Hybrid estimators for diffusion processes with small noises from reduced data

[1]Yusuke Kaino and [1,2]Masayuki Uchida
[1]Graduate School of Engineering Science and
[2]Center for Mathematical Modeling and Data Science (MMDS), Osaka University
Toyonaka, Osaka 560-8531, Japan

September 7, 2017

Abstract. We deal with the Bayes type estimators and the maximum likelihood type estimators of both drift and volatility parameters for small diffusion processes defined by stochastic differential equations with small perturbations from high frequency data. From the viewpoint of numerical analysis, initial Bayes type estimators for both drift and volatility parameters based on reduced data are required, and adaptive maximum likelihood type estimators with the initial Bayes type estimators, which are called hybrid estimators, are proposed. The asymptotic properties of the initial Bayes type estimators based on reduced data are derived and it is shown that the hybrid estimators have asymptotic normality and convergence of moments. Furthermore, a concrete example and simulation results are given.

Key words and phrases: Bayes type estimator, convergence of moments, diffusion process, discrete time observations, maximum likelihood type estimator, small dispersion parameters

AMS 2000 subject classification: Primary 62F12, 62M05, Secondary 60J60

Abbreviated Title: Hybrid estimator for small diffusions

1 Introduction

We treat a d-dimensional small diffusion process defined by the following stochastic differential equation

$$\begin{cases} dX_t = a(X_t, \alpha)dt + \epsilon b(X_t, \beta)dw_t, \quad t \in [0, T], \quad \epsilon \in (0, 1], \\ X_0 = x_0, \end{cases} \tag{1}$$

where ϵ and T are known constants, x_0 is a deterministic initial condition, w is an r-dimensional standard Wiener process, $\theta = (\alpha, \beta) \in \Theta = \Theta_\alpha \times \Theta_\beta$ with Θ_α and Θ_β being compact convex subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, $a : \mathbb{R}^d \times \Theta_\alpha \to \mathbb{R}^d$ and $b : \mathbb{R}^d \times \Theta_\beta \to \mathbb{R}^d \otimes \mathbb{R}^r$, and $\theta^* = (\alpha^*, \beta^*) \in \mathrm{Int}(\Theta)$ is the true value of θ. Suppose that the parameter spaces Θ_α and Θ_β have locally Lipschitz boundaries, see Adam and Fournier (2003). The data are discrete observations $\mathbb{X}_n = (X_{t_i})_{0 \le i \le n}$ with $t_i = ih_n$, $h_n = T/n$. We will consider the case when $\epsilon \to 0$, $n \to \infty$, $\frac{1}{\epsilon\sqrt{n}} = O(1)$, and there exists $\gamma \in (0, 1]$ satisfying that $\epsilon(\sqrt{n})^\gamma = O(1)$.

A family of small diffusion processes defined by (1) is an important class of dynamical systems with small perturbations. For dynamical systems with small perturbations, see Azencott (1982) and Freidlin and Wentzell (1998). For applications of small diffusion processes to mathematical finance and mathematical biology, see Yoshida (1992b), Kunitomo and Takahashi (2001), Takahashi and Yoshida (2004), Uchida and Yoshida (2004b), Fuchs (2013), Guy et. al. (2014, 2015) and references therein. Statistical inference for continuously observed small diffusion processes is well developed by Kutoyants (1984, 1994), Yoshida (1992a, 2003), Iacus (2000), Iacus and Kutoyants (2001), Uchida and Yoshida (2004a), Brouste et. al. (2014) and references therein. Furthermore, there are a number of researches on parametric inference for discretely observed small diffusion processes, see Genon-Catalot (1990), Laredo (1990), Sørensen (2000, 2012), Sørensen and Uchida (2003), Uchida (2003, 2004, 2006, 2008), Gloter and Sørensen (2009), Guy et. al. (2014) and Nomura and Uchida (2016). For statistical inference of dynamical systems with small Lévy noises, see Long et. al. (2013), Ma and Yang (2014), Long et. al. (2016) and Yang (2017). Furthermore, from the viewpoint of numerical analysis, it is also crucial to investigate adaptive maximum likelihood type estimators of both drift and volatility parameters for discretely observed diffusion processes. For discretely observed ergodic/non-ergodic diffusion type processes, various types of adaptive estimators have been studied; see Yoshida (1992c, 2011), Kessler (1995), Uchida and Yoshida (2012, 2014), Kamatani and Uchida (2015), Kamatani et. al. (2016) and Nomura and Uchida (2016).

In order to explain the goal of this paper, we first review the joint estimation of both drift and volatility parameters for discretely observed small diffusion processes. Set $A^{\otimes 2} = AA^\star$ and $C[A] = \mathrm{tr}(CA^\star)$ for matrices A and C of the same size, where $\star$ means the transpose. Let $B(x,\beta) = bb^\star(x,\beta)$, $\Delta X_i = X_{t_i} - X_{t_{i-1}}$, $a_{i-1}(\alpha) = a(X_{t_{i-1}},\alpha)$ and $B_{i-1}(\beta) = B(X_{t_{i-1}},\beta)$. The quasi-log likelihood function is defined as

$$U_{\epsilon,n}(\alpha,\beta) \quad = \quad -\frac{1}{2}\sum_{i=1}^{n}\left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\right\}.$$

The joint maximum likelihood (ML) type estimators $\hat{\alpha}_{\epsilon,n}^{(J)}$ and $\hat{\beta}_{\epsilon,n}^{(J)}$ are defined as

$$U_{\epsilon,n}(\hat{\alpha}_{\epsilon,n}^{(J)},\hat{\beta}_{\epsilon,n}^{(J)}) \quad = \quad \sup_{\alpha\in\Theta_\alpha,\beta\in\Theta_\beta} U_{\epsilon,n}(\alpha,\beta).$$

Sørensen and Uchida (2003) showed that under some regularity conditions, as $\epsilon \to 0$, $n \to \infty$ and $\frac{1}{\epsilon\sqrt{n}} = O(1)$,

$$\left(\epsilon^{-1}(\hat{\alpha}_{\epsilon,n}^{(J)} - \alpha^*), \sqrt{n}(\hat{\beta}_{\epsilon,n}^{(J)} - \beta^*)\right) \xrightarrow{d} (\zeta_1,\zeta_2) \sim N_{p+q}(0, I(\theta^*)^{-1}), \tag{2}$$

where $\xrightarrow{d}$ means convergence in distribution, $N_{p+q}(0, I(\theta^*)^{-1})$ is the normal random variable with mean zero and the covariance matrix $I(\theta^*)^{-1}$ and $I(\theta^*)$ is the asymptotic Fisher information matrix, see Section 2 below. From the viewpoint of numerical analysis, the joint ML type estimator is unstable when the dimension of Θ is large. For that reason, we consider the adaptive ML type estimators. In the same way as Uchida and Yoshida (2012) for ergodic diffusion models, the quasi-log likelihood functions are defined as

$$V_{\epsilon,n}^{(1)}(\beta) \quad = \quad -\frac{1}{2}\sum_{i=1}^{n}\left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B_{i-1}^{-1}(\beta)\left[(\Delta X_i)^{\otimes 2}\right]\right\},$$

$$V_{\epsilon,n}^{(2)}(\alpha,\beta) \quad = \quad -\frac{1}{2}\sum_{i=1}^{n}(\epsilon^2 h_n)^{-1} B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right].$$

The adaptive ML type estimators $\hat{\alpha}_{\epsilon,n}^{(E)}$ and $\hat{\beta}_{\epsilon,n}^{(E)}$ are defined as

$$V_{\epsilon,n}^{(1)}(\hat{\beta}_{\epsilon,n}^{(E)}) = \sup_{\beta\in\Theta_\beta} V_{\epsilon,n}^{(1)}(\beta), \tag{3}$$

$$V_{\epsilon,n}^{(2)}(\hat{\alpha}_{\epsilon,n}^{(E)}, \hat{\beta}_{\epsilon,n}^{(E)}) = \sup_{\alpha\in\Theta_\alpha} V_{\epsilon,n}^{(2)}(\alpha, \hat{\beta}_{\epsilon,n}^{(E)}). \tag{4}$$

Then, under some regularity conditions, as $\epsilon \to 0$, $n \to \infty$ and $\frac{1}{\epsilon^2\sqrt{n}} = o(1)$,

$$\left(\epsilon^{-1}(\hat{\alpha}_{\epsilon,n}^{(E)} - \alpha^*), \sqrt{n}(\hat{\beta}_{\epsilon,n}^{(E)} - \beta^*)\right) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p+q}(0, I(\theta^*)^{-1}).$$

In the case of small diffusion process, the adaptive ML type estimators (3) and (4), which are obtained by the same method as the case of the ergodic diffusion processes, are worse than the joint ML type estimators $\hat{\alpha}_{\epsilon,n}^{(J)}$ and $\hat{\beta}_{\epsilon,n}^{(J)}$ since the stronger condition $\frac{1}{\epsilon^2\sqrt{n}} = o(1)$ is needed to get the same asymptotic properties as (2). Therefore, the aim of this paper is to propose the adaptive ML type estimator which has the same asymptotic normality as (2) under $\frac{1}{\epsilon\sqrt{n}} = O(1)$ from the viewpoint of numerical analysis.

In order to compute the adaptive ML type estimators, it is indispensable to get a suitable initial estimator for optimization of quasi-log likelihood function. Nomura and Uchida (2016) obtained the initial Bayes type estimator from full data of small diffusion processes. They considered the hybrid estimator with the initial Bayes type estimator and showed that the hybrid estimator has asymptotic normality and convergence of moments. However, it takes much time to compute the initial Bayes type estimator when the sample size is large. Kutoyants (2017) considered the multi-step ML type estimation procedure for ergodic diffusion processes from continuous path data on $[0, T]$. He proposed the multi-step estimator with the initial estimator derived from the reduced continuous path data on $[0, T_0]$ for $T_0 \leq T$ and showed asymptotic efficiency of the multi-step ML type estimator as $T_0 \to \infty$. Kaino et al. (2017) studied the initial Bayes type estimator based on reduced sampled data for a discretely observed ergodic diffusion processes and they showed asymptotic normality and convergence of moments for the adaptive ML type estimator with the initial Bayes type estimator.

In this paper, we consider the initial Bayes type estimator based on reduced sampled data for a discretely observed small diffusion process by applying the initial estimator with reduced data for a ergodic diffusion process in Kutoyants (2017) and Kaino et al. (2017) to the initial Bayes type estimator for a small diffusion model from the viewpoint of numerical analysis. The adaptive ML type estimator with the initial Bayes type estimator, which is called the hybrid estimator with the initial Bayes type estimator, is proposed for a small diffusion process. Moreover, it is shown that the proposed hybrid estimator has asymptotic normality and convergence of moments by applying the Ibragimov-Has'minskii program (1972a,b, 1981) and the polynomial type large deviation inequality for statistical random field in Yoshida (2011) to the case of discretely observed small diffusion processes. Needless to say, the convergence of moments and the polynomial type large deviation inequality of statistical random field play an important part in the proof of the mathematical validity of asymptotic expansions and asymptotic unbiasedness of information criteria for model selection, see Yoshida (1992a, 1992b), Uchida and Yoshida (2001, 2004a, 2004b, 2006) and Uchida (2010).

This paper is organized as follows. In section 2, we consider an initial Bayes type estimator from reduced data. Moreover, by using both the adaptive ML type estimation from full data and the initial Bayes type estimator from reduced data, we propose hybrid estimators with the initial Bayes type estimator for small diffusion processes and obtain asymptotic results of consistency, asymptotic normality and convergence of moments of hybrid estimators. Section 3 gives an example of a non-linear small diffusion model and simulation studies of the estimators for the model. In section 4, the asymptotic results presented in section 2 are proved.

2 Initial Bayes type estimators with reduced data and hybrid estimators

Although the data are discrete observations $\mathbb{X}_n = (X_{t_i})_{0\le i\le n}$ with $t_i = ih_n$, $h_n = T/n$, from the viewpoint of numerical analysis, we consider initial estimators with reduced data $\mathbb{Y}_{n_0} = (X_{t_i})_{0\le i\le n_0}$, where $n_0 = \left[\frac{n}{c}\right]$ for $c \ge 1$.

For a matrix A, we define $\|A\| = \mathrm{tr}(AA^\star)^{1/2}$ and $|\cdot|$ denotes the Euclidian norm. Let $\xrightarrow{p}$ and $\xrightarrow{d}$ be the convergence in probability and convergence in distribution, respectively. Let X_t^0 be the solution of the ordinary differential equation corresponding to $\epsilon = 0$, i.e., $dX_t^0 = a(X_t^0, \alpha^*)dt$, $X_0^0 = x_0$. Let $C_\uparrow^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ denote the space of all functions f satisfying the following conditions:

(i) $f(x,\theta)$ is an $\mathbb{R}^d$-valued function on $\mathbb{R}^d \times \Theta$ and is continuously differentiable with respect to x and θ up to order k and l, respectively.

(ii) for $|\boldsymbol{n}| = 0,1,\ldots,k$ and $|\boldsymbol{\nu}| = 0,1,\ldots,l$, there exists $C>0$ such that $\sup_{\theta\in\Theta}|\delta^{\boldsymbol{\nu}}\partial^{\boldsymbol{n}} f| \le C(1+|x|)^C$ for all x.

Here, $\boldsymbol{n} = (n_1,\ldots,n_d)$ and $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_l)$ are multi-indices, $l = \dim(\Theta)$, $|\boldsymbol{n}| = n_1 + \cdots + n_d$, $|\boldsymbol{\nu}| = \nu_1 + \cdots + \nu_l$, $\partial^{\boldsymbol{n}} = \partial_1^{n_1}\cdots\partial_d^{n_d}$, $\partial_i = \partial/\partial x_i$, $i = 1,\ldots,d$, $\delta^{\boldsymbol{\nu}} = \delta_1^{\nu_1}\cdots\delta_l^{\nu_l}$, $\delta_j = \partial/\partial\theta_j$, $j = 1,\ldots,l$.

In this paper, we make the assumptions as follows.

[**A1**] **(i)** There exists $K > 0$ such that for all $x, y \in \mathbb{R}^d$,
$$\sup_{\alpha\in\Theta_\alpha}|a(x,\alpha) - a(y,\alpha)| + \sup_{\beta\in\Theta_\beta}\|b(x,\beta) - b(y,\beta)\| \le K|x-y|.$$

(ii) $\inf_{x,\beta}\det B(x,\beta) > 0$.

[**A2**] $a(x,\alpha) \in C_\uparrow^{6,4}(\mathbb{R}^d \times \Theta_\alpha; \mathbb{R}^d)$, $b(x,\beta) \in C_\uparrow^{6,4}(\mathbb{R}^d \times \Theta_\beta; \mathbb{R}^d \otimes \mathbb{R}^r)$.

The quasi log-likelihood functions $U_{\epsilon,n}^{(1)}(\alpha)$ and $U_{\epsilon,n}^{(2)}(\alpha,\beta)$ with reduced data $\mathbb{Y}_{n_0}$, and the quasi log-likelihood functions $U_{\epsilon,n}^{(3)}(\alpha,\beta)$ and $U_{\epsilon,n}^{(4)}(\alpha,\beta)$ with full data $\mathbb{X}_n$ are defined as follows.

$$\begin{aligned}
U_{\epsilon,n_0}^{(1)}(\alpha) &= -\frac{1}{2\epsilon^2 h_n}\sum_{i=1}^{n_0}|\Delta X_i - h_n a_{i-1}(\alpha)|^2, \\
U_{\epsilon,n_0}^{(2)}(\alpha,\beta) &= -\frac{1}{2\epsilon^4 h_n^2}\sum_{i=1}^{n_0}\left\|(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2} - (\epsilon^2 h_n)B_{i-1}(\beta)\right\|^2, \\
U_{\epsilon,n}^{(3)}(\alpha,\beta) &= -\frac{1}{2\epsilon^2 h_n}\sum_{i=1}^{n} B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right], \\
U_{\epsilon,n}^{(4)}(\alpha,\beta) &= -\frac{1}{2}\sum_{i=1}^{n}\left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1}B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\right\}.
\end{aligned}$$

Note that under [$A1$]-[$A2$], as $\frac{1}{\epsilon\sqrt{n}} = O(1)$, uniformly in $\theta \in \Theta$,

$$\begin{aligned}
&\epsilon^2\left\{U_{\epsilon,n_0}^{(1)}(\alpha) - U_{\epsilon,n_0}^{(1)}(\alpha^*)\right\} \xrightarrow{p} \mathbb{Y}^{(1)}(\alpha), \\
&h_n\left\{U_{\epsilon,n_0}^{(2)}(\alpha^*,\beta) - U_{\epsilon,n_0}^{(2)}(\alpha^*,\beta^*)\right\} \xrightarrow{p} \mathbb{Y}^{(2)}(\beta), \\
&\epsilon^2\left\{U_{\epsilon,n}^{(3)}(\alpha,\beta^*) - U_{\epsilon,n}^{(3)}(\alpha^*,\beta^*)\right\} \xrightarrow{p} \mathbb{Y}^{(3)}(\alpha), \\
&h_n\left\{U_{\epsilon,n}^{(4)}(\alpha^*,\beta) - U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)\right\} \xrightarrow{p} \mathbb{Y}^{(4)}(\beta),
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &= -\frac{1}{2}\frac{c}{T}\int_0^{T/c} \left|a(X_t^0,\alpha)-a(X_t^0,\alpha^*)\right|^2 dt,\\
\mathbb{Y}^{(2)}(\beta) &= -\frac{1}{2}\frac{c}{T}\int_0^{T/c} \left\|B(X_t^0,\beta)-B(X_t^0,\beta^*)\right\|^2 dt,\\
\mathbb{Y}^{(3)}(\alpha) &= -\frac{1}{2}\frac{1}{T}\int_0^{T} B(X_t^0,\beta^*)^{-1}\left[\left(a(X_t^0,\alpha)-a(X_t^0,\alpha^*)\right)^{\otimes 2}\right] dt,\\
\mathbb{Y}^{(4)}(\beta) &= -\frac{1}{2}\frac{1}{T}\int_0^{T} \left\{\operatorname{tr}\left[B(X_t^0,\beta)^{-1}B(X_t^0,\beta^*)-I_d\right]+\log\frac{\det B(X_t^0,\beta)}{\det B(X_t^0,\beta^*)}\right\} dt.
\end{aligned}$$

[**A3**] There exist positive constants $\chi^{(1)}, \chi^{(2)}, \chi^{(3)}$ and $\chi^{(4)}$ such that for all $\alpha \in \Theta_\alpha$ and $\beta \in \Theta_\beta$,

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &\leq -\chi^{(1)}|\alpha-\alpha^*|^2,\\
\mathbb{Y}^{(2)}(\beta) &\leq -\chi^{(2)}|\beta-\beta^*|^2,\\
\mathbb{Y}^{(3)}(\alpha) &\leq -\chi^{(3)}|\alpha-\alpha^*|^2,\\
\mathbb{Y}^{(4)}(\beta) &\leq -\chi^{(4)}|\beta-\beta^*|^2.
\end{aligned}$$

[**A4**] We assume that $\frac{1}{\epsilon\sqrt{n}} = O(1)$ as $\epsilon \to 0$ and $n \to \infty$, and there exists $\gamma \in (0,1]$ satisfying that $\epsilon(\sqrt{n})^\gamma = O(1)$. Moreover, $r_2 \leq 2r_1\gamma$ for $r_1, r_2 \in (0,1]$.

