Glasnik Matematički

SERIJA III

www.math.hr/glasnik

Miljenko Huzak, Snježana Lubura Strunjak and Andreja Vlahek Štrok

An approximate maximum likelihood estimator of drift parameters in a multidimensional diffusion model

AN APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATOR OF DRIFT PARAMETERS IN A MULTIDIMENSIONAL DIFFUSION MODEL

Miljenko Huzak, Snježana Lubura Strunjak and Andreja Vlahek Štrok

University of Zagreb, Croatia

ABSTRACT. For a fixed T and $k \geq 2$, a k-dimensional vector stochastic differential equation $dX_t = \mu(X_t, \theta)\, dt + \nu(X_t)\, dW_t$, is studied over a time interval $[0, T]$. Vector of drift parameters θ is unknown. The dependence in θ is in general nonlinear. We prove that the difference between approximate maximum likelihood estimator of the drift parameter $\bar{\theta}_n \equiv \bar{\theta}_{n,T}$ obtained from discrete observations $(X_{i\Delta_n}, 0 \leq i \leq n)$ and maximum likelihood estimator $\hat{\theta} \equiv \hat{\theta}_T$ obtained from continuous observations $(X_t, 0 \leq t \leq T)$, when $\Delta_n = T/n$ tends to zero, converges stably in law to the mixed normal random vector with covariance matrix that depends on $\hat{\theta}$ and on path $(X_t, 0 \leq t \leq T)$. The uniform ellipticity of diffusion matrix $S(x) = \nu(x)\nu(x)^T$ emerges as the main assumption on the diffusion coefficient function.

1. INTRODUCTION

Let $X = (X_t, t \geq 0)$ be a multidimensional diffusion with values in an open and convex state space $E \subseteq \mathbb{R}^k$ which satisfies Itô stochastic differential equation (SDE)

$$(1.1) \qquad dX_t = \mu(X_t, \theta)\, dt + \nu(X_t)\, dW_t, \quad X_0 = x_0, \quad t \geq 0,$$

where $W = (W_t, t \geq 0)$ is a k-dimensional Brownian motion and $x_0 \in E$ is a known nonrandom initial state of X. Componentwise, it is a system of differential equations of the following form

$$X_t^i = x_0^i + \int_0^t \mu_i(X_s, \theta)\, ds + \sum_{j=1}^k \int_0^t \nu_{ij}(X_s)\, dW_s^j, \quad \text{for } i = 1, 2, \ldots, k.$$

For $d = 1$, let $\theta \in \Theta \subseteq \mathbb{R}^d$ be a drift parameter, or a vector of drift parameters if $d \geq 2$. Let us denote by θ_0 its true value, and $\mathbb{P} \equiv \mathbb{P}_{\theta_0}$. Function $\mu(x, \theta)$ is called drift function and $S(x) = \nu(x)\nu(x)^T$ diffusion matrix. Diffusion coefficient function $\nu(x)$ may contain parameters that we consider known.

Let $T > 0$ be fixed. For a fixed $n \in \mathbb{N}$, let $0 =: t_0 < t_1 < \cdots < t_n := T$ be such a subdivision of the time interval $[0, T]$ that for $i = 1, \ldots, n$, $t_i = i\Delta_n$ and $\Delta_n = T/n$. For a given discrete observation $(X_{t_i}, 0 \leq i \leq n)$ of X over time interval $[0, T]$, our goal is to estimate the vector of parameters θ belonging to Θ. We assume that Θ is an open, relatively compact and convex set in $\mathbb{R}^d$. In this paper, we use the Euler approximation of Riemann and Itô integrals in (1.1) to obtain the set of difference equations whose solution is an approximation of the diffusion X. A similar estimation technique has been used in the case of a one-dimensional diffusion ([17, 26]). In this one-dimensional case it has been proved the existence of a sequence of so-called approximate maximum likelihood estimator (AMLE) which converges in probability to the maximum likelihood estimator with respect to a continuous observation (MLE) when n goes to infinity [16]. This result was obtained under smoothness assumptions of log-likelihood functions of the initial process and the approximation process, and under conditions that these log-likelihood functions as well as their first and second derivatives were in some sense close. The same result can also be stated in a multidimensional framework.

In this paper we consider multidimensional diffusion ($k \geq 2$). We aim to prove that the difference between AMLE based on a discrete observation $\overline{\theta}_n$ and MLE based on continuous observations $\hat{\theta}$ over $[0, T]$ scaled by $\sqrt{\Delta_n}$ is asymptotically mixed normal when Δ_n tends to zero. The covariance matrix of the resulting random vector depends on the MLE and the path $(X_t, t \in [0, T])$.

The same result has been proved in [26] for one-dimensional diffusion. In the multidimensional domain, the idea of the proof is similar, but the techniques are different. Advanced knowledge of linear algebra simplifies the notation and analysis. The components of k-dimensional Brownian motion are independent by definition, which often comes into play in the calculations in the proofs. There is also a new regularity condition for the diffusion matrix. It is a so-called uniform ellipticity condition ([3, 22, 13, 14]). This condition actually means that the noise in the system is non-degenerate in any dimension and the process is somehow controlled ([10], [15]). Most of the main results are first proved for a compact state space E and then for an open $E \subseteq \mathbb{R}^k$.

There are also other approaches to obtain an approximate maximum likelihood estimator in a multidimensional setting ([2], [23], [32]). In these works, the initial expression is a log-likelihood function for a discretely observed process defined by a transition density function. Its maximum is called the maximum likelihood estimator and is usually not computable in practise. The

main task is to approximate the transition density function using certain expansions. In the work of Aït-Sahalia [2], a Hermite polynomial expansion is used. This is a generalized version of the approach used in the one-dimensional case ([1], [9]). In this way, the approximated log-likelihood function is obtained, whose point of maximum is called an approximate maximum likelihood estimator. Such an estimator behaves asymptotically in the same way as the maximum likelihood estimator and has good properties with respect to the real value of the parameters under certain conditions (Theorem 3 in [2]). Similar results are obtained in [23] and [32] using Taylor-like expansions of the transition density function. Our approach is slightly different, but it is equivalent to a simple approximation of the transition density. In fact, in such a way obtained criterion function can also be interpreted as a simple approximation of the log-likelihood function based on the continuously observed process. This simplicity makes analysis of local mixed normality of the corresponding AMLE somehow straightforward and clear. We believe that our approach can also be applied for analyzing local asymptotic behavior of more complex approximate maximum likelihood estimators such as those analyzed in e.g. [2], [32] and [23].

Notice that it can be shown that a recurrent ergodic diffusion such that assumptions (A1-2) and (A4) of Section 4 hold, and that is parameterized in a way that is regular enough and identifiability holds, also satisfies assumption (A6) for large enough $T > 0$. A diffusion is regular enough if (A3) and (A5) are satisfied, and there exists a positive definite Fisher information matrix. Then it can be shown that MLE is consistent and asymptotically efficient estimator when $T \to +\infty$ (see e.g. [5]). This implies (A6).

The paper is organized as follows. In the next section we introduce the notation and give necessary definitions and known auxiliary results. Section 3 describes the estimation method. The assumptions and main theorems are stated in Section 4. The proofs of the main results are in Section 5, and an example and simulations are presented in Section 6. The Appendix consists of proofs of the lemmas.

2. Preliminaries

For two matrices A and B of the same dimension $m \times n$, the *Hadamard product* $A \circ B$ is the matrix of the same dimension, with elements given by $(A \circ B)_{ij} = A_{ij}B_{ij}$ [12]. We denote by $\langle \cdot | \cdot \rangle$ the scalar product in Euclidean space $\mathbb{R}^k$ and by $\|\cdot\|_2$ the induced norm. We say that the Frobenius norm of a square matrix $A \in \mathbb{R}^k$, $\|A\|_F$, is consistent with the Euclidean norm $\|x\|_2$ if $\|Ax\|_2 \leq \|A\|_F \|x\|_2$. Let $f : E \times \Theta \to \mathbb{R}^k$ be a vector valued function. We denote by $D_{\mathbf{j}}^m f(x, \theta)$ a partial derivative of m-th order of $f(x, \theta)$ with respect

to the vector of parameters θ, ie.

$$D_{\mathbf{j}}^m f(x, \theta) := \left[\frac{\partial^m f_1}{\partial \theta_1^{j_1} \cdots \theta_d^{j_d}}(x, \theta), \cdots, \frac{\partial^m f_k}{\partial \theta_1^{j_1} \cdots \theta_d^{j_d}}(x, \theta) \right]^T$$

where $\mathbf{j} = [j_1, j_2, \ldots, j_d]^T$ and $m = j_1 + \cdots + j_d$. $\nabla_x f(x, \theta)$ denotes a k-dimensional matrix of partial derivatives of f with respect to the space vector x. If function f is such that $f : E \to \mathbb{R}^k$, ∇f is the Jacobian matrix of f, and for every component function f_i, $\nabla (\nabla f_i)$ denotes a matrix of second partial derivative of f_i. If function f is such that $f : \Theta \to \mathbb{R}$, then $Df(\theta)$ denotes a d-dimensional vector of partial derivatives with respect to θ and $D^2 f(\theta)$ a matrix of second partial derivatives with respect to θ. Briefly, $\partial_j f(\theta)$ denotes partial derivative $\partial f(\theta)/\partial \theta_j$ for $j = 1, \ldots, d$. Vector e_p is a unit vector that has one at p-th place and zeros elsewhere.

LEMMA 2.1. *For C^2-function $F : \mathbb{R}^k \to \mathbb{R}^k$ and random process $(X_t)_{0 \le t \le T}$ that satisfies (1.1), Itô formula on $[s_1, s_2]$ yields* [21]:

$$F(X_{s_2}) - F(X_{s_1}) =$$
$$= \int_{s_1}^{s_2} \left(\nabla F(X_u) \mu(X_u, \theta_0) + \frac{1}{2} \nabla_2 F(X_u) \right) du +$$
$$+ \int_{s_1}^{s_2} \nabla F(X_u) \nu(X_u) \, dW_u,$$

where $\nabla_2 F$ denotes vector $\left[\text{Tr} \left[S(\cdot) \nabla (\nabla F_1(\cdot)) \right], \ldots, \text{Tr} \left[S(\cdot) \nabla (\nabla F_k(\cdot)) \right] \right]^T$.

For a scalar function F the previous lemma has the following form.

LEMMA 2.2. *For C^2-function $F : \mathbb{R}^k \to \mathbb{R}$ and random process $(X_t)_{0 \le t \le T}$ that satisfies (1.1) Itô formula yields:*

$$F(X_{s_2}) - F(X_{s_1}) =$$
$$= \int_{s_1}^{s_2} \left(\langle \nabla F(X_u) \,|\, \mu(X_u, \theta_0) \rangle + \frac{1}{2} \text{Tr} \left[S(X_u) \nabla (\nabla F(X_u)) \right] \right) du$$
$$+ \int_{s_1}^{s_2} \langle \nabla F(X_u) \,|\, \nu(X_u) \, dW_u \rangle.$$

In the proofs of this paper we often use Itô isometry [27, Corollary 3.1.7], but for vector and matrix random processes. For this reason we formulate it here.

LEMMA 2.3. *If $M : [0, T] \times \Omega \to \mathbb{R}^{k \times k}$ is a matrix random process and $V : [0, T] \times \Omega \to \mathbb{R}^k$ is a vector random process, both adapted to the natural filtration of Brownian motion W, then the following holds.*

(i) $\mathbb{E} \left[\left\| \int_0^T M_t \, dW_t \right\|_2^2 \right] = \mathbb{E} \left[\int_0^T \|M_t\|_F^2 \, dt \right]$,

(ii) $\mathbb{E}\left[\left(\int_0^T V_t^T \, dW_t\right)^2\right] = \mathbb{E}\left[\int_0^T \|V_t\|_2^2 \, dt\right].$

We also use the generalized version of the mean value theorem given in the following lemma.

LEMMA 2.4. *Let $n, m \in \mathbb{N}$. Let $U \subseteq \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}^m$ be a continuously differentiable function. Let $x \in U$ and $h \in \mathbb{R}^n$ be vectors such that for all $t \in [0, 1]$ the line segment $x + th$ lies inside the set U. It follows that*

$$f(x + h) - f(x) = \left(\int_0^1 \nabla f(x + th) \, dt\right) \cdot h,$$

where the integral of a matrix is taken componentwise.

For fixed $T > 0$, let $(\Omega, \mathcal{F}_T, (\mathcal{F})_{0 \leq t \leq T}, \mathbb{P})$ be a given filtered probability space and $\left(\tilde{\Omega}, \tilde{\mathcal{F}}_T, (\tilde{\mathcal{F}})_{0 \leq t \leq T}, \tilde{\mathbb{P}}\right)$ be an *extension* of this space. The extension is called *very good* if all martingales on the initial space are also martingales on the extension [18]. Let D be a Polish space. In this paper D will be the Skorokhod space $D([0, T], \mathbb{R}^d)$ or d-dimensional Euclidean space $\mathbb{R}^d$. Let $(Z_n)_{n \in \mathbb{N}}$ be a sequence of random vectors with values in D, defined on the initial space $(\Omega, \mathcal{F}_T, (\mathcal{F})_{0 \leq t \leq T}, \mathbb{P})$, and let Z be a random vector also with values in D defined on the extension. We say that $(Z_n)_n$ *converges stably in law* to Z, $Z_n \overset{st}{\Rightarrow} Z$, if

$$\lim_{n \to +\infty} \mathbb{E}\left[Y f(Z_n)\right] = \tilde{\mathbb{E}}\left[Y f(Z)\right]$$

for all bounded continuous functions $f : D \to \mathbb{R}$ and all bounded random variable Y on $(\Omega, \mathcal{F}_T, (\mathcal{F})_{0 \leq t \leq T}, \mathbb{P})$ [19].

We say that an $\mathbb{R}^d$-valued random vector Y has *mixed normal distribution* with $\mathcal{F}_T$-measurable random covariance matrix $C = (C_{jl})$ and write $Y \sim MN(0, C)$ if

$$\mathbb{E}\left[e^{i\langle t \,|\, Y \rangle}|\mathcal{F}_T\right] = e^{-\frac{1}{2} \sum_{j,l=1,\ldots,d} t_j t_l C_{jl}}.$$

Moreover, if $Y \sim MN(0, C)$, then Y has the same distribution as $\sqrt{C}Z$ where $\sqrt{C}$ is a square symmetric root of C and Z is a standard normal random vector independent of $\mathcal{F}_T$.

Let us denote by $A_n^t = \max\{j; 0 \leq j \leq n \text{ and } t_j \leq t\}$ and $\mathcal{F}_{n,i} := \mathcal{F}_{t_i}, i = 0, 1, \ldots, n$.

