial_x H)^\star\big)_{n,j-1}(\theta_1^*, \theta_3^*) \cdot I^{ML}_{n,j}\Big)\Big] \\
&\quad - \frac{1}{n^{1/2}} \sum_{j=1}^n 6\big((\partial_x H)^\star V^{-1}\big)_{n,j-1}(\theta_1^*, \theta_3^*)\big[h_n^{-1/2}(B_{n,j-1}(\theta_1^*) \cdot I^M_{n,j}) \otimes \partial_{\theta_3} H_{n,j-1}(\theta_3^*)\big] \\
&\quad + O_n(h_n^{1/2}).
\end{aligned}
$$

Finally, we have the desired result from the central limit theorem. □

9 Proof of Proposition 6.2

9.1 Estimation of $\mathbb{Y}_n^{1',k}(\theta_1|\bar{\theta}), \partial_{\theta_1}\mathbb{Y}_n^{1',k}(\theta_1^*|\bar{\theta}), \partial_{\theta_1}^2\mathbb{Y}_n^{1',k}(\theta_1|\bar{\theta})$, $\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta}), \partial_{\theta_1}\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta}), \partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta})$

Recall the definition of $\mathcal{D}_{n,j}^{k-1}$. We use the following estimation:

$$
\begin{aligned}
(\mathcal{D}_{n,j}^{k-1,x})^{\otimes 2}(\bar{\theta}) &= \Big(h_n^{-1/2}(X_{n,j} - \sum_{l=0}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}))\Big)^{\otimes 2} \\
&= \Big(\mathcal{D}_{n,j}^{k-1,x}(\theta^*) - h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)\Big)^{\otimes 2} \\
&= (\mathcal{D}_{n,j}^{k-1,x})^{\otimes 2}(\theta^*) - 2h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*) \odot \mathcal{D}_{n,j}^{k-1,x}(\theta^*) \\
&\quad + \Big(h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)\Big)^{\otimes 2} \\
&= (\mathcal{D}_{n,j}^{k-1,x})^{\otimes 2}(\theta^*) \\
&\quad - 2h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*) \odot h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\
&\quad + O_n^{\bar{\theta}}(h_n\bar{\delta}_2^2 + h_n^2\bar{\delta}_2\bar{\delta} + h_n^3\bar{\delta}^2) \\
&\quad \text{(by Lemma 8.3 and Lemma 7.4)},
\end{aligned}
$$

$$
\begin{aligned}
(\mathcal{D}_{n,j}^{k-1})^{\otimes 2}(\bar{\theta}) &= \begin{pmatrix} h_n^{-1/2}(X_{n,j} - \sum_{l=0}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta})) \\ h_n^{-3/2}(Y_{n,j} - Y_{n,j-1} - hH_{n,j-1}(\bar{\theta}_3) - \sum_{l=2}^{k} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar{\theta})) \end{pmatrix}^{\otimes 2} \\
&= \left(\mathcal{D}_{n,j}^{k-1}(\theta^*) - \begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*) + \sum_{l=2}^{k} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \end{pmatrix}\right)^{\otimes 2} \\
&= (\mathcal{D}_{n,j}^{k-1})^{\otimes 2}(\theta^*) \\
&\quad - 2\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*) + \sum_{l=2}^{k} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \end{pmatrix} \odot \mathcal{D}_{n,j}^{k}(\theta^*) \\
&\quad + \begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*) + \sum_{l=2}^{k} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \end{pmatrix}^{\otimes 2} \\
&= (\mathcal{D}_{n,j}^{k-1})^{\otimes 2}(\theta^*) \\
&\quad - 2\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l \pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*) + \sum_{l=2}^{k} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \end{pmatrix} \\
&\qquad \odot \begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix} \\
&\quad + O_n^{\bar{\theta}}(h_n^{-1}\bar{\delta}_3^2 + \bar{\delta}_3\bar{\delta} + h_n\bar{\delta}^2 + (h_n^{-1/2}\bar{\delta}_3 + h_n^{1/2}\bar{\delta})h_n^{(p/2)\wedge(k-1/2)}) \\
&\quad \text{(by Lemma 8.3 and Lemma 7.4)}.
\end{aligned}
$$