The statistical random fields $\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha)$ and $\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\alpha,\beta)$ with reduced data $\mathbb{Y}_{n_0}$ are given by

$$\begin{aligned}
&\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha) = \epsilon^{2-2r_1}U^{(1)}_{\epsilon,n_0}(\alpha),\\
&\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\alpha,\beta) = \frac{1}{(\sqrt{n_0})^{2-2r_2}}U^{(2)}_{\epsilon,n_0}(\alpha,\beta).
\end{aligned}$$

We assume that the prior densities $\pi_1(\alpha)$ and $\pi_2(\beta)$ are continuous and satisfy that $0 < \inf_{\alpha\in\Theta_\alpha}\pi_1(\alpha) \leq \sup_{\alpha\in\Theta_\alpha}\pi_1(\alpha) < \infty$ and $0 < \inf_{\beta\in\Theta_\beta}\pi_2(\beta) \leq \sup_{\beta\in\Theta_\beta}\pi_2(\beta) < \infty$. The initial Bayes type estimators $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$ with reduced data $\mathbb{Y}_{n_0}$ are defined by

$$\begin{aligned}
\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1} &= \frac{\int_{\Theta_\alpha}\alpha\exp\left\{\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha))\right\}\pi_1(\alpha)d\alpha}{\int_{\Theta_\alpha}\exp\left\{\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha)\right\}\pi_1(\alpha)d\alpha},\\
\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} &= \frac{\int_{\Theta_\beta}\beta\exp\left\{\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta)\right\}\pi_2(\beta)d\beta}{\int_{\Theta_\beta}\exp\left\{\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta)\right\}\pi_2(\beta)d\beta}.
\end{aligned}$$

The hybrid estimators $\hat{\alpha}_{\epsilon,n}$ and $\hat{\beta}_{\epsilon,n}$ with full data $\mathbb{X}_n$ are defined by

$$\begin{aligned}
U^{(3)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) &= \sup_{\alpha\in\Theta_\alpha}U^{(3)}_{\epsilon,n}(\alpha,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}),\\
U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\hat{\beta}_{\epsilon,n}) &= \sup_{\beta\in\Theta_\beta}U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta).
\end{aligned}$$

Theorem 1. *Let $r_1, r_2 \in (0,1]$. Assume [A1]–[A4]. Then, for all $M > 0$, as $\epsilon \to 0$ and $n \to \infty$,*

(i) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*)\right|^M\right] < \infty.$

(ii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|(\sqrt{n_0})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)\right|^M\right] < \infty.$

(iii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*)\right|^M\right] < \infty.$

(iv) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*)\right|^M\right] < \infty.$

Remark 1. *It follows from Theorem 1 that when $r_1, r_2 \in (0,1)$, the convergence rates of the initial Bayes type estimators $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$ are ϵ^{r_1} and $\frac{1}{(\sqrt{n_0})^{r_2}}$, respectively, which means that the initial Bayes type estimators do not have optimal rates. However, the hybrid estimators $\hat{\alpha}_{\epsilon,n}$ and $\hat{\beta}_{\epsilon,n}$ have optimal rates, ϵ and $\frac{1}{\sqrt{n}}$, respectively.*

Let

$$\begin{aligned}
I(\theta^*) &= \begin{pmatrix} \left(I_a^{ij}(\theta^*)\right)_{1\le i,j\le p} & 0 \\ 0 & \left(I_b^{ij}(\beta^*)\right)_{1\le i,j\le q} \end{pmatrix}, \\
I_a^{ij}(\theta^*) &= \int_0^T \left(\partial_{\alpha_i} a(X_t^0, \alpha^*)\right)^\star B(X_t^0, \beta^*)\partial_{\alpha_j} a(X_t^0, \alpha^*)dt, \\
I_b^{ij}(\beta^*) &= \frac{1}{2}\frac{1}{T}\int_0^T \mathrm{tr}\left\{B^{-1}(\partial_{\beta_i}B)B^{-1}(\partial_{\beta_j}B)(X_t^0, \beta^*)\right\} dt.
\end{aligned}$$

Theorem 2. *Assume [A1]–[A4]. Then, as $\epsilon \to 0$ and $n \to \infty$,*

$$\left(\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*)\right) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p+q}(0, I(\theta^*)^{-1})$$

and

$$E_{\theta^*}\left[f\left(\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*)\right)\right] \to \mathbb{E}\left[f(\zeta_1, \zeta_2)\right]$$

for all continuous functions f of at most polynomial growth.

3 Example and simulation results

Consider the three-dimensional small diffusion process defined by

$$\begin{aligned}
dX_t &= a(X_t, \alpha) + \epsilon b(X_t, \beta)dw_t, \quad t \in [0,T], \quad \epsilon \in (0,1], \\
X_0 &= (1,1,1)^\star,
\end{aligned}$$

where

$$\begin{aligned}
a(X_t, \alpha) &= \begin{pmatrix} 1 - X_{t,1} - 10\sin(\alpha_1 X_{t,2} + \alpha_2 X_{t,2}^2) \\ 2 - \alpha_3 X_{t,2} - 10\sin(\alpha_4 X_{t,3}^2) \\ 3 - \alpha_5 X_{t,3} - 10\sin(\alpha_6 X_{t,1}^2) \end{pmatrix}, \\
b(X_t, \beta) &= \begin{pmatrix} \sqrt{\beta_1(2+\cos(X_{t,3}^2))} & 0.01 & 0 \\ 0.01 & \sqrt{\beta_2(2+\cos(X_{t,1}^2))} & 0 \\ 0 & 0 & \sqrt{\beta_3(2+\cos(X_{t,2}^2))} \end{pmatrix}.
\end{aligned}$$

Furthermore, $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$, and $\beta = (\beta_1, \beta_2, \beta_3)$ are unknown parameters, and their true values $(\alpha_1^*, \alpha_2^*, \alpha_3^*, \alpha_4^*, \alpha_5^*, \alpha_6^*, \beta_1^*, \beta_2^*, \beta_3^*) = (3, 7, 5, 2, 4, 6, 1, 2, 3)$. The parameter space is assumed to be $\Theta = [0.01, 50]^9$.

The simulations were done for $T = 1$, $h = 10^{-5}$, which means that $n = 10^5$. Let $c = 10$ and $n_0 = n/10 = 10^4$. We set $\epsilon = 0.05, 0.01$. The initial Bayes type estimator $\tilde{\theta}_B = (\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}, \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})$ with reduced data $\mathbb{Y}_{n_0}$ is defined by

$$\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha))\right\} \pi_1(\alpha) d\alpha}{\int_{\Theta_\alpha} \exp\left\{\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha)\right\} \pi_1(\alpha) d\alpha},$$

$$\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} = \frac{\int_{\Theta_\beta} \beta \exp\left\{\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}, \beta)\right\} \pi_2(\beta) d\beta}{\int_{\Theta_\beta} \exp\left\{\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}, \beta)\right\} \pi_2(\beta) d\beta},$$

where $\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha)$ and $\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\alpha, \beta)$ with reduced data $\mathbb{Y}_{n_0}$ are given by

$$\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha) = \epsilon^{2-2r_1} U^{(1)}_{\epsilon,n_0}(\alpha),$$
$$\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\alpha, \beta) = \frac{1}{(\sqrt{n_0})^{2-2r_2}} U^{(2)}_{\epsilon,n_0}(\alpha, \beta).$$

It follows from Theorem 1 that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1} - \alpha^*)\right|^M\right] < \infty,$$
$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|(\sqrt{n_0})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)\right|^M\right] < \infty.$$

The Bayes type estimators are calculated with MpCN method proposed by Kamatani (2014). MpCN algorithm is as follows.

- Choose $x \in \mathbf{R}^d$ and $\mu \in \mathbf{R}^d$.
- Generate r from the gamma distribution with the shape parameter $d/2$ and the scale parameter $\| x - \mu \|^2 /2$.
- Generate $x^* = \mu + \rho^{1/2}(x - \mu) + (1-\rho)^{1/2} r^{-1/2} \omega$ where w follows the standard normal distribution.
- Accept x^* as x with probability $\min\left\{1, \frac{p(x^*)\|x-\mu\|^2}{p(x)\|x^*-\mu\|^2}\right\}$. Otherwise, discard x^*.

In practice, it is advisable to take the two-stage procedure.

- Choose $x \in \mathbf{R}^d$ and $\mu \in \mathbf{R}^d$. Run MpCN algorithm. Let $(x_1, \ldots, x_M)$ be the output.
- Set $x = x_M$, $\mu = \sum_{m=1}^M \frac{x_m}{M}$ and run MpCN algorithm again.

In this paper, we set $\rho = 0.8$. We used 10^7 and 10^6 Markov chains and 10^6 and 10^5 burn-in iterations for estimation of α and β, respectively.

The adaptive ML type estimator $(\hat{\alpha}^{(3)}_{A,n}, \hat{\beta}^{(4)}_{A,n})$ is defined as

$$\hat{\alpha}^{(1)}_{A,n} = \arg\sup_{\alpha \in \Theta_\alpha} U^{(1)}_{\epsilon,n}(\alpha),$$
$$\hat{\beta}^{(2)}_{A,n} = \arg\sup_{\beta \in \Theta_\beta} U^{(2)}_{\epsilon,n}(\hat{\alpha}^{(1)}_{A,n}, \beta),$$
$$\hat{\alpha}^{(3)}_{A,n} = \arg\sup_{\alpha \in \Theta_\alpha} U^{(3)}_{\epsilon,n}(\alpha, \hat{\beta}^{(2)}_{A,n}),$$
$$\hat{\beta}^{(4)}_{A,n} = \arg\sup_{\beta \in \Theta_\beta} U^{(4)}_{\epsilon,n}(\hat{\alpha}^{(3)}_{A,n}, \beta).$$

In order to compute the ML type estimator, we used **optim()** with the "L-BFGS-B" method in the R Language.

The hybrid estimators $\hat{\alpha}_{\epsilon,n}$ and $\hat{\beta}_{\epsilon,n}$ with full data $\mathbb{X}_n$ are computed as follows.

$$\begin{aligned} U_{\epsilon,n}^{(3)}(\hat{\alpha}_{\epsilon,n}, \hat{\beta}_{\epsilon,n}^{(2)}) &= \sup_{\alpha\in\Theta_\alpha} U_{\epsilon,n}^{(3)}(\alpha, \hat{\beta}_{\epsilon,n}^{(2)}), \\ U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n}, \hat{\beta}_{\epsilon,n}) &= \sup_{\beta\in\Theta_\beta} U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n}, \beta), \end{aligned}$$

where $\hat{\alpha}_{\epsilon,n}^{(1)}$ is obtained by using **optim()** for $U_{\epsilon,n}^{(1)}(\alpha)$ with the initial Bayes type estimator $\tilde{\alpha}_{\epsilon,n_0,r_1}^{(1)}$, and $\hat{\beta}_{\epsilon,n}^{(2)}$ is given by using **optim()** for $U_{\epsilon,n}^{(2)}(\hat{\alpha}_{\epsilon,n}^{(1)}, \beta)$ with the initial Bayes type estimator $\tilde{\beta}_{\epsilon,n_0,r_2}^{(2)}$.

For the true model, 100 independent sample paths are generated by the Euler-Maruyama scheme, and the mean and the standard deviation (s.d.) for the estimators are computed. Tables 1-8 and 9-16 are simulation results for $\epsilon = 0.01$ and 0.05, respectively. The time in each table is the computation time of estimation for one sample path. The personal computer with Intel i7-5930K (3.5GHz base clock) was used for simulations.

Tables 1-2 and 9-10 show the simulation results of the adaptive ML type estimator $(\hat{\alpha}_{A,n}^{(3)}, \hat{\beta}_{A,n}^{(4)})$ when the initial values are the true value and the uniform random number on Θ. We see from Tables 1 and 9 that all estimators have good behavior. Tables 2 and 10 are the simulation results of the adaptive ML type estimator $(\hat{\alpha}_{A,n}^{(3)}, \hat{\beta}_{A,n}^{(4)})$ with the initial value being the uniform random number on Θ. All estimators have considerable biases, which means that the optimization fails since the initial value may be far from the true value. As we know very well, it is quite important to choose the initial value for optimization. Tables 3-4 and 11-12 show the simulation results of the initial Bayes type estimator $\tilde{\theta}_B = (\tilde{\alpha}_{\epsilon,n_0,r_1}^{(1)}, \tilde{\beta}_{\epsilon,n_0,r_2}^{(2)})$ when the sample size of the reduced data $n_0 = 10^4$ and the tuning parameters $(r_1, r_2) = (1, 0, 1.0)$, $(0.7, 1.0)$, $(0.5, 1.0)$, $(0.3, 0.6)$ and (0.1, 0.2). In Table 3, the initial Bayes type estimators with $r_1 = 1.0, 0.7, 0.5, 0.3$ have good behavior. On the other hand, all of the initial Bayes type estimators in Table 11 have biases. In Tables 4 and 12, most of the initial Bayes type estimators have good performance. Tables 5-6 and 13-14 show the results of the hybrid estimators $\hat{\theta}_n = (\hat{\alpha}_{\epsilon,n}, \hat{\beta}_{\epsilon,n})$ with the initial Bayes type estimators in Tables 3-4 and 11-12, respectively. In Tables 5-6, the hybrid estimators of α with the tuning parameters $(r_1, r_2) = (1, 0, 1.0)$, $(0.7, 1.0)$, $(0.5, 1.0)$, $(0.3, 0.6)$ have good behavior and the hybrid estimators of β with the tuning parameters $(r_1, r_2) = (1, 0, 1.0)$, $(0.7, 1.0)$, $(0.5, 1.0)$, $(0.3, 0.6)$ are unbiased. In Tables 13-14, the hybrid estimator of (α, β) with $(r_1, r_2) = (0.3, 0.6)$ is best among the competing hybrid estimators.

Next, in order to compare with the hybrid estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ based on the initial Bayes type estimator from reduced data, we consider the following two kinds of initial estimators $(\hat{\alpha}_{G,n_0}^{(1)}, \hat{\beta}_{G,n_0}^{(2)})$ and $(\hat{\alpha}_{U,n_0}^{(1)}, \hat{\beta}_{U,n_0}^{(2)})$. Let $n_0 = 10^4$.

Method G. For 18^6 points $\bar{\alpha}_{0,m}$ $(m = 1, \ldots, 18^6)$ with 18 equally spaced points on each axis on $[0.01, 50]^6$, the initial estimator $\hat{\alpha}_{G,n_0}^{(1)}$ is defined as

$$U_{\epsilon,n_0}^{(1)}(\hat{\alpha}_{G,n_0}^{(1)}) = \max\left\{U_{\epsilon,n_0}^{(1)}(\bar{\alpha}_{0,1}), U_{\epsilon,n_0}^{(1)}(\bar{\alpha}_{0,2}), \ldots, U_{\epsilon,n_0}^{(1)}(\bar{\alpha}_{0,18^6})\right\}.$$

Next, for 130^3 points $\bar{\beta}_{0,m}$ $(m = 1, \ldots, 130^3)$ with 130 equally spaced points on each axis on $[0.01, 50]^3$, the initial estimator $\hat{\beta}_{G,n_0}^{(2)}$ is defined as

$$U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{G,n_0}^{(1)}, \hat{\beta}_{G,n_0}^{(2)}) = \max\left\{U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{G,n_0}^{(1)}, \bar{\beta}_{0,1}), U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{G,n_0}^{(1)}, \bar{\beta}_{0,2}), \ldots, U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{G,n_0}^{(1)}, \bar{\beta}_{0,130^3})\right\}.$$

Method U. Using 58000 uniform random numbers $\alpha_{0,m}$ $(m = 1, \dots, 58000)$ on $[0.01, 50]^6$, we compute

$$\hat{\alpha}_m^{(1)} = \arg\sup_{\alpha} U_{\epsilon,n_0}^{(1)}(\alpha)$$

by means of **optim()** with each initial value $\alpha_{0,m}$. The initial estimator $\hat{\alpha}_{U,n_0}^{(1)}$ is defined as

$$U_{\epsilon,n_0}^{(1)}(\hat{\alpha}_{U,n_0}^{(1)}) = \max\left\{U_{\epsilon,n_0}^{(1)}(\hat{\alpha}_1^{(1)}), U_{\epsilon,n_0}^{(1)}(\hat{\alpha}_2^{(1)}), \dots, U_{\epsilon,n_0}^{(1)}(\hat{\alpha}_{58000}^{(1)})\right\}.$$

Next, using 40000 uniform random numbers $\beta_{0,m}$ $(m = 1, \dots, 40000)$ on $[0.01, 50]^3$, we compute

$$\hat{\beta}_m^{(2)} = \arg\sup_{\beta} U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{U,n_0}^{(1)}, \beta)$$

by means of **optim()** with each initial value $\beta_{0,m}$. The initial estimator $\hat{\beta}_{U,n_0}^{(2)}$ is defined as

$$U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{U,n_0}^{(1)}, \hat{\beta}_{U,n_0}^{(2)}) = \max\left\{U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{U,n_0}^{(1)}, \hat{\beta}_1^{(2)}), U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{U,n_0}^{(1)}, \hat{\beta}_2^{(2)}), \dots, U_{\epsilon,n_0}^{(2)}(\hat{\alpha}_{U,n_0}^{(1)}, \hat{\beta}_{40000}^{(2)})\right\}.$$

Let $k = G, U$. The hybrid estimator $(\bar{\alpha}_{k,n}^{(3)}, \bar{\beta}_{k,n}^{(4)})$ is computed as follows.

$$\begin{aligned}\bar{\alpha}_{k,n}^{(3)} &= \arg\sup_{\alpha\in\Theta_\alpha} U_{\epsilon,n}^{(3)}(\alpha, \bar{\beta}_{k,n}^{(2)}),\\ \bar{\beta}_{k,n}^{(4)} &= \arg\sup_{\beta\in\Theta_\beta} U_{\epsilon,n}^{(4)}(\bar{\alpha}_{k,n}^{(3)}, \beta),\end{aligned}$$

where $\bar{\alpha}_{k,n}^{(1)}$ is obtained by using **optim()** for $U_{\epsilon,n}^{(1)}(\alpha)$ with the initial estimator $\hat{\alpha}_{k,n_0}^{(1)}$, and $\bar{\beta}_{k,n}^{(2)}$ is given by using **optim()** for $U_{\epsilon,n}^{(2)}(\bar{\alpha}_{k,n}^{(1)}, \beta)$ with the initial estimator $\hat{\beta}_{k,n_0}^{(2)}$. Let $\hat{\theta}_B = (\hat{\alpha}_n, \hat{\beta}_n)$ with the initial Bayes type estimator $\tilde{\theta}_B$. Let $\hat{\theta}_G = (\bar{\alpha}_{G,n}^{(3)}, \bar{\beta}_{G,n}^{(4)})$ with $\tilde{\theta}_G = (\hat{\alpha}_{G,n_0}^{(1)}, \hat{\beta}_{G,n_0}^{(2)})$ and $\hat{\theta}_U = (\bar{\alpha}_{U,n}^{(3)}, \bar{\beta}_{U,n}^{(4)})$ with $\tilde{\theta}_U = (\hat{\alpha}_{U,n_0}^{(1)}, \hat{\beta}_{U,n_0}^{(2)})$.

Tables 7-8 and 15-16 show the results of the three kinds of hybrid estimators $\hat{\theta}_B$, $\hat{\theta}_G$, $\hat{\theta}_U$. In Tables 7-8 and 15, the hybrid estimator $\hat{\theta}_G$ has bad performance since the optimization fails. In Table 15, the hybrid estimators $\hat{\alpha}_U$ for α_1, α_2 and α_6 have considerable biases. On the other hand, $\hat{\alpha}_B$ with $(r_1, r_2) = (1.0, 1.0)$ in Table 7 and $\hat{\alpha}_B$ with $(r_1, r_2) = (0.3, 0.6)$ in Table 15 have good behavior. In this example, we see that the hybrid estimator with the initial Bayes type estimator based on reduced data is best among the competing estimators.

3.1 $\epsilon = 0.01$

Table 1: adaptive ML type estimator with the initial value being the true value

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(sec.)
	2.999	7.001	5.002	2.000	4.002	5.999	1.000	2.000	3.000	
true	(0.007)	(0.008)	(0.047)	(0.004)	(0.026)	(0.004)	(0.005)	(0.010)	(0.014)	40

Table 2: adaptive ML type estimator with the initial value being the uniform random number on Θ

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(sec.)
	22.700	18.913	3.336	25.073	2.898	23.781	3.351	4.290	5.383	
unif	(20.578)	(19.646)	(2.859)	(21.192)	(1.466)	(15.470)	(1.237)	(1.603)	(1.182)	40

Table 3: initial Bayes type estimator $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ with $n_0 = 10^4$.

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	time(h.)
	3.006	6.989	5.120	2.006	3.986	6.000	
$r_1 = 1.0$	(0.064)	(0.085)	(0.502)	(0.027)	(0.119)	(0.011)	8
	2.994	7.003	5.072	2.003	3.997	6.000	
$r_1 = 0.7$	(0.069)	(0.084)	(0.323)	(0.023)	(0.105)	(0.012)	8
	3.000	6.997	5.112	2.005	3.985	5.998	
$r_1 = 0.5$	(0.052)	(0.069)	(0.247)	(0.016)	(0.090)	(0.010)	8
	2.981	7.011	5.793	2.042	3.997	5.996	
$r_1 = 0.3$	(0.048)	(0.063)	(0.347)	(0.024)	(0.091)	(0.010)	8
	3.437	6.267	10.489	2.914	4.198	5.971	
$r_1 = 0.1$	(0.160)	(0.070)	(0.717)	(0.407)	(0.151)	(0.033)	8

Table 4: initial Bayes type estimator $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$ with $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $n_0 = 10^4$.