For a $\mathbb{R}^d$-valued process $(Y_t)_{1 \leq t \leq T}$ we say that it is a *centered Gaussian process* if for every $0 \leq t_1 < t_2 < \cdots < t_l \leq T$ the random vector $(Y_{t_1}, \ldots, Y_{t_l}) \in \mathbb{R}^{dl}$ has dl-dimensional normal distribution, and $\mathbb{E}\left[Y_t\right] = \mathbf{0}_d, t \in [0, T]$ [27].

The decisive role in the future analysis is played by Theorem 3-2 in [18], which has been adapted in its notation for our paper.

THEOREM 2.5. *Let W be a k-dimensional Brownian motion on $[0, T]$ and χ_i^n be $\mathcal{F}_{n,i}$-measurable square-integrable random vectors in $\mathbb{R}^d$. Assume that $C = (C_{jl})$ is a continuous adapted process defined on $(\Omega, \mathcal{F}_T, \mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$, and $(C_{jl}(t))$ is a positive semidefinite symmetric $d \times d$ matrix for every $t \in [0, T]$. Assume also that:*

(i) $\sup_{0 \le t \le T} \left\| \sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^n | \mathcal{F}_{n,i-1} \right] \right\|_2 \xrightarrow{\mathbb{P}} 0$,

(ii) $\sum_{i=1}^{A_n^t} \left(\mathbb{E}\left[\chi_i^{n,j} \chi_i^{n,l} | \mathcal{F}_{n,i-1} \right] - \mathbb{E}\left[\chi_i^{n,j} | \mathcal{F}_{n,i-1} \right] \mathbb{E}\left[\chi_i^{n,l} | \mathcal{F}_{n,i-1} \right] \right) \xrightarrow{\mathbb{P}} C_{jl}(t)$, $\forall t \in [0, T], \quad j, l = 1, \dots, d$,

(iii) $\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^n \left(W_{t_i} - W_{t_{i-1}} \right)^T | \mathcal{F}_{n,i-1} \right] \xrightarrow{\mathbb{P}} \mathbf{0}_{d \times k}, \forall t \in [0, T]$,

(iv) $\sum_{i=1}^n \mathbb{E}\left[\|\chi_i^n\|_2^2 \mathbb{1}_{\|\chi_i^n\|_2 > \epsilon} | \mathcal{F}_{n,i-1} \right] \xrightarrow{\mathbb{P}} 0, \forall \epsilon > 0$,

(v) $\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^n \left(N_{t_i} - N_{t_{i-1}} \right) | \mathcal{F}_{n,i-1} \right] \xrightarrow{\mathbb{P}} \mathbf{0}_d, \forall t \in [0, T]$,

where N is a bounded $\mathcal{F}_t$-martingale orthogonal to W, i.e. $N_0 = 0$ and for all $m = 1, 2, \dots, k$, $[N, W^m]_t \equiv 0$.

Then there is a very good extension $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{0 \le t \le T}, \tilde{\mathbb{P}} \right)$ of the space $(\Omega, \mathcal{F}_T, \mathbb{F} = (\mathcal{F}_t)_{0 \le t \le T}, \mathbb{P})$ and a continuous process Y defined on that extension that is, conditionally on $\mathcal{F}_T$, centered Gaussian process with independent increments satisfying $\tilde{\mathbb{E}}\left[Y_t^j Y_t^l | \mathcal{F}_T \right] = C_{jl}(t)$, $t \in [0, T]$, $j, l = 1, \dots, d$, and such that

$$\sum_{i=1}^{A_n^t} \chi_i^n \stackrel{st}{\Rightarrow} Y \quad on \ D([0, T], \mathbb{R}^d).$$

Let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers and let $(Y_n)_n$ be a sequence of random vectors defined on some probability space. We say that $(Y_n)_n$ is of order $O_{\mathbb{P}}(\gamma_n)$, and write $Y_n = O_{\mathbb{P}}(\gamma_n)$, if the sequence $(Y_n / \gamma_n)_{n \in \mathbb{N}}$ is bounded in probability, i.e. if

$$\lim_{A \to +\infty} \overline{\lim_n} \mathbb{P}\{ \gamma_n^{-1} \|Y_n\|_2 > A \} = 0.$$

3. ESTIMATION METHOD

Using the Euler discretization scheme, for $l = 1, 2, \dots, k$ we obtain the system of stochastic difference equations [21]:

$$Z_{t_i}^l - Z_{t_{i-1}}^l = \mu_l(Z_{t_{i-1}}, \theta)(t_i - t_{i-1}) + \sum_{j=1}^k \nu_{lj}(Z_{t_{i-1}}) \left(W_{t_i}^j - W_{t_{i-1}}^j \right).$$

If we denote by $\Delta W_i^j := W_{t_i}^j - W_{t_{i-1}}^j \sim N(0, \Delta_n)$, then $\Delta W_{i_1}^{j_1}$ and $\Delta W_{i_2}^{j_2}$ are independent for every pair of indices j_1 and j_2 such that $j_1 \ne j_2$, and for every i_1 and i_2 such that $i_1 \ne i_2$ in case when $j_1 = j_2$. If matrix S is a regular

matrix and Hermitian, then the log-likelihood function of the process Z is

$$\ln L_n(\theta) := -\ln \left(\prod_{i=1}^{n} \sqrt{2\pi\Delta_n}^{k} \sqrt{\det(S(Z_{t_{i-1}}))} \right) -$$

$$- \frac{1}{2\Delta_n} \sum_{i=1}^{n} \left\langle S^{-1}(Z_{t_{i-1}}) \left(Z_{t_i} - Z_{t_{i-1}} - \Delta_n \mu(Z_{t_{i-1}}, \theta) \right) \mid \right.$$

$$\left. Z_{t_i} - Z_{t_{i-1}} - \Delta_n \mu(Z_{t_{i-1}}, \theta) \right\rangle.$$

If we substitute Z_i with $X_i \equiv X_{t_i}$, for every $i = 0, 1, \ldots, n$, use the fact that matrix S is Hermitian, and omit the constant that does not depend on θ, we get an approximate log-likelihood function of the form

$$(3.2) \qquad \ell_n(\theta) := \sum_{i=1}^{n} \left\langle S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) \mid X_i - X_{i-1} \right\rangle -$$

$$- \frac{1}{2}\Delta_n \sum_{i=1}^{n} \left\langle S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) \mid \mu(X_{i-1}, \theta) \right\rangle.$$

If there is a point of global maximum $\overline{\theta}_n$ of function ℓ_n, we call it an approximate maximum likelihood estimator (AMLE) of parameter θ.

4. Assumptions and main theorems

Let $O^\Theta \subseteq \mathbb{R}^d$ be an open set such that $Cl(\Theta) \subseteq O^\Theta$. We assume the following hypothesis:

A1 (uniform ellipticity on compact sets) For every compact set $K \subseteq E$ there exists a real constant $\lambda_K > 0$ such that for all $x \in K$ and $\xi \in \mathbb{R}^k$

$$\langle S(x)\xi \mid \xi \rangle \geq \lambda_K \|\xi\|_2^2.$$

In particular, matrix $S(x)$ is positive definite (and so regular) for every $x \in E$.

A2 For all $p, r = 1, 2, \ldots, k$ and all $(x, \theta) \in E \times O^\Theta$ there exist partial derivatives $\frac{\partial}{\partial x_p}\mu(x, \theta)$ and $\frac{\partial^2}{\partial x_p \partial x_r}\mu(x, \theta)$. Moreover, functions $(x, \theta) \mapsto \mu(x, \theta)$, $(x, \theta) \mapsto \frac{\partial}{\partial x_p}\mu(x, \theta)$, $(x, \theta) \mapsto \frac{\partial^2}{\partial x_p \partial x_r}\mu(x, \theta)$ are continuous functions on $E \times Cl(\Theta)$, for all $p, r = 1, 2, \ldots, k$.

A3 For all $i, j = 1, 2, \ldots, d$, all $p, r = 1, 2, \ldots, k$ and all $(x, \theta) \in E \times O^\Theta$ there exist partial derivatives $\frac{\partial}{\partial \theta_i}\mu(x, \theta)$, $\frac{\partial^2}{\partial x_p \partial \theta_i}\mu(x, \theta)$, $\frac{\partial^3}{\partial x_r \partial x_p \partial \theta_i}\mu(x, \theta)$, and $\frac{\partial^2}{\partial \theta_i \partial \theta_j}\mu(x, \theta)$, $\frac{\partial^3}{\partial x_p \partial \theta_i \partial \theta_j}\mu(x, \theta)$. Moreover, functions $(x, \theta) \mapsto \frac{\partial}{\partial \theta_i}\mu(x, \theta)$, $(x, \theta) \mapsto \frac{\partial^2}{\partial x_p \partial \theta_i}\mu(x, \theta)$, $(x, \theta) \mapsto \frac{\partial^3}{\partial x_r \partial x_p \partial \theta_i}\mu(x, \theta)$, and $(x, \theta) \mapsto \frac{\partial^2}{\partial \theta_i \partial \theta_j}\mu(x, \theta)$, $(x, \theta) \mapsto \frac{\partial^3}{\partial x_p \partial \theta_i \partial \theta_j}\mu(x, \theta)$ are continuous on $E \times Cl(\Theta)$, for all $p, r = 1, 2, \ldots, k$ and all $i, j = 1, 2, \ldots, d$.

A4 For all $p, r = 1, 2, \ldots, k$ and all $x \in E$ there exist partial derivatives $\frac{\partial}{\partial x_p}\nu(x)$, and $\frac{\partial^2}{\partial x_r \partial x_p}\nu(x)$. Moreover, functions $x \mapsto \nu(x)$, $x \mapsto \frac{\partial}{\partial x_p}\nu(x)$, and $x \mapsto \frac{\partial^2}{\partial x_r \partial x_p}\nu(x)$ are continuous on E, for all $p, r = 1, 2, \ldots, k$.

Assumptions (A1-2) and (A4) imply that there exists a strong and by-path-unique solution to the SDE (1.1) on time-interval $[0, T]$ (see [20]), and (A2-4) enable applications of the Itô formula on the drift and diffusion coefficient functions. Notice that (A2-3) implies that all other partial derivatives of orders less or equal 3 exists and among them all partial derivatives with respect to the same set of variable directions are mutually equal. For example, it turns out that

$$\frac{\partial^3}{\partial\theta_i \partial x_p \partial\theta_j}\mu(x, \theta) = \frac{\partial^3}{\partial\theta_i \partial\theta_j \partial x_p}\mu(x, \theta) = \frac{\partial^3}{\partial x_p \partial\theta_i \partial\theta_j}\mu(x, \theta)$$

for all $(x, \theta) \in E \times Cl(\Theta)$ and all $p = 1, \ldots, k$ and all $i, j = 1, 2, \ldots, d$, and all these functions are continuous on $E \times Cl(\Theta)$.

Since matrix $S(x)$ is regular by (A1), the log-likelihood function of (1.1) is given by [29, Theorem 6.4.3]

$$\ell(\theta) := \int_0^T \left\langle S^{-1}(X_s)\mu(X_s, \theta) \mid dX_s \right\rangle -$$
$$- \frac{1}{2}\int_0^T \left\langle \mu(X_s, \theta) \mid S^{-1}(X_s)\mu(X_s, \theta)\right\rangle ds.$$

By the next assumption we will be able to prove that log-likelihood function ℓ is two-times continuously differentiable on Θ, i.e. $\ell \in C^2(\Theta)$ (see Theorem 4.1 below).

A5 For all $(x, \theta) \in E \times O^\Theta$ and $m \leq d + 3$ there exist $D_{\mathbf{j}}^m\mu(x, \theta)$ and $\frac{\partial}{\partial x_p}D_{\mathbf{j}}^m\mu(x, \theta)$, for all $p = 1, \ldots, k$ and all $\mathbf{j} = [j_1, j_2, \ldots, j_d]^T$ such that $m = j_1 + \cdots + j_d$. Moreover, these functions are continuous on $E \times Cl(\Theta)$.

THEOREM 4.1. *Assume that (A1) and (A4-5) hold. Then the log-likelihood function $\ell \in C^2(\Theta)$, and*

$$\frac{\partial}{\partial \theta_i} \ell(\theta) = \int_0^T \left\langle S^{-1}(X_s) \frac{\partial}{\partial \theta_i} \mu(X_s, \theta) \,|\, dX_s \right\rangle -$$
$$- \int_0^T \left\langle \frac{\partial}{\partial \theta_i} \mu(X_s, \theta) \,|\, S^{-1}(X_s) \mu(X_s, \theta) \right\rangle ds$$
$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta) = \int_0^T \left\langle S^{-1}(X_s) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mu(X_s, \theta) \,|\, dX_s \right\rangle -$$
$$- \int_0^T \left\langle \frac{\partial^2}{\partial \theta_i \partial \theta_j} \mu(X_s, \theta) \,|\, S^{-1}(X_s) \mu(X_s, \theta) \right\rangle ds -$$
$$- \int_0^T \left\langle \frac{\partial}{\partial \theta_i} \mu(X_s, \theta) \,|\, S^{-1}(X_s) \frac{\partial}{\partial \theta_j} \mu(X_s, \theta) \right\rangle ds$$

for all $i, j = 1, 2, \ldots, d$.

Now, we can formulate the last assumption:

A6 For all $\omega \in \Omega$, the log-likelihood function $\ell(\theta) \equiv \ell_T(\theta, \omega)$ based on continuous observations of X over $[0, T]$, as a function $\ell : \Theta \to \mathbb{R}$, has a unique point of global maximum $\hat{\theta} \equiv \hat{\theta}_T(\omega)$, and $D^2\ell(\hat{\theta})$ is a negatively definite matrix.

Since Θ is a convex space, using the mean value theorem, we get that for all $\theta \in \Theta$

$$D\ell_n(\theta) = D\ell_n(\hat{\theta}) + \int_0^1 D^2\ell_n(\hat{\theta} + (\theta - \hat{\theta})z) \, dz \cdot (\theta - \hat{\theta}).$$

In particular, on event $\{D\ell_n(\overline{\theta}_n) = 0\}$, we have

$$\frac{1}{\sqrt{\Delta_n}} \int_0^1 D^2\ell_n(\hat{\theta} + (\overline{\theta}_n - \hat{\theta})z) \, dz \cdot (\overline{\theta}_n - \hat{\theta}) =$$
(4.3)
$$= \frac{1}{\sqrt{\Delta_n}} \left(D\ell(\hat{\theta}) - D\ell_n(\hat{\theta}) \right)$$

since $D\ell(\hat{\theta}) = 0$ by (A6). We are interested in the behavior of the difference between the derivative of the log-likelihood function for continuous observations and its discretized version.