Thus, from Lemma 8.3,

$$
\begin{aligned}
&(\mathcal{D}^{k-1,x}_{n,j})^{\otimes 2}(\bar\theta) - S^{0,x}_{n,j-1}(\theta_1) - \sum_{l=1}^{k-1} h_n^l S^{l,x}_{n,j-1}(\bar\theta) \\
&\quad= \left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right)^{\otimes 2} - E_{n,j-1}\left[\left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right)^{\otimes 2}\right] \\
&\qquad - 2\left(h_n^{-1/2}\textstyle\sum_{l=1}^{k-1} h_n^l (L_0^l\pi_x)_{n,j-1}(\bar\theta\backslash\theta^*)\right) \\
&\qquad\qquad \odot \left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right) \\
&\qquad + 2E_{n,j-1}[\mathcal{D}^{k-1,x}_{n,j}(\theta^*)] \odot \left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right) \\
&\qquad + O_{n,j}(h_n^{(p/2)\wedge k} + h_n\bar\delta) \\
&\quad= \left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right)^{\otimes 2} - E_{n,j-1}\left[\left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right)^{\otimes 2}\right] \\
&\qquad - 2\left(h_n^{-1/2}\textstyle\sum_{l=1}^{k-1} h_n^l (L_0^l\pi_x)_{n,j-1}(\bar\theta)\right) \\
&\qquad\qquad \odot \left(h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\right) \\
&\qquad + O^{\theta_1,\bar\theta}_{n,j}(h_n^{(p/2)\wedge k} + h_n\bar\delta), \\
&(\mathcal{D}^{k-1}_{n,j})^{\otimes 2}(\bar\theta) - S^{0}_{n,j-1}(\theta_1,\theta_3) - \sum_{l=1}^{k-1} h_n^l S^{l}_{n,j-1}(\bar\theta) \\
&\quad= \begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right] \\
&\qquad - 2\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l\pi_x)_{n,j-1}(\bar\theta\backslash\theta^*) \\ h_n^{-3/2}(hH_{n,j-1}(\bar\theta_3\backslash\theta_3^*) + \sum_{l=2}^{k} h_n^l (L_0^l\pi_y)_{n,j-1}(\bar\theta\backslash\theta^*)) \end{pmatrix} \\
&\qquad\qquad \odot \begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix} \\
&\qquad + 2E_{n,j-1}[\mathcal{D}^{k-1}_{n,j}(\theta^*)] \odot \begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix} \\
&\qquad + O_{n,j}(h_n^{(p/2)\wedge k} + h_n^{-1}\bar\delta_3^2 + \bar\delta_3 + \delta_1 + h_n\bar\delta + (h_n^{-1/2}\bar\delta_3)h_n^{(p/2)\wedge(k-1/2)}) \\
&\quad= \begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix}^{\otimes 2}\right] \qquad (17) \\
&\qquad - 2\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l (L_0^l\pi_x)_{n,j-1}(\bar\theta) \\ h_n^{-3/2}(hH_{n,j-1}(\bar\theta_3) + \sum_{l=2}^{k} h_n^l (L_0^l\pi_y)_{n,j-1}(\bar\theta)) \end{pmatrix} \\
&\qquad\qquad \odot \begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}]) \end{pmatrix} \\
&\qquad + O^{\theta_1,\theta_3,\bar\theta}_{n,j}(h_n^{(p/2)\wedge k} + h_n^{-1}\bar\delta_3^2 + \bar\delta_3\bar\delta + \delta_1 + \delta_3 + h_n\bar\delta + (h_n^{-1/2}\bar\delta_3)h_n^{(p/2)\wedge(k-1/2)}).
\end{aligned}
$$

The argument for $\mathbb{Y}_n^{1',k}$ is almost identical to that of $\mathbb{Y}_n^{1,k}$, so we will only prove $\mathbb{Y}_n^{1,k}$ and omit the proof of $\mathbb{Y}_n^{1',k}$.

9.1.1 Estimation of $\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta})$

We have

$$
\begin{aligned}
&\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta})\\
&= -\frac{1}{2n}\sum_{j=1}^{n}(S^0)^{-1}_{n,j-1}(\theta_1\backslash\theta_1^*,\bar{\theta}_3)[(\mathcal{D}^{k-1}_{n,j})^{\otimes 2}(\bar{\theta}) - \sum_{l=1}^{k-1} h_n^l S^l_{n,j-1}(\bar{\theta})] + \log\det S^0_{n,j-1}(\theta_1\backslash\theta_1^*,\bar{\theta}_3)
\end{aligned}
$$

$$
\begin{aligned}
= \ & -\frac{1}{2n}\sum_{j=1}^{n}(S^0)^{-1}_{n,j-1}(\theta_1\backslash\theta_1^*,\theta_3^*)[S^0_{n,j-1}(\theta_1^*,\theta_3^*)] + \log\det S^0_{n,j-1}(\theta_1\backslash\theta_1^*,\theta_3^*)\\
&-\frac{1}{2n}\sum_{j=1}^{n}(S^0)^{-1}_{n,j-1}(\theta_1\backslash\theta_1^*,\theta_3^*)\\
&\qquad\cdot\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2}\right]\right]\\
&+\frac{1}{n}\sum_{j=1}^{n}(S^0)^{-1}_{n,j-1}(\theta_1\backslash\theta_1^*,\theta_3^*)\left[\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1}h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta}))\\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3)+\sum_{l=2}^{k}h_n^l(L_0^l\pi_y)_{n,j-1}(\bar{\theta}))\end{pmatrix}\right.\\
&\qquad\qquad\left.\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right]\\
&+O_n^{\theta_1,\bar{\theta}}(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\bar{\delta}_3\bar{\delta}+h_n\bar{\delta}+(h_n^{-1/2}\bar{\delta}_3)h_n^{(p/2)\wedge(k-1/2)})\\
&\text{(by (17))}
\end{aligned}
$$