	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(h.)
	0.999	2.000	3.002	
$r_1 = 1.0, r_2 = 1.0$	(0.014)	(0.028)	(0.043)	1.5
	0.999	2.000	3.002	
$r_1 = 0.7, r_2 = 1.0$	(0.014)	(0.028)	(0.043)	1.5
	0.999	1.999	3.002	
$r_1 = 0.5, r_2 = 1.0$	(0.014)	(0.028)	(0.043)	1.5
	0.999	2.000	3.002	
$r_1 = 0.3, r_2 = 0.6$	(0.014)	(0.028)	(0.043)	1.5
	0.999	2.217	3.002	
$r_1 = 0.1, r_2 = 0.2$	(0.014)	(1.035)	(0.043)	1.5

Table 5: hybrid estimator $\hat{\alpha}$ with the initial Bayes estimator $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	time(sec.)
	2.999	7.001	5.002	2.000	4.002	5.999	
$r_1 = 1.0$, $r_2 = 1.0$	(0.007)	(0.008)	(0.047)	(0.004)	(0.025)	(0.004)	40
	2.999	7.001	5.002	2.000	4.002	5.999	
$r_1 = 0.7$, $r_2 = 1.0$	(0.007)	(0.008)	(0.047)	(0.004)	(0.026)	(0.004)	40
	2.999	7.001	5.002	2.000	4.002	5.999	
$r_1 = 0.5$, $r_2 = 1.0$	(0.007)	(0.008)	(0.047)	(0.004)	(0.026)	(0.004)	40
	2.999	7.001	5.002	2.000	4.003	5.999	
$r_1 = 0.3$, $r_2 = 0.6$	(0.007)	(0.008)	(0.047)	(0.004)	(0.026)	(0.004)	40
	2.999	7.001	4.967	1.961	4.002	5.999	
$r_1 = 0.1$, $r_2 = 0.2$	(0.007)	(0.008)	(0.551)	(0.592)	(0.025)	(0.004)	40

Table 6: hybrid estimator $\hat{\beta}$ with $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$

	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(sec.)
	1.000	2.000	3.000	
$r_1 = 1.0$, $r_2 = 1.0$	(0.005)	(0.010)	(0.014)	40
	1.000	2.000	3.000	
$r_1 = 0.7$, $r_2 = 1.0$	(0.005)	(0.010)	(0.014)	40
	1.000	2.000	3.000	
$r_1 = 0.5$, $r_2 = 1.0$	(0.005)	(0.010)	(0.014)	40
	1.000	2.000	3.000	
$r_1 = 0.3$, $r_2 = 0.6$	(0.005)	(0.010)	(0.014)	40
	1.000	2.124	3.000	
$r_1 = 0.1$, $r_2 = 0.2$	(0.005)	(0.521)	(0.014)	40

Table 7: hybrid estimators $\hat{\alpha}_B$, $\hat{\alpha}_G$ and $\hat{\alpha}_U$ with $\tilde{\alpha}_B$, $\tilde{\alpha}_G$ and $\tilde{\alpha}_U$, respectively.

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	time(h.)
	2.999	7.001	5.002	2.000	4.002	5.999	
$\hat{\alpha}_B$	(0.007)	(0.008)	(0.047)	(0.004)	(0.026)	(0.004)	8
	6.885	11.753	4.620	1.735	3.670	15.031	
$\tilde{\alpha}_G$	(13.889)	(11.500)	(0.989)	(0.666)	(0.750)	(16.523)	12
	2.999	7.001	5.002	2.000	4.002	5.999	
$\tilde{\alpha}_U$	(0.007)	(0.008)	(0.047)	(0.004)	(0.026)	(0.004)	9

Table 8: hybrid estimators $\hat{\beta}_B$, $\hat{\beta}_G$ and $\hat{\beta}_U$ with $\tilde{\theta}_B$, $\tilde{\theta}_G$ and $\tilde{\theta}_U$, respectively.

	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(h.)
	1.000	2.000	3.000	
$\hat{\beta}_B$	(0.005)	(0.010)	(0.014)	1.5
	2.355	2.210	4.075	
$\tilde{\beta}_G$	(1.343)	(0.549)	(1.575)	1.5
	1.000	2.000	3.000	
$\tilde{\beta}_U$	(0.005)	(0.010)	(0.014)	1.5

3.2 $\epsilon = 0.05$

Table 9: adaptive ML type estimator with the initial value being the true value

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(sec.)
	3.003	6.996	5.055	2.001	4.021	5.995	1.000	2.000	3.000	
true	(0.029)	(0.049)	(0.326)	(0.019)	(0.163)	(0.024)	(0.005)	(0.010)	(0.014)	40

Table 10: adaptive ML type estimator with the initial value being the uniform random number on Θ

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(sec.)
	22.750	23.984	3.069	25.590	3.111	31.623	1.093	2.096	3.110	
unif	(21.828)	(20.621)	(2.855)	(22.029)	(0.999)	(20.306)	(0.063)	(0.064)	(0.044)	40

Table 11: initial Bayes type estimator $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ with $n_0 = 10^4$.

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	time(h.)
	3.587	6.654	5.317	2.383	3.966	5.984	
$r_1 = 1.0$	(2.638)	(1.389)	(1.457)	(3.751)	(0.483)	(0.081)	8
	3.388	7.183	5.386	2.443	3.962	5.986	
$r_1 = 0.7$	(2.564)	(3.408)	(1.403)	(4.505)	(0.483)	(0.082)	8
	4.432	7.235	5.632	2.0279	3.976	5.976	
$r_1 = 0.5$	(5.982)	(4.564)	(1.379)	(0.091)	(0.497)	(0.082)	8
	3.506	6.607	7.412	2.134	4.020	5.974	
$r_1 = 0.3$	(1.088)	(0.536)	(1.956)	(0.139)	(0.544)	(0.142)	8
	5.734	6.873	10.529	4.623	4.344	6.369	
$r_1 = 0.1$	(2.384)	(0.980)	(2.878)	(1.311)	(0.757)	(0.441)	8

Table 12: initial Bayes type estimator $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$ with $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $n_0 = 10^4$.

	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(h.)
	1.001	2.002	3.002	
$r_1 = 1.0, r_2 = 1.0$	(0.015)	(0.034)	(0.043)	1.5
	1.002	2.002	3.002	
$r_1 = 0.7, r_2 = 1.0$	(0.024)	(0.028)	(0.044)	1.5
	1.003	2.000	3.002	
$r_1 = 0.5, r_2 = 1.0$	(0.020)	(0.028)	(0.043)	1.5
	1.001	2.000	3.002	
$r_1 = 0.3, r_2 = 0.6$	(0.019)	(0.028)	(0.043)	1.5
	1.033	2.052	3.006	
$r_1 = 0.1, r_2 = 0.2$	(0.063)	(0.101)	(0.052)	1.5

Table 13: hybrid estimator $\hat{\alpha}$ with the initial Bayes estimator $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	time(sec.)
	3.433	6.754	5.009	2.435	4.020	5.995	
$r_1 = 1.0,\, r_2 = 1.0$	(2.249)	(1.199)	(0.603)	(4.336)	(0.161)	(0.024)	40
	3.383	7.359	5.064	2.478	4.020	5.995	
$r_1 = 0.7,\, r_2 = 1.0$	(2.766)	(4.360)	(0.332)	(4.775)	(0.163)	(0.024)	40
	4.180	7.336	5.057	2.001	3.993	5.974	
$r_1 = 0.5,\, r_2 = 1.0$	(5.580)	(4.668)	(0.328)	(0.019)	(0.327)	(0.213)	40
	3.082	6.976	5.062	2.001	4.021	5.995	
$r_1 = 0.3,\, r_2 = 0.6$	(0.795)	(0.207)	(0.333)	(0.019)	(0.164)	(0.024)	40
	4.791	6.821	4.864	3.334	3.887	5.835	
$r_1 = 0.1,\, r_2 = 0.2$	(3.486)	(2.060)	(2.157)	(2.017)	(0.660)	(1.051)	40

Table 14: hybrid estimator $\hat{\beta}$ with $\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}$

	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(sec.)
	1.004	2.002	3.000	
$r_1 = 1.0,\, r_2 = 1.0$	(0.019)	(0.019)	(0.014)	40
	1.002	2.001	3.000	
$r_1 = 0.7,\, r_2 = 1.0$	(0.015)	(0.015)	(0.014)	40
	1.006	2.000	3.001	
$r_1 = 0.5,\, r_2 = 1.0$	(0.027)	(0.010)	(0.015)	40
	1.001	2.000	3.000	
$r_1 = 0.3,\, r_2 = 0.6$	(0.007)	(0.010)	(0.014)	40
	1.024	2.038	3.004	
$r_1 = 0.1,\, r_2 = 0.2$	(0.041)	(0.057)	(0.025)	40

Table 15: hybrid estimators $\hat{\alpha}_B$, $\hat{\alpha}_G$ and $\hat{\alpha}_U$ with $\tilde{\alpha}_B$, $\tilde{\alpha}_G$ and $\tilde{\alpha}_U$, respectively.

	$\hat{\alpha}_1(3)$	$\hat{\alpha}_2(7)$	$\hat{\alpha}_3(5)$	$\hat{\alpha}_4(2)$	$\hat{\alpha}_5(4)$	$\hat{\alpha}_6(6)$	time(h.)
	3.082	6.976	5.062	2.001	4.021	5.995	
$\hat{\alpha}_B$	(0.795)	(0.207)	(0.333)	(0.019)	(0.164)	(0.024)	8
	6.972	8.008	4.463	1.661	3.862	10.098	
$\tilde{\alpha}_G$	(9.493)	(8.365)	(1.480)	(0.823)	(0.596)	(11.453)	12
	3.486	6.770	5.057	2.001	4.007	6.152	
$\tilde{\alpha}_U$	(2.436)	(1.126)	(0.327)	(0.019)	(0.232)	(1.102)	9

Table 16: hybrid estimators $\hat{\beta}_B$, $\hat{\beta}_G$ and $\hat{\beta}_U$ with $\tilde{\theta}_B$, $\tilde{\theta}_G$ and $\tilde{\theta}_U$, respectively.

	$\hat{\beta}_1(1)$	$\hat{\beta}_2(2)$	$\hat{\beta}_3(3)$	time(h.)
	1.001	2.000	3.000	
$\hat{\beta}_B$	(0.007)	(0.010)	(0.014)	1.5
	1.058	2.013	3.024	
$\tilde{\beta}_G$	(0.051)	(0.029)	(0.056)	1.5
	1.003	2.000	3.002	
$\tilde{\beta}_U$	(0.014)	(0.010)	(0.020)	1.5

4 Proofs

Let $P_i(\alpha) = X_{t_i} - X_{t_{i-1}} - h_n a(X_{t_{i-1}}, \alpha)$ and $\mathcal{G}_i = \sigma(w_s; s \leq t_i)$. $P_i^j(\alpha)$ denotes the j-th element of $P_i(\alpha)$. Set $X_i = X_{t_i}$. Let R denote a function $(0,1] \times \mathbb{R}^d \to \mathbb{R}$ satisfying that there exists a constant C such that $|R(a,x)| \leq aC(1+|x|)^C$ for all a, x. We simply write $\|f\|_M \lesssim \|g\|_M$ when there exists a constant $C_M > 0$ such that $\|f\|_M \leq C_M \|g\|_M$ for $f, g \in L^M(P)$. First of all, we state the following three lemmas in order to show Theorems 1 and 2.

Lemma 1 (Sørensen and Uchida (2003))**.** *Suppose that [A1] and [A2] hold true. Then, as* $\frac{1}{\epsilon\sqrt{n}} = O(1)$,

(i) $E_{\theta^*}\left[P_i^j(\alpha^*)\Big|\mathcal{G}_{i-1}\right] = R(h_n^2, X_{i-1})$.

(ii) $E_{\theta^*}\left[P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)\Big|\mathcal{G}_{i-1}\right] = \epsilon^2 h_n B_{i-1}^{j_1 j_2}(\beta^*) + R(\epsilon^2 h_n^2, X_{i-1})$.

(iii) $E_{\theta^*}\left[P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)P_i^{j_3}(\alpha^*)\Big|\mathcal{G}_{i-1}\right] = R\left(\epsilon^4 h_n^2, X_{i-1}\right)$.

(iv) $E_{\theta^*}\left[\prod_{k=1}^4 P_i^{j_k}(\alpha^*)\Big|\mathcal{G}_{i-1}\right] = \epsilon^4 h_n^2 \left\{B_{i-1}^{j_1 j_2} B_{i-1}^{j_3 j_4}(\beta^*) + B_{i-1}^{j_1 j_3} B_{i-1}^{j_2 j_4}(\beta^*) + B_{i-1}^{j_1 j_4} B_{i-1}^{j_2 j_3}(\beta^*)\right\} + R\left(\epsilon^4 h_n^3, X_{i-1}\right)$.

Lemma 2. *Let* $n_0 = \left[\frac{n}{c}\right]$ *for* $c \geq 1$. *Let* $f \in C_{\uparrow}^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$. *Assume [A1] and [A2]. Then, for any* $M > 0$,

(i) $\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \left|h_n \sum_{i=1}^{n_0} f(X_{i-1},\theta)\right|\right)^M\right] < \infty$.

(ii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}\left|h_n \sum_{i=1}^{n_0} f(X_{i-1},\theta) - \int_0^{\frac{T}{c}} f(X_t^0,\theta)dt\right|\right)^M\right] < \infty$.

(iii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^j(\alpha^*)\right|\right)^M\right] < \infty$.

(iv) $\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^j(\alpha)\right|\right)^M\right] < \infty$.

Proof. (i) One has that

$$\left(\sup_\theta \left|h_n \sum_{i=1}^{n_0} f(X_{i-1},\theta)\right|\right)^M \leq (TC(1+\sup_{t\in[0,T]}|X_t|^C))^M,$$

which completes the proof of (i).

(ii) Let

$$\begin{aligned} G_n(\theta) &:= h_n \sum_{i=1}^{n_0} f(X_{i-1},\theta) - \int_0^{T/c} f(X_t^0,\theta)dt \\ &= \sum_{i=1}^{n_0}\int_{t_{i-1}}^{t_i} f(X_{i-1},\theta)dt - \sum_{i=1}^{n_0}\int_{t_{i-1}}^{t_i} f(X_t,\theta)dt \\ &\quad + \int_0^{n_0 h_n} f(X_t,\theta)dt - \int_0^{T/c} f(X_t,\theta)dt + \int_0^{T/c}\left(f(X_t,\theta) - f(X_t^0,\theta)\right)dt \\ &=: G_n^{(1)}(\theta) + G_n^{(2)}(\theta) + G_n^{(3)}(\theta). \end{aligned}$$

Since the standard estimates yields that for all $M > 0$,

$$E_{\theta^*}\left[\left(\sup_{0\le t\le T}\epsilon^{-1}|X_t - X_t^0|\right)^M\right] < \infty$$

and

$$G_n^{(3)}(\theta) = \int_0^{T/c}\left(f(X_t,\theta) - f(X_t^0,\theta)\right)dt = \int_0^{T/c}\int_0^1 \partial_x f(X_t^0 + u(X_t - X_t^0),\theta)du[X_t - X_t^0]dt,$$

we have that

$$\begin{aligned}\sup_\theta E_{\theta^*}\left[|G_n^{(3)}(\theta)|^M\right] &\lesssim \epsilon^M E_{\theta^*}\left[\left(\sup_{0\le t\le T}\epsilon^{-1}|X_t - X_t^0|\right)^{2M}\right]^{1/2} = O(\epsilon^M),\\ \sup_\theta E_{\theta^*}\left[|\partial_\theta G_n^{(3)}(\theta)|^M\right] &= O(\epsilon^M).\end{aligned}$$

Sobolev's inequality implies that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}|G_n^{(3)}(\theta)|\right)^M\right] < \infty.$$

Next, noting that

$$\begin{aligned}G_n^{(1)}(\theta) &= \sum_{i=1}^n\int_{t_{i-1}}^{t_i}\left(f(X_{i-1},\theta) - f(X_t,\theta)\right)dt\\ &= \sum_{i=1}^n\int_{t_{i-1}}^{t_i}\int_0^1 \partial_x f(X_t + u(X_{i-1} - X_t),\theta)du[X_{i-1} - X_t]dt\end{aligned}$$

and for $i = 1,\ldots,n$

$$\sup_{t_{i-1}\le t\le t_i} E_{\theta^*}\left[\left((\epsilon h_n^{\frac{1}{2}})^{-1}|X_t - X_{i-1}|\right)^M\right] < \infty,$$

one has that

$$\begin{aligned}\sup_\theta E_{\theta^*}\left[|G_n^{(1)}(\theta)|^M\right] &\lesssim (\epsilon h_n^{\frac{1}{2}})^M n^M \frac{1}{n}\sum_{i=1}^n h_n^{M-1}\int_{t_{i-1}}^{t_i} E_{\theta^*}\left[\left((\epsilon h_n^{\frac{1}{2}})^{-1}|X_t - X_{i-1}|\right)^{2M}\right]^{1/2} dt\\ &\lesssim (\epsilon h_n^{\frac{1}{2}})^M n^M \frac{1}{n}\sum_{i=1}^n h_n^M \sup_{t_{i-1}\le t\le t_i} E_{\theta^*}\left[\left((\epsilon h_n^{\frac{1}{2}})^{-1}|X_t - X_{i-1}|\right)^{2M}\right]^{1/2}\\ &= O((\epsilon h_n^{\frac{1}{2}})^M),\\ \sup_\theta E_{\theta^*}\left[|\partial_\theta G_n^{(1)}(\theta)|^M\right] &= O((\epsilon h_n^{\frac{1}{2}})^M).\end{aligned}$$

Thus, it follows from the Sobolev inequality that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}|G_n^{(1)}(\theta)|\right)^M\right] < \infty.$$

Moreover, since $G_n^{(2)}(\theta) = \int_0^{n_0 h_n} f(X_t,\theta)dt - \int_0^{T/c} f(X_t,\theta)dt$ and $\left|n_0 h_n - \frac{T}{c}\right| \le h_n$, one has that

$$E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}|G_n^{(2)}(\theta)|\right)^M\right] \le \left(\epsilon^{-1}\left(\frac{T}{c} - n_0 h_n\right)\right)^M E_{\theta^*}\left[\sup_{t\in[0,T],\theta}|f(X_t,\theta)|^M\right] < \infty,$$

which completes the proof of (ii).