Let us fix $\theta \in \Theta$. For $1 \le j \le d$, we have

(4.4)
$$\frac{1}{\sqrt{\Delta_n}} \left(D\ell(\theta) - D\ell_n(\theta)\right)_j = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle S^{-1}(X_s)\partial_j\mu(X_s,\theta) - \right.$$

(4.5)
$$\left. - S^{-1}(X_{i-1})\partial_j\mu(X_{i-1},\theta) \mid \mu(X_s,\theta_0) \right\rangle ds -$$

$$- \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left(\left\langle \partial_j\mu(X_s,\theta) \mid S^{-1}(X_s)\mu(X_s,\theta) \right\rangle - \right.$$

(4.6)
$$\left. - \left\langle \partial_j\mu(X_{i-1},\theta) \mid S^{-1}(X_{i-1})\mu(X_{i-1},\theta) \right\rangle \right) ds +$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle S^{-1}(X_s)\partial_j\mu(X_s,\theta) - \right.$$

(4.7)
$$\left. - S^{-1}(X_{i-1})\partial_j\mu(X_{i-1},\theta) \mid \nu(X_t) \, dW_s \right\rangle.$$

We introduce the following notation:

(4.8)
$$g_j(x,\theta) := S^{-1}(x)\partial_j\mu(x,\theta), \quad (x,\theta) \in E \times \Theta.$$

First, the difference in (4.4) can be decomposed into two parts. One part is negligible in probability and the other part determines the resulting limit.

LEMMA 4.2. *Assume that (A1-5) hold and that E is a compact set. For arbitrary $\theta \in \Theta$ the standardized difference*

$$\frac{1}{\sqrt{\Delta_n}} \left(D\ell(\theta) - D\ell_n(\theta)\right)$$

is equal to

$$V_n(\theta) + \frac{1}{\sqrt{\Delta_n}} \begin{bmatrix} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla_x g_1(X_u,\theta)\nu(X_u) \, dW_u \mid \nu(X_s) \, dW_s \right\rangle \\ \vdots \\ \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla_x g_d(X_u,\theta)\nu(X_u) \, dW_u \mid \nu(X_s) \, dW_s \right\rangle \end{bmatrix},$$

where $V_n(\theta) \overset{\mathbb{P}}{\to} \mathbf{0}_d$ when $n \to +\infty$.

THEOREM 4.3. *Assume that (A1-5) hold. For arbitrary $\theta \in \Theta$ the standardized difference*

$$\frac{1}{\sqrt{\Delta_n}} \left(D\ell(\theta) - D\ell_n(\theta)\right) \overset{st}{\Rightarrow} Y(\theta)$$

where $Y(\theta) \sim MN(0, \Sigma(\theta))$ and for $j,l = 1, \ldots, d$,

$$\Sigma_{jl}(\theta) = \frac{1}{2} \int_0^T \sum_{p,r=1}^{k} S_{pr}(X_s) \left\langle e_r^T \nabla_x g_j(X_s,\theta)\nu(X_s) \mid e_p^T \nabla_x g_l(X_s,\theta)\nu(X_s) \right\rangle ds.$$

Compactness of $Cl(\Theta)$ implies that is possible to pass MLE $\hat{\theta}$ in the limiting equation in the preceding theorem.

THEOREM 4.4. *Assume that (A1-6) holds. Assume that E is an open convex set and Θ is a convex and relatively compact set. Then, we have*

$$\frac{1}{\sqrt{\Delta_n}}\left(D\ell(\hat{\theta}) - D\ell_n(\hat{\theta})\right) \overset{st}{\Rightarrow} Y(\hat{\theta})$$

where Y is the same as in Theorem 4.3.

To analyze the difference between the AMLE and MLE, we need to obtain the uniform bound on the difference between the derivative of the log-likelihood and the discretized log-likelihood. The following lemma serves this purpose.

LEMMA 4.5. *Assume that (A1-5) hold. Let Θ be a relatively compact set. Then for $r = 0, 1, 2$*

$$\sup_{\theta \in \Theta} \|D^r \ell_n(\theta) - D^r \ell(\theta)\|_2 = O_{\mathbb{P}}(\sqrt{\Delta_n}), \quad n \in \mathbb{N}.$$

Using Lemma 4.5 and assumption (A6), it can be proved (exactly as in [16]) that there exists a sequence of AMLEs. Namely, the following theorem holds.

THEOREM 4.6. *Assume that (A1-6) hold. Then there exists a sequence $(\bar{\theta}_n)_{n \in \mathbb{N}} \subseteq \Theta$ of $\mathcal{F}_T$-measurable random vectors such that*

(i) $\lim_{n \to +\infty} \mathbb{P}\left(D\ell_n(\bar{\theta}_n) = 0\right) = 1$,

(ii) $\bar{\theta}_n \overset{\mathbb{P}}{\to} \hat{\theta}$, $n \to +\infty$,

(iii) *if $(\tilde{\theta}_n)_{n \in \mathbb{N}}$ is another sequence of random vectors which satisfies (i) and (ii), then $\mathbb{P}(\tilde{\theta}_n = \bar{\theta}_n) = 1$,*

(iv) $\left(\frac{1}{\sqrt{\Delta_n}}\left\|\bar{\theta}_n - \hat{\theta}\right\|_2\right)_n$ *is bounded in probability.*

The core of our paper is the following theorem, which gives the limit of the standardized difference between the AMLE and MLE in terms of stable convergence. From now on we assume that the vector $\bar{\theta}_n$ is such that (i)-(iv) from Theorem 4.6 hold.

THEOREM 4.7. *Let Θ be a convex and relatively compact set. Assume (A1-6). Then*

$$\frac{1}{\sqrt{\Delta_n}}\left(\bar{\theta}_n - \hat{\theta}\right) \overset{st}{\Rightarrow} MN\left(0, \left(D^2\ell(\hat{\theta})\right)^{-1} \Sigma(\hat{\theta}) \left(D^2\ell(\hat{\theta})\right)^{-1}\right).$$

5. PROOFS

For proving Theorem 4.1 and Lemma 4.5 we need help of Fourier analysis. Let $U \subset O^\Theta$ be an open and bounded set such that the following holds:

$$U \subset Cl(U) \subset U' \subset Cl(U') \subset O^\Theta$$

for some d-dimensional rectangular $U' = \prod_{i=1}^{d}\langle a_i, b_i \rangle$. Since $O^\Theta \subset \mathbb{R}^d$ is an open and locally compact set, for every $\theta \in O^\Theta$ there exists such an open neighborhood U of θ in O^Θ. On the other hand, if $\mathcal{K}$ is a relatively compact open set in Θ, then it can be covered by a finite number of such open sets U_j (j from a finite set). If $f : E \times O^\Theta \to \mathbb{R}^k$ is a bounded function such that $f(x, \cdot) \in C^m(Cl(\Theta))$ for some m and all $x \in E$, then there exists a bounded function $\tilde{f} : E \times \mathbb{R}^d \to \mathbb{R}^k$ and a regular linear mapping $A : \mathbb{R}^d \to \mathbb{R}^d$ such that $\tilde{f}(x, \cdot) \in C^m(\mathbb{R}^d)$ for all $x \in E$ and $A(Cl(U')) = [-\pi, \pi]^d$, and

$$(\forall (x, \theta) \in E \times Cl(\mathcal{K}))\ \ f(x, \theta) = \tilde{f}(x, A\theta)$$

$$(\forall \theta \in \partial([-\pi, \pi]^d))\, (\forall x \in E)\ \ \tilde{f}(x, \theta) = 0$$

where $\partial \mathcal{S}$ denotes boundary of a set $\mathcal{S}$. Such a function $\tilde{f}$ can be constructed from f by use of an appropriate test-function (see [17]). For all of above reasons it is enough to prove the following two technical theorems for neighborhoods and relative compacts of the form $\mathcal{K}_0 := \langle -\pi, \pi \rangle^d$ and functions that satisfy assumptions (P1) and (P2) below.

Let $\mathbf{k} = [k_1, \ldots, k_d]^T \in \mathbb{Z}^d$, $\mathbf{j} = [j_1, \ldots, j_d]^T \in \mathbb{N}_0^d$ and $m = |\mathbf{j}| := j_1 + \cdots + j_d \leq d + 1$. For function $f : E \times \mathbb{R}^d \to \mathbb{R}^k$ we assume the following:

P1 For all $\mathbf{j} \in \mathbb{N}_0^d$ such that $|\mathbf{j}| = m \leq d + 1$, and all $(x, \theta) \in E \times \mathbb{R}^d$ there exist $\frac{\partial f}{\partial x_p}(x, \theta), D_{\mathbf{j}}^m f(x, \theta)$ and $\frac{\partial}{\partial x_p} D_{\mathbf{j}}^m f(x, \theta)$ for $p = 1, \ldots, k$. Additionally, f, $\frac{\partial f}{\partial x_p}(x, \theta)$, $D_{\mathbf{j}}^m f(x, \theta)$ and $\frac{\partial}{\partial x_p} D_{\mathbf{j}}^m f(x, \theta)$, $m \leq d + 1$, $p = 1, \ldots, k$, are continuous and uniformly bounded functions on $E \times \mathbb{R}^d$.

P2 For all $x \in E$ and all $\mathbf{j} \in \mathbb{N}_0^d$ such that $|\mathbf{j}| = m \leq d + 1$, $f(x, \cdot) \equiv \mathbf{0}_{k \times k}$ and $D_{\mathbf{j}}^m f(x, \cdot) \equiv \mathbf{0}_k$ on $\partial \mathcal{K}_0$.

Fourier coefficients of function f are defined (componentwise) as vectors

$$C_{\mathbf{k}}(x) := \frac{1}{(2\pi)^d} \int_{Cl(\mathcal{K}_0)} f(x, \theta) e^{-i\langle \mathbf{k}|\theta \rangle}\, d\theta,$$

$$C_{\mathbf{k}}^{(\mathbf{j})}(x) := \frac{1}{(2\pi)^d} \int_{Cl(\mathcal{K}_0)} D_{\mathbf{j}}^m f(x, \theta) e^{-i\langle \mathbf{k}|\theta \rangle}\, d\theta.$$

Let us denote $\mathbf{k}^{\mathbf{j}} := k_1^{j_1} \cdots k_d^{j_d}$. Under assumptions (P1) and (P2) using partial integration it is easy to see that for all $x \in E$, we have [30]

$$(5.9) \qquad\qquad C_{\mathbf{k}}^{(\mathbf{j})}(x) = i^m \mathbf{k}^{\mathbf{j}} C_{\mathbf{k}}(x).$$

REMARK 5.1. If assumption (P2) does not hold for function f, we can easily calculate the relation similar to (5.9), and it is of the form

$$F_{\mathbf{k}}(x) + C_{\mathbf{k}}^{(\mathbf{j})}(x) = i^m \mathbf{k}^{\mathbf{j}} C_{\mathbf{k}}(x),$$

where $F_{\mathbf{k}}(x)$ is a linear combination of functions $D_\theta^{m-1} f(x, \theta)$, for $1 \le m \le d+1$, on $\partial \mathcal{K}_0$. Since these functions are bounded on $E \times \mathbb{R}^d$ by (P1), expression $F_{\mathbf{k}}(x)$ can be bounded by a constant so it may be ignored in further considerations (see the proof of Lemma 5.3).

The following is a list of technical statements essential for proving the uniform (in θ) L^2-bounds of sum of certain Riemann and Itô integrals as well as absolute convergence of Fourier series

(5.10)
$$\sum_{\mathbf{k} \in \mathbb{Z}^d} C_{\mathbf{k}}(x) e^{i \langle \mathbf{k} | \theta \rangle}$$

for $(x, \theta) \in E \times Cl(\mathcal{K}_0)$.

LEMMA 5.2. Let functions $\mu(\cdot, \theta_0)$ and ν are bounded and continuous on E. There exists a constant $K > 0$ such that for $s_1, s_2, 0 \le s_1 \le s_2 \le T$

$$\mathbb{E}\left[\|X_{s_2} - X_{s_1}\|_2^2 \right] \le 2K \left((s_2 - s_1)^2 + (s_2 - s_1) \right).$$

LEMMA 5.3. Let $E \subseteq \mathbb{R}^k$ be an open and convex set and $x, y \in E$, and let $f : E \times \mathbb{R}^d \to \mathbb{R}^k$ be a function that satisfies (P1) and (P2). Then for every $\mathbf{k} \in \mathbb{Z}^d$ there exist constants $k_1, k_2 > 0$ such that

(5.11)
$$\|C_{\mathbf{k}}(x) - C_{\mathbf{k}}(y)\|_2 \le k_1 K_m(\mathbf{k}) \|x - y\|_2,$$

(5.12)
$$\|C_{\mathbf{k}}(x)\|_2 \le k_2 K_m(\mathbf{k}),$$

where $K_m(\mathbf{k}) = \left(\frac{d+1}{1 + |k_1| + \cdots + |k_d|} \right)^m$.

LEMMA 5.4. $\sum_{\mathbf{k} \in \mathbb{Z}^d} \frac{1}{(1 + |k_1| + \cdots + |k_d|)^{d+1}} < +\infty$.

If f satisfies (P1) and (P2), then

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \left\| C_{\mathbf{k}}(x) e^{i \langle \mathbf{k} | \theta \rangle} \right\|_2 \le \sum_{\mathbf{k} \in \mathbb{Z}^d} \|C_{\mathbf{k}}(x)\|_2 \le k_2 \sum_{\mathbf{k} \in \mathbb{Z}^d} K_{d+1}(\mathbf{k}) =$$

$$= k_2 (d+1)^{d+1} \sum_{\mathbf{k} \in \mathbb{Z}^d} \frac{1}{(1 + |k_1| + \cdots + |k_d|)^{d+1}} < +\infty,$$

by Lemma 5.3 and Lemma 5.4, i.e. Fourier series (5.10) absolutely converges and hence it converges to $f(x, \theta)$ since $f(x, \cdot)$ is continuous by (P1) (see [30], pp. 197-206).

THEOREM 5.5. *Let $\Theta \subseteq \mathbb{R}^d$ be an open set and let $f : E \times \Theta \to \mathbb{R}$ be a uniformly bounded function such that for all $x \in E \subseteq \mathbb{R}^k$, $f(x, \cdot) \in C^{d+2}(\Theta)$, and all partial derivatives of f of the orders $m \leq d+2$ are uniformly bounded. Moreover, let $(X_t)_{1 \leq t \leq T}$ be a random process that satisfies (1.1) and let B be a standard Brownian motion defined on the same filtered probability space. Then there exists random function $F : \Omega \times \Theta \to \mathbb{R}$ such that for all $\omega \in \Omega$, $F(\omega, \cdot) \in C^1(\Theta)$, and for all $\theta \in \Theta$ (with $F(\theta) \equiv F(\cdot, \theta)$)*

$$F(\theta) = \int_0^T f(X_t, \theta) \, dB_t \text{ a.s.}$$

$$\frac{\partial}{\partial \theta_j} F(\theta) = \int_0^T \frac{\partial}{\partial \theta_j} f(X_t, \theta) \, dB_t \text{ a.s. for } j = 1, 2, \ldots, d.$$

PROOF OF THEOREM 5.5. Let $\theta' \in \Theta$ be an arbitrary point. Since Θ is an open set, there exists open neighborhood $U' \subset \Theta$ of θ' which is d-dimensional rectangular such that $Cl(U') \subset \Theta$. Without loss of generality, let us assume that $U' = \mathcal{K}_0$, and that f can be extended on $E \times \mathbb{R}^d$ such that it satisfies (P1) and (P2).