$$
\begin{aligned}
= \ & -\frac{1}{2n}\sum_{j=1}^{n}(S^0)^{-1}_{n,j-1}(\theta_1\backslash\theta_1^*,\theta_3^*)[S^0_{n,j-1}(\theta_1^*,\theta_3^*)] + \log\det S^0_{n,j-1}(\theta_1\backslash\theta_1^*,\theta_3^*)\\
&+O_n^{\theta_1,\bar{\theta}}\Big(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\bar{\delta}_3\bar{\delta}+h_n\bar{\delta}+(h_n^{-1/2}\bar{\delta}_3)h_n^{(p/2)\wedge(k-1/2)}+\frac{1}{(nh)^{1/2}}\Big)\\
&\text{(by Lemma 7.6)}\\
=\ &\mathbb{Y}^1(\theta_1)+O_n^{\theta_1,\bar{\theta}}(h_n^{-1}\bar{\delta}_3^2+\bar{\delta}_3+h_n^{\epsilon})\\
&\text{(by Lemma 7.7).}
\end{aligned}
$$

9.1.2 Estimation of $\partial_{\theta_1}\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta})$

Note that

$$
\partial_{\theta_1}\log\det S^0_{n,j-1}(\theta_1^*,\bar{\theta}_3) = -(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\bar{\theta}_3)[S^0_{n,j-1}(\theta_1^*,\bar{\theta}_3)].
$$

From this, we obtain

$$
\begin{aligned}
&\partial_{\theta_1}\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta})\\
&= -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\bar{\theta}_3)[(\mathcal{D}^{k-1}_{n,j})^{\otimes 2}(\bar{\theta}) - S^0_{n,j-1}(\theta_1^*,\bar{\theta}_3) - \sum_{l=1}^{k-1}h_n^l S^l_{n,j-1}(\bar{\theta})]
\end{aligned}
$$

$$
\begin{aligned}
= & -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\theta_3^*) \\
& \qquad \cdot \left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2}\right]\right] \\
& + \frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\theta_3^*)\left[E_{n,j-1}[\mathcal{D}_{n,j}^{k-1}(\theta^*)]\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right] \\
& + \frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\theta_3^*)\left[\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1}h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*)+\sum_{l=2}^{k}h_n^l(L_0^l\pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*))\end{pmatrix}\right. \\
& \qquad\qquad \left.\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right] \\
& + O_n^{\bar{\theta}}(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\bar{\delta}_3+h_n\bar{\delta}) \\
& \text{(by (17))}
\end{aligned}
$$

$$
\begin{aligned}
= & \frac{1}{n^{1/2}}M_n^1 \\
& + \frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\theta_3^*)\left[\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1}h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*)+\sum_{l=2}^{k}h_n^l(L_0^l\pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*))\end{pmatrix}\right. \\
& \qquad\qquad \left.\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right] \\
& + O_n^{\bar{\theta}}(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\bar{\delta}_3+h_n\bar{\delta}) \\
& \text{(by Lemma 8.3 and Lemma 7.6).}
\end{aligned}
$$

Here, it follows from Lemma 7.6 that

$$
\partial_{\bar{\theta}}^l\left(\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\theta_3^*)\left[\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1}h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta}) \\ h_n^{-3/2}(\sum_{l=2}^{k}h_n^l(L_0^l\pi_y)_{n,j-1}(\bar{\theta}))\end{pmatrix}\right. \\ & \qquad\qquad \left.\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right]\end{aligned}\right) = O_n^{\bar{\theta}}\Big(\frac{h_n^{1/2}}{n^{1/2}}\Big),
$$

$$
\partial_{\bar{\theta}_3}^l\left(\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\theta_3^*)\left[\begin{pmatrix} O \\ h_n^{-1/2}H_{n,j-1}(\bar{\theta}_3)\end{pmatrix}\right. \\ & \qquad\qquad \left.\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right]\end{aligned}\right) = O_n^{\bar{\theta}_3}\Big(\frac{h_n^{-1/2}}{n^{1/2}}\Big)
$$

for $l = 0, 1, 2$. By Lemma 7.4,

$$\frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*, \theta_3^*)\left[\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1} h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3\backslash\theta_3^*) + \sum_{l=2}^{k} h_n^l(L_0^l\pi_y)_{n,j-1}(\bar{\theta}\backslash\theta^*)) \end{pmatrix} \odot \begin{pmatrix} h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}]) \\ h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}]) \end{pmatrix}\right]$$

$$= O_n^{\bar{\theta}}\Big(\frac{h_n^{-1/2}\bar{\delta}_3 + h_n^{1/2}\bar{\delta}}{n^{1/2}}\Big).$$