(iii) We have

$$\begin{aligned} X_i - E_{\theta^*}\left[X_i|\mathcal{G}_{i-1}\right] &= X_i - X_{i-1} - E_{\theta^*}\left[X_i - X_{i-1}|\mathcal{G}_{i-1}\right] \\ &= \int_{t_{i-1}}^{t_i}\left(a(X_s,\alpha^*) - E_{\theta^*}\left[a(X_s,\alpha^*)|\mathcal{G}_{i-1}\right]\right)ds + \epsilon\int_{t_{i-1}}^{t_i} b(X_s,\beta^*)dw_s. \end{aligned}$$

Therefore,

$$\begin{aligned} E_{\theta^*}\left[|X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}]|^M\right] &\lesssim h_n^{M-1}\int_{t_{i-1}}^{t_i} E_{\theta^*}\left[|a_s - E_{\theta^*}[a_s|\mathcal{G}_{i-1}]|^M\right]ds + \epsilon^M h_n^{\frac{M}{2}-1}\int_{t_{i-1}}^{t_i} E_{\theta^*}[b_s^M]ds \\ &= O(h_n^M) + O(\epsilon^M h_n^{\frac{M}{2}}) = O(\epsilon^M h_n^{\frac{M}{2}}). \end{aligned} \tag{5}$$

Noting that

$$\begin{aligned} P_i^j(\alpha^*) &= (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^j + (E_{\theta^*}[X_i|\mathcal{G}_{i-1}] - X_{i-1} - h_n a_{i-1}(\alpha^*))^j \\ &= (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^j + R(h_n^2, X_{i-1}) =: M_{1,i}^j + R(h_n^2, X_{i-1}), \end{aligned}$$

one has

$$\begin{aligned} \sup_\theta E_{\theta^*}\left[\left(\epsilon^{-1}\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)M_{1,i}^j\right|\right)^M\right] &\lesssim \epsilon^{-M}\sup_\theta E_{\theta^*}\left[\left(\sum_{i=1}^{n_0}\left|f(X_{i-1},\theta)M_{1,i}^j\right|^2\right)^{\frac{M}{2}}\right] \\ &\lesssim \epsilon^{-M}\sup_\theta n_0^{\frac{M}{2}}\frac{1}{n_0}\sum_{i=1}^{n_0} E_{\theta^*}\left[|f(X_{i-1},\theta)|^M\,|M_{1,i}^j|^M\right] \\ &\lesssim \epsilon^{-M} n_0^{\frac{M}{2}}\frac{1}{n_0}\sum_{i=1}^{n_0}\sup_\theta E_{\theta^*}\left[|f(X_{i-1},\theta)|^{2M}\right]^{1/2} E_{\theta^*}\left[|M_{1,i}^j|^{2M}\right]^{1/2} \\ &= \epsilon^{-M} n_0^{\frac{M}{2}} O(\epsilon^M h_n^{\frac{M}{2}}) < \infty \end{aligned}$$

and similarly,

$$\sup_\theta E_{\theta^*}\left[\left(\epsilon^{-1}\left|\sum_{i=1}^{n_0}\partial_\theta f(X_{i-1},\theta)M_{1,i}^j\right|\right)^M\right] = \epsilon^{-M} n_0^{\frac{M}{2}} O(\epsilon^M h_n^{\frac{M}{2}}) < \infty.$$

Therefore,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)M_{1,i}^j\right|\right)^M\right] < \infty.$$

Moreover, one has that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \epsilon^{-1}\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)R(h_n^2, X_{i-1})\right|\right)^M\right] < \infty,$$

which completes the proof of (iii).

(iv) Since

$$\begin{aligned} P_i(\alpha) &= X_i - X_{i-1} - h_n a_{i-1}(\alpha^*) - h_n(a_{i-1}(\alpha) - a_{i-1}(\alpha^*)) \\ &= P_i(\alpha^*) - h_n(a_{i-1}(\alpha) - a_{i-1}(\alpha^*)), \end{aligned}$$

and

$$\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^j(\alpha) = \sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^j(\alpha^*) - h_n \sum_{i=1}^{n_0} f(X_{i-1},\theta)(a_{i-1}(\alpha) - a_{i-1}(\alpha^*)),$$

it follows from Lemma 2–(i) and (iii) that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^j(\alpha)\right|\right)^M\right] < \infty,$$

which completes the proof of (iv).

Lemma 3. *Let $n_0 = [\frac{n}{c}]$ for $c \geq 1$. Let $f \in C_\uparrow^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})$. Assume [A1] and [A2]. Then, for any $M > 0$,*

(i) $\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \sqrt{n_0}\left|\epsilon^{-2}\sum_{i=1}^{n_0} f(X_{i-1},\theta)\left(P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*) - \epsilon^2 h_n B^{j_1j_2}(X_{i-1},\beta^*)\right)\right|\right)^M\right] < \infty.$

(ii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta \left|\epsilon^{-2}\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^{j_1}(\alpha)P_i^{j_2}(\alpha)\right|\right)^M\right] < \infty.$

Proof. (i) We have

$$\begin{aligned} P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*) &= \left(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}] + R(h_n^2, X_{i-1})\right)^{j_1}\left(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}] + R(h_n^2, X_{i-1})\right)^{j_2} \\ &= (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1}(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2} \\ &\quad + R(h_n^2, X_{i-1})\left\{(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1} + (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2}\right\} \\ &\quad + R(h_n^4, X_{i-1}) \\ &= (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1}(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2} - V_{\theta^*}[X_i|\mathcal{G}_{i-1}]^{j_1j_2} \\ &\quad + R(h_n^2, X_{i-1})\left\{(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1} + (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2}\right\} \\ &\quad + V_{\theta^*}[X_i|\mathcal{G}_{i-1}]^{j_1j_2} - \epsilon^2 h_n B_{i-1}^{j_1j_2}(\beta^*) + \epsilon^2 h_n B_{i-1}^{j_1j_2}(\beta^*) + R(h_n^4, X_{i-1}), \end{aligned}$$

$$\begin{aligned} V_{\theta^*}[X_i|\mathcal{G}_{i-1}]^{j_1j_2} &= E_{\theta^*}\left[(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1}(X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2}|\mathcal{G}_{i-1}\right] \\ &= E_{\theta^*}\left[\left\{(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{j_1} - (E_{\theta^*}[X_i|\mathcal{G}_{i-1}] - X_{i-1} - h_n a_{i-1}(\alpha^*))^{j_1}\right\}\right. \\ &\quad \left.\times\left\{(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{j_2} - (E_{\theta^*}[X_i|\mathcal{G}_{i-1}] - X_{i-1} - h_n a_{i-1}(\alpha^*))^{j_2}\right\}|\mathcal{G}_{i-1}\right] \\ &= E_{\theta^*}\left[P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)|\mathcal{G}_{i-1}\right] \\ &\quad - (E_{\theta^*}[X_i|\mathcal{G}_{i-1}] - X_{i-1} - h_n a_{i-1}(\alpha^*))^{j_1}(E_{\theta^*}[X_i|\mathcal{G}_{i-1}] - X_{i-1} - h_n a_{i-1}(\alpha^*))^{j_2} \\ &= E_{\theta^*}\left[P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)|\mathcal{G}_{i-1}\right] - R(h_n^4, X_{i-1}) \\ &= \epsilon^2 h_n B_{i-1}^{j_1j_2}(\beta^*) + R(\epsilon^2 h_n^2, X_{i-1}). \end{aligned}$$

Therefore,

$$P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*) - \epsilon^2 h_n B_{i-1}^{j_1 j_2}(\beta^*) = M_i^{(1)} + M_i^{(2)} + R(\epsilon^2 h_n^2, X_{i-1}),$$

where

$$\begin{aligned} M_i^{(1)} &= (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1} (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2} - V_{\theta^*}[X_i|\mathcal{G}_{i-1}], \\ M_i^{(2)} &= R(h_n^2, X_{i-1}) \left\{ (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1} + (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2} \right\}. \end{aligned}$$

One has that

$$\begin{aligned} E_{\theta^*}\left[\sup_\theta \left| \epsilon^{-2}\sqrt{n_0} \sum_{i=1}^{n_0} f(X_{i-1},\theta) R(\epsilon^2 h_n^2, X_{i-1}) \right|^M \right] &\lesssim E_{\theta^*}\left[\sup_\theta \left| (\sqrt{n_0} h_n) h_n \sum_{i=1}^{n_0} R(1, X_{i-1}) \right|^M \right] \\ &\lesssim (\sqrt{n_0} h_n)^M < \infty. \end{aligned}$$

Burkholder's inequality implies that for $k = 1, 2$,

$$\begin{aligned} E_{\theta^*}\left[\left| \epsilon^{-2}\sqrt{n_0} \sum_{i=1}^{n_0} f(X_{i-1},\theta) M_i^{(k)}(\theta^*) \right|^M \right] &\lesssim \epsilon^{-2M} (\sqrt{n_0})^M E_{\theta^*}\left[\left(\sum_{i=1}^{n_0} \left| f_{i-1}(\theta) M_i^{(k)}(\theta^*) \right|^2 \right)^{\frac{M}{2}} \right] \\ &\lesssim \epsilon^{-2M} {n_0}^M \frac{1}{n_0} \sum_{i=1}^{n_0} E_{\theta^*}\left[\left| f_{i-1}(\theta) M_i^{(k)}(\theta^*) \right|^M \right]. \end{aligned}$$

By (5),

$$\begin{aligned} & E_{\theta^*}\left[\left| \epsilon^{-2}\sqrt{n_0} \sum_{i=1}^{n_0} f(X_{i-1},\theta) M_i^{(1)}(\theta^*) \right|^M \right] \\ \lesssim\ & \epsilon^{-2M} {n_0}^M \frac{1}{n_0} \sum_{i=1}^{n_0} E_{\theta^*}\left[|f_{i-1}(\theta)|^{2M} \right]^{1/2} \\ & \times E_{\theta^*}\left[\left| (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1} \right|^{4M} \right]^{1/4} E_{\theta^*}\left[\left| (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_2} \right|^{4M} \right]^{1/4} \\ =\ & \epsilon^{-2M} {n_0}^M O((\epsilon^M h^{\frac{M}{2}})^2) = O(1) < \infty. \end{aligned}$$

Moreover,

$$\begin{aligned} & E_{\theta^*}\left[\left| \epsilon^{-2}\sqrt{n_0} \sum_{i=1}^{n_0} f(X_{i-1},\theta) M_i^{(2)}(\theta^*) \right|^M \right] \\ \lesssim\ & \epsilon^{-2M} {n_0}^M \frac{1}{n_0} \sum_{i=1}^{n_0} E_{\theta^*}\left[\left| R(h_n^2, X_{i-1}) \right|^{2M} \right]^{1/2} E_{\theta^*}\left[\left| (X_i - E_{\theta^*}[X_i|\mathcal{G}_{i-1}])^{j_1} \right|^{2M} \right]^{1/2} \\ \lesssim\ & \epsilon^{-2M} {n_0}^M h^{2M} \epsilon^M h^{\frac{M}{2}} = O((\epsilon^{-M} h^{\frac{M}{2}}) h^M) < \infty. \end{aligned}$$

Thus,

$$\sup_{\epsilon,n} \sup_\theta E_{\theta^*}\left[\left| \epsilon^{-2}\sqrt{n_0} \sum_{i=1}^{n_0} f(X_{i-1},\theta) \left(M_i^{(1)}(\theta^*) + M_i^{(2)}(\theta^*) \right) \right|^M \right] < \infty.$$

In the similar way,

$$\sup_{\epsilon,n}\sup_{\theta} E_{\theta^*}\left[\left|\epsilon^{-2}\sqrt{n_0}\sum_{i=1}^{n_0}\partial_\theta f(X_{i-1},\theta)\left(M_i^{(1)}(\theta^*)+M_i^{(2)}(\theta^*)\right)\right|^M\right]<\infty.$$

Consequently,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_{\theta}\left|\epsilon^{-2}\sqrt{n_0}\sum_{i=1}^{n_0} f(X_{i-1},\theta)\left(M_i^{(1)}(\theta^*)+M_i^{(2)}(\theta^*)\right)\right|\right)^M\right]<\infty,$$

which completes the proof of (i).

(ii) We have

$$\begin{aligned}
P_i^{j_1}(\alpha)P_i^{j_2}(\alpha) &= (X_i-X_{i-1}-h_n a_{i-1}(\alpha))^{j_1}(X_i-X_{i-1}-h_n a_{i-1}(\alpha))^{j_2}\\
&= (X_i-X_{i-1}-h_n a_{i-1}(\alpha^*)-h_n(a_{i-1}(\alpha)-a_{i-1}(\alpha^*)))^{j_1}\\
&\quad\times\ (X_i-X_{i-1}-h_n a_{i-1}(\alpha^*)-h_n(a_{i-1}(\alpha)-a_{i-1}(\alpha^*)))^{j_2}\\
&= (X_i-X_{i-1}-h_n a_{i-1}(\alpha^*))^{j_1}(X_i-X_{i-1}-h_n a_{i-1}(\alpha^*))^{j_2}\\
&\quad + h_n\left\{(X_i-X_{i-1}-h_n a_{i-1}(\alpha^*))^{j_1}(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_2}\right.\\
&\quad \left.+ (X_i-X_{i-1}-h_n a_{i-1}(\alpha^*))^{j_2}(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_1}\right\}\\
&\quad + h_n^2\left\{(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_1}(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_2}\right\},
\end{aligned}$$

and

$$\begin{aligned}
\epsilon^{-2}\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)\right| &\le \left|\epsilon^{-2}\sum_{i=1}^{n_0} f(X_{i-1},\theta)\left(P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)-\epsilon^2 h_n B^{j_1j_2}(X_{i-1},\beta^*)\right)\right|\\
&\quad + \left|h_n\sum_{i=1}^{n_0} f(X_{i-1},\theta)B^{j_1j_2}(X_{i-1},\beta^*)\right|.
\end{aligned}$$

By Lemma 2-(i) and Lemma 3–(i),

$$E_{\theta^*}\left[\left(\sup_{\theta}\epsilon^{-2}\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)\right|\right)^M\right]<\infty.$$

By Lemma 2–(ii),

$$E_{\theta^*}\left[\left(\sup_{\theta}\epsilon^{-2}h_n\left|\sum_{i=1}^{n_0} f(X_{i-1},\theta)(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_1}P_i^{j_2}(\alpha^*)\right|\right)^M\right]<\infty.$$

By Lemma 2-(i),

$$E_{\theta^*}\left[\left(\sup_{\theta}\epsilon^{-2}h_n^2\sum_{i=1}^{n_0} f(X_{i-1},\theta)(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_1}(a_{i-1}(\alpha)-a_{i-1}(\alpha^*))^{j_2}\right)^M\right]<\infty,$$

which completes the proof of (ii).

Proof of Theorem 1. (i). Set for $u_1 \in \mathbb{U}^{(1)}_{\epsilon,r_1}(\alpha^*) := \{u_1 \in \mathbb{R}^p | \alpha^* + \epsilon^{r_1} u_1 \in \Theta_\alpha\}$,

$$\begin{aligned}\mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*) &= \exp\Big\{\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha^*+\epsilon^{r_1}u_1) - \mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha^*)\Big\} \\ &= \exp\Big\{\epsilon^{2-2r_1}\Big(U^{(1)}_{\epsilon,n_0}(\alpha^*+\epsilon^{r_1}u_1) - U^{(1)}_{\epsilon,n_0}(\alpha^*)\Big)\Big\}.\end{aligned}$$

Let $\mathbb{V}^{(1)}_{\epsilon,r_1}(r) = \Big\{u_1 \in \mathbb{U}^{(1)}_{\epsilon,r_1}(\alpha^*) \mid r \le |u_1|\Big\}$. One has that

$$\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1} - \alpha^*) = \frac{\int_{\mathbb{U}^{(1)}_{\epsilon,r_1}(\alpha^*)} u_1 \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1}{\int_{\mathbb{U}^{(1)}_{\epsilon,r_1}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1},$$

and

$$\begin{aligned}K^{(1)}_{\epsilon,n_0,r_1}(u_1) &:= \log \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*) \\ &= \epsilon^{2-r_1}\partial_\alpha U^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1] + \frac{1}{2}\epsilon^2\partial^2_\alpha U^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1^{\otimes 2}] \\ &\quad + \frac{1}{2}\epsilon^{2+r_1}\sum_{i,j,k=1}^{p}\int_0^1 (1-t)^2\partial_{\alpha_i}\partial_{\alpha_j}\partial_{\alpha_k}U^{(1)}_{\epsilon,n_0}(\alpha^*+t\epsilon^{r_1}u_1)dt u_{1,i}u_{1,j}u_{1,k}.\end{aligned}$$

Let

$$\begin{aligned}&\mathbb{Y}^{(1)}_{\epsilon,n_0}(\alpha) = \epsilon^{2r_1}\Big\{\mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha) - \mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha^*)\Big\} = \epsilon^2\Big\{U^{(1)}_{\epsilon,n_0}(\alpha) - U^{(1)}_{\epsilon,n_0}(\alpha^*)\Big\}, \\ &\Delta^{(1)}_{\epsilon,n_0,r_1}(\alpha^*)[u_1] = \epsilon^{r_1}\partial_\alpha \mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha^*)[u_1] = \epsilon^{2-r_1}\partial_\alpha U^{(1)}_{\epsilon,n}(\alpha^*)[u_1] =: \epsilon^{1-r_1}\Delta^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1], \\ &\Gamma^{(1)}_{\epsilon,n}(\alpha^*)[u_1,u_1] = -\epsilon^{2r_1}\partial^2_\alpha \mathbb{H}^{(1)}_{\epsilon,n_0,r_1}(\alpha^*)[u_1,u_1] = -\epsilon^2\partial^2_\alpha U^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1,u_1], \\ &\Gamma^{(1)}(\alpha^*)[u_1,u_1] = \int_{\mathbb{R}^d}\Big[\partial_\alpha a(X^0_t,\alpha^*)[u_1], \partial_\alpha a(X^0_t,\alpha^*)[u_1]\Big]dt,\end{aligned}$$

for $u_1 \in \mathbb{R}^p$. We will show that there exists $\eta_1 \in (0,1)$ such that that for all $M > 0$,

$$\sup_{\epsilon,n} E_{\theta^*}\Big[|\Delta^{(1)}_{\epsilon,n_0,r_1}(\alpha^*)|^M\Big] = \sup_{\epsilon,n} E_{\theta^*}\Big[|\epsilon^{1-r_1}\Delta^{(1)}_{\epsilon,n_0}(\alpha^*)|^M\Big] < \infty, \tag{6}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\alpha \epsilon^{-\eta_1}|\mathbb{Y}^{(1)}_{\epsilon,n_0}(\alpha) - \mathbb{Y}^{(1)}(\alpha)|\right)^M\right] < \infty, \tag{7}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\epsilon^{-\eta_1}|\Gamma^{(1)}_{\epsilon,n_0}(\alpha^*) - \Gamma^{(1)}(\alpha)|\right)^M\right] < \infty, \tag{8}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\epsilon^2\sup_\alpha |\partial^3_\alpha U^{(1)}_{\epsilon,n_0}(\alpha)|\right)^M\right] < \infty \tag{9}$$

for some $\eta_1 > 0$.

Proof of (6). We have

$$\partial_\alpha U^{(1)}_{\epsilon,n_0}(\alpha)[u_1] = -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0} E_d\left[-2h_n\partial_\alpha a_{i-1}(\alpha)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha)\right],$$

where E_d is a d-dimensional identity matrix. Thus,

$$\Delta^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1] = \epsilon\partial_\alpha U^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1] = \epsilon^{-1}\sum_{i=1}^{n_0} E_d\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha^*)\right].$$

By Lemma 2–(iii),

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|\Delta^{(1)}_{\epsilon,n_0}(\alpha^*)\right|^M\right] < \infty.$$

Proof of (7). One has that

$$\begin{aligned}
U^{(1)}_{\epsilon,n_0}(\alpha) - U^{(1)}_{\epsilon,n_0}(\alpha^*) &= -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0} E_d\left[(X_i - X_{i-1} - h_n a_{i-1}(\alpha))^{\otimes 2} - (X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{\otimes 2}\right] \\
&= \epsilon^{-2}\sum_{i=1}^{n_0}\{E_d\left[X_i - X_{i-1} - h_n a_{i-1}(\alpha^*), a_{i-1}(\alpha) - a_{i-1}(\alpha^*)\right]\} \\
&\quad - \frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0} E_d\left[(a_{i-1}(\alpha) - a_{i-1}(\alpha^*))^{\otimes 2}\right] h_n^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\epsilon^2\left\{U^{(1)}_{\epsilon,n_0}(\alpha) - U^{(1)}_{\epsilon,n_0}(\alpha^*)\right\} &= \sum_{i=1}^{n_0} E_d\left[X_i - X_{i-1} - h_n a_{i-1}(\alpha^*), a_{i-1}(\alpha) - a_{i-1}(\alpha^*)\right] \\
&\quad - \frac{1}{2}h_n\sum_{i=1}^{n_0} E_d\left[(a_{i-1}(\alpha) - a_{i-1}(\alpha^*))^{\otimes 2}\right] \\
&=: A^{(1)}_{n_0}(\alpha) - B^{(1)}_{n_0}(\alpha).
\end{aligned}$$

By Lemma 2–(iii),

$$E_{\theta^*}\left[\left(\sup_\alpha\left|A^{(1)}_{n_0}(\alpha)\right|\right)^M\right] \lesssim O(\epsilon^M)$$

and

$$E_{\theta^*}\left[\left(\sup_\alpha \epsilon^{-\eta_1}\left|A^{(1)}_{n_0}(\alpha)\right|\right)^M\right] \lesssim O(\epsilon^{(1-\eta_1)M}).$$

By Lemma 2–(ii),

$$E_{\theta^*}\left[\left(\sup_\alpha \epsilon^{-\eta_1}\left|B^{(1)}_{n_0}(\alpha) - \mathbb{Y}^{(1)}(\alpha)\right|\right)^M\right] < \infty,$$

which completes the proof of (7).