Let $C_{\mathbf{k}}(x) = (1/(2\pi)^d) \int_{Cl(\mathcal{K}_0)} f(x, \theta) e^{-i\langle \mathbf{k} | \theta \rangle} d\theta$ be a Fourier coefficient of a scalar function f for $\mathbf{k} \in \mathbb{Z}^d$ such that $|\mathbf{k}| \leq m = d + 1 < d + 2$. Then $C_{\mathbf{k}}(\cdot)$ is a bounded and continuous function on E by Lemma 5.3 that implies that stochastic integral $\int_0^T C_{\mathbf{k}}(X_t) \, dB_t$ is well defined. It follows that

$$\mathbb{E}\left(\sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \int_0^T C_{\mathbf{k}}(X_t) \, dB_t \right| \right) \leq \sum_{\mathbf{k} \in \mathbb{Z}^d} \left\| \int_0^T C_{\mathbf{k}}(X_t) \, dB_t \right\|_{L^2} =$$

$$= \sum_{\mathbf{k} \in \mathbb{Z}^d} \left(\mathbb{E} \int_0^T C_{\mathbf{k}}^2(X_t) \, dt \right)^{\frac{1}{2}} \leq$$

$$\leq \sqrt{T} k_2 \sum_{\mathbf{k} \in \mathbb{Z}^d} K_{d+1}(\mathbf{k}) < +\infty$$

by Lemmas 5.3 and 5.4 implying that

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \int_0^T C_{\mathbf{k}}(X_t) \, dB_t \cdot e^{i\langle \mathbf{k} | \theta \rangle} \right| \leq \sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \int_0^T C_{\mathbf{k}}(X_t) \, dB_t \right| < +\infty \text{ a.s.}$$

Hence series $\sum_{\mathbf{k} \in \mathbb{Z}^d} \int_0^T C_{\mathbf{k}}(X_t) \, dB_t \cdot e^{i\langle \mathbf{k} | \theta \rangle}$ converges absolutely and uniformly for all $\theta \in \mathcal{K}_0$ on event $\Omega_0 := \{ \sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \int_0^T C_{\mathbf{k}}(X_t) \, dB_t \right| < +\infty \}$ of probability 1. Then for all fixed $\omega \in \Omega_0$, its sum

$$F(\omega, \theta) := \sum_{\mathbf{k} \in \mathbb{Z}^d} \int_0^T C_{\mathbf{k}}(X_t) \, dB_t(\omega) \cdot e^{i\langle \mathbf{k} | \theta \rangle}, \ \theta \in \mathcal{K}_0,$$

is a continuous function on $\mathcal{K}_0$. Moreover,

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} \left| \frac{\partial}{\partial\theta_j} \left(\int_0^T C_{\mathbf{k}}(X_t)\, dB_t \cdot e^{i\langle\mathbf{k}|\theta\rangle} \right) \right| = \sum_{\mathbf{k}\in\mathbb{Z}^d} \left| \int_0^T iC_{\mathbf{k}}(X_t)k_j\, dB_t \right| \stackrel{(5.9)}{=}$$

$$= \sum_{\mathbf{k}\in\mathbb{Z}^d} \left| \int_0^T C_{\mathbf{k}}^{(e_j)}(X_t)\, dB_t \right|$$

for all $j = 1, 2, \ldots, d$, and hence functional series $\sum_{\mathbf{k}\in\mathbb{Z}^d} \frac{\partial}{\partial\theta_j}(\int_0^T C_{\mathbf{k}}(X_t)\, dB_t \cdot e^{i\langle\mathbf{k}|\theta\rangle})$ converges absolutely and uniformly on event

$$\Omega_j := \left\{ \sum_{\mathbf{k}\in\mathbb{Z}^d} \left| \int_0^T C_{\mathbf{k}}^{(e_j)}(X_t)\, dB_t \right| < +\infty \right\}.$$

Since $C_{\mathbf{k}}^{(e_j)}(x) = (1/(2\pi)^d) \int_{Cl(\mathcal{K}_0)} \frac{\partial}{\partial\theta_j} f(x,\theta) e^{-i\langle\mathbf{k}|\theta\rangle} d\theta$ and $\frac{\partial}{\partial\theta_j} f$ is a bounded function, it follows that $\mathbb{P}(\Omega_j) = 1$, and hence $F(\omega, \cdot) \in C^1(\mathcal{K}_0)$ and

$$\frac{\partial}{\partial\theta_j} F(\omega, \theta) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \int_0^T C_{\mathbf{k}}^{(e_j)}(X_t)\, dB_t(\omega) \cdot e^{i\langle\mathbf{k}|\theta\rangle}$$

for all $\omega \in \Omega_0 \cap \bigcap_{j=1}^d \Omega_j$ that is an event of probability 1.

Let $S_N^{(0)}(x,\theta) := \sum_{|\mathbf{k}|\le N} C_{\mathbf{k}}(x) e^{i\langle\mathbf{k}|\theta\rangle}$ and $S_N^{(j)}(x,\theta) := \sum_{|\mathbf{k}|\le N} C_{\mathbf{k}}^{(e_j)}(x) e^{i\langle\mathbf{k}|\theta\rangle}$ ($N \in \mathbb{N}$) be partial sums of corresponding Fourier series ($j = 1, 2, \ldots, d$). Since $S_N^{(0)}(X_t, \theta) - f(X_t, \theta)$ and $S_N^{(j)}(X_t, \theta) - \frac{\partial}{\partial\theta_j} f(X_t, \theta)$ are uniformly bounded with constants by Lemmas 5.3 and 5.4, and

$$\lim_{N\to+\infty} S_N^{(0)}(X_t, \theta) = f(X_t, \theta), \text{ and } \lim_{N\to+\infty} S_N^{(j)}(X_t, \theta) = \frac{\partial}{\partial\theta_j} f(X_t, \theta),$$

for all $(t, \theta) \in [0, T] \times \mathcal{K}_0$, it follows that

$$\int_0^T S_N^{(0)}(X_t, \theta)\, dB_t \xrightarrow{\mathbb{P}} \int_0^T f(X_t, \theta)\, dB_t,\ N \to +\infty,$$

$$\int_0^T S_N^{(j)}(X_t, \theta)\, dB_t \xrightarrow{\mathbb{P}} \int_0^T \frac{\partial}{\partial\theta_j} f(X_t, \theta)\, dB_t,\ N \to +\infty,$$

by the dominated convergence theorem for stochastic integrals [28, Theorem 2.12]. On the other hand, since $\int_0^T S_N^{(0)}(X_t, \theta)\, dB_t$ and $\int_0^T S_N^{(j)}(X_t, \theta)\, dB_t$ are partial sums of functional series that a.s. converge to $F(\cdot, \theta)$ and $\frac{\partial}{\partial\theta_j} F(\cdot, \theta)$ respectively, it follows that

$$F(\cdot, \theta) = \int_0^T f(X_t, \theta)\, dB_t \text{ a.s., and } \frac{\partial}{\partial\theta_j} F(\cdot, \theta) = \int_0^T \frac{\partial}{\partial\theta_j} f(X_t, \theta)\, dB_t \text{ a.s.}$$

We can extend definition of F on Θ in the following way. Let $\{U_j : j \in \mathbb{N}\}$ be a countable open covering of Θ such that $U_j \subset Cl(U_j) \subset \Theta$ be an

open rectangular. If F_{U_j} is the function defined by Fourier series like above, then $F(\omega, \theta) := F_{U_j}(\omega, \theta)$ if $\theta \in U_j$ and for ω from the probability-one-event obtained as countable intersections of probability-one-events like above. If there exists $j' \neq j$ such that $\theta \in U_j \cap U_{j'}$, then $F_{U_{j'}}(\theta) = \int_0^T f(X_t, \theta) \, dB_t = F_{U_j}(\theta)$ a.s. This implies that F is correctly defined. $\square$

PROOF OF THEOREM 4.1. The conclusions follow directly by applying (1.1) for $\theta = \theta_0$ and then Theorem 5.5 on Itô stochastic integrals with respect to components of k-dimensional Brownian motion W, and from the fact that integrals with respect to the time variable and partial derivatives with respect to θ commute. $\square$

THEOREM 5.6. *Let $(X_t)_{1 \leq t \leq T}$ be a random process that satisfies (1.1). Let $E \subseteq \mathbb{R}^k$ and $\Theta \subseteq \mathbb{R}^d$ be open and convex sets. Moreover, let $f : E \times \Theta \to \mathbb{R}^k$ be a function that satisfies (P1). We assume that $\mu(\cdot, \theta_0)$ and ν are bounded and continuous on E.*
Let $a : E \to \mathbb{R}^k$ be a bounded vector valued function. For every relatively compact set $\mathcal{K} \subset \Theta$ there exist constants $C_1, C_2 > 0$ such that:

$$(5.13) \quad \left\| \sup_{\theta \in \Theta} \Big| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \big\langle f(X_t, \theta) - f(X_{i-1}, \theta) \,\big|\, a(X_t) \big\rangle \, dt \Big| \right\|_{L^2} \leq C_1 \sqrt{\Delta_n}$$

$$(5.14) \quad \left\| \sup_{\theta \in \mathcal{K}} \Big| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \big\langle f(X_t, \theta) - f(X_{i-1}, \theta) \,\big|\, \nu(X_t) \, dW_t \big\rangle \Big| \right\|_{L^2} \leq C_2 \sqrt{\Delta_n}.$$

PROOF OF THEOREM 5.6. Let $c_{20}, \ldots, c_{24}$ denote positive constants occurring in the proof. Using Cauchy-Schwarz inequality for integrals and vectors, Lemma 2.4 and boundedness of functions $\nabla_x f$ and a, we have that there exists a constant c_{20} such that

$$\mathbb{E}\left[\left(\sup_{\theta \in \Theta} \Big| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \big\langle f(X_t, \theta) - f(X_{i-1}, \theta) \,\big|\, a(X_t) \big\rangle \, dt \Big| \right)^2 \right] \leq$$

$$\leq c_{20} \mathbb{E}\left[\sum_{i=1}^n \int_{t_{i-1}}^{t_i} \| X_t - X_{i-1} \|_2^2 \, dt \right].$$

For final step, we use Lemma 5.2.

$$c_{20} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \mathbb{E}\left[\|X_t - X_{i-1}\|_2^2\right] dt \le c_{20} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} 2K\big((t-t_{i-1})^2 +$$
$$+ (t-t_{i-1})\big) dt \le$$
$$\le c_{20} K \left(\frac{2}{3} T^2 + T\right) \Delta_n =$$

(5.15)
$$=: C_1^2 \Delta_n$$

so inequality (5.13) is proved.

Next, we prove inequality (5.14) by following the arguments from part of the proof of Theorem 6.1 in [17]. The first main difference is that Brownian motion W is multidimensional, so the first part of the proof is derived componentwise, and then the continuous mapping theorem is used. Also, the observed time interval is fixed compared to the proof of Theorem 6.1 in [17], where this is not the case. For this reason, it is easier to justify the application of the dominated convergence theorem for stochastic integrals to the integrals defined below. Moreover, for the same reason, it is easier to obtain the L^2-bound of the right-hand side of the inequality in (5.17). To clarify the arguments presented and to distinguish details due to multidimensionality, we provide the entire proof.

Without loss of generality, it is sufficient to prove it for $\mathcal{K} = \mathcal{K}_0$.
Let $S_N(x, \theta) = \sum_{|\mathbf{k}| \le N} C_{\mathbf{k}}(x) e^{i\langle \mathbf{k} | \theta \rangle}$ be N-th partial sum of Fourier's series of functions $f(x, \theta)$. For fixed $\theta \in \mathcal{K}_0$ and $N \in \mathbb{N}$, let us define the following random processes

$$V_t := \sum_{l=1}^{k} \sum_{i=1}^{n} \big\langle f(X_t, \theta) - f(X_{i-1}, \theta) \,\big|\, \nu(X_t) e_l \big\rangle \mathbb{1}_{\langle t_{i-1}, t_i]}(t), \quad t \in [0, T],$$

$$V_t^{(N)} := \sum_{l=1}^{k} \sum_{i=1}^{n} \big\langle S_N(X_t, \theta) - S_N(X_{i-1}, \theta) \,\big|\, \nu(X_t) e_l \big\rangle \mathbb{1}_{\langle t_{i-1}, t_i]}(t), \quad t \in [0, T].$$

Moreover, let us define $M_N := \sup_{x \in E, \theta \in Cl(\mathcal{K}_0)} \|S_N(x, \theta) - f(x, \theta)\|_2$ and $M := \sup_{x \in E} \sum_{\mathbf{k} \in \mathbb{Z}^d} \|C_{\mathbf{k}}(x)\|_2$. For $m = d + 1$, using (5.12) in Lemma 5.3 and Lemma 5.4 we conclude that M is finite.
For every $N \in \mathbb{N}$

$$\left| V_t^{(N)} \right| \le \sum_{l=1}^{k} \sum_{i=1}^{n} \left\| \sum_{|\mathbf{k}| \le N} (C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1})) \right\|_2 \|\nu(X_t) e_l\|_2 \, \mathbb{1}_{\langle t_{i-1}, t_i]}(t) \le$$

(5.16) $$\le 2M c_{21}$$

where $c_{21} = \max_{l=1,\dots,k} \|\nu(X_t) e_l\|_2$.
Because of smoothness of function f, it can be shown that for fixed $x \in E$,

$\lim_{N \to +\infty} \sup_{\theta \in Cl(\mathcal{K}_0)} \|S_N(x, \theta) - f(x, \theta)\|_2 = 0$ holds (see [30]). Moreover, using (5.12) in Lemma 5.3 and Lemma 5.4, for all $x \in E$ we have

$$\|S_N(x, \theta) - f(x, \theta)\|_2 \le \sum_{\mathbf{k} > N} \|C_{\mathbf{k}}(x)\|_2 \le k_2 \sum_{\mathbf{k} > N} K_{d+1}(\mathbf{k}) < +\infty.$$