Henceforth, this calculation will simply be referred to as "by Lemma 7.6 and Lemma 7.4." Thus,

$$\begin{aligned}\partial_{\theta_1}\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta}) &= \frac{1}{n^{1/2}}M_n^1 + O_n^{\bar{\theta}}\Big(h_n^{-1}\bar{\delta}_3^2 + \bar{\delta}_3 + h_n\bar{\delta} + h_n^{(p/2)\wedge k} + \frac{h^{1/2}}{n^{1/2}}\bar{\delta}\Big) \\ &\quad \text{(by Lemma 7.6).} \\ &= O_n^{\bar{\theta}}\Big(h_n^{-1}\bar{\delta}_3^2 + \bar{\delta}_3 + h_n\bar{\delta} + h_n^{(p/2)\wedge k} + \frac{1}{n^{1/2}}\Big).\end{aligned}$$

9.1.3 Estimation of $\partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta})$

We have

$$\begin{aligned}&\partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta}) \\ &= -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1, \bar{\theta}_3)[\partial_{\theta_1}S_{n,j-1}^0(\theta_1, \bar{\theta}_3)] \\ &\quad -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}^2(S^0)^{-1})_{n,j-1}(\theta_1, \bar{\theta}_3)[(\mathcal{D}_{n,j}^{k-1})^{\otimes 2}(\bar{\theta}) - S_{n,j-1}^0(\theta_1, \bar{\theta}_3) - \sum_{l=1}^{k-1} h_n^l S_{n,j-1}^l(\bar{\theta})].\end{aligned}$$

Then,

$$\begin{aligned}&-\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1, \bar{\theta}_3)[\partial_{\theta_1}S_{n,j-1}^0(\theta_1, \bar{\theta}_3)] \\ &\qquad = -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*, \bar{\theta}_3)[\partial_{\theta_1}S_{n,j-1}^0(\theta_1^*, \bar{\theta}_3)] + O_n^{\theta_1,\bar{\theta}}(\delta_1) \\ &\qquad\quad \text{(by Lemma 7.4)} \\ &\qquad = -\Gamma^1(\theta_1^*|\bar{\theta}_3) + O_n^{\theta_1,\bar{\theta}}(h_n^\epsilon + \delta_1) \\ &\qquad\quad \text{(by Lemma 7.7).}\end{aligned}$$

Furthermore, we have

$$-\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}^2(S^0)^{-1})_{n,j-1}(\theta_1, \bar{\theta}_3)[(\mathcal{D}_{n,j}^{k-1})^{\otimes 2}(\bar{\theta}) - S_{n,j-1}^0(\theta_1, \bar{\theta}_3) - \sum_{l=1}^{k-1} h_n^l S_{n,j-1}^l(\bar{\theta})]$$

$$
\begin{aligned}
&= -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}^2(S^0)^{-1})_{n,j-1}(\theta_1,\bar{\theta}_3)\\
&\qquad\cdot\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2} - E_{n,j-1}\left[\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}^{\otimes 2}\right]\right]\\
&\quad+\frac{1}{n}\sum_{j=1}^{n}(\partial_{\theta_1}^2(S^0)^{-1})_{n,j-1}(\theta_1,\bar{\theta}_3)\left[\begin{pmatrix} h_n^{-1/2}\sum_{l=1}^{k-1}h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta}))\\ h_n^{-3/2}(hH_{n,j-1}(\bar{\theta}_3)+\sum_{l=2}^{k}h_n^l(L_0^l\pi_y)_{n,j-1}(\bar{\theta}))\end{pmatrix}\right.\\
&\qquad\left.\odot\begin{pmatrix} h_n^{-1/2}(X_{n,j}-E_{n,j-1}[X_{n,j}])\\ h_n^{-3/2}(Y_{n,j}-E_{n,j-1}[Y_{n,j}])\end{pmatrix}\right]\\
&\quad+O_{n,j}^{\theta_1,\bar{\theta}}\left(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\delta_1+\bar{\delta}_3+h_n\bar{\delta}\right)\\
&\quad\text{(by (17))}\\
&=O_{n,j}^{\theta_1,\bar{\theta}}\left(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\delta_1+\bar{\delta}_3+h_n\bar{\delta}+\frac{1}{n^{1/2}}\right)\\
&\quad\text{(by Lemma 7.6).}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta}) &= -\frac{1}{2n}\sum_{j=1}^{n}(\partial_{\theta_1}(S^0)^{-1})_{n,j-1}(\theta_1^*,\bar{\theta}_3)[\partial_{\theta_1}S_{n,j-1}^0(\theta_1^*,\bar{\theta}_3)]+O_n^{\theta_1,\bar{\theta}}(\delta_1)\\
&\quad+O_{n,j}^{\theta_1,\bar{\theta}}\left(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\delta_1+\bar{\delta}_3+h_n\bar{\delta}+\frac{1}{n^{1/2}}\right)\\
&=\Gamma^1(\theta_1^*|\bar{\theta}_3)+O_n^{\theta_1,\bar{\theta}}\left(h_n^{\epsilon}+\delta_1+\bar{\delta}_3+h_n^{-1}\bar{\delta}_3\right)
\end{aligned}
$$