Proof of (8). We have

$$\begin{aligned}
\partial_\alpha^2 U^{(1)}_{\epsilon,n_0}(\alpha)[u_1,u_1] &= -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0} E_d\left[-2h_n\partial_\alpha^2 a_{i-1}(\alpha)[u_1,u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha)\right] \\
&\quad - \frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0} E_d\left[-2h_n\partial_\alpha a_{i-1}(\alpha)[u_1], -h_n\partial_\alpha a_{i-1}(\alpha)[u_1]\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
-\epsilon^2\partial_\alpha^2 U^{(1)}_{\epsilon,n_0}(\alpha^*)[u_1,u_1] &= -\sum_{i=1}^{n_0} E_d\left[\partial_\alpha^2 a_{i-1}(\alpha^*)[u_1,u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha^*)\right] \\
&\quad + h_n\sum_{i=1}^{n_0} E_d\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], \partial_\alpha a_{i-1}(\alpha^*)[u_1]\right] \\
&=: C^{(1)}_{n_0}(\alpha^*)[u_1,u_1] + D^{(1)}_{n_0}(\alpha^*)[u_1,u_1].
\end{aligned}$$

By Lemma 2–(iii),

$$E_{\theta^*}\left[\left(\epsilon^{-\eta_1}\left|C_{n_0}^{(1)}(\alpha^*)\right|\right)^M\right] \lesssim \epsilon^{-\eta_1 M}O(\epsilon^M) \lesssim \epsilon^{(1-\eta_1)M}.$$

By Lemma 2–(ii),

$$E_{\theta^*}\left[\left(\epsilon^{-\eta_1}\left|D_{n_0}^{(1)}(\alpha^*) - \Gamma^{(1)}(\alpha^*)\right|\right)^M\right] < \infty,$$

which completes the proof of (8).

Proof of (9). We have

$$\begin{aligned}
\partial_\alpha^3 U_{\epsilon,n_0}^{(1)}(\alpha)[u_1,u_1,u_1] \;&=\; -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0} E_d\Big\{\Big[-2h_n\partial_\alpha^3 a_{i-1}(\alpha)[u_1,u_1,u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha)\Big] \\
&\quad + \Big[-6h_n\partial_\alpha^2 a_{i-1}(\alpha)[u_1,u_1], -h_n\partial_\alpha a_{i-1}(\alpha)[u_1]\Big]\Big\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\epsilon^2\partial_\alpha^3 U_{\epsilon,n_0}^{(1)}(\alpha)[u_1,u_1,u_1] \;&=\; \sum_{i=1}^{n_0} E_d\Big\{\Big[\partial_\alpha^3 a_{i-1}(\alpha)[u_1,u_1,u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha)\Big] \\
&\quad - \Big[\partial_\alpha^2 a_{i-1}(\alpha)[u_1,u_1], \partial_\alpha a_{i-1}(\alpha)[u_1]\Big]\, 3h_n\Big\} \\
&=: E_{n_0}^{(1)}(\alpha) + F_{n_0}^{(1)}(\alpha).
\end{aligned}$$

By Lemma 2–(iv),

$$E_{\theta^*}\left[\left(\sup_\alpha\left|E_{n_0}^{(1)}(\alpha)\right|\right)^M\right] < \infty.$$

By Lemma 2–(i),

$$E_{\theta^*}\left[\left(\sup_\alpha\left|F_{n_0}^{(1)}(\alpha)\right|\right)^M\right] < \infty,$$

which completes the proof of (9).

Theorem 3 of Yoshida (2011) together with (6)–(9) yields that for any $L > 0$, there exists $C_L > 0$ such that for all $n \in \mathbb{N}$, $\epsilon \in (0,1]$ and $r > 0$,

$$P_{\theta^*}\left[\sup_{u_1 \in \mathbb{V}_{\epsilon,r_1}^{(1)}(r)} \mathbb{Z}_{\epsilon,n_0,r_1}^{(1)}(u_1;\alpha^*) \geq e^{-r}\right] \leq \frac{C_L}{r^L}. \tag{10}$$

It follows from (6), (8) and (9) that for some $M > p$, $\delta > 0$ and $C_0 > 0$,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|K_{\epsilon,n_0,r_1}^{(1)}(u_1)\right|^M\right] \leq C_0|u_1|^M$$

for $u_1 \in \mathbb{W}_{\epsilon,r_1}^{(1)}(\alpha^*) := \Big\{u_1 \in \mathbb{U}_{\epsilon,r_1}^{(1)}(\alpha^*) \big| |u_1| \leq \delta\Big\}$. It follows from Lemma 2 of Yoshida (2011) that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\int_{\mathbb{W}_{\epsilon,r_1}^{(1)}(\alpha^*)} e^{K_{\epsilon,n_0,r_1}^{(1)}(u_1)} du_1\right)^{-1}\right] < \infty.$$

Thus,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(1)}_{\epsilon,r_1}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)du_1\right)^{-1}\right] < \infty. \tag{11}$$

In the similar way to the proof of Theorem 8 of Yoshida (2011), for all $n \in \mathbb{N}$ and $\epsilon \in (0,1]$,

$$\begin{aligned}
& E_{\theta^*}\left[\left|\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*)\right|^M\right] \\
\leq\ & E_{\theta^*}\left[\left\{\left(\int_{\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right)^{-1}\int_{\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)}|u_1|\mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right\}^M\right] \\
\leq\ & C\sum_{r=0}^{\infty}(r+1)^M E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right)^{-1}\right. \\
& \left.\times\int_{\{u_1|r\leq|u_1|\leq r+1\}\cap\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right] \\
\leq\ & C\sum_{r=0}^{\infty}(r+1)^M\left\{E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right)^{-1}\right.\right. \\
& \left.\times\int_{\{u_1|r\leq|u_1|\leq r+1\}\cap\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1 1_{\left\{\sup_{u_1\in\mathbb{V}^{(1)}_{\epsilon}(r)}\mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\geq e^{-r}\right\}}\right] \\
& \left.+e^{-r}\int_{r\leq|u_1|\leq r+1}du_1 E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right)^{-1}\right]\right\} \\
\leq\ & C\sum_{r=0}^{\infty}(r+1)^M\left\{P_{\theta^*}\left[\sup_{u_1\in\mathbb{V}^{(1)}_{\epsilon}(r)}\mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\geq e^{-r}\right]\right. \\
& \left.+e^{-r}\int_{r\leq|u_1|\leq r+1}du_1 E_{\theta^*}\left[\left(\int_{\mathbb{W}^{(1)}_{\epsilon}(\alpha^*)} \mathbb{Z}^{(1)}_{\epsilon,n_0,r_1}(u_1;\alpha^*)\pi_1(\alpha^*+\epsilon^{r_1}u_1)du_1\right)^{-1}\right]\right\},
\end{aligned}$$

which together with (6) and (7) completes the proof of (i).

(ii). Set $\mathbb{U}^{(2)}_{n_0,r_2}(\beta^*) = \left\{u_2\in\mathbb{R}^q|\beta^*+\frac{u_2}{(\sqrt{n_0})^{r_2}}\in\Theta_\beta\right\}$ and $\mathbb{V}^{(2)}_{n_0,r_2}(r) = \left\{u_2\in\mathbb{U}^{(2)}_{n_0,r_2}(\beta^*)|r\leq|u_2|\right\}$. Let

$$\begin{aligned}
\mathbb{Z}^{(2)}_{\epsilon,n_0,r_2}(u_2;\beta^*) &= \exp\left\{\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n,r_1},\beta^*+\frac{u_2}{(\sqrt{n_0})^{r_2}})-\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)\right\} \\
&= \exp\left[\frac{1}{(\sqrt{n_0})^{2-2r_2}}\left\{U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*+\frac{u_2}{(\sqrt{n_0})^{r_2}})-U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)\right\}\right],
\end{aligned}$$

for $u_2\in\mathbb{U}^{(2)}_{n_0,r_2}(\beta^*)$. We have that

$$(\sqrt{n_0})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}-\beta^*) = \frac{\int_{\mathbb{U}^{(2)}_{n_0,r_2}(\beta^*)}u_2\mathbb{Z}^{(2)}_{\epsilon,n_0,r_2}(u_2;\beta^*)\pi_2(\beta^*+\frac{u_2}{(\sqrt{n_0})^{r_2}})du_2}{\int_{\mathbb{U}^{(2)}_{n_0,r_2}(\beta^*)}\mathbb{Z}^{(2)}_{\epsilon,n_0,r_2}(u_2;\beta^*)\pi_2(\beta^*+\frac{u_2}{(\sqrt{n_0})^{r_2}})du_2},$$

and

$$\begin{aligned}K^{(2)}_{\epsilon,n_0,r_2}(u_2) &:= \log \mathbb{Z}^{(2)}_{\epsilon,n_0,r_2}(u_2;\beta^*)\\ &= \frac{1}{(\sqrt{n_0})^{2-r_2}}\partial_\beta U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)[u_2] + \frac{1}{2}\frac{1}{n_0}\partial^2_\beta U^{(2)}_{\epsilon,n}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)[u_2^{\otimes 2}]\\ &\quad + \frac{1}{2}\frac{1}{n_0}\frac{1}{n_0^{\frac{r_2}{2}}}\sum_{i,j,k=1}^{q}\int_0^1 (1-t)^2\partial_{\beta_i}\partial_{\beta_j}\partial_{\beta_k}U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*+t\frac{u_2}{(\sqrt{n_0})^{r_2}})dt u_{2,i}u_{2,j}u_{2,k}.\end{aligned}$$

Set

$$\begin{aligned}\mathbb{Y}^{(2)}_{\epsilon,n_0,r_2}(\beta) &= \frac{1}{(\sqrt{n_0})^{2r_2}}\left\{\mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta) - \mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)\right\}\\ &= \frac{1}{n_0}\left\{U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta) - U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)\right\},\\ \Delta^{(2)}_{\epsilon,n_0,r_2}(\beta^*)[u_2] &= \frac{1}{(\sqrt{n_0})^{r_2}}\partial_\beta \mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)[u_2]\\ &= \frac{1}{(\sqrt{n_0})^{2-r_2}}\partial_\beta U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)[u_2]\\ &= \frac{1}{(\sqrt{n_0})^{1-r_2}}\frac{1}{\sqrt{n_0}}\partial_\beta U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta^*)[u_2]\\ &\quad + \epsilon^{r_1}\frac{1}{(\sqrt{n_0})^{-r_2}}\frac{\epsilon^2}{\epsilon^2 n_0}\partial_\alpha\partial_\beta U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta^*)[\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*),u_2]\\ &\quad + \epsilon^{2r_1}\frac{1}{(\sqrt{n_0})^{-r_2}}\frac{\epsilon^2}{\epsilon^2 n_0}\int_0^1 \partial^2_\alpha\partial_\beta U^{(2)}_{\epsilon,n_0}(\alpha^*+t(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*),\beta^*)dt\\ &\quad \times [(\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*))^{\otimes 2},u_2],\\ \Gamma^{(2)}_{\epsilon,n_0,r_2}(\beta^*)[u_2,u_2] &= -\frac{1}{n_0^{2r_2}}\partial^2_\beta \mathbb{H}^{(2)}_{\epsilon,n_0,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)[u_2,u_2]\\ &= -\frac{1}{n_0}\partial^2_\beta U^{(2)}_{\epsilon,n_0}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1},\beta^*)[u_2,u_2],\\ \Gamma^{(2)}_{ij}(\beta^*)[u_2,u_2] &= \frac{1}{2}\frac{c}{T}\int_0^{T/c}\operatorname{tr}\left\{(\partial_{\beta_i}B)(X^0_t,\beta^*)[u_2](\partial_{\beta_j}B)(X^0_t,\beta^*)[u_2]\right\}dt\end{aligned}$$

for $i,j=1,\dots,q$.

We have that

$$\begin{aligned}\partial^3_\beta U^{(2)}_{\epsilon,n_0}(\alpha,\beta)[u_2,u_2,u_2] &= -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0}\partial^3_\beta B^{-1}_{i-1}(\beta)[u_2,u_2,u_2][(X_i - X_{i-1} - h_n a_{i-1}(\alpha))^{\otimes 2}]\\ &\quad - \frac{1}{2}\sum_{i=1}^{n_0}\partial^3_\beta \log\det B_{i-1}(\beta)[u_2,u_2,u_2]\\ &=: (\epsilon^2 h_n)^{-1}\sum_{i=1}^{n_0}E^{(2)}_{i-1}(\alpha,\beta) + \sum_{i=1}^{n_0}F^{(2)}_{i-1}(\beta).\end{aligned}$$

By Lemma 3–(ii),

$$E_{\theta^*}\left[\left(\sup_\theta(\epsilon^2 h_n)^{-1}\left|\frac{1}{n_0}\sum_{i=1}^{n_0}E^{(2)}_{i-1}(\alpha,\beta)\right|\right)^M\right] \lesssim (\epsilon^2 h_n)^{-M}\frac{1}{n_0^M}O(\epsilon^{2M}) = O(1).$$

By Lemma 2–(i),

$$E_{\theta^*}\left[\left(\sup_\beta\left|\frac{1}{n_0}\sum_{i=1}^{n_0}F_{i-1}^{(2)}(\beta)\right|\right)^M\right] \lesssim O(1).$$

Thus, for all $M>0$,

$$\sup_{\epsilon,n}E_{\theta^*}\left[\left(\frac{1}{n_0}\sup_\theta|\partial_\beta^3 U_{\epsilon,n_0}^{(2)}(\alpha,\beta)|\right)^M\right]<\infty. \tag{12}$$

We will show that for all $M>0$, under $\frac{1}{\epsilon\sqrt{n}}=O(1)$ and $\epsilon(\sqrt{n})^\gamma=O(1)$,

$$\sup_{\epsilon,n}E_{\theta^*}\left[|\Delta_{\epsilon,n_0,r_2}^{(2)}(\beta^*)|^M\right]<\infty, \tag{13}$$

$$\sup_{\epsilon,n}E_{\theta^*}\left[\left(\sup_\beta(\sqrt{n_0})^{r_1\gamma}|\mathbb{Y}_{\epsilon,n_0,r_2}^{(2)}(\beta)-\mathbb{Y}^{(2)}(\beta)|\right)^M\right]<\infty, \tag{14}$$

$$\sup_{\epsilon,n}E_{\theta^*}\left[\left((\sqrt{n_0})^{r_1\gamma}|\Gamma_{\epsilon,n_0,r_2}^{(2)}(\beta^*)-\Gamma^{(2)}(\beta^*)|\right)^M\right]<\infty. \tag{15}$$

Proof of (13). We have

$$\begin{aligned}\partial_\beta U_{\epsilon,n_0}^{(2)}(\alpha,\beta)[u_2] &= -\frac{1}{2}(\epsilon^2h_n)^{-1}\sum_{i=1}^{n_0}\Big\{\partial_\beta B_{i-1}^{-1}(\beta)[u_2][(X_i-X_{i-1}-h_na_{i-1}(\alpha))^{\otimes 2}]\\ &\quad+\epsilon^2h_n\partial_\beta\log\det B_{i-1}(\beta)[u_2]\Big\}\\ &= -\frac{1}{2}(\epsilon^2h_n)^{-1}\sum_{i=1}^{n_0}\partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left[(X_i-X_{i-1}-h_na_{i-1}(\alpha))^{\otimes 2}-\epsilon^2h_nB_{i-1}(\beta)\right].\end{aligned}$$

By Lemma 3–(i),

$$E_{\theta^*}\left[\left|\frac{1}{\sqrt{n_0}}\partial_\beta U_{\epsilon,n_0}^{(2)}(\alpha^*,\beta^*)\right|^M\right]=O(1)<\infty.$$

Next,

$$\partial_\alpha\partial_\beta U_{\epsilon,n_0}^{(2)}(\alpha,\beta)[u_1,u_2] = h_n(\epsilon^2h_n)^{-1}\sum_{i=1}^{n_0}\partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left[\partial_\alpha a_{i-1}(\alpha)[u_1],X_i-X_{i-1}-h_na_{i-1}(\alpha)\right].$$

By Lemma 2–(iii),

$$\sup_{\epsilon,n}E_{\theta^*}\left[\left|\epsilon\partial_\alpha\partial_\beta U_{\epsilon,n_0}^{(2)}(\alpha^*,\beta^*)\right|^M\right]<\infty.$$

Moreover,

$$\begin{aligned}&\partial_\alpha^2\partial_\beta U_{\epsilon,n_0}^{(2)}(\alpha,\beta)[u_1,u_1,u_2]\\ =\ & h_n(\epsilon^2h_n)^{-1}\sum_{i=1}^{n_0}\partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left[\partial_\alpha^2a_{i-1}(\alpha)[u_1,u_1],X_i-X_{i-1}-h_na_{i-1}(\alpha)\right]\\ & -2h_n^2(\epsilon^2h_n)^{-1}\sum_{i=1}^{n_0}\partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left[\partial_\alpha^2a_{i-1}(\alpha)[u_1,u_1],\partial_\alpha a_{i-1}(\alpha)[u_1]\right].\end{aligned}$$

By Lemma 2–(i) and (iv),

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_{\alpha}\left|\epsilon^2\partial_\alpha^2\partial_\beta U^{(2)}_{\epsilon,n_0}(\alpha,\beta^*)\right|\right)^M\right]<\infty.$$

Note that

$$\begin{aligned}\epsilon^{2r_1}(\sqrt{n_0})^{r_2} &= \epsilon^{2r_1-\frac{r_2}{\gamma}}(\epsilon(\sqrt{n_0})^{\gamma})^{\frac{r_2}{\gamma}}<\infty,\\ \epsilon^{1+r_1}(\sqrt{n_0})^{r_2} &= \epsilon^{1-r_1}\epsilon^{2r_1}(\sqrt{n_0})^{r_2}<\infty,\end{aligned}$$

which completes the proof of (13).

Proof of (14). We have that

$$\begin{aligned}\mathbb{Y}^{(2)}_{\epsilon,n_0,r_2}(\beta) &= \frac{1}{n_0}\left\{U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta)-U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta^*)\right\}\\ &\quad+\epsilon^{r_1}\frac{1}{n_0}\int_0^1\left\{\partial_\alpha U^{(2)}_{\epsilon,n_0}(\alpha^*+t(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*),\beta)\right.\\ &\quad\left.-\partial_\alpha U^{(2)}_{\epsilon,n}(\alpha^*+t(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*),\beta^*)\right\}dt[\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n_0,r_1}-\alpha^*)]\end{aligned}$$

and

$$\begin{aligned}&U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta)-U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta^*)\\ =\;&\frac{1}{\epsilon^2h_n}\sum_{i=1}^{n_0}tr\left(\left((\Delta_i-h_na_{i-1}(\alpha^*))^{\otimes 2}-\epsilon^2h_nB_{i-1}(\beta^*)\right)(B_{i-1}(\beta)-B_{i-1}(\beta^*))\right)\\ &-\frac{1}{2}\sum_{i=1}^{n_0}tr\left((B_{i-1}(\beta)-B_{i-1}(\beta^*))^2\right).\end{aligned}$$

By noting that $0<r_1\gamma\leq 1$, it follows from Lemma 2-(ii) and Lemma 3-(i) that

$$\sup_{\epsilon,n}E_{\theta^*}\left[\left(\sup_{\beta}(\sqrt{n_0})^{r_1\gamma}\left|\frac{1}{n_0}\left\{U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta)-U^{(2)}_{\epsilon,n_0}(\alpha^*,\beta^*)\right\}-\mathbb{Y}^{(2)}(\beta)\right|\right)^M\right]<\infty.$$

Next, one has that

$$\begin{aligned}\partial_\alpha U^{(2)}_{\epsilon,n_0}(\alpha,\beta) &= -\frac{1}{2\epsilon^4h_n^2}\sum_{i=1}^{n_0}\partial_\alpha\left\|(\Delta X_i-h_na_{i-1}(\alpha))^{\otimes 2}-(\epsilon^2h_n)B_{i-1}(\beta)\right\|^2\\ =\;&\frac{2}{\epsilon^4h_n}\sum_{i=1}^{n_0}\left((\Delta X_i-h_na_{i-1}(\alpha))^{\otimes 2}-(\epsilon^2h_n)B_{i-1}(\beta)\right)[\Delta X_i-h_na_{i-1}(\alpha),\partial_\alpha a_{i-1}(\alpha)[u_1]].\end{aligned}$$

Since for all $M>0$,

$$E_{\theta^*}\left[\left|P_i^{j_1}(\alpha^*)P_i^{j_2}(\alpha^*)P_i^{j_3}(\alpha^*)\right|^M\right]=R\left(h_n^{3M},X_{i-1}\right)+R\left(\epsilon^{3M}h_n^{\frac{3M}{2}},X_{i-1}\right),$$

we have that

$$\sup_{\epsilon,n}E_{\theta^*}\left[\left(\sup_{\theta}\;(\epsilon(\sqrt{n_0})^{\gamma})^{r_1}\left|\frac{1}{n_0}\partial_\alpha U^{(2)}_{\epsilon,n_0}(\alpha,\beta)\right|\right)^M\right]<\infty,$$

by Lemma 2-(i), (iii) and Lemma 3-(i). This completes the proof of (14).