From the previous inequality we can easily determine that $M_N \to 0$.
In the same manner as in (5.16), we obtain

$$\sup_{t \in [0, T]} |V_t^{(N)} - V_t| \le 2M_N c_{21} \to 0, \quad N \to +\infty.$$

The dominated convergence theorem for stochastic integrals [28, Theorem 2.12] applied on Itô integrals with respect to each component of Brownian motion $(W_t)_t$ yields

$$\int_0^T V_t^{(N)} \, dW_t^l \xrightarrow{\mathbb{P}} \int_0^T V_t \, dW_t^l, \text{ for } l = 1, \dots, k,$$

when $N \to +\infty$. Let us denote

$$I_N(\theta) \coloneqq \sum_{l=1}^k \int_0^T V_t^{(N)} \, dW_t^l = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \langle S_N(X_t, \theta) - S_N(X_{i-1}, \theta) \,|\, \nu(X_t) \, dW_t \rangle,$$

$$I(\theta) \coloneqq \sum_{l=1}^k \int_0^T V_t \, dW_t^l = \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \langle f(X_t, \theta) - f(X_{i-1}, \theta) \,|\, \nu(X_t) \, dW_t \rangle.$$

Using continuous mapping theorem [31, Theorem 2.3], it follows that $I_N(\theta) \xrightarrow{\mathbb{P}} I(\theta)$ when $N \to +\infty$.
Hence, for every $\theta \in \mathcal{K}_0 \cap \mathbb{Q}^d$ exists a subsequence $(N_p) \equiv (N_p(\theta))$ and an event $A(\theta)$, $\mathbb{P}(A) = 1$, such that for all $\omega \in A(\theta)$, $\lim_{p \to +\infty} I_{N_p}(\theta)(\omega) = I(\theta)(\omega)$.
We define $A_0 \coloneqq \cap_{\theta \in \mathcal{K}_0 \cap \mathbb{Q}^d} A(\theta)$ and it is also the events of probability one. On event A_0, for $\theta \in \mathcal{K}_0 \cap \mathbb{Q}^d$ we have that

$$\left| I(\theta) \right| \le \left| I(\theta) - I_{N_p}(\theta) \right| +$$

$$+ \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left\langle \sum_{|\mathbf{k}| \le N_p} (C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1})) e^{-i\langle \mathbf{k} | \theta \rangle} \,\Big|\, \nu(X_t) \, dW_t \right\rangle \right| \le$$

$$\le \left| I(\theta) - I_{N_p}(\theta) \right| + \sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \,|\, \nu(X_t) \, dW_t \rangle \right|.$$

After letting $p \to +\infty$, it follows

$$\left| I(\theta) \right| \le \sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \,|\, \nu(X_t) \, dW_t \rangle \right|.$$

For $g(X_t, \theta) = \sum_{i=1}^n (f(X_t, \theta) - f(X_{i-1}, \theta)) \mathbb{1}_{\langle t_{i-1}, t_i]}(t)$ using Theorem 5.5 the mapping $\theta \mapsto I(\theta)$ is almost surely a continuous function. There is an

event B_0 of probability one such that for every $\omega \in B_0$, $\sup_{\theta \in \mathcal{K}_0} |I(\theta)(\omega)| = \sup_{\theta \in \mathcal{K}_0 \cap \mathbb{Q}^d} |I(\theta)(\omega)|$. Accordingly, on event $C_0 := A_0 \cap B_0$ of probability one, we have that

$$(5.17) \qquad \sup_{\theta \in \mathcal{K}_0} |I(\theta)| \leq \sum_{\mathbf{k} \in \mathbb{Z}^d} \left| \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \,\middle|\, \nu(X_t)\, dW_t \right\rangle \right|.$$

In analysing the squared L^2-norm of the right-hand side in (5.17), we use Lemma 2.3 (ii), the norm consistency, (5.11) from Lemma 5.3 ($m = d + 1$), and Lemma 5.2. The last inequality is obtained in the same way as in (5.15).

$$\mathbb{E}\left[\left| \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \,\middle|\, \nu(X_t)\, dW_t \right\rangle \right|^2 \right] =$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} (C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}))^T \nu(X_t)\, dW_t \right)^2 \right]$$

$$+ 2 \sum_{1 \leq i < l \leq n} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left\langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \,\middle|\, \nu(X_t)\, dW_t \right\rangle \cdot \right.$$

$$\left. \int_{t_{l-1}}^{t_l} \left\langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{l-1}) \,\middle|\, \nu(X_t)\, dW_t \right\rangle \right] =$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left\| \nu(X_t)^T (C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1})) \right\|_2^2 dt \right] \leq$$

$$\leq c_{22} \mathbb{E}\left[\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\| C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \right\|_2^2 dt \right] \leq$$

$$\leq c_{22} \mathbb{E}\left[\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} k_1^2 \left(\frac{d+1}{1 + |k_1| + \cdots + |k_d|} \right)^{2(d+1)} \|X_t - X_{i-1}\|_2^2 \, dt \right] \leq$$

$$\leq c_{23} \left(\frac{d+1}{1 + |k_1| + \cdots + |k_d|} \right)^{2(d+1)} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \mathbb{E}\left[\|X_t - X_{i-1}\|_2^2 \right] dt \leq$$

$$\leq c_{24}(d+1)^{2(d+1)} \frac{1}{(1 + |k_1| + \cdots + |k_d|)^{2(d+1)}} \Delta_n$$

Since series $\sum_{\mathbf{k} \in \mathbb{Z}^d} \left(\frac{1}{1 + |k_1| + \cdots + |k_d|} \right)^{d+1}$ converges, we have

$$\sum_{k \in \mathbb{Z}^d} \left\| \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle C_{\mathbf{k}}(X_t) - C_{\mathbf{k}}(X_{i-1}) \,\middle|\, \nu(X_t)\, dW_t \right\rangle \right\|_{L^2} < +\infty.$$

Therefore, the series on the right hand side of (5.17) converges in L^2 and almost sure to the same limits that are equal almost sure [8, Proposition

2.10.1]. Finally, we obtain

$$\left\| \sup_{\theta \in \mathcal{K}_0} |I(\theta)| \right\|_{L^2} \leq C_2 \sqrt{\Delta_n}.$$

□

PROOF OF THEOREM 4.3. Let us assume that E is a compact set. Let $c_8, \ldots, c_{19}$ be positive constants occurring in the proof.

Using abbreviation in (4.8) we define a random vector $\chi_i^n := [\chi_i^{n,1}, \ldots, \chi_i^{n,d}]^T$ where each component is equal to

$$\chi_i^{n,j} := \frac{1}{\sqrt{\Delta_n}} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^s \nabla g_j(X_u) \nu(X_u) \, dW_u \, \middle| \, \nu(X_s) \, dW_s \right\rangle.$$

We also introduce matrix functions $R^{(j)}$ and column vectors $J^{(i,j)} = \left[J_1^{(i,j)}, \ldots, J_k^{(i,j)} \right]^T$ in the following form.

$$R^{(j)}(t, \theta) := \nabla g_j(X_t) \nu(X_t), \quad j = 1, 2, \ldots, d$$

$$J^{(i,j)}(s) := \int_{t_{i-1}}^s R^{(j)}(u, \theta) \, dW_u$$

Let $C = (C_{jl}(t))_{0 \leq t \leq T}$ be a continuous adapted process given by

$$C_{jl}(t) = \frac{1}{2} \int_0^t \sum_{p,r=1}^k S_{pr}(X_s) \left\langle e_r^T R^{(j)}(s, \theta) \, \middle| \, e_p^T R^{(l)}(s, \theta) \right\rangle \, ds, \quad j, l = 1, \ldots, d.$$

We will prove that all conditions of Theorem 2.5 are fulfilled. Since E is a compact set, the matrix function $R^{(j)}(t, \theta)$ is bounded for every $j = 1, 2, \ldots, d$, and for every $i = 1, 2, \ldots, n$ vectors χ_i^n are square-integrable random vectors. Using the definition of the scalar product, it follows that for all $i = 1, 2, \ldots, n$, χ_i^n is a sum of Itô integrals, hence it is a martingale. The equality $\mathbb{E}\left[\chi_i^{n,j} | \mathcal{F}_{n,i-1} \right] = 0$ trivially implies the statement of condition (i).

To satisfy condition (iii), it is sufficient to prove that for $1 \leq j \leq d$ and $1 \leq m \leq k$

$$\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^{n,j} \left(W_{t_i}^m - W_{t_{i-1}}^m \right) \, \middle| \, \mathcal{F}_{n,i-1} \right] \xrightarrow{\mathbb{P}} 0.$$

Using the product formula, we have

$$\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^{n,j} \left(W_{t_i}^m - W_{t_{i-1}}^m \right) | \mathcal{F}_{n,i-1} \right] =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \sum_{p=1}^k \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \nu(X_s) e_p \, dW_s^p \cdot \int_{t_{i-1}}^{t_i} dW_s^m \, \middle| \, \mathcal{F}_{n,i-1} \right] =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \sum_{p=1}^{k} \mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left(J^{(i,j)}(u) \right)^T \nu(X_u) e_p \, dW_u^p \, dW_s^m +$$

$$+ \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} dW_u^m \left(J^{(i,j)}(s) \right)^T \nu(X_s) e_p \, dW_s^p +$$

$$+ \left\langle \int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \nu(X_s) e_p \, dW_s^p, \int_{t_{i-1}}^{t_i} dW_s^m \right\rangle \Big| \mathcal{F}_{n,i-1} \Big].$$

Using the martingale property and the independence of the components of Brownian motion, we have

$$\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^{n,j} \left(W_{t_i}^m - W_{t_{i-1}}^m \right) \Big| \mathcal{F}_{n,i-1} \right] =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \nu(X_s) e_m \, ds \Big| \mathcal{F}_{n,i-1} \right] =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \left(\nu(X_s) - \nu(X_{i-1}) \right) e_m \, ds \Big| \mathcal{F}_{n,i-1} \right] +$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \nu(X_{i-1}) e_m \, ds \Big| \mathcal{F}_{n,i-1} \right].$$

We denote by $\tilde{\nu}_m(x)$ the m-th column of matrix $\nu(x)$. Using Lemma 2.1 for function $\tilde{\nu}_m$ on interval $[t_{i-1}, s]$, we have

$$\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^{n,j} \left(W_{t_i}^m - W_{t_{i-1}}^m \right) \Big| \mathcal{F}_{n,i-1} \right] =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \mathbb{E}\left[\left(J^{(i,j)}(s) \right)^T \left(\tilde{\nu}_m(X_s) - \tilde{\nu}_m(X_{i-1}) \right) ds \Big| \mathcal{F}_{n,i-1} \right] =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \int_{t_{i-1}}^{s} \Big(\nabla \tilde{\nu}_m(X_u) \mu(X_u, \theta_0) +$$

(5.18)

$$+ \frac{1}{2} \nabla_2 \tilde{\nu}_m(X_u) \Big) du \, ds \Big| \mathcal{F}_{n,i-1} \Big] +$$

(5.19)

$$+ \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(J^{(i,j)}(s) \right)^T \int_{t_{i-1}}^{s} \nabla \tilde{\nu}_m(X_u) \nu(X_u) \, dW_u \, ds \Big| \mathcal{F}_{n,i-1} \right].$$

Next, we analyze the L^1-norm of (5.18). There exists a constant c_8 such that

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{A_n^t}\mathbb{E}\left[\int_{t_{i-1}}^{t_i}\left\langle J^{(i,j)}(s)\,\bigg|\,\int_{t_{i-1}}^{s}\left(\nabla\tilde{\nu}_m(X_u)\mu(X_u,\theta_0)+\right.\right.\right.\right.$$

$$\left.\left.\left.\left.+\frac{1}{2}\nabla_2\tilde{\nu}_m(X_u)\right)du\right\rangle ds\big|\mathcal{F}_{n,i-1}\right]\right\|\right]\leq$$

$$\leq c_8\sqrt{\Delta_n}T\mathbb{E}[M],$$

where $M := \sup_{0\leq s\leq T}\left\|\int_0^s R^{(j)}(u,\theta)\,dW_u\right\|_2$. Using Doob's maximal inequality for vector martingale, we have that (5.18) converges to 0 in L^1 norm, so it also converges to 0 in probability.

Let us denote $H_m(x) := \nabla\tilde{\nu}_m(x)\nu(x)$. Using Itô formula for function $F(x,y) = \langle x\,|\,y\rangle$, we have

$$\frac{1}{\Delta_n}\sum_{i=1}^{A_n^t}\mathbb{E}\left[\int_{t_{i-1}}^{t_i}\left\langle\int_{t_{i-1}}^{s}R^{(j)}(u,\theta)\,dW_u\,\bigg|\,\int_{t_{i-1}}^{s}H_m(X_u)\,dW_u\right\rangle ds\big|\mathcal{F}_{n,i-1}\right]=$$

$$=\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{A_n^t}\mathbb{E}\Big[\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\left\langle\int_{t_{i-1}}^{u}H_m(X_v)\,dW_v\,\bigg|\,R^{(j)}(u,\theta)\,dW_u\right\rangle ds\big|\mathcal{F}_{n,i-1}\Big]+$$

$$+\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{A_n^t}\mathbb{E}\Big[\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\left\langle\int_{t_{i-1}}^{u}R^{(j)}(v,\theta)\,dW_v\,\bigg|\,H_m(X_u)\,dW_u\right\rangle ds\big|\mathcal{F}_{n,i-1}\Big]+$$

$$+\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{A_n^t}\sum_{p,r=1}^{k}\mathbb{E}\Big[\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\left(R^{(j)}(u,\theta)\circ H_m(X_u)\right)_{pr}du\,ds\big|\mathcal{F}_{n,i-1}\Big].$$

Using the well-known characterization of L^1 random variables, it is easy to prove that the integral sign and expectation can be interchanged so that the first two summands are zero. For the third one there is a constant c_9 such that

$$\left|\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{A_n^t}\sum_{p,r=1}^{k}\mathbb{E}\Big[\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\left(R^{(j)}(u,\theta)\circ H_m(X_u)\right)_{pr}du\,ds\big|\mathcal{F}_{n,i-1}\Big]\right|\leq$$

$$\leq c_9 T\sqrt{\Delta_n}.$$

We conclude that expression (5.19) converges almost surely to 0, so it also converges to zero in probability.