holds. Finally, we have

$$
\begin{aligned}
\partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta}) &= \partial_{\theta_1}^2\mathbb{Y}_n^{1,k}(\theta_1^*|\bar{\theta})+O_{n,j}^{\theta_1,\bar{\theta}}\left(h_n^{(p/2)\wedge k}+h_n^{-1}\bar{\delta}_3^2+\delta_1+\bar{\delta}_3+h_n\bar{\delta}+\frac{1}{n^{1/2}}\right)\\
&=\Gamma^1(\theta_1^*|\bar{\theta}_3)+O_n^{\theta_1,\bar{\theta}}\left(h_n^{\epsilon}+\delta_1+\bar{\delta}_3+h_n^{-1}\bar{\delta}_3\right).
\end{aligned}
$$

9.2 Estimation of $\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}), \partial_{\theta_2}\mathbb{Y}_n^{2,k}(\theta_2^*|\bar{\theta}), \partial_{\theta_2}^2\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta})$

Recall

$$
\mathcal{D}_{n,j}^{k,x}(\theta_2,\bar{\theta}) = h_n^{-1/2}\left(X_{n,j}-X_{n,j-1}-h_nA_{n,j-1}(\theta_2)-\sum_{l=2}^{k}h_n^l(L_0^l\pi_x)_{n,j-1}(\bar{\theta})\right).
$$

9.2.1 Estimation of $\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta})$

From the definition of $\mathcal{D}_{n,j}^{x,k}$, we can easily derive the following:

$$
\begin{aligned}
\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}) &= -\frac{1}{2nh}\sum_{j=1}^{n}C_{n,j-1}^{-1}(\bar{\theta}_1)[(\mathcal{D}_{n,j}^{k,x})^{\otimes 2}(\theta_2\backslash\theta_2^*,\bar{\theta})]\\
&= -\frac{1}{2n}\sum_{j=1}^{n}C_{n,j-1}^{-1}(\theta_1^*)[A_{n,j-1}^{\otimes 2}(\theta_2^*\backslash\theta_2)]+O_n^{\theta_2,\bar{\theta}}(h_n^{1/2})\\
&\quad\text{(by Lemma 7.4)}
\end{aligned}
$$

$$= \mathbb{Y}^2(\theta_2) + O_n^{\theta_2}(\bar{\delta}_1 + h_n^{\epsilon})$$
(by Lemma 7.7).

9.2.2 Estimation of $\partial_{\theta_2}\mathbb{Y}_n^{2,k}(\theta_2^*|\bar{\theta})$

It is obvious that

$$\mathcal{D}_{n,j}^{k,x}(\theta_2^*, \theta^*) - E_{n,j-1}[\mathcal{D}_{n,j}^{k,x}(\theta_2^*, \theta^*)] = h_n^{-1/2}(X_{n,j} - E_{j-1}[X_{n,j}]).$$

We have

$$\begin{aligned}
\partial_{\theta_2}\mathbb{Y}_n^{2,k}(\theta_2^*|\bar{\theta}) &= -\frac{1}{nh}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}\mathcal{D}_{n,j}^{k,x}(\theta_2^*, \bar{\theta}) \odot \mathcal{D}_{n,j}^{k,x}(\theta_2^*, \bar{\theta})] \\
&= \frac{1}{nh^{1/2}}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}A_{n,j-1}(\theta_2^*) \odot \mathcal{D}_{n,j}^{k,x}(\theta_2^*, \bar{\theta})] \\
&= \frac{1}{nh^{1/2}}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}A_{n,j-1}(\theta_2^*) \odot \mathcal{D}_{n,j}^{k,x}(\theta_2^*, \theta^*)] + O_n^{\bar{\theta}}(h_n\bar{\delta}) \\
&\quad \text{(by Lemma 7.4)} \\
&= \frac{1}{nh^{1/2}}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}A_{n,j-1}(\theta_2^*) \odot h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}])] \\
&\quad + O_n^{\bar{\theta}}(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta}) \\
&\quad \text{(by Lemma 8.3)} \\
&= \frac{1}{(nh)^{1/2}}M_n^2 + O_n^{\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{\bar{\delta}_1}{(nh)^{1/2}}\Big) \\
&\quad \text{(by Lemma 7.6 and Lemma 7.4)} \\
&= O_n^{\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{1}{(nh)^{1/2}}\Big) \\
&\quad \text{(by Lemma 7.6).}
\end{aligned}$$