Proof of (15). We have that for all $M > 0$,

$$\begin{aligned}\frac{1}{n_0}\partial_\beta^2 U_{\epsilon,n_0}^{(2)}(\tilde{\alpha}_{\epsilon,n_0,r_1}^{(1)},\beta^*) - \Gamma^{(2)}(\beta^*) &= \frac{1}{n_0}\partial_\beta^2 U_{\epsilon,n_0}^{(2)}(\alpha^*,\beta^*) - \Gamma^{(2)}(\beta^*)\\ &+ \frac{\epsilon^{r_1}}{n_0}\int_0^1 \left\{\partial_\alpha\partial_\beta^2 U_{\epsilon,n_0}^{(2)}(\alpha^* + t(\tilde{\alpha}_{\epsilon,n_0,r_1}^{(1)} - \alpha^*),\beta^*)\right\} dt[\epsilon^{-r_1}(\tilde{\alpha}_{\epsilon,n_0,r_1}^{(1)} - \alpha^*)],\end{aligned}$$

and for $j, k = 1, \ldots, q$,

$$\begin{aligned}\partial_{\beta_j}\partial_{\beta_k} U_{\epsilon,n_0}^{(2)}(\alpha,\beta) &= \frac{1}{\epsilon^2 h_n}\sum_{i=1}^{n_0} tr\left(\left((\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2} - (\epsilon^2 h_n)B_{i-1}(\beta)\right)\left(\partial_{\beta_j}\partial_{\beta_k} B_{i-1}(\beta)\right)\right)\\ &\quad - \sum_{i=1}^{n_0} tr\left(\left(\partial_{\beta_j} B_{i-1}(\beta)\right)(\partial_{\beta_k} B_{i-1}(\beta))\right),\\ \partial_{\alpha_l}\partial_{\beta_j}\partial_{\beta_k} U_{\epsilon,n_0}^{(2)}(\alpha,\beta) &= \frac{-2}{\epsilon^2}\sum_{i=1}^{n_0}\left(\partial_{\beta_j}\partial_{\beta_k} B_{i-1}(\beta)\right)\left[\Delta X_i - h_n a_{i-1}(\alpha), \partial_{\alpha_l} a_{i-1}(\alpha)\right].\end{aligned}$$

Since

$$\epsilon^{r_1}(\sqrt{n})^{r_1\gamma} = (\epsilon(\sqrt{n})^\gamma)^{r_1} < \infty,$$

it follows from Lemma 2-(ii) and (iv) and Lemma 3-(i) that

$$\begin{aligned}&\sup_{\epsilon,n} E_{\theta^*}\left[\left((\sqrt{n_0})^\gamma\left|h_n\partial_\beta^2 U_{\epsilon,n_0}^{(2)}(\alpha^*,\beta^*) - \Gamma^{(2)}(\beta^*)\right|\right)^M\right] < \infty,\\ &\sup_{\epsilon,n_0} E_{\theta^*}\left[\left(\sup_\alpha\left|\frac{1}{n_0}\partial_\alpha\partial_\beta^2 U_{\epsilon,n_0}^{(2)}(\alpha,\beta^*)\right|\right)^M\right] < \infty,\end{aligned}$$

which completes the proof of (15).

By (12)–(15) and Theorem 3 of Yoshida (2011), one has that for any $L > 0$, there exists $C_L > 0$ such that for all $n \in \mathbb{N}$, $\epsilon \in (0, 1]$ and $r > 0$,

$$P_{\theta^*}\left[\sup_{u_2\in\mathbb{V}_{n_0,r_2}^{(2)}(r)} \mathbb{Z}_{\epsilon,n_0,r_2}^{(2)}(u_2;\beta^*) \geq e^{-r}\right] \leq \frac{C_L}{r^L}. \tag{16}$$

By (12), (13) and standard estimates, for some $M > q$, $\delta > 0$ and $C_0 > 0$,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|K_{\epsilon,n_0,r_2}^{(2)}(u_2)\right|^M\right] \leq C_0|u_2|^M$$

for $u_2 \in \mathbb{W}_{n_0,r_2}^{(2)}(\beta^*) := \left\{u_2 \in \mathbb{U}_{n_0,r_2}^{(2)}(\beta^*) | |u_2| \leq \delta\right\}$. It follows from Lemma 2 of Yoshida (2011) that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\int_{\mathbb{W}_{n_0,r_2}^{(2)}(\beta^*)} e^{K_{\epsilon,n,r_2}^{(2)}(u_2)} du_2\right)^{-1}\right] < \infty.$$

Hence,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\int_{\mathbb{U}^{(2)}_{n_0,r_2}(\beta^*)} \mathbb{Z}^{(2)}_{\epsilon,n_0,r_2}(u_2;\beta^*)du_2\right)^{-1}\right] < \infty.$$

In the similar way to the proof of Theorem 1–(i), for all $M>0$,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|(\sqrt{n_0})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}-\beta^*)\right|^M\right] < \infty, \tag{17}$$

which completes the proof of (ii).

(iii). Set

$$\begin{aligned}
&\mathbb{Y}^{(3)}_{\epsilon,n}(\alpha) = \epsilon^2\left\{U^{(3)}_{\epsilon,n}(\alpha,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) - U^{(3)}_{\epsilon,n}(\alpha^*,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})\right\},\\
&\Delta^{(3)}_{\epsilon,n}(\alpha^*)[u_1] = \epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})[u_1],\\
&\Gamma^{(3)}_{\epsilon,n}(\alpha^*)[u_1,u_1] = -\epsilon^2\partial^2_\alpha U^{(3)}_{\epsilon,n}(\alpha^*,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})[u_1,u_1],\\
&\Gamma^{(3)}(\theta^*)[u_1,u_1] = I_a(\theta^*)[u_1,u_1],\\
&\mathbb{Z}^{(3)}_{\epsilon,n}(u_1;\alpha^*) = \exp\left\{U^{(3)}_{\epsilon,n}(\alpha^*+\epsilon u_1,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) - U^{(3)}_{\epsilon,n}(\alpha^*,\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})\right\}.
\end{aligned}$$

One has that

$$\begin{aligned}
\epsilon^2\partial^3_\alpha U^{(3)}_{\epsilon,n}(\alpha,\beta)[u_1,u_1,u_1] \;&=\; \sum_{i=1}^n B^{-1}_{i-1}(\beta)\left[X_i - X_{i-1} - h_n a_{i-1}(\alpha), \partial^3_\alpha a_{i-1}(\alpha)[u_1,u_1,u_1]\right]\\
&\quad - 3h_n\sum_{i=1}^n B^{-1}_{i-1}(\beta)\left[\partial_\alpha a_{i-1}(\alpha)[u_1], \partial^2_\alpha a_{i-1}(\alpha)[u_1,u_1]\right].
\end{aligned}$$

By Lemma 2–(i) and (iii),

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\theta |\epsilon^2\partial^3_\alpha U^{(3)}_{\epsilon,n}(\alpha,\beta)|\right)^M\right] < \infty. \tag{18}$$

It will be shown that there exists $\eta_3\in(0,r_2)$ such that for all $M>0$,

$$\sup_{\epsilon,n} E_{\theta^*}\left[|\Delta^{(3)}_{\epsilon,n}(\alpha^*)|^M\right] < \infty, \tag{19}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\alpha \epsilon^{-\eta_3}|\mathbb{Y}^{(3)}_{\epsilon,n}(\alpha) - \mathbb{Y}^{(3)}(\alpha)|\right)^M\right] < \infty, \tag{20}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\epsilon^{-\eta_3}|\Gamma^{(3)}_{\epsilon,n}(\alpha^*) - \Gamma^{(3)}(\alpha^*)|\right)^M\right] < \infty. \tag{21}$$

Proof of (19). We have that

$$\begin{aligned}
\Delta^{(3)}_{\epsilon,n}(\alpha^*)[u_1] \;&=\; \epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*,\beta^*)[u_1]\\
&\quad + \epsilon\frac{1}{(\sqrt{n})^{r_2}}\int_0^1 \partial_\alpha\partial_\beta U^{(3)}_{\epsilon,n}(\alpha^*,\beta^*+t(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}-\beta^*))dt[(\sqrt{n})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}-\beta^*)].
\end{aligned}$$

Since

$$\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*,\beta^*)[u_1] = \epsilon^{-2}\sum_{i=1}^n B^{-1}_{i-1}(\beta^*)\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha^*)\right],$$

it follows from Lemma 2–(iii) that

$$E_{\theta^*}\left[\left|\epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\theta^*)\right|^M\right] < \infty.$$

Next, by noting that

$$\partial_\alpha\partial_\beta U^{(3)}_{\epsilon,n}(\theta)[u_1,u_2] = \epsilon^{-2}\sum_{i=1}^n \partial_\beta B^{-1}_{i-1}(\beta)[u_2]\left[\partial_\alpha a_{i-1}(\alpha)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha)\right],$$

and by using Lemma 2–(iv),

$$E_{\theta^*}\left[\left(\sup_\theta\left|\frac{\epsilon}{\sqrt{n}}\partial_\alpha\partial_\beta U^{(3)}_{\epsilon,n}(\theta)\right|\right)^M\right] \lesssim \left(\frac{1}{\epsilon\sqrt{n}}\right)^M O(1) < \infty.$$

Consequently,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|\Delta^{(3)}_{\epsilon,n}(\alpha^*)\right|^M\right] < \infty,$$

which completes the proof of (19).

Proof of (20). One has that

$$\begin{aligned}
\mathbb{Y}^{(3)}_{\epsilon,n}(\alpha) &= \sum_{i=1}^n B^{-1}_{i-1}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})\left[X_i - X_{i-1} - h_n a_{i-1}(\alpha^*), a_{i-1}(\alpha) - a_{i-1}(\alpha^*)\right] \\
&\quad - \frac{1}{2}h_n\sum_{i=1}^n B^{-1}_{i-1}(\beta^*)\left[a_{i-1}(\alpha) - a_{i-1}(\alpha^*), a_{i-1}(\alpha) - a_{i-1}(\alpha^*)\right] \\
&\quad - \frac{1}{2}\frac{1}{(\sqrt{n})^{r_2}}h_n\sum_{i=1}^n\int_0^1 \partial_\beta B^{-1}_{i-1}(\beta^* + t(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*))dt \\
&\quad \times [(\sqrt{n})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)][(a_{i-1}(\alpha) - a_{i-1}(\alpha^*))^{\otimes 2}] \\
&=: A_n + B_n(\alpha) + \frac{1}{(\sqrt{n})^{r_2}}C_n[(\sqrt{n})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)].
\end{aligned}$$

By Lemma 2-(iii), one has that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}A_n\right|^M\right] < \infty.$$

Let

$$G_{i-1}(\alpha) := B^{-1}_{i-1}(\beta^*)\left[a_{i-1}(\alpha) - a_{i-1}(\alpha^*), a_{i-1}(\alpha) - a_{i-1}(\alpha^*)\right].$$

By Lemma 2–(ii),

$$\sup_{\epsilon,n} E\left[\left(\sup_\alpha \epsilon^{-\eta_3}\left|B_n(\alpha) - \mathbb{Y}^{(3)}(\alpha)\right|\right)^M\right] = E_{\theta^*}\left[\left(\sup_\alpha \epsilon^{-\eta_3}\left|-\frac{1}{2}h_n\sum_{i=1}^n G_{i-1}(\alpha) - \mathbb{Y}^{(3)}(\alpha)\right|\right)^M\right] < \infty.$$

It follows from Lemma 2–(i) and $\frac{1}{\epsilon\sqrt{n}} = O(1)$ that

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\frac{1}{(\epsilon\sqrt{n})^{r_2}}|C_n|\right)^M\right] = E_{\theta^*}\left[\left(\sup_\theta \frac{1}{(\epsilon\sqrt{n})^{r_2}}h_n\left|\sum_{i=1}^n \partial_\beta B^{-1}_{i-1}(\beta)[(a_{i-1}(\alpha) - a_{i-1}(\alpha^*))^{\otimes 2}]\right|\right)^M\right] < \infty,$$

which completes the proof of (20).

Proof of (21). We have that

$$\begin{aligned}\Gamma^{(3)}_{\epsilon,n}(\alpha^*)[u_1,u_1] &= -\sum_{i=1}^{n} B^{-1}_{i-1}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})\left[X_i - X_{i-1} - h_n a_{i-1}(\alpha^*), \partial^2_\alpha a_{i-1}(\alpha)[u_1,u_1]\right] \\ &\quad + h_n \sum_{i=1}^{n} B^{-1}_{i-1}(\beta^*)\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], \partial_\alpha a_{i-1}(\alpha^*)[u_1]\right] \\ &\quad + \frac{1}{(\sqrt{n})^{r_2}} h_n \sum_{i=1}^{n} \int_0^1 \partial_\beta B^{-1}_{i-1}(\beta^* + t(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*))dt \\ &\quad \times [(\sqrt{n})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)]\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], \partial_\alpha a_{i-1}(\alpha^*)[u_1]\right] \\ &=: A^{(3)}_n + B^{(3)}_n + \frac{1}{(\sqrt{n})^{r_2}} C^{(3)}_n[(\sqrt{n})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)].\end{aligned}$$

By Lemma 2–(iii),

$$E_{\theta^*}\left[\left(\sup_\beta \epsilon^{-\eta_3}\left|A^{(3)}_n(\beta)\right|\right)^M\right] \lesssim \epsilon^{-\eta_3 M} O(\epsilon^M) \lesssim O(\epsilon^{(1-\eta_3)M}).$$

By Lemma 2–(ii),

$$E_{\theta^*}\left[\left(\epsilon^{-\eta_3}\left|B^{(3)}_n - \Gamma^{(3)}(\theta^*)\right|\right)^M\right] < \infty. \tag{22}$$

By Lemma 2–(i) and $\frac{1}{\epsilon\sqrt{n}} = O(1)$,

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\frac{1}{(\epsilon\sqrt{n})^{r_2}}\left|C^{(3)}_n\right|\right)^M\right] < \infty,$$

Hence,

$$E_{\theta^*}\left[\left|\epsilon^{-\eta_3}\left(\Gamma^{(3)}_{\epsilon,n}(\alpha^*) - \Gamma^{(3)}(\theta^*)\right)\right|^M\right] < \infty,$$

which completes the proof of (21).

Let $\mathbb{V}^{(3)}_{\epsilon,n}(r) = \left\{u_1 \in \mathbb{U}^{(3)}_{\epsilon,n}(\alpha^*) \mid r \le |u_1|\right\}$, where $\mathbb{U}^{(3)}_{\epsilon,n}(\alpha^*) := \{u_1 \in \mathbb{R}^p | \alpha^* + \epsilon u_1 \in \Theta_\alpha\}$. It follows from (18)–(21) and Theorem 3 of Yoshida (2011) that for any $L > 0$, there exists $C_L > 0$ such that for all $n \in \mathbb{N}$, $\epsilon \in (0,1]$ and $r > 0$,

$$P_{\theta^*}\left[\sup_{u_1 \in \mathbb{V}^{(3)}_{\epsilon,n}(r)} \mathbb{Z}^{(3)}_{\epsilon,n}(u_1;\alpha^*) \ge e^{-r}\right] \le \frac{C_L}{r^L}, \tag{23}$$

and

$$P_{\theta^*}\left[\left|\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*)\right| > r\right] \le \frac{C_L}{r^L}, \tag{24}$$

which completes the proof of (iii).

(iv). Let

$$
\begin{aligned}
&\mathbb{Y}^{(4)}_{\epsilon,n}(\beta) = \frac{1}{n}\left\{U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta) - U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta^*)\right\},\\
&\Delta^{(4)}_{\epsilon,n}(\beta^*)[u_2] = \frac{1}{\sqrt{n}}\partial_\beta U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta^*)[u_2],\\
&\Gamma^{(4)}_{\epsilon,n}(\beta^*)[u_2,u_2] = -\frac{1}{n}\partial^2_\beta U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta^*)[u_2,u_2],\\
&\Gamma^{(4)}(\beta^*)[u_2,u_2] = I_b(\beta^*)[u_2,u_2],\\
&\mathbb{Z}^{(4)}_{\epsilon,n}(u_2;\beta^*) = \exp\left\{U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta^* + \frac{u_2}{\sqrt{n}}) - U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n},\beta^*)\right\}.
\end{aligned}
$$

We will show that there exists $\eta_4 \in (0,\gamma)$ such that for all $M > 0$,

$$\sup_{\epsilon,n} E_{\theta^*}\left[|\Delta^{(4)}_{\epsilon,n}(\beta^*)|^M\right] < \infty, \tag{25}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\sup_\beta (\sqrt{n})^{\eta_4}|\mathbb{Y}^{(4)}_{\epsilon,n}(\beta) - \mathbb{Y}^{(4)}(\beta)|\right)^M\right] < \infty, \tag{26}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left((\sqrt{n})^{\eta_4}|\Gamma^{(4)}_{\epsilon,n}(\beta^*) - \Gamma^{(4)}(\beta^*)|\right)^M\right] < \infty, \tag{27}$$

$$\sup_{\epsilon,n} E_{\theta^*}\left[\left(\frac{1}{n}\sup_\theta |\partial^3_\beta U^{(4)}_{\epsilon,n}(\theta)|\right)^M\right] < \infty. \tag{28}$$

Proof of (25). We have that

$$
\begin{aligned}
\partial_\beta U^{(4)}_{\epsilon,n}(\alpha,\beta)[u_2] &= -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \Big\{\partial_\beta B^{-1}_{i-1}(\beta)[u_2][(X_i - X_{i-1} - h_n a_{i-1}(\alpha))^{\otimes 2}]\\
&\quad + \epsilon^2 h_n \partial_\beta \log\det B_{i-1}(\beta)[u_2]\Big\}\\
&= -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \Big\{\partial_\beta B^{-1}_{i-1}(\beta)[u_2]\left[(X_i - X_{i-1} - h_n a_{i-1}(\alpha))^{\otimes 2} - \epsilon^2 h_n B_{i-1}(\beta)\right]\\
&\quad + \epsilon^2 h_n \partial_\beta B^{-1}_{i-1}(\beta)[u_2][B_{i-1}(\beta)] + \epsilon^2 h_n \partial_\beta \log\det B_{i-1}(\beta)[u_2]\Big\}
\end{aligned}
$$

and

$$\frac{1}{\sqrt{n}}\partial_\beta U^{(2)}_{\epsilon,n}(\alpha^*,\beta^*)[u_2] = -\frac{1}{2\sqrt{T}}(\epsilon^2 h_n^{\frac{1}{2}})^{-1}\sum_{i=1}^n \partial_\beta B^{-1}_{i-1}(\beta)[u_2]\left[(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{\otimes 2} - \epsilon^2 h_n B_{i-1}(\beta^*)\right].$$

By Lemma 3–(i),

$$E_{\theta^*}\left[\left|\frac{1}{\sqrt{n}}\partial_\beta U^{(2)}_{\epsilon,n}(\alpha^*,\beta^*)\right|^M\right] = O(1) < \infty.$$

Next,

$$
\begin{aligned}
\Delta^{(4)}_{\epsilon,n}(\beta^*)[u_2] &= \frac{1}{\sqrt{n}}\partial_\beta U^{(2)}_{\epsilon,n}(\alpha^*,\beta^*)[u_2]\\
&\quad + \frac{\epsilon}{\sqrt{n}}\int_0^1 \partial_\alpha\partial_\beta U^{(2)}_{\epsilon,n}(\alpha^* + t(\hat{\alpha}_{\epsilon,n} - \alpha^*),\beta^*)dt[u_2,\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*)].
\end{aligned}
$$

Let

$$A_n^{(4)}(\alpha) := \frac{1}{\epsilon\sqrt{n}} \sum_{i=1}^{n} \partial_\beta B_{i-1}^{-1}(\beta^*)[u_2]\left[\partial_\alpha a_{i-1}(\alpha)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha)\right].$$

By Lemma 2–(iv),

$$E_{\theta^*}\left[\left(\sup_\alpha \left|A_n^{(4)}(\alpha)\right|\right)^M\right] \lesssim O\left(\left(\frac{1}{\epsilon\sqrt{n}}\right)^M\right) < \infty.$$

Therefore,

$$E_{\theta^*}\left[\left|\Delta_{\epsilon,n}^{(4)}(\beta^*)\right|^M\right] < \infty,$$

which completes the proof of (25).