To prove (v), it is enough to show that for arbitrary j, $1\leq j,\leq d$

$$\sum_{i=1}^{A_n^t}\mathbb{E}\left[\chi_i^{n,j}\left(N_{t_i}-N_{t_{i-1}}\right)\big|\mathcal{F}_{n,i-1}\right]\xrightarrow{\mathbb{P}}0,\quad\forall N\in\mathcal{M}_b\left(W^\perp\right).$$

In a similar way as before, we have

$$
\sum_{i=1}^{A_n^t} \mathbb{E}\left[\chi_i^{n,j}\left(N_{t_i} - N_{t_{i-1}} \right) \middle| \mathcal{F}_{n,i-1} \right] =
$$

$$
= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left\langle J^{(i,j)}(s,\theta) \middle| \nu(X_s)\, dW_s \right\rangle \cdot \int_{t_{i-1}}^{t_i} dN_s \middle| \mathcal{F}_{n,i-1} \right] =
$$

$$
= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \sum_{p=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left\langle J^{(i,j)}(u,\theta) \middle| \nu(X_u)e_p \right\rangle dW_u^p\, dN_s \middle| \mathcal{F}_{n,i-1} \right] +
$$

$$
+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \sum_{p=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} dN_u \left\langle J^{(i,j)}(s,\theta) \middle| \nu(X_s)e_p \right\rangle dW_s^p \middle| \mathcal{F}_{n,i-1} \right] +
$$

$$
+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \sum_{p=1}^{k} \mathbb{E}\left[\left\langle \int_{t_{i-1}}^{t_i} \left\langle J^{(i,j)}(u,\theta) \middle| \nu(X_u)e_p \right\rangle dW_u^p, \int_{t_{i-1}}^{t_i} dN_u \right\rangle \middle| \mathcal{F}_{n,i-1} \right] =
$$

$$
= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{A_n^t} \sum_{p=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left\langle J^{(i,j)}(u,\theta) \middle| \nu(X_u)e_p \right\rangle d\,[W^p, N]_u \middle| \mathcal{F}_{n,i-1} \right] =
$$

$$
= 0.
$$

We use the fact that the conditional expectation of the martingale difference is zero [28, Theorem 2.2]. In the last equality we use the orthogonality of the process N on the components of Brownian motion.

Next, we prove (iv). Let $\epsilon > 0$. Using Cauchy-Schwarz and Markov inequality [11, Theorem 1.6.4], we have

$$
\mathbb{E}\left| \sum_{i=1}^{n} \mathbb{E}\left[\|\chi_i^n\|_2^2 \cdot \mathbb{1}_{\{\|\chi_i^n\|_2 > \epsilon\}} \middle| \mathcal{F}_{n,i-1} \right] \right| \leq \sum_{i=1}^{n} \mathbb{E}\left[\|\chi_i^n\|_2^4 \right]^{\frac{1}{2}} \cdot \mathbb{P}\left(\|\chi_i^n\|_2 > \epsilon \right)^{\frac{1}{2}} \leq
$$

$$
\leq \sum_{i=1}^{n} \mathbb{E}\left[\|\chi_i^n\|_2^4 \right]^{\frac{1}{2}} \left(\frac{\mathbb{E}\left[\|\chi_i^n\|_2^4 \right]}{\epsilon^4} \right)^{\frac{1}{2}} =
$$

$$
= \frac{1}{\epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[\|\chi_i^n\|_2^4 \right] =
$$

$$
\leq \frac{d}{\epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[\sum_{j=1}^{d} (\chi_i^{n,j})^4 \right].
$$

To satisfy condition (iv), it is sufficient to prove that

$$\lim_{n \to +\infty} \frac{1}{\epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(\chi_i^{n,j}\right)^4\right] = 0.$$

If we denote by

$$H^{(i,j)}(s) := \int_{t_{i-1}}^{s} \left\langle J^{(i,j)}(u) \,\middle|\, \nu(X_u)\,dW_u \right\rangle,$$

there exist a constant c_{10} such that

$$\frac{1}{\epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(\chi_i^{n,j}\right)^4\right] =$$

$$= \frac{1}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(H^{(i,j)}(t_i)\right)^4\right] =$$

$$= \frac{1}{\Delta^2 \epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[4\int_{t_{i-1}}^{t_i} \left(H^{(i,j)}(s)\right)^3 \left(J^{(i,j)}(s)\right)^T \nu(X_s)\,dW_s\right] +$$

$$+ \frac{1}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \mathbb{E}\left[6\int_{t_{i-1}}^{t_i} \left(H^{(i,j)}(s)\right)^2 d\left\langle H^{(i,j)}\right\rangle_s\right] =$$

$$= \frac{6}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n}\sum_{p=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(H^{(i,j)}(s)\right)^2 \left\langle J^{(i,j)}(s) \,|\, \nu(X_s)e_p\right\rangle^2 ds\right] \leq$$

(5.20) $$\leq \frac{c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n}\sum_{r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left(H^{(i,j)}(s)\right)^2 \cdot \left(J_r^{(i,j)}(s)\right)^2 ds\right].$$

To analyze the expression under the integral sign, we apply Itô formula for the function $F(x,y) = x^2 y^2$. The independence of the components of Brownian motion is essential for the calculation of the quadratic covariance and variances.

$$\left(H^{(i,j)}\right)^2 \cdot \left(J_r^{(i,j)}\right)^2 =$$

$$= 2\int_{t_{i-1}}^{s} H^{(i,j)}(u) \left(J_r^{(i,j)}(u)\right)^2 dH^{(i,j)}(u) +$$

$$+ 2\int_{t_{i-1}}^{s} \left(H^{(i,j)}(u)\right)^2 J_r^{(i,j)}(u)\,dJ_r^{(i,j)}(u) +$$

$$+ \frac{1}{2}\int_{t_{i-1}}^{s} 2\left(J_r^{(i,j)}(u)\right)^2 d\left\langle H^{(i,j)}, H^{(i,j)}\right\rangle_u +$$

$$+ \frac{1}{2} \cdot 2\int_{t_{i-1}}^{s} 4H^{(i,j)}(u)J_r^{(i,j)}(u)\,d\left\langle H^{(i,j)}, J_r^{(i,j)}\right\rangle_u +$$

$$+ \frac{1}{2} \int_{t_{i-1}}^{s} 2 \left(H^{(i,j)} \right)^2 d \left\langle J_r^{(i,j)}, J_r^{(i,j)} \right\rangle_u =$$

$$= 2 \int_{t_{i-1}}^{s} H_r^{(i,j)}(u) \left(J_r^{(i,j)}(u) \right)^2 \left(J^{(i,j)}(u) \right)^T \nu(X_u) \, dW_u +$$

$$+ 2 \int_{t_{i-1}}^{s} \left(H^{(i,j)}(u) \right)^2 J_r^{(i,j)}(u) e_r^T R^{(i,j)}(u, \theta) \, dW_u +$$

$$+ \int_{t_{i-1}}^{s} \left(J_r^{(i,j)}(u) \right)^2 \left\| \nu(X_u)^T J^{(i,j)}(u) \right\|_2^2 \, du +$$

$$+ 4 \int_{t_{i-1}}^{s} H^{(i,j)}(u) J_r^{(i,j)}(u) \left\langle \left(J^{(i,j)}(u) \right)^T \nu(X_u) \,\middle|\, e_r^T R^{(j)}(u, \theta) \right\rangle \, du +$$

$$(5.21) \quad + \int_{t_{i-1}}^{s} \left(H^{(i,j)}(u) \right)^2 \left\| e_r^T R^{(j)}(u, \theta) \right\|_2^2 \, du$$

Considering equality (5.21), we conclude that

$$\frac{1}{\epsilon^2} \sum_{i=1}^{n} \mathbb{E} \left[\left(\chi_i^{n,j} \right)^4 \right] \le$$

$$\le \frac{12 c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \sum_{r=1}^{k} \mathbb{E} \left[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left(J_r^{(i,j)}(u) \right)^2 \left\| \nu(X_u)^T J^{(i,j)}(u) \right\|_2^2 \, du \, ds \right] +$$

$$+ \frac{48 c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \sum_{r=1}^{k} \mathbb{E} \Big[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} H^{(i,j)}(u) J_r^{(i,j)}(u) \cdot$$

$$\cdot \left\langle \left(J^{(i,j)}(u) \right)^T \nu(X_u) \,\middle|\, e_r^T R^{(j)}(u, \theta) \right\rangle \, du \, ds \Big] +$$

$$+ \frac{12 c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \sum_{r=1}^{k} \mathbb{E} \left[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left(H^{(i,j)}(u) \right)^2 \left\| e_r^T R^{(j)}(u, \theta) \right\|_2^2 \, du \, ds \right].$$

By a similar reasoning as before (using norm consistency, Itô formula, Lemma 2.3 (i) and (ii)), there exist constants c_{12} and c_{14} such that

$$\frac{12 c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \sum_{r=1}^{k} \mathbb{E} \left[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left(J_r^{(i,j)}(u) \right)^2 \left\| \nu(X_u)^T J^{(i,j)}(u) \right\|_2^2 \, du \, ds \right] \le$$

$$\le \frac{12 c_{11}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^{n} \sum_{p,r=1}^{k} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \int_{t_{i-1}}^{u} \left[\mathbb{E} \left[\left(J_p^{(i,j)}(v) \right)^2 \right] + 4 \mathbb{E} \left[J_p^{(i,j)}(v) J_r^{(i,j)}(v) \right] + \right.$$

$$+ \mathbb{E} \left[\left(J_r^{(i,j)}(v) \right)^2 \right] \Big] \, dv \, du \, ds \le c_{12} \Delta_n T,$$

and

$$\frac{12c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^n \sum_{r=1}^k \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^s \mathbb{E}\left[\left(H^{(i,j)}(u)\right)^2 \left\|e_r^T R^{(j)}(u,\theta)\right\|_2^2\right] du\, ds \le$$

$$\le \frac{c_{13}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^s \mathbb{E}\left[\left(H^{(i,j)}(u)\right)^2\right] du\, ds \le c_{14}\Delta_n T.$$

Using the previous two bounds, there is a constant c_{16} such that

$$\left|\frac{48c_{10}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^n \sum_{r=1}^k \mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^s \quad H^{(i,j)}(u) J_r^{(i,j)}(u)\cdot\right.$$

$$\left.\cdot \left\langle \left(J^{(i,j)}(u)\right)^T \nu(X_u) \,\Big|\, e_r^T R^{(j)}(u,\theta) \right\rangle du\, ds\Big]\right| \le$$

$$\le \frac{c_{15}}{\Delta_n^2 \epsilon^2} \sum_{i=1}^n \sum_{r=1}^k \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^s \mathbb{E}\Big[\left(J_r^{(i,j)}(u)\right)^2 \left\|\left(J^{(i,j)}(u)\right)^T \nu(X_u)\right\|_2^2\Big]^{\frac{1}{2}}\cdot$$

$$\cdot \mathbb{E}\Big[\left(H^{(i,j)}\right)^2 \left\|e_r^T R^{(j)}(u,\theta)\right\|_2^2\Big]^{\frac{1}{2}} du\, ds \le$$

$$\le c_{16}\Delta_n T.$$

Hence, expression in (5.20) is bounded above by $c_{17}\Delta_n T$ so it converges to zero when $n \to +\infty$.

Lastly, we show that assumption (ii) is fulfilled. For arbitrary and fixed j, l such that $1 \le j \le l \le d$ we consider

$$\sum_{i=1}^{A_n^t} \left(\mathbb{E}\left[\chi_i^{n,j}\chi_i^{n,l}|\mathcal{F}_{n,i-1}\right] - \mathbb{E}\left[\chi_i^{n,j}|\mathcal{F}_{n,i-1}\right] \cdot \mathbb{E}\left[\chi_i^{n,l}|\mathcal{F}_{n,i-1}\right]\right) =$$

$$= \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^k \mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \left\langle J^{(i,j)}(s) \,|\, \nu(X_s)e_p \right\rangle dW_s^p\cdot$$

$$\cdot \int_{t_{i-1}}^{t_i} \left\langle J^{(i,l)}(s) \,|\, \nu(X_s)e_r \right\rangle dW_s^r |\mathcal{F}_{n,i-1}\Big].$$

Using the product formula and independence of the components of Brownian motion, we have

$$\sum_{i=1}^{A_n^t} \left(\mathbb{E}\left[\chi_i^{n,j}\chi_i^{n,l}|\mathcal{F}_{n,i-1}\right] - \mathbb{E}\left[\chi_i^{n,j}|\mathcal{F}_{n,i-1}\right] \cdot \mathbb{E}\left[\chi_i^{n,l}|\mathcal{F}_{n,i-1}\right]\right) =$$

(5.22)

$$= \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p=1}^k \mathbb{E}\left[\int_{t_{i-1}}^{t_i} \left\langle J^{(i,j)}(s) \,|\, \nu(X_s)e_p \right\rangle \cdot \left\langle J^{(i,l)}(s) \,|\, \nu(X_s)e_p \right\rangle ds |\mathcal{F}_{n,i-1}\right].$$

Using simple relations of linear algebra it results that (5.22) is equal to

$$\frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) J_r^{(i,j)}(s) J_p^{(i,l)}(s)\, ds \big| \mathcal{F}_{n,i-1} \right].$$

Again, the main tool is Itô formula and independence of the components of Brownian motion.

$$\frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) J_r^{(i,j)}(s) J_p^{(i,l)}(s)\, ds \big| \mathcal{F}_{n,i-1} \right] =$$

(5.23)

$$= \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} J_r^{(i,j)}(u) e_p^T R^{(l)}(u,\theta)\, dW_u\, ds \big| \mathcal{F}_{n,i-1} \right] +$$

(5.24)

$$+ \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} J_p^{(i,l)}(u) e_r^T R^{(j)}(u,\theta)\, dW_u\, ds \big| \mathcal{F}_{n,i-1} \right] +$$

(5.25)

$$+ \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} \left\langle e_r^T R^{(j)}(u,\theta) | e_p^T R^{(l)}(u,\theta) \right\rangle du\, ds \big| \mathcal{F}_{n,i-1} \right]$$

Since the matrix S is symmetric, it is enough to consider one of the terms (5.23) and (5.24). Let $\tilde{h}_{r,p}^{(i,j,l)}(u,\theta) := J_r^{(i,j)}(u) e_p^T R^{(l)}(u,\theta)$. Using Jensen's inequality for conditional expectation, Cauchy-Schwarz inequality and some simple inequalities, and Lemma 2.2, it can be shown that there is a constant c_{18} such that

$$\mathbb{E}\left[\left(\frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} \tilde{h}_{r,p}^{(i,j,l)}(u,\theta)\, dW_u\, ds \big| \mathcal{F}_{n,i-1} \right] \right)^2 \right] =$$

$$= \frac{1}{\Delta_n^2} \mathbb{E}\left[\sum_{i=1}^{A_n^t} \left(\mathbb{E}\left[\sum_{p,r=1}^{k} \int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} \tilde{h}_{r,p}^{(i,j,l)}(u,\theta)\, dW_u\, ds \big| \mathcal{F}_{n,i-1} \right] \right)^2 \right] +$$

$$+ \frac{1}{\Delta_n^2} \mathbb{E}\Big[\sum_{i,m=1,i\neq m}^{A_n^t} \left(\sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} \tilde{h}_{r,p}^{(i,j,l)}(u,\theta)\, dW u\, ds \big| \mathcal{F}_{n,i-1} \right] \right) \cdot$$

$$\cdot \left(\sum_{p,r=1}^{k} \mathbb{E}\left[\int_{t_{m-1}}^{t_m} S_{pr}(X_s) \int_{t_{m-1}}^{s} \tilde{h}_{r,p}^{(m,j,l)}(u,\theta)\, dW_u\, ds \big| \mathcal{F}_{n,m-1} \right] \right) \Big] \leq$$

$$\leq c_{18}(T\Delta_n + T^2\Delta_n),$$

so (5.23) converges to zero in L^2. Hence, it converges to zero in probability, too.