9.2.3 Estimation of $\partial_{\theta_2}^2\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta})$

We have

$$\begin{aligned}
\partial_{\theta_2}^2\mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}) = &-\frac{1}{nh}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}^2\mathcal{D}_{n,j}^{k,x}(\theta_2, \bar{\theta}) \odot \mathcal{D}_{n,j}^{k,x}(\theta_2, \bar{\theta})] \\
&-\frac{1}{nh}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[(\partial_{\theta_2}D^{k,x})_{n,j}^{\otimes 2}(\theta_2, \bar{\theta})] \\
= &\frac{1}{nh^{1/2}}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}^2 A_{n,j-1}(\theta_2) \odot \mathcal{D}_{n,j}^{k,x}(\theta_2, \bar{\theta})] \\
&-\frac{1}{n}\sum_{j=1}^{n} C_{n,j-1}^{-1}(\bar{\theta}_1)[(\partial_{\theta_2}A)_{n,j-1}^{\otimes 2}(\theta_2)].
\end{aligned}$$

By mimicking the argument of $\partial_{\theta_1}^2 \mathbb{Y}_n^{1,k}(\theta_1|\bar{\theta})$, we derive

$$
\begin{aligned}
&-\frac{1}{n}\sum_{j=1}^n C_{n,j-1}^{-1}(\bar{\theta}_1)[(\partial_{\theta_2}A)_{n,j-1}^{\otimes 2}(\theta_2)] = -\frac{1}{n}\sum_{j=1}^n C_{n,j-1}^{-1}(\bar{\theta}_1)[(\partial_{\theta_2}A)_{n,j-1}^{\otimes 2}(\theta_2^*)] + O_n^{\theta_2,\bar{\theta}}(\delta_2) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad = \Gamma^2(\theta_2^*|\bar{\theta}) + O_n^{\theta_2,\bar{\theta}}(h_n^\epsilon + \delta_2), \\
&\frac{1}{nh^{1/2}}\sum_{j=1}^n C_{n,j-1}^{-1}(\bar{\theta}_1)[\partial_{\theta_2}^2 A_{n,j-1}(\theta_2) \odot \mathcal{D}_{n,j}^{k,x}(\theta_2,\bar{\theta})] \\
&\qquad = O_n^{\theta_2,\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{1}{(nh)^{1/2}} + \delta_2\Big).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\partial_{\theta_2}^2 \mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}) = &\ -\frac{1}{n}\sum_{j=1}^n C_{n,j-1}^{-1}(\bar{\theta}_1)[(\partial_{\theta_2}A)_{n,j-1}^{\otimes 2}(\theta_2^*)] \\
&+ O_n^{\theta_2,\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{1}{(nh)^{1/2}} + \delta_2\Big). \\
= &\ \Gamma^2(\theta_2^*|\bar{\theta}) + O_n^{\theta_2,\bar{\theta}}(h_n^\epsilon + \delta_2)
\end{aligned}
$$

and

$$
\begin{aligned}
\partial_{\theta_2}^2 \mathbb{Y}_n^{2,k}(\theta_2|\bar{\theta}) &= \partial_{\theta_2}^2 \mathbb{Y}_n^{2,k}(\theta_2^*|\bar{\theta}) \\
&\quad + O_n^{\theta_2,\bar{\theta}}\Big(h_n^{((p-1)/2)\wedge(k+1/2)} + h_n\bar{\delta} + \frac{1}{(nh)^{1/2}} + \delta_2\Big) \\
&= \Gamma^2(\theta_2^*|\bar{\theta}) + O_n^{\theta_2,\bar{\theta}}(h_n^\epsilon + \delta_2).
\end{aligned}
$$

9.3 Estimation of $\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta}), \partial_{\theta_3}\mathbb{Y}_n^{3,k}(\theta_3^*|\bar{\theta}), \partial_{\theta_3}^2 \mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta})$

Recall

$$
\mathcal{D}_{n,j}^{k,y}(\theta_3,\bar{\theta}_3) = h_n^{-3/2}\Big(Y_{n,j} - Y_{n,j-1} - hH_{n,j-1}(\theta_3) - \sum_{l=2}^{k+1} h_n^l (L_0^l \pi_y)_{n,j-1}(\bar{\theta})\Big).
$$

9.3.1 Estimation of $\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta})$

It is trivial that

$$
\begin{aligned}
\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta}) = &\ -\frac{h_n}{2n}\sum_{j=1}^n 12V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[(\mathcal{D}_{n,j}^{k,y})^{\otimes 2}(\theta_3\backslash\theta_3^*,\bar{\theta})] \\
&+ \frac{h_n}{2n}\sum_{j=1}^n 12(\partial_x H)^\star V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\mathcal{D}_{n,j}^{k,x}(\bar{\theta}) \otimes \mathcal{D}_{n,j}^{k,y}(\theta_3\backslash\theta_3^*,\bar{\theta})] \\
= &\ -\frac{1}{2n}\sum_{j=1}^n 12V_{n,j-1}^{-1}(\theta_1^*,\bar{\theta}_3)[H_{n,j-1}^{\otimes 2}(\theta_3\backslash\theta_3^*)] + O_n^{\theta_3,\bar{\theta}}(\bar{\delta}_1 + h_n^{1/2}) \\
= &\ \mathbb{Y}^3(\theta_3|\bar{\theta}_3) + O_n^{\theta_3,\bar{\theta}}(\bar{\delta}_1 + h_n^\epsilon) \\
&\ \text{(by Lemma 7.7).}
\end{aligned}
$$