Proof of (26). One has that

$$\begin{aligned} & U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n}, \beta) - U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n}, \beta^*) \\ = \; & U_{\epsilon,n}^{(4)}(\alpha^*, \beta) - U_{\epsilon,n}^{(4)}(\alpha^*, \beta^*) \\ & + \left(\int_0^1 \partial_\alpha U_{\epsilon,n}^{(4)}(\alpha^* + t(\hat{\alpha}_{\epsilon,n} - \alpha^*), \beta)dt - \int_0^1 \partial_\alpha U_{\epsilon,n}^{(4)}(\alpha^* + t(\hat{\alpha}_{\epsilon,n} - \alpha^*), \beta^*)dt\right)[\hat{\alpha}_{\epsilon,n} - \alpha^*] \end{aligned}$$

and

$$\begin{aligned} & U_{\epsilon,n}^{(4)}(\alpha^*, \beta) - U_{\epsilon,n}^{(4)}(\alpha^*, \beta^*) \\ = \; & -\frac{1}{2}\sum_{i=1}^{n}\Big\{(\epsilon^2 h_n)^{-1}\left(B_{i-1}^{-1}(\beta) - B_{i-1}^{-1}(\beta^*)\right)\left[(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{\otimes 2} - \epsilon^2 h_n B_{i-1}(\beta^*)\right] \\ & + \left(B_{i-1}^{-1}(\beta) - B_{i-1}^{-1}(\beta^*)\right)[B_{i-1}(\beta^*)] + \log\frac{\det B_{i-1}(\beta)}{\det B_{i-1}(\beta^*)}\Big\} \\ =: \; & \sum_{i=1}^{n} A_{i-1}^{(4)}(\beta) + \sum_{i=1}^{n} B_{i-1}^{(4)}(\beta). \end{aligned}$$

By Lemma 3–(i),

$$\begin{aligned} & \sup_\beta E_{\theta^*}\left[\left((\sqrt{n})^{\eta_4}\left|\frac{1}{n}\sum_{i=1}^{n} A_{i-1}^{(4)}(\beta)\right|\right)^M\right] \\ \lesssim \; & (\epsilon^2 h_n)^{-M}(\sqrt{n})^{\eta_4 M}\left(\frac{1}{n}\right)^M E_{\theta^*}\left[\left|\sum_{i=1}^{n} f_{i-1}(\beta)\left[(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{\otimes 2} - \epsilon^2 h_n B_{i-1}(\beta^*)\right]\right|^M\right] \\ \lesssim \; & (\epsilon^2 h_n)^{-M}\left(\frac{(\sqrt{n})^{\eta_4}}{n}\right)^M (\epsilon^2 h_n^{\frac{1}{2}})^M \lesssim ((\sqrt{n})^{\eta_4} h_n^{\frac{1}{2}})^M < \infty \end{aligned}$$

and

$$\sup_\beta E_{\theta^*}\left[\left((\sqrt{n})^{\eta_4}\left|\frac{1}{n}\sum_{i=1}^{n}\partial_\beta A_{i-1}^{(2)}(\beta)\right|\right)^M\right] < \infty.$$

We can show that by Lemma 2–(ii) and $\epsilon(\sqrt{n})^{\eta_4} \le \epsilon(\sqrt{n})^{\gamma} = O(1)$,

$$E_{\theta^*}\left[\left(\sup_{\beta}(\sqrt{n})^{\eta_4}\left|\frac{1}{n}\sum_{i=1}^{n}B_{i-1}^{(4)}(\beta)-\mathbb{Y}^{(4)}(\beta)\right|\right)^M\right]<\infty,$$

which completes the proof of (26).

Proof of (27). We have

$$\begin{aligned}\frac{1}{n}\partial_\beta^2 U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n},\beta^*)-\Gamma^{(4)}(\alpha^*,\beta^*) &= \frac{1}{n}\partial_\beta^2 U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)-\Gamma^{(4)}(\alpha^*,\beta^*)\\ &\quad+\frac{\epsilon}{n}\int_0^1 \partial_\alpha\partial_\beta^2 U_{\epsilon,n}^{(4)}(\alpha^*+t(\hat{\alpha}_{\epsilon,n}-\alpha^*),\beta^*)dt[\epsilon^{-1}(\hat{\alpha}_{\epsilon,n}-\alpha^*)]\end{aligned}$$

and

$$\begin{aligned}&\partial_\beta^2 U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)[u_2,u_2]\\ =\ & -\frac{1}{2}(\epsilon^2h_n)^{-1}\sum_{i=1}^{n}\left(\partial_\beta^2 B_{i-1}^{-1}(\beta^*)[u_2,u_2]\right)\left[(X_i-X_{i-1}-h_na_{i-1}(\alpha^*))^{\otimes 2}-\epsilon^2h_nB_{i-1}(\beta^*)\right]\\ & -\frac{1}{2}\sum_{i=1}^{n}\left\{\left(\partial_\beta^2 B_{i-1}^{-1}(\beta^*)[u_2,u_2]\right)[B_{i-1}(\beta^*)]+\partial_\beta^2\log\det B_{i-1}(\beta^*)[u_2,u_2]\right\}\\ =\ & -(\epsilon^2h_n)^{-1}\sum_{i=1}^{n}C_{i-1}^{(4)}(\theta^*)-\sum_{i=1}^{n}D_{i-1}^{(4)}(\theta^*).\end{aligned}$$

Therefore,

$$-\frac{1}{n}\partial_\beta^2 U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)[u_2,u_2]=(\epsilon^2h_n)^{-1}\frac{1}{n}\sum_{i=1}^{n}C_{i-1}^{(4)}(\theta^*)+\frac{1}{n}\sum_{i=1}^{n}D_{i-1}^{(4)}(\theta^*).$$

By Lemma 3–(i),

$$\begin{aligned}E_{\theta^*}\left[\left|(\sqrt{n})^{\eta_4}(\epsilon^2h_n)^{-1}\frac{1}{n}\sum_{i=1}^{n}C_{i-1}^{(4)}(\theta^*)\right|^M\right] &\lesssim (\sqrt{n})^{\eta_4 M}\epsilon^{-2M}O\left((\epsilon^2h_n^{\frac{1}{2}})^M\right)\\ &\lesssim \left(\frac{(\sqrt{n})^{\eta_4}}{n^{\frac{1}{2}}}\right)^M<\infty.\end{aligned}$$

By Lemma 2–(ii),

$$E_{\theta^*}\left[\left|(\sqrt{n})^{\eta_4}\left(\frac{1}{n}\sum_{i=1}^{n}D_{i-1}^{(4)}(\theta^*)-\Gamma^{(4)}(\theta^*)\right)\right|^M\right]<\infty. \tag{29}$$

Therefore,

$$E_{\theta^*}\left[\left((\sqrt{n})^{\eta_4}\left|\frac{1}{n}\partial_\beta^2 U_{\epsilon,n}^{(4)}(\theta^*)-\Gamma^{(4)}(\theta^*)\right|\right)^M\right]<\infty.$$

Next,

$$\partial_\alpha\partial_\beta^2 U_{\epsilon,n}^{(2)}(\alpha,\beta^*)[u_1,u_2,u_2]=-\frac{1}{2}(\epsilon^2h_n)^{-1}\sum_{i=1}^{n}\partial_\beta^2 B_{i-1}^{-1}(\beta)[u_2,u_2]\left[-2h_n\partial_\alpha a_{i-1}(\alpha)[u_1],X_i-X_{i-1}-h_na_{i-1}(\alpha)\right].$$

By Lemma 2–(iv),

$$E_{\theta^*}\left[\left((\sqrt{n})^{\eta_4}\frac{\epsilon}{n}\sup_{\alpha}\left|\partial_\alpha\partial_\beta^2 U_{\epsilon,n}^{(4)}(\alpha,\beta^*)\right|\right)^M\right] \lesssim (\sqrt{n})^{\eta_4 M}\left(\frac{\epsilon}{n}\right)^M \epsilon^{-2M}O(1)$$
$$\lesssim (\sqrt{n})^{\eta_4 M}\left(\frac{1}{\epsilon n}\right)^M O(1) = \left(\frac{1}{\epsilon\sqrt{n}}\right)^M \left(\frac{(\sqrt{n})^{\eta_2}}{\sqrt{n}}\right)^M O(1),$$

which completes the proof of (27).

Proof of (28). We have

$$\begin{aligned}\partial_\beta^3 U_{\epsilon,n}^{(4)}(\alpha,\beta)[u_2,u_2,u_2] &= -\frac{1}{2}(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \partial_\beta^3 B_{i-1}^{-1}(\beta)[u_2,u_2,u_2][(X_i - X_{i-1} - h_n a_{i-1}(\alpha))^{\otimes 2}] \\ &\quad - \frac{1}{2}\sum_{i=1}^n \partial_\beta^3 \log\det B_{i-1}(\beta)[u_2,u_2,u_2] \\ &= (\epsilon^2 h_n)^{-1}\sum_{i=1}^n E_{i-1}^{(4)}(\alpha,\beta) + \sum_{i=1}^n F_{i-1}^{(4)}(\beta).\end{aligned}$$

By Lemma 3–(ii),

$$E_{\theta^*}\left[\left(\sup_{\theta}(\epsilon^2 h_n)^{-1}\left|\frac{1}{n}\sum_{i=1}^n E_{i-1}^{(4)}(\alpha,\beta)\right|\right)^M\right] \lesssim (\epsilon^2 h_n)^{-M}\frac{1}{n^M}O(\epsilon^{2M}) = O(1).$$

By Lemma 2–(i),

$$E_{\theta^*}\left[\left(\sup_{\beta}\left|\frac{1}{n}\sum_{i=1}^n F_{i-1}^{(4)}(\beta)\right|\right)^M\right] \lesssim O(1),$$

which completes the proof of (28).

Let $\mathbb{U}_{\epsilon,n}^{(4)} = \left\{u_2 \in \mathbb{R}^q | \beta^* + \frac{1}{\sqrt{n}}u_2 \in \Theta_\beta\right\}$ and $\mathbb{V}_{\epsilon,n}^{(4)}(r) = \left\{u_2 \in \mathbb{U}_{\epsilon,n}^{(4)} | r \le |u_2|\right\}$. By (25)–(28) and Theorem 3 of Yoshida (2011), one has that for any $L > 0$, there exists $C_L > 0$ such that for all $n \in \mathbb{N}$, $\epsilon \in (0,1)$ and $r > 0$,

$$P_{\theta^*}\left[\sup_{u_2 \in \mathbb{V}_{\epsilon,n}^{(4)}(r)} \mathbb{Z}_{\epsilon,n}^{(4)}(u_2;\beta^*) \ge e^{-r}\right] \le \frac{C_L}{r^L}, \tag{30}$$

and

$$P_{\theta^*}\left[\left|\sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*)\right| > r\right] \le \frac{C_L}{r^L}, \tag{31}$$

which completes the proof of (iv).

Proof of Theorem 2. Since

$$\frac{1}{\sqrt{n}}\partial_\beta U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n},\hat{\beta}_{\epsilon,n}) - \frac{1}{\sqrt{n}}\partial_\beta U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n},\beta^*) = \frac{1}{n}\int_0^1 \partial_\beta^2 U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n},\beta^* + t(\hat{\beta}_{\epsilon,n} - \beta^*))dt[\sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*)],$$

and

$$\epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n}, \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) - \epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) = \epsilon^2 \int_0^1 \partial^2_\alpha U^{(3)}_{\epsilon,n}(\alpha^* + t(\hat{\alpha}_{\epsilon,n} - \alpha^*), \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})dt[\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*)],$$

one has that

$$\begin{aligned}
&\begin{pmatrix} -\epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) \\ -\frac{1}{\sqrt{n}}\partial_\beta U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n}, \beta^*) \end{pmatrix} \\
= &\begin{pmatrix} \epsilon^2 \int_0^1 \partial^2_\alpha U^{(3)}_{\epsilon,n}(\alpha^* + t(\hat{\alpha}_{\epsilon,n} - \alpha^*), \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})dt & 0 \\ 0 & \frac{1}{n}\int_0^1 \partial^2_\beta U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n}, \beta^* + t(\hat{\beta}_{\epsilon,n} - \beta^*))dt \end{pmatrix} \begin{pmatrix} \epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*) \\ \sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*) \end{pmatrix}.
\end{aligned}$$

We will prove that as $(\epsilon\sqrt{n})^{-1} = O(1)$

$$\epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2}) - \epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \beta^*) \xrightarrow{p} 0, \tag{32}$$

$$\int_0^1 \epsilon^2\partial^2_\alpha U^{(3)}_{\epsilon,n}(\alpha^* + t(\hat{\alpha}_{\epsilon,n} - \alpha^*), \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})dt - \epsilon^2\partial^2_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \beta^*) \xrightarrow{p} 0, \tag{33}$$

$$\frac{1}{\sqrt{n}}\partial_\beta U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n}, \beta^*) - \frac{1}{\sqrt{n}}\partial_\beta U^{(4)}_{\epsilon,n}(\alpha^*, \beta^*) \xrightarrow{p} 0, \tag{34}$$

$$\int_0^1 \frac{1}{n}\partial^2_\beta U^{(4)}_{\epsilon,n}(\hat{\alpha}_{\epsilon,n}, \beta^* + t(\hat{\beta}_{\epsilon,n} - \beta^*))dt - \frac{1}{n}\partial^2_\beta U^{(4)}_{\epsilon,n}(\alpha^*, \beta^*) \xrightarrow{p} 0, \tag{35}$$

$$\left(-\epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\theta^*), -\frac{1}{\sqrt{n}}\partial_\beta U^{(4)}_{\epsilon,n}(\theta^*)\right) \xrightarrow{d} N_{p+q}(0, I(\theta^*)), \tag{36}$$

$$\left(\epsilon^2\partial^2_\alpha U^{(3)}_{\epsilon,n}(\theta^*), \frac{1}{n}\partial^2_\beta U^{(4)}_{\epsilon,n}(\theta^*)\right) \xrightarrow{p} (-I_a(\theta^*), -I_b(\beta^*)). \tag{37}$$

Proof of (32). One has that

$$\epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \tilde{\beta}^{(2)}_{\epsilon,n_0,r_2})[u_1] - \epsilon\partial_\alpha U^{(3)}_{\epsilon,n}(\alpha^*, \beta^*)[u_1] \quad = \quad \int_0^1 \epsilon\partial_\alpha\partial_\beta U^{(3)}_{\epsilon,n}(\alpha^*, \beta^* + t(\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*))dt[u_1, (\tilde{\beta}^{(2)}_{\epsilon,n_0,r_2} - \beta^*)]$$

and

$$\begin{aligned}
&\partial_\alpha\partial_\beta U^{(3)}_{\epsilon,n}(\alpha^*, \beta)[u_1, u_2] \\
= &\left(-\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \partial_\beta B^{-1}_{i-1}(\beta)[u_2]\left[-2h_n\partial_\alpha a_{i-1}(\alpha^*)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha^*)\right] \\
= &\epsilon^{-2}\sum_{i=1}^n \partial_\beta B^{-1}_{i-1}(\beta)[u_2]\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], X_i - X_{i-1} - h_n a_{i-1}(\alpha^*)\right].
\end{aligned}$$

By Lemma 2–(iii),

$$\begin{aligned}
E_{\theta^*}\left[\left(\epsilon\sup_\beta\left|\partial_\alpha\partial_\beta U^{(3)}_{\epsilon,n}(\alpha^*, \beta)\right|\right)^M\right] &\lesssim \epsilon^{-M}E_{\theta^*}\left[\left|\sum_{i=1}^n R(1, X_{i-1})(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))\right|^M\right] \\
&\lesssim \epsilon^{-M}O(\epsilon^M) < \infty.
\end{aligned}$$

And $\hat{\beta}_{\epsilon,n} - \beta^* \xrightarrow{p} 0$, which completes the proof of (32).

Proof of (33). We have that

$$\left|\epsilon^2\left(\partial_\alpha^2 U_{\epsilon,n}^{(3)}(\alpha^*+t(\hat{\alpha}_{\epsilon,n}-\alpha^*),\tilde{\beta}_{\epsilon,n_0,r_2}^{(2)})-\partial_\alpha^2 U_{\epsilon,n}^{(3)}(\alpha^*,\beta^*)\right)\right| \le \epsilon^2\sup_\theta\left|\partial_\theta\partial_\alpha^2 U_{\epsilon,n}^{(3)}(\theta)\right|\cdot\left|\begin{pmatrix} t(\hat{\alpha}_{\epsilon,n}-\alpha^*) \\ (\tilde{\beta}_{\epsilon,n_0,r_2}^{(2)}-\beta^*)\end{pmatrix}\right|.$$

Furthermore,

$$\begin{aligned}
&\epsilon^2\partial_\beta\partial_\alpha^2 U_{\epsilon,n}^{(3)}(\theta)[u_1,u_1,u_2] \\
=\ & \epsilon^2\left(-\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left\{\left[-2h_n\partial_\alpha^2 a_{i-1}(\alpha)[u_1,u_1],X_i-X_{i-1}-h_n a_{i-1}(\alpha)\right]\right. \\
& \left.+\left[-2h_n\partial_\alpha a_{i-1}(\alpha)[u_1],-h_n\partial_\alpha a_{i-1}(\alpha)[u_1]\right]\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\epsilon^2\partial_\alpha^3 U_{\epsilon,n}^{(3)}(\theta)[u_1,u_1,u_1] \\
=\ & \epsilon^2\left(-\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \partial_\beta B_{i-1}^{-1}(\beta)\left\{\left[-2h_n\partial_\alpha^3 a_{i-1}(\alpha)[u_1,u_1,u_1],X_i-X_{i-1}-h_n a_{i-1}(\alpha)\right]\right. \\
& \left.+\left[-4h_n\partial_\alpha^2 a_{i-1}(\alpha)[u_1,u_1],-h_n\partial_\alpha a_{i-1}(\alpha)[u_1]\right]\right\}.
\end{aligned}$$

By Lemma 2–(i) and (iv),

$$\begin{aligned}
E_{\theta^*}\left[\left(\sup_\theta\left|\epsilon^2\partial_\theta\partial_\alpha^2 U_{\epsilon,n}^{(3)}(\theta)\right|\right)^M\right] &\lesssim E_{\theta^*}\left[\left(\sup_\theta\left|\sum_{i=1}^n R(1,X_{i-1})(X_i-X_{i-1}-h_n a_{i-1}(\alpha))\right|\right)^M\right] \\
&\quad + E_{\theta^*}\left[\left(\sup_\theta\left|h_n\sum_{i=1}^n f_{i-1}(\theta)\right|\right)^M\right] \\
&< \infty.
\end{aligned}$$

Since $\hat{\alpha}_{\epsilon,n}\xrightarrow{p}\alpha^*$ and $\hat{\beta}_{\epsilon,n}\xrightarrow{p}\beta^*$, the proof of (33) is completed.