To analyze (5.25) we introduce the following notation

$$D_t^{(i,j,l)} := \sum_{p,r=1}^{k} \int_{t_{i-1}}^{t} S_{pr}(X_s) \int_{t_{i-1}}^{s} \left\langle e_r^T R^{(j)}(u,\theta) \,|\, e_p^T R^{(l)}(u,\theta) \right\rangle du\, ds.$$

Then, we have that (5.25) equals

$$(5.26) \quad \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \mathbb{E}\left[D_{t_i}^{(i,j,l)} | \mathcal{F}_{n,i-1} \right] = \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \left(\mathbb{E}\left[D_{t_i}^{(i,j,l)} | \mathcal{F}_{n,i-1} \right] - D_{t_i}^{(i,j,l)} \right)$$

$$(5.27) \quad\quad\quad\quad\quad\quad\quad\quad\quad + \frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} D_{t_i}^{(i,j,l)}.$$

It can be shown that (5.26) converges to 0 in L^2 because there exists a constant c_{19} such that

$$\mathbb{E}\left[\left(\frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \left(\mathbb{E}\left[D_{t_i}^{(i,j,l)} | \mathcal{F}_{n,i-1} \right] - D_{t_i}^{(i,j,l)} \right) \right)^2 \right] \le c_{19} T \Delta_n.$$

There is only left to show that (5.27) converges in probability to

$$C_{jl}(t) = \frac{1}{2} \sum_{p,r=1}^{k} \int_0^t S_{pr}(X_s) \left\langle e_r^T R^{(j)}(s,\theta) \,|\, e_p^T R^{(l)}(s,\theta) \right\rangle ds.$$

It is sufficient to show that for arbitrary $p, r = 1, 2, \ldots, k$ the following convergence holds

$$\frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \int_{t_{i-1}}^{t_i} S_{pr}(X_s) \int_{t_{i-1}}^{s} \left\langle e_r^T R^{(j)}(u,\theta) \,|\, e_p^T R^{(l)}(u,\theta) \right\rangle du\, ds \overset{\mathbb{P}}{\to}$$

$$\frac{1}{2} \int_0^t S_{pr}(X_s) \left\langle e_r^T R^{(j)}(s,\theta) \,|\, e_p^T R^{(l)}(s,\theta) \right\rangle ds,$$

when $n \to +\infty$.

Let us denote $m(u,s)(\omega) := S_{pr}(X_s(\omega)) \left\langle e_r^T R^{(j)}(u,\theta) \,|\, e_p^T R^{(l)}(u,\theta) \right\rangle$. For fixed $\omega \in \Omega$, using assumptions (A3), (A4) and compactness of E function $m(u,s)(\omega)$ is bounded and continuous on $[0,T] \times [0,T]$. Hence, there exist $u_i^*(\omega), s_i^*(\omega) \in [t_{i-1}, t_i]$ such that $u_i^*(\omega) \le s_i^*(\omega)$ and

$$\int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} m(u,s)\, du\, ds = \frac{\Delta_n^2}{2} m(u_i^*, s_i^*), \quad i = 1, 2, \ldots, n.$$

We have

$$\frac{1}{\Delta_n} \sum_{i=1}^{A_n^t} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} m(u,s)\, du\, ds - \frac{1}{2} \int_0^t m(s,s)\, ds =$$

(5.28)
$$= \frac{\Delta_n}{2} \sum_{i=1}^{A_n^t} m(u_i^*, u_i^*) - \frac{1}{2} \int_0^t m(s,s)\, ds +$$

$$+ \frac{\Delta_n}{2} \sum_{i=1}^{A_n^t} \left(m(u_i^*, s_i^*) - m(u_i^*, u_i^*) \right).$$

Since function given by $(u,s) \mapsto m(u,s)$ is continuous, there exists $t^* \in \left[t_{A_n^t}, t \right]$ such that $\frac{1}{2} \int_{t_{A_n^t}}^t m(s,s)\, ds = \frac{1}{2} \left(t - t_{A_n^t} \right) m(t^*, t^*)$. Hence,

$$\frac{\Delta_n}{2} \sum_{i=1}^{A_n^t} m(u_i^*, u_i^*) - \frac{1}{2} \int_0^t m(s,s)\, ds =$$

$$= \frac{\Delta_n}{2} \sum_{i=1}^{A_n^t} m(u_i^*, u_i^*) + \frac{1}{2} \left(t - t_{A_n^t} \right) m(t^*, t^*) - \frac{1}{2} \int_0^t m(s,s)\, ds -$$

(5.29)
$$- \frac{1}{2} \int_{t_{A_n^t}}^t m(s,s)\, ds.$$

The first two terms of (5.29) forms Riemann integral sum so they converge almost surely to $\frac{1}{2} \int_0^t m(s,s)\, ds$. Boundedness of function m assures that $\frac{1}{2} \int_{A_n^t}^t m(s,s)\, ds$ converges almost surely to zero so (5.28) converges almost surely to zero.

Let $\epsilon > 0$. The function given by $t \mapsto S_{pr}(X_t)$ is continuous on $[0,T]$. Moreover, it is uniformly continuous so there exists $\delta > 0$ such that for all $s, t \in [0,T]$, $|s - t| < \delta$ implies that $|S_{pr}(X_s) - S_{pr}(X_t)| < \epsilon$. Since Δ_n tends to 0, there exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, $\Delta_n < \delta$ holds. Therefore, for $n \geq n_0$ we have that

$$\left| \frac{\Delta_n}{2} \sum_{i=1}^{A_n^t} \left(m(u_i^*, s_i^*) - m(u_i^*, u_i^*) \right) \right| \leq \epsilon \frac{T}{2}.$$

So (5.27) converges almost surely to $C_{jl}(t)$ so it converges in probability.

Finally, we proved that conditions of Theorem 2.5 are satisfied, hence the conclusion of theorem holds. We denote by π_T the projection function $\pi_T : D([0,T], \mathbb{R}^d) \to \mathbb{R}^d$, defined by $\pi_T((X_t, t \in [0,T])) := X_T$. Then by [6,

Theorem 12.5] it is a continuous function. Since we have the following relation

$$\frac{1}{\sqrt{\Delta_n}} \left(D\ell(\theta) - D\ell_n(\theta)\right) = V_n(\theta) + \pi_T \left(\left(\sum_{i=1}^{A_n^t} \chi_i^n\right)_{1 \le t \le T}\right),$$

the statement of Theorem 4.3 holds in the case of compact set E.

In general, when $E \subseteq \mathbb{R}^k$ is open, there exists a set of open and bounded sets $(E_N)_{N \in \mathbb{N}}$ such that for every $N \in \mathbb{N}$, $Cl(E_N) \subset E_{N+1}$ and $E = \cup_{N=1}^{+\infty} E_N$. Without loss of generality, let $x_0 \in E_1$. There exists a sequence of $C^\infty(E)$-functions $(\phi_N)_{N \in \mathbb{N}}$ such that $\phi_N(x) \in [0,1]$ for all $x \in E$, $\phi_N(x) = 1$ on $Cl(E_N)$ and $\phi_N(x) = 0$, on $Cl(E_{N+1})^C$[7]. Let us define $\mu^{(N)}(x, \theta) := \phi_N(x)\mu(x,\theta)$. Let $\nu^{(N)}$ be continuous functions on E such that $\nu^{(N)}(x) := \nu(x)$ on $Cl(E_N)$ and $\nu^{(N)}(x) = K$ on $E \setminus Cl(E_{N+1})$ where K is a square root of some constant positive definite $k \times k$ matrix. Moreover, for $N \in \mathbb{N}$ we define $T_N := \inf\{t \ge 0 : X_t \in E_N^C\}$. Since X is a continuous process, $(T_N)_{N \in \mathbb{N}}$ is an increasing sequence of stopping times and $T_N \nearrow +\infty$ a.s.
Let $N \in \mathbb{N}$ be fixed. Let process $X^{(N)} = (X_t^N; 0 \le t \le T)$ be a diffusion process given as a strong solution of the following system of stochastic differential equations

$$(5.30) \quad X_t^{(N)} = x_0 + \int_0^t \mu^{(N)}(X_s^{(N)}, \theta_0)\, ds + \int_0^t \nu^{(N)}(X_s^{(N)})\, dW_s, \quad 0 \le t \le T.$$

Under assumptions (A2) and (A4) functions $\mu^{(N)}(\cdot, \theta_0)$ and $\nu^{(N)}(\cdot)$ are bounded on E. They are also Lipschitz continuous. Using [29, Corollary 5.1.2] diffusion process (5.30) exists and it is unique a.s. Let $\frac{1}{\sqrt{\Delta_n}}\left(D\ell^{(N)}(\theta) - D\ell_n^{(N)}(\theta)\right)$ be the term from the statement of the theorem for diffusion (5.30). Then, first part of the proof implies that $\frac{1}{\sqrt{\Delta_n}}\left(D\ell^{(N)}(\theta) - D\ell_n^{(N)}(\theta)\right) \overset{st}{\Rightarrow} Y^{(N)}(\theta)$, $n \to +\infty$ where $Y^{(N)}(\theta) \sim MN(0, \Sigma^{(N)}(\theta))$, and $\Sigma^{(N)}$ is a random matrix $\Sigma(\theta)$ that is applied on process $X^{(N)}$ and functions $\mu^{(N)}$ and $\nu^{(N)}$. Let us denote $V_n(\theta) := \frac{1}{\sqrt{\Delta_n}}\left(D\ell(\theta) - D\ell_n(\theta)\right)$ and $V_n^{(N)}(\theta) := \frac{1}{\sqrt{\Delta_n}}\left(D\ell^{(N)}(\theta) - D\ell_n^{(N)}(\theta)\right)$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded and continuous function, and let U be arbitrary bounded $\mathcal{F}_T$-measurable random variable. For almost all $\omega \in \Omega$ and $t \in \left[0, T^{(N)}\right]$, processes $X_t(\omega)$ and $X_t^{(N)}(\omega)$ are equal so we have:

$$\left| \mathbb{E}\left[f(V_n(\theta))U\right] - \tilde{\mathbb{E}}\left[f(Y(\theta))U\right] \right| \le$$

$$\left| \mathbb{E}\left[f(V_n^{(N)}(\theta))U \mathbb{1}_{\{T^{(N)} > T\}}\right] - \tilde{\mathbb{E}}\left[f(Y^{(N)}(\theta))U \mathbb{1}_{\{T^{(N)} > T\}}\right] \right| +$$

$$(5.31) \quad \left| \mathbb{E}\left[f(V_n(\theta))U \mathbb{1}_{\{T^{(N)} \le T\}}\right] - \tilde{\mathbb{E}}\left[f(Y(\theta))U \mathbb{1}_{\{T^{(N)} \le T\}}\right] \right|.$$

Using boundedness (5.31) is bounded by $B \mathbb{P}\left(T^{(N)} \leq T\right)$ where B is some positive constant. Using first part of the proof we have that

$$\overline{\lim_n} \left| \mathbb{E}\left[f(V_n(\theta))U\right] - \tilde{\mathbb{E}}\left[f(Y(\theta))U\right] \right| \leq B \mathbb{P}\left(T^{(N)} \leq T\right).$$

Letting $N \to +\infty$ we have

$$\overline{\lim_n} \left| \mathbb{E}\left[f(V_n(\theta))U\right] - \tilde{\mathbb{E}}\left[f(Y(\theta))U\right] \right| = 0$$

that implies the statement of the theorem. □

PROOF OF THEOREM 4.4. The idea of the proof is similar to the case when $k = 1$ (see [26, Theorem 5.4]). We denote by $Z_n(\theta) \equiv Z_n(\omega, \theta) := \frac{1}{\sqrt{\Delta_n}}\left(D\ell(\theta) - D\ell_n(\theta)\right)$. Since we assumed that certain functions are smooth enough, using [16, Lemma 4.1], we have that function $(\omega, \theta) \mapsto Z_n(\omega, \theta)$ is $\mathcal{F}_T \otimes \mathcal{B}(\Theta)$, for all $n \in \mathbb{N}$. Since MLE is a $\mathcal{F}_T$-measurable random vector [16], functions $\omega \mapsto Z_n(\hat{\theta}) \equiv Z_n(\omega, \hat{\theta}(\omega))$ are $\mathcal{F}_T$-measurable.