9.3.2 Estimation of $\partial_{\theta_3}\mathbb{Y}_n^{3,k}(\theta_3^*|\bar{\theta})$

As mentioned about $\mathcal{D}_{n,j}^{k,x}$, it also holds that

$$\mathcal{D}_{n,j}^{k,y}(\theta_3^*,\theta^*) - E_{n,j-1}[\mathcal{D}_{n,j}^{k,y}(\theta_3^*,\theta^*)] = h_n^{-3/2}(Y_{n,j} - E_{j-1}[Y_{n,j}]).$$

Considering this, we have

$$\begin{aligned}
\partial_{\theta_3}\mathbb{Y}_n^{3,k}(\theta_3^*|\bar{\theta}) &= \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\partial_{\theta_3}H_{n,j-1}(\theta_3^*)\odot\mathcal{D}_{n,j}^{k,y}(\theta_3^*,\bar{\theta})] \\
&\quad - \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 6(\partial_x H)^\star V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\mathcal{D}_{n,j}^{k,x}(\bar{\theta})\otimes\partial_{\theta_3}H_{n,j-1}(\theta_3^*)] \\
&= \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\partial_{\theta_3}H_{n,j-1}(\theta_3^*)\odot\mathcal{D}_{n,j}^{k,y}(\theta_3^*,\theta^*)] \\
&\quad - \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 6(\partial_x H)^\star V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\mathcal{D}_{n,j}^{k,x}(\theta^*)\otimes\partial_{\theta_3}H_{n,j-1}(\theta_3^*)] \\
&\quad + O_n^{\bar{\theta}}(h_n(\bar{\delta}_2 + \bar{\delta}1_{\{k\geq 2\}})) \\
&\quad \text{(by Lemma 7.4)} \\
&= \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\partial_{\theta_3}H_{n,j-1}(\theta_3^*)\odot h_n^{-3/2}(Y_{n,j} - E_{n,j-1}[Y_{n,j}])] \\
&\quad - \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 6(\partial_x H)^\star V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[h_n^{-1/2}(X_{n,j} - E_{n,j-1}[X_{n,j}])\otimes\partial_{\theta_3}H_{n,j-1}(\theta_3^*)] \\
&\quad + O_n^{\bar{\theta}}(h_n(\bar{\delta}_2 + \bar{\delta}1_{\{k\geq 2\}}) + h_n^{((p+1)/2)\wedge(k+3/2)}) \\
&\quad \text{(by Lemma 8.3)} \\
&= M_n^3 + O_n^{\bar{\theta}}\Big(h_n(\bar{\delta}_2 + \bar{\delta}1_{\{k\geq 2\}}) + h_n^{((p+1)/2)\wedge(k+3/2)} + \frac{h_n^{1/2}}{n^{1/2}}\bar{\delta}\Big) \\
&\quad \text{(by Lemma 7.6 and Lemma 7.4)} \\
&= O_n^{\bar{\theta}}\Big(h_n(\bar{\delta}_2 + \bar{\delta}1_{\{k\geq 2\}}) + h_n^{((p+1)/2)\wedge(k+3/2)} + \frac{h_n^{1/2}}{n^{1/2}}\Big).
\end{aligned}$$