Proof of (34). We have that

$$\begin{aligned}
&\frac{\epsilon}{\sqrt{n}}\partial_\alpha\partial_\beta U_{\epsilon,n}^{(4)}(\alpha,\beta)[u_1,u_2] \\
=\ & \frac{\epsilon}{\sqrt{n}}\left(-\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left[-2h_n\partial_\alpha a_{i-1}(\alpha)[u_1],X_i-X_{i-1}-h_n a_{i-1}(\alpha)\right] \\
=\ & \frac{1}{\sqrt{n}}\epsilon^{-1}\sum_{i=1}^n \partial_\beta B_{i-1}^{-1}(\beta)[u_2]\left[\partial_\alpha a_{i-1}(\alpha)[u_1],X_i-X_{i-1}-h_n a_{i-1}(\alpha)\right]
\end{aligned}$$

and

$$\begin{aligned}
-\frac{1}{\sqrt{n}}\partial_\beta U_{\epsilon,n}^{(4)}(\hat{\alpha}_{\epsilon,n},\beta^*)[u_2] &= -\frac{1}{\sqrt{n}}\partial_\beta U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)[u_2]-\frac{\epsilon}{\sqrt{n}}\partial_\alpha\partial_\beta U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)[u_2,\epsilon^{-1}(\hat{\alpha}_{\epsilon,n}-\alpha^*)] \\
&\quad -\frac{\epsilon^2}{\sqrt{n}}\int_0^1(1-t)\partial_\alpha^2\partial_\beta U_{\epsilon,n}^{(4)}(\alpha^*+t(\hat{\alpha}_{\epsilon,n}-\alpha^*),\beta^*)dt[u_2,(\epsilon^{-1}(\hat{\alpha}_{\epsilon,n}-\alpha^*))^{\otimes 2}].
\end{aligned}$$

By Lemma 2–(iii),

$$\begin{aligned}
E_{\theta^*}\left[\left|\frac{\epsilon}{\sqrt{n}}\partial_\alpha\partial_\beta U_{\epsilon,n}^{(4)}(\theta^*)\right|^M\right] &\lesssim \left(\frac{1}{\epsilon\sqrt{n}}\right)^M E_{\theta^*}\left[\left|\sum_{i=1}^n R(1,X_{i-1})(X_i-X_{i-1}-h_n a_{i-1}(\alpha^*))\right|^M\right] \\
&\lesssim \left(\frac{1}{\epsilon\sqrt{n}}\right)^M O(\epsilon^M)\to 0.
\end{aligned}$$

By Lemma 2–(i) and (iv),

$$
\begin{aligned}
E_{\theta^*}\left[\left(\frac{\epsilon^2}{\sqrt{n}}\sup_{\alpha}\left|\partial_\alpha^2\partial_\beta U_{\epsilon,n}^{(4)}(\alpha,\beta^*)\right|\right)^M\right] &\lesssim \left(\frac{1}{\sqrt{n}}\right)^M E_{\theta^*}\left[\sup_{\alpha}\left|\sum_{i=1}^n R(1,X_{i-1})(X_i-X_{i-1}-h_n a_{i-1}(\alpha))\right|^M\right] \\
&\quad + \left(\frac{1}{\sqrt{n}}\right)^M E_{\theta^*}\left[\left|\sup_{\alpha}\frac{1}{n}\sum_{i=1}^n R(1,X_{i-1})\right|^M\right] \\
&\lesssim \left(\frac{1}{\sqrt{n}}\right)^M O(1) \to 0,
\end{aligned}
$$

which completes the proof of (34).

Proof of (35). We have

$$
\left|\frac{1}{n}\partial_\beta^2 U_{\epsilon,n}^{(4)}(\hat\alpha_{\epsilon,n},\beta^*+t(\hat\beta_{\epsilon,n}-\beta^*)) - \frac{1}{n}\partial_\beta^2 U_{\epsilon,n}^{(4)}(\alpha^*,\beta^*)\right| \le \frac{1}{n}\sup_\theta\left|\partial_\theta\partial_\beta^2 U_{\epsilon,n}^{(4)}(\theta)\right|\cdot\left|\begin{pmatrix}\hat\alpha_{\epsilon,n}-\alpha^* \\ t(\hat\beta_{\epsilon,n}-\beta^*)\end{pmatrix}\right|.
$$

It follows that

$$
E_{\theta^*}\left[\left(\frac{1}{n}\sup_\theta\left|\partial_\theta\partial_\beta^2 U_{\epsilon,n}^{(4)}(\theta)\right|\right)^M\right] < \infty
$$

and

$$
\hat\alpha_{\epsilon,n}-\alpha^* \xrightarrow{p} 0,\ \hat\beta_{\epsilon,n}-\beta^* \xrightarrow{p} 0,
$$

which completes the proof of (35).

Proof of (36). We have

$$
\begin{aligned}
& -\frac{1}{\sqrt{n}}\partial_\beta U_{\epsilon,n}^{(4)}(\theta^*)[u_2] \\
= & -\frac{1}{\sqrt{n}}\left(-\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^n \partial_\beta B_{i-1}^{-1}(\beta^*)[u_2]\left[(X_i-X_{i-1}-h_n a_{i-1}(\alpha^*))^{\otimes 2}-\epsilon^2 h_n B_{i-1}(\beta^*)\right] =: \sum_{i=1}^n \xi_i.
\end{aligned}
$$

By Lemma 1–(ii),

$$
\sum_{i=1}^n E_{\theta^*}\left[\xi_i|\mathcal{G}_{i-1}\right] = \sum_{i=1}^n \partial_\beta B_{i-1}^{-1}(\beta^*)[u_2]\left[R\left(\epsilon^2 h_n^2, X_{i-1}\right)\right]\frac{1}{2\sqrt{n}}(\epsilon^2 h_n)^{-1} \lesssim \frac{1}{n}\sum_{i=1}^n R\left(\frac{1}{\sqrt{n}}, X_{i-1}\right) \xrightarrow{p} 0.
$$

Next,

$$
-\epsilon\partial_\alpha U_{\epsilon,n}^{(3)}(\alpha^*,\beta^*)[u_1] = \epsilon^{-1}\sum_{i=1}^n B_{i-1}^{-1}(\beta^*)\left[\partial_\alpha a_{i-1}(\alpha^*)[u_1], X_i-X_{i-1}-h_n a_{i-1}(\alpha^*)\right] =: \sum_{i=1}^n \eta_i.
$$

By Lemma 1–(i),

$$
\sum_{i=1}^n E_{\theta^*}\left[\eta_i|\mathcal{G}_{i-1}\right] = \sum_{i=1}^n \epsilon^{-1}R\left(h_n^2, X_{i-1}\right) = \frac{1}{\epsilon n}\frac{1}{n}\sum_{i=1}^n R\left(1, X_{i-1}\right) \xrightarrow{p} 0.
$$

By Lemma 1–(ii) and (iv),

$$
\begin{aligned}
& \sum_{i=1}^{n} E_{\theta^*}\left[\xi_i^{(j)}\xi_i^{(k)}|\mathcal{G}_{i-1}\right] \\
= \ & \frac{1}{4n}(\epsilon^4 h_n^2)^{-1}\sum_{i=1}^{n}\Big[\epsilon^4 h_n^2\Big\{2\mathrm{tr}(\partial_{\beta_j}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*)\partial_{\beta_k}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*)) \\
& + \mathrm{tr}(\partial_{\beta_j}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*))\mathrm{tr}(\partial_{\beta_k}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*))\Big\} + R\left(\epsilon^4 h_n^3, X_{i-1}\right) \\
& - (\epsilon^2 h_n)^2\mathrm{tr}(\partial_{\beta_j}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*))\mathrm{tr}(\partial_{\beta_k}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*)) + R\left(\epsilon^2 h_n^2, X_{i-1}\right)\epsilon^2 h_n \\
& - (\epsilon^2 h_n)^2\mathrm{tr}(\partial_{\beta_k}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*))\mathrm{tr}(\partial_{\beta_j}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*)) + R\left(\epsilon^2 h_n^2, X_{i-1}\right)\epsilon^2 h_n \\
& + (\epsilon^2 h_n)^2\mathrm{tr}(\partial_{\beta_j}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*))\mathrm{tr}(\partial_{\beta_k}B_{i-1}^{-1}(\beta^*)B_{i-1}(\beta^*))\Big] \\
\xrightarrow{p} \ & I_b(\beta^*).
\end{aligned}
$$

By Lemma 1–(i) and (iii),

$$
\begin{aligned}
\sum_{i=1}^{n} E_{\theta^*}\left[\xi_i^{(j)}\eta_i^{(k)}|\mathcal{G}_{i-1}\right] &= \frac{1}{\epsilon\sqrt{n}}\left(\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n}E_{\theta^*}\left[R(1,X_{i-1})(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))^{\otimes 3}|\mathcal{G}_{i-1}\right] \\
&\quad + \frac{1}{\epsilon\sqrt{n}}\left(\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n}(\epsilon^2 h_n)E_{\theta^*}\left[R(1,X_{i-1})(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))|\mathcal{G}_{i-1}\right] \\
&\lesssim \frac{1}{\epsilon\sqrt{n}}\left(\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n}R\left(\epsilon^4 h_n^2, X_{i-1}\right) + \frac{1}{\epsilon\sqrt{n}}\left(\frac{1}{2}\right)(\epsilon^2 h_n)^{-1}\sum_{i=1}^{n}R\left(h_n^2, X_{i-1}\right)(\epsilon^2 h_n) \\
&\lesssim \frac{1}{\sqrt{n}}\frac{\epsilon}{n}\sum_{i=1}^{n}R\left(1, X_{i-1}\right) + \frac{1}{\epsilon\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}R\left(\frac{1}{n}, X_{i-1}\right) \\
&\xrightarrow{p} 0.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\sum_{i=1}^{n} E_{\theta^*}\left[\xi_i^{(j)}|\mathcal{G}_{i-1}\right]E_{\theta^*}\left[\eta_i^{(k)}|\mathcal{G}_{i-1}\right] &= \sum_{i=1}^{n}R\left(\frac{1}{n\sqrt{n}}, X_{i-1}\right)R\left(\frac{1}{\epsilon n^2}, X_{i-1}\right) \xrightarrow{p} 0, \\
\sum_{i=1}^{n} E_{\theta^*}\left[\xi_i^{(j)}|\mathcal{G}_{i-1}\right]E_{\theta^*}\left[\xi_i^{(k)}|\mathcal{G}_{i-1}\right] &= \sum_{i=1}^{n}R\left(\frac{1}{n^3}, X_{i-1}\right) \xrightarrow{p} 0, \\
\sum_{i=1}^{n} E_{\theta^*}\left[\eta_i^{(j)}|\mathcal{G}_{i-1}\right]E_{\theta^*}\left[\eta_i^{(k)}|\mathcal{G}_{i-1}\right] &= \sum_{i=1}^{n}R\left(\frac{1}{\epsilon^2 n^4}, X_{i-1}\right) \xrightarrow{p} 0.
\end{aligned}
$$

We obtain that

$$
\begin{aligned}
\sum_{i=1}^{n} E_{\theta^*}\left[\eta_i^{(j)}\eta_i^{(k)}|\mathcal{G}_{i-1}\right] &= \epsilon^{-2}\sum_{i=1}^{n}E_{\theta^*}\Big[\partial_{\alpha_j}a_{i-1}B_{i-1}^{-1}(\beta^*)(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*)) \\
&\quad \times \partial_{\alpha_k}a_{i-1}B_{i-1}^{-1}(\beta^*)(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))|\mathcal{G}_{i-1}\Big] \\
&= h_n\sum_{i=1}^{n}\left\{B_{i-1}^{-1}(\beta^*)[\partial_{\alpha_j}a_{i-1}(\alpha^*), \partial_{\alpha_k}a_{i-1}(\alpha^*)] + R\left(h_n, X_{i-1}\right)\right\} \\
&\xrightarrow{p} I_a(\theta^*).
\end{aligned}
$$

Moreover,

$$\begin{aligned}\sum_{i=1}^{n} E_{\theta^*}\left[|\xi_i|^4|\mathcal{G}_{i-1}\right] &\lesssim \frac{1}{n^2}(\epsilon^2 h_n)^{-4}\sum_{i=1}^{n}\left\{E_{\theta^*}\left[|f_{i-1}(\theta)(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))|^8|\mathcal{G}_{i-1}\right] + (\epsilon^2 h_n)^4\right\} \\ &\lesssim \frac{1}{n^2}(\epsilon^2 h_n)^{-4}\sum_{i=1}^{n}\left\{R\left(h_n^8, X_{i-1}\right) + R\left(\epsilon^8 h_n^4, X_{i-1}\right) + (\epsilon^2 h_n)^4\right\} \\ &\lesssim \frac{1}{n^2}(\epsilon^2 h_n)^{-4}\sum_{i=1}^{n} R\left((\epsilon^2 h_n)^4\right) \xrightarrow{p} 0\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{n} E_{\theta^*}\left[|\eta_i|^4|\mathcal{G}_{i-1}\right] &\lesssim \epsilon^{-4}\sum_{i=1}^{n} E_{\theta^*}\left[|f_{i-1}(\theta)(X_i - X_{i-1} - h_n a_{i-1}(\alpha^*))|^4|\mathcal{G}_{i-1}\right] \\ &\lesssim \epsilon^{-4}\sum_{i=1}^{n} R\left(\epsilon^4 h_n^2\right) \xrightarrow{p} 0,\end{aligned}$$

which completes the proof of (36).

Proof of (37). By (22) and (29), we can show (37).

Therefore, $\left(\epsilon^{-1}(\hat{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\hat{\beta}_{\epsilon,n} - \beta^*)\right) \xrightarrow{d} (\zeta_1, \zeta_2)$ as $(\epsilon\sqrt{n})^{-1} = O(1)$. This together with Theorem 1 completes the proof of Theorem 2.

References

Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev spaces.* Second edition. Pure and Applied Mathematics (Amsterdam), 140. Elsevier/Academic Press, Amsterdam.

Azencott, R. (1982). Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynmann. *Séminaire de Probabilités XVI; Supplément: Géométrie Différentielle Stochastique. Lecture Notes In Math.* **921**, 237-285 Springer Verlag, Berlin.

Brouste, A., Fukasawa, M., Hino, H., Iacus, S., Kamatani, K., Koike, Y., Masuda, H., Nomura, R., Shimuzu, Y., Uchida, M., and Yoshida, N. (2014). The YUIMA Project : a Computational Framework for Simulation and Inference of Stochastic Differential Equations, *Journal of Statistical Software*, Vol. 57, Issue 4, Apr 2014.

Freidlin, M. I. and Wentzell, A. D. (1998). *Random perturbations of dynamical systems.* 2nd. ed. Springer-Verlag, New York.

Fuchs, C. (2013). *Inference for Diffusion Processes.* Springer-Verlag Berlin Heidelberg.

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics* **21**, 99–116.

Gloter A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, **119**, 679-699.

Guy, R., Laredo, C. and Vergu, E. (2014). Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. *Stochastic Processes and their Applications*, **124**, 51–80.

Guy, R., Laredo, C. and Vergu, E. (2015). Approximation of epidemic models by diffusion processes and their statistical inference. *Journal of Mathematical Biology*, **70**, 621–646.

Iacus, S. (2000). Semiparametric estimation of the state of a dynamical system with small noise. *Statistical Inference for Stochastic Processes*, **3**, 277–288.

Iacus S. and Kutoyants Yu. A. (2001). Semiparametric hypotheses testing for dynamical systems with small noise . *Mathematical Methods of Statistics*, **10**, 105–120.

Ibragimov, I. A. and Has 'minskii, R. Z. (1972a). The asymptotic behavior of certain statistical estimates in the smooth case. I. Investigation of the likelihood ratio. (Russian) Teorija Verojatnostei i ee Primenenija, **17**, 469–486.

Ibragimov, I. A. and Has 'minskii, R. Z. (1972b). Asymptotic behavior of certain statistical estimates. II. Limit theorems for a posteriori density and for Bayesian estimates. (Russian) Teorija Verojatnostei i ee Primenenija, **18**, 78–93.

Ibragimov, I. A. and Has 'minskii, R. Z. (1981). *Statistical estimation: Asymptotic theory.* New York, Springer.

Kaino, Y., Uchida, M. and Yoshida, Y. (2017). Hybrid type estimators for a diffusion process with ergodicity from reduced data. Preprint.
(`http://www.sigmath.es.osaka-u.ac.jp/~uchida/Paper/ergodic_reduced_KUY.pdf`)

Kamatani, K. (2014). Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. arXiv:1412.6231.

Kamatani, K. and Uchida, M. (2015). Hybrid multi-step estimators for ergodic diffusion processes from discrete observations. *Statistical Inference for Stochastic Processes,* **18**, 177-204.

Kamatani, K., Nogita, A. and Uchida, M. (2016). Hybrid multi-step estimation of the volatility for stochastic regression models. *Bulletin of Informatics and Cybernetics*, **48**, 19–35.

Kessler, M. (1995). Estimation des paramètres d'une diffusion par des contrastes corrigés. *Comptes Rendus de l'Académie des Sciences, Series I, Mathematics,* **320**, 359–362.

Kunitomo, N. and Takahashi, A. (2001). The asymptotic expansion approach to the valuation of interest rate contingent claims. *Mathematical Finance* **11**, 117-151.

Kutoyants, Yu. A. (1984). Parameter estimation for stochastic processes. Prakasa Rao, B.L.S. (ed.) Heldermann, Berlin.

Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise.* Kluwer, Dordrecht.

Kutoyants, Yu. A. (2017). On the multi-step MLE-process for ergodic diffusion. *Stochastic Processes and their Applications,* **127**, 2243–2261.

Laredo, C. F. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *Annals of Statistics,* **18**, 1158–1171.

Long, H., Shimizu, Y. and Sun, W. (2013). Least squares estimators for discretely observed stochastic processes driven by small Lévy noises. *Journal of Multivariate Analysis*, **116**, 422–439.

Long, H., Ma, C. and Shimizu, Y. (2016). Least squares estimators for stochastic differential equations driven by small Lévy noises. To appear in Stochastic Processes and their Applications.

Ma, C. and Yang, X. (2014). Small noise fluctuations of the CIR model driven by α-stable noises. *Statistics & Probability Letters*, **94**, 1–11.

Nomura, R. and Uchida, M. (2016). Adaptive Bayes estimators and hybrid estimators for small diffusion processes based on sampled data. *Journal of the Japan Statistical Society*, **46**, 129–154.

Sørensen, M. (2000). Small dispersion asymptotics for diffusion martingale estimating functions. Preprint No. 2000-2, Department of Statistics and Operations Research, University of Copenhagen.
http://www.math.ku.dk/~michael/smalld.pdf

Sørensen, M. (2012). Estimating functions for diffusion-type processes. 1–107. In Kessler, M., Lindner, A. and Sørensen, M. (eds.): Statistical Methods for Stochastic Differential Equations, CRC Press, Chapmann and Hall.

Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051–1069.

Takahashi, A. and Yoshida, N. (2004). An asymptotic expansion scheme for optimal investment problems. *Statistical Inference for Stochastic Processes,* **7**, 153-188.

Uchida, M. (2003). Estimation for dynamical systems with small noise from discrete observations. *Journal of the Japan Statistical Society,* **33**, 157–167.

Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scandinavian Journal of Statistics,* **31**, 553-566.

Uchida, M. (2006). Martingale estimating functions based on eigenfunctions for discretely observed small diffusions. *Bulletin of Informatics and Cybernetics*, **38**, 1-13.

Uchida, M. (2008). Approximate martingale estimating functions for stochastic differential equations with small noises. *Stochastic Processes and their Applications*, **118**, 1706-1721.

Uchida, M. (2010). Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Annals of the Institute of Statistical Mathematics*, **62**, 161–187.

Uchida, M. and Yoshida, N. (2001). Information criteria in model selection for mixing processes. *Statistical Inference for Stochastic Processes*, **4**, 73–98.

Uchida, M. and Yoshida, N. (2004a). Information criteria for small diffusions via the theory of Malliavin-Watanabe. *Statistical Inference for Stochastic Processes*, **7**, 35-67.

Uchida, M. and Yoshida, N. (2004b). Asymptotic expansion for small diffusions applied to option pricing. *Statistical Inference for Stochastic Processes*, **7**, 189-223.

Uchida, M. and Yoshida, N. (2006). Asymptotic expansion and information criteria. Dedicated to Professor Minoru Siotani on his 80th birthday. *SUT Journal of Mathematics*, **42**, 31–58.

Uchida, M. and Yoshida, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Processes and their Applications*, **122**, 2885–2924.

Uchida, M. and Yoshida, N. (2014). Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. *Statistical Inference for Stochastic Processes*, **17**, 181-219.

Yang, X. (2017). Maximum likelihood type estimation for discretely observed CIR model with small α-stable noises. *Statistics & Probability Letters*, **120**, 18–27.

Yoshida, N. (1992a). Asymptotic expansion of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. *Probability Theory and Related Fields*, **92**, 275-311.

Yoshida, N. (1992b). Asymptotic expansion for statistics related to small diffusions. *Journal of the Japan Statistical Society*, **22**, 139-159.

Yoshida, N. (1992c). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, **41**, 220–242.

Yoshida, N. (2003). Conditional expansions and their applications. *Stochastic Processes and their Applications*, **107**, 53-81.

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