9.3.3 Estimation of $\partial_{\theta_3}^2\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta})$

By mimicking the above argument, it is easy to see that

$$\begin{aligned}
\partial_{\theta_3}^2\mathbb{Y}_n^{3,k}(\theta_3|\bar{\theta}) = &-\frac{1}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[(\partial_{\theta_3}H)_{n,j-1}^{\otimes 2}(\theta_3)] \\
&+ \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\partial_{\theta_3}^2 H_{n,j-1}(\theta_3)\odot\mathcal{D}_{n,j}^{k,y}(\theta_3,\bar{\theta})] \\
&- \frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 6(\partial_x H)^\star V_{n,j-1}^{-1}(\bar{\theta}_1,\bar{\theta}_3)[\mathcal{D}_{n,j}^{k,x}(\bar{\theta})\otimes\partial_{\theta_3}^2 H_{n,j-1}(\theta_3)]
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[(\partial_{\theta_3}H)_{n,j-1}^{\otimes 2}(\theta_3^*)] \\
&\quad -\frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[\partial_{\theta_3}^2 H_{n,j-1}(\theta_3)\odot \mathcal{D}_{n,j}^{k,y}(\theta_3^*,\bar\theta)] \\
&\quad -\frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 3(\partial_x H)^\star V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[\mathcal{D}_{n,j}^{k,x}(\bar\theta)\otimes \partial_{\theta_3}^2 H_{n,j-1}(\theta_3)] \\
&\quad + O_n^{\theta_3,\bar\theta}(\delta_3) \\
&= -\frac{1}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[(\partial_{\theta_3}H)_{n,j-1}^{\otimes 2}(\theta_3^*)] \\
&\quad -\frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[\partial_{\theta_3}^2 H_{n,j-1}(\theta_3)\odot \mathcal{D}_{n,j}^{k,y}(\theta^*)] \\
&\quad -\frac{h_n^{1/2}}{n}\sum_{j=1}^{n} 3(\partial_x H)^\star V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[\mathcal{D}_{n,j}^{k,x}(\theta^*)\otimes \partial_{\theta_3}^2 H_{n,j-1}(\theta_3)] \\
&\quad + O_n^{\theta_3,\bar\theta}(\delta_3+h_n\bar\delta) \\
&= -\frac{1}{n}\sum_{j=1}^{n} 12V_{n,j-1}^{-1}(\bar\theta_1,\bar\theta_3)[(\partial_{\theta_3}H)_{n,j-1}^{\otimes 2}(\theta_3^*)] \\
&\quad + O_n^{\theta_3,\bar\theta}\Big(\delta_3+h_n\bar\delta+h_n^{(p+1)/2\wedge(k+1)}+\frac{h_n^{1/2}}{n^{1/2}}\Big) \\
&= -\Gamma^3(\theta_3^*|\bar\theta_1,\bar\theta_3)+O_n^{\theta_3,\bar\theta}(h_n^\epsilon+\delta_3) \\
&\quad \text{(by Lemma 7.7).}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\partial_{\theta_3}^2\mathbb{Y}_n^{3,k}(\theta_3|\bar\theta) &= \partial_{\theta_3}^2\mathbb{Y}_n^{3,k}(\theta_3^*|\bar\theta)+O_n^{\theta_3,\bar\theta}\Big(\delta_3+h_n\bar\delta+h_n^{(p+1)/2\wedge(k+1)}+\frac{h_n^{1/2}}{n^{1/2}}\Big) \\
&= -\Gamma^3(\theta_3^*|\bar\theta_1,\bar\theta_3)+O_n^{\theta_3,\bar\theta}(h_n^\epsilon+\delta_3).
\end{aligned}
$$

References

Susanne Ditlevsen and Adeline 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.

Daniéle Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. Statistics, 20(4):547–557, 1989. ISSN 0233-1888.

Arnaud Gloter and Nakahiro Yoshida. Adaptive and non-adaptive estimation for degenerate diffusion processes. arXiv preprint arXiv:2002.10164, 2020.

Arnaud Gloter and Nakahiro Yoshida. Adaptive estimation for degenerate diffusion processes. Electronic Journal of Statistics, 15(1):1424–1472, 2021.

Arnaud Gloter and Nakahiro Yoshida. Non-adaptive estimation for degenerate diffusion processes. Theory of Probability and Mathematical Statistics, 110:75–99, 2024a.

Arnaud Gloter and Nakahiro Yoshida. Quasi-likelihood analysis for adaptive estimation of a degenerate diffusion process. arXiv preprint arXiv:2402.15256, 2024b.

Yuga Iguchi and Alexandros Beskos. Parameter inference for hypo-elliptic diffusions under a weak design condition. arXiv preprint arXiv:2312.04444, 2023.

Yuga Iguchi, Alexandros Beskos, and Matthew M Graham. Parameter inference for degenerate diffusion processes. Stochastic Processes and their Applications, 174:104384, 2024.

Yuga Iguchi, Alexandros Beskos, and Matthew Graham. Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions. Bernoulli, 31(1):333–358, 2025.

Kengo Kamatani and Masayuki Uchida. Hybrid multi-step estimators for stochastic differential equations based on sampled data. Statistical Inference for Stochastic Processes, 18(2):177–204, 2014.

Mathieu Kessler. Estimation of an ergodic diffusion from discrete observations. Scand. J. Statist., 24(2):211–229, 1997. ISSN 0303-6898.

B. L. S. Prakasa Rao. Asymptotic theory for nonlinear least squares estimator for diffusion processes. Math. Operationsforsch. Statist. Ser. Statist., 14(2):195–209, 1983. ISSN 0323-3944.

B. L. S. Prakasa Rao. Statistical inference from sampled data for stochastic processes. In Statistical inference from stochastic processes (Ithaca, NY, 1987), volume 80 of Contemp. Math., pages 249–284. Amer. Math. Soc., Providence, RI, 1988.

Adeline Samson, Massimiliano Tamborrino, and Irene Tubikanec. Inference for the stochastic fitzhugh-nagumo model from real action potential data via approximate bayesian computation. Computational Statistics & Data Analysis, 204:108095, 2025.

Masayuki Uchida and Nakahiro Yoshida. Adaptive estimation of an ergodic diffusion process based on sampled data. Stochastic Process. Appl., 122(8): 2885–2924, 2012. ISSN 0304-4149. doi: 10.1016/j.spa.2012.04.001. URL http://dx.doi.org/10.1016/j.spa.2012.04.001.

Nakahiro Yoshida. Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. Probab. Theory Related Fields, 92(3): 275–311, 1992. ISSN 0178-8051.

Nakahiro Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. Annals of the Institute of Statistical Mathematics, 63(3):431–479, 2011.