Let $v \in \mathbb{R}^d$ be an arbitrary fixed vector and let U be an arbitrary and almost surely bounded $\mathcal{F}_T$-measurable random variable. Let $B > 0$ be a constant such that $|U| \leq B$ a.s. Using [25, Lemma 4.3], it is sufficient to prove that

$$\lim_{n \to +\infty} \left| \mathbb{E}\left[e^{i\langle v \,|\, Z_n(\hat{\theta})\rangle}U\right] - \mathbb{E}\left[e^{-\frac{1}{2}\sum_{p,r=1}^d v_p v_r \Sigma_{pr}(\hat{\theta})}U\right] \right| = 0.$$

For every $n \in \mathbb{N}$ we define functions: $h_{n,1}(\omega, \theta) := \cos\left(\langle v \,|\, Z_n(\omega, \theta)\rangle\right)$, $h_{n,2}(\omega, \theta) := \sin\left(\langle v \,|\, Z_n(\omega, \theta)\rangle\right)$ and $h_{n,3}(\omega, \theta) := e^{-\frac{1}{2}\sum_{p,r=1}^d v_p v_r \Sigma_{pr}(\omega, \theta)}$. For $m = 1, \ldots, d$ we denote by $\partial_m h_{n,l}$ m-th partial derivative of $h_{n,l}$ with respect to θ_m for $l = 1, 2, 3$. Under assumption (A3), for every $\omega \in \Omega$, functions $\theta \mapsto h_{n,1}(\theta) \equiv h_{n,1}(\omega, \theta)$, $\theta \mapsto h_{n,2}(\theta) \equiv h_{n,2}(\omega, \theta)$ and $\theta \mapsto h_{n,3} \equiv h_{n,3}(\omega, \theta)$ are of class $C^1(\Theta)$. Using mean value theorem, we obtain

$$(5.32) \qquad |h_{n,l}(\theta_1) - h_{n,l}(\theta_2)| \leq \left(\sum_{m=1}^d \sup_{\theta \in \Theta} |\partial_m h_{n,l}(\theta)|\right) \|\theta_2 - \theta_1\|_2.$$

It is easy to see that in order to bound $\partial_m h_{n,l}$, we have to bound $\partial_m Z_{n,j}$ (for $h_{n,1}$ and $h_{n,2}$) and $\partial_m \Sigma_{bc}$ (for $h_{n,3}$). After a tedious calculation, we have

$$\partial_m Z_{n,j}(\theta) = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left\langle S^{-1}(X_s)\partial_m\partial_j\mu(X_s, \theta) - \right.$$

$$(5.33) \qquad \left. - S^{-1}(X_{i-1})\partial_m\partial_j\mu(X_{i-1}, \theta) \,\middle|\, \mu(X_s, \theta_0)\right\rangle ds -$$

$$- \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left(\left\langle \partial_j\mu(X_s, \theta) \,\middle|\, S^{-1}(X_s)\partial_m\mu(X_s, \theta)\right\rangle - \right.$$

$$(5.34) \qquad \left. - \left\langle \partial_j\mu(X_{i-1}, \theta) \,\middle|\, S^{-1}(X_{i-1})\partial_m\mu(X_{i-1}, \theta)\right\rangle\right) ds -$$

$$- \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \Big(\big\langle \partial_m \partial_j \mu(X_s, \theta) \,\big|\, S^{-1}(X_s)\mu(X_s, \theta) \big\rangle -$$

$$(5.35) \qquad - \big\langle \partial_m \partial_j \mu(X_{i-1}, \theta) \,\big|\, S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) \big\rangle \Big) \, ds +$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \big\langle S^{-1}(X_s)\partial_m \partial_j \mu(X_s, \theta) -$$

$$(5.36) \qquad - S^{-1}(X_{i-1})\partial_m \partial_j \mu(X_s, \theta) \,\big|\, \nu(X_s) \, dW_s \big\rangle.$$

Under assumption (A3), there exist $\partial_m \nabla g_j(x)$ for every $m, j = 1, \ldots, d$ so there exists

$$\partial_m \Sigma_{bc}(\theta) = \frac{1}{2} \int_0^T \sum_{p,r=1}^{k} S_{pr}(X_s) \partial_m \big(\big\langle e_r^T \nabla g_b(X_s)\nu(X_s) \,\big|$$

$$(5.37) \qquad\qquad\qquad\qquad e_p^T \nabla g_c(X_s)\nu(X_s) \big\rangle \big) \, ds.$$

Assume for the moment that E is a compact set, so all partial derivatives that appear in lines (5.33)-(5.37) are bounded functions on E.

Using (5.13) in Theorem 5.6, there exists a constant $K_{1,j,m} > 0$ such that

$$\mathbb{E}\Big[\sup_{\theta \in \Theta} \big| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \big\langle S^{-1}(X_s)\partial_m \partial_j \mu(X_s, \theta) -$$

$$- S^{-1}(X_{i-1})\partial_m \partial_j \mu(X_{i-1}, \theta) \,\big|\, \mu(X_s, \theta_0) \big\rangle \, ds \big| \Big] \leq K_{1,j,m}.$$

Then, let us denote $\tilde{f}(x, \theta) := \big\langle \partial_j \mu(x, \theta) \,\big|\, S^{-1}(x)\partial_m \mu(x, \theta) \big\rangle$.

We construct a vector function $f = [\tilde{f}, \cdots, \tilde{f}]^T$ and a vector function $a(x) \equiv [\frac{1}{k}, \cdots, \frac{1}{k}]^T$. Using (5.13) in Theorem 5.6, there exists a constant $K_{2,j,m} > 0$ such that

$$\mathbb{E}\Big[\sup_{\theta \in \Theta} \big| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \Big(\big\langle \partial_j \mu(X_s, \theta) \,\big|\, S^{-1}(X_s)\partial_m \mu(X_s, \theta) \big\rangle -$$

$$- \big\langle \partial_j \mu(X_{i-1}, \theta) \,\big|\, S^{-1}(X_{i-1})\partial_m \mu(X_{i-1}, \theta) \big\rangle \Big) \, ds \big| \Big] \leq K_{2,j,m}.$$

In the same manner, we conclude that there exists a constant $K_{3,j,m} > 0$ such that

$$\mathbb{E}\Big[\sup_{\theta \in \Theta} \big| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \Big(\big\langle \partial_m \partial_j \mu(X_s, \theta) \,\big|\, S^{-1}(X_s)\mu(X_s, \theta) \big\rangle -$$

$$- \big\langle \partial_m \partial_j \mu(X_{i-1}, \theta) \,\big|\, S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) \big\rangle \Big) \, ds \big| \leq K_{3,j,m}.$$

Using (5.14) in Theorem 5.6, there exists a constant $K_{4,j,m} > 0$ such that

$$\mathbb{E}\Big[\sup_{\theta \in \Theta} \big| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \big\langle S^{-1}(X_s)\partial_m \partial_j \mu(X_s, \theta) -$$

$$- S^{-1}(X_{i-1})\partial_m\partial_j\mu(X_s,\theta)\,\big|\,\nu(X_s)\,dW_s\big\rangle\big|\Big] \le K_{4,j,m}.$$

Hence, for every $j,m = 1,2,\ldots,d$ we have that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\partial_m Z_{n,j}(\theta)\right] \le K_{1,j,m} + K_{2,j,m} + K_{3,j,m} =: K_{j,m}.$$

Moreover, there exists a constant $L_{m,b,c} > 0$ such that

$$\mathbb{E}\left[\sup_{\theta\in\Theta}\partial_m\Sigma_{bc}(\theta)\right] \le L_{m,b,c}.$$

For $l = 1,2,3$ let us denote $H_{n,l} := \sum_{m=1}^{d}\sup_{\theta\in\Theta}|\partial_m h_{n,l}(\theta)|$ and $H_n := H_{n,1} + H_{n,2} + H_{n,3}$. Let us also denote $F_n := h_{n,1} + ih_{n,2} - h_{n,3}$. Using (5.32), we can easily see that for every $n\in\mathbb{N}$ we have

$$\begin{aligned}
|F_n(\theta_1) - F_n(\theta_2)| &\le |h_{n,1}(\theta_1) - h_{n,1}(\theta_2)| + |h_{n,2}(\theta_1) - h_{n,2}(\theta_2)| +\\
&\quad + |h_{n,3}(\theta_1) - h_{n,3}(\theta_2)| \le\\
&\le (H_{n,1} + H_{n,2} + H_{n,3})\,\|\theta_1 - \theta_2\|_2 =\\
&= H_n\,\|\theta_1 - \theta_2\|_2\,,
\end{aligned}$$

where

$$\mathbb{E}\left[H_n\right] \le 2\,\|v\|_2\sum_{m=1}^{d}\sum_{j=1}^{d}K_{j,m} + \frac{1}{2}\,\|v\|_2^2\sum_{m=1}^{d}\sum_{b,c=1}^{d}L_{m,b,c} := K$$

and constant K does not depend on n.

Using previously introduced notation, it is sufficient to prove that

$$\lim_{n\to+\infty}\left|\mathbb{E}\left[F_n(\hat\theta)U\right]\right| = 0.$$

Let $\varepsilon > 0$ be arbitrary and fixed. We define $\delta := \frac{\varepsilon}{2KB}$. Since $Cl(\Theta)$ is compact set, there exists its finite cover, i.e. there exist finitely many balls $K(\theta_l,\delta)$, $l = 1,\ldots,N$, such that $\theta_l\in\Theta$ and $\Theta\subseteq Cl(\Theta)\subseteq\cup_{l=1}^{N}K(\theta_l,\delta)$. We define a finite partition $\{K_1,\ldots,K_N\}$ of set Θ in the following way:

$$\begin{aligned}
K_1 &:= K(\theta_1,\delta)\cap\Theta,\\
K_2 &:= K(\theta_2,\delta)\cap\Theta\cap K_1^{C},\\
&\;\;\vdots\\
K_N &:= K(\theta_N,\delta)\cap\Theta\cap K_1^{C}\cdots\cap K_{N-1}^{C}.
\end{aligned}$$

$$(5.38)\qquad \left|\mathbb{E}\left[F_n(\hat\theta)U\right]\right| \le \left|\sum_{l=1}^{N}\mathbb{E}\left[(F_n(\hat\theta) - F_n(\theta_l))U\mathbb{1}_{\{\hat\theta\in K_l\}}\right]\right| +$$

$$+ \left|\sum_{l=1}^{N}\mathbb{E}\left[F_n(\theta_l)U\mathbb{1}_{\{\hat\theta\in K_l\}}\right]\right|$$

On the event $\{\hat{\theta} \in K_l\}$ holds $\left\|\hat{\theta} - \theta_l\right\|_2 < \delta$ so for the first summand in (5.38) we have that

$$\left| \sum_{l=1}^{N} \mathbb{E}\left[(F_n(\hat{\theta}) - F_n(\theta_l))U \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] \right| \leq \sum_{l=1}^{N} \mathbb{E}\left[|F_n(\hat{\theta}) - F_n(\theta_l)||U| \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] \leq$$

$$\leq B \sum_{l=1}^{N} \mathbb{E}\left[H_n \left\|\hat{\theta} - \theta_l\right\|_2 \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] <$$

$$< B\delta \sum_{l=1}^{N} \mathbb{E}\left[H_n \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] = BK\delta = \frac{\varepsilon}{2}.$$

Since $\hat{\theta}$ is a $\mathcal{F}_T$-measurable random vector, $U \mathbb{1}_{\{\hat{\theta} \in K_l\}}$ is bounded a $\mathcal{F}_T$-measurable random variable for every $l = 1, \ldots, N$. For the second summand in (5.38) Theorem 4.3 and [25, Lemma 4.3] imply

$$\lim_{n \to +\infty} \mathbb{E}\left[F_n(\theta_l)U \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] = 0.$$

Moreover,

$$\lim_{n \to +\infty} \left| \sum_{l=1}^{N} \mathbb{E}\left[F_n(\theta_l)U \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] \right| = 0$$

holds.

Finally, now we may choose $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \geq n_0$

$$\left| \sum_{l=1}^{N} \mathbb{E}\left[F_n(\theta_l)U \mathbb{1}_{\{\hat{\theta} \in K_l\}} \right] \right| < \frac{\varepsilon}{2}.$$

For all $n \geq n_0$ it follows that

$$\left| \mathbb{E}\left[F_n(\hat{\theta})U \right] \right| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The statement of the theorem holds when E is compact. Using the same construction with stopping times $(T_N)_{N \in \mathbb{N}}$ as in Theorem 4.3, we achieve the general statement. $\qquad\square$

PROOF OF LEMMA 4.5. Without loss of generality, we prove the assertion for $r = 0$. The difference between ℓ_n and ℓ equals

$$
\ell_n(\theta) - \ell(\theta) = \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \big\langle S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) -
$$

(5.39)
$$
- S^{-1}(X_s)\mu(X_s, \theta) \,\big|\, dX_s \big\rangle -
$$

$$
- \frac{1}{2} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \Big(\big\langle \mu(X_{i-1}, \theta) \,\big|\, S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) \big\rangle -
$$

(5.40)
$$
- \big\langle \mu(X_s, \theta) \,\big|\, S^{-1}(X_s)\mu(X_s, \theta) \big\rangle \Big) \, ds.
$$

Let us assume for the moment that E is compact set. Relations (5.13) and (5.14) in Theorem 5.6 and (1.1) imply that for (5.39) there exists a constant $C > 0$ such that

$$
\left\| \sup_{\theta \in \Theta} \Big| \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} \big\langle S^{-1}(X_{i-1})\mu(X_{i-1}, \theta) - S^{-1}(X_s)\mu(X_s, \theta) \,\big|\, dX_s \big\rangle \Big| \right\|_{L^2}
$$

is bounded above by $C\sqrt{\Delta_n}$. Constructing a vector function in the same way as in the proof of Theorem 4.4, we may use Theorem 5.6 again for (5.40). Finally, there exists a constant C_X (which depends on process X) such that

$$
\left\| \sup_{\theta \in \Theta} |\ell_n(\theta) - \ell(\theta)| \right\|_{L^2} \leq C_X \sqrt{\Delta_n}.
$$

In the case when E is an open set or functions that appear in calculations are not bounded, the same construction from proof of Theorem 4.3 has been imposed. For an open set E there exists a sequence of open and bounded sets $(E_N)_{N \in \mathbb{N}}$ such that for all $N \in \mathbb{N}$, $Cl(E_N) \subset E_{N+1}$, $E = \cup_{N=1}^{+\infty} E_N$, and $x_0 \in E_1$. Also, there exists a sequence of C^∞-functions $(\phi_N)_{N \in \mathbb{N}}$ such that $\phi_N(x) \in [0, 1]$ for all $x \in E$, and $\phi_N(x) = 1$ on $Cl(E_N)$ and $\phi_N(x) = 0$ on $Cl(E_{N+1})^C$.

As before, we define functions $\mu^{(N)}(x, \theta) := \phi_N(x)\mu(x, \theta)$ and $\nu^{(N)}(x) := \nu(x)$ on $Cl(E_N)$ and $\nu^{(N)}(x) = K$ on $E \backslash Cl(E_{N+1})$ where K is a square root of some constant positive definite $k \times k$ matrix. Also, we define a sequence of stopping times $(T_N)_{N \in \mathbb{N}}$ and observe the diffusion process $X^{(N)}$ that is connected to T_N.

By the same reasoning as in the proof of Theorem 4.3, the diffusion process (5.30) exists and it is unique a.s. Let us denote the diffusion matrix of $X^{(N)}$ by $S^{(N)}(x) := \nu^{(N)}(x)\nu^{(N)}(x)^T$. Under assumptions (A2-4) function $(x, \theta) \mapsto (S^{(N)})^{-1}(x)\mu^{(N)}(x, \theta)$ satisfies the assumptions of Theorem 5.6. Using first part of this proof, it follows that

$$
\left\| \sup_{\theta \in \Theta} |\ell_n^{(N)}(\theta) - \ell^{(N)}(\theta)| \right\|_{L^2} \leq C_{X^{(N)}} \sqrt{\Delta_n},
$$

where constant $C_{X^{(N)}}$ depends on process $\left(X_t^{(N)}\right)_{0 \leq t \leq T}$.

Let $P_{x_0}^{(N)}$ be a distribution of solution (5.30), and P_{x_0} a distribution of solution (1.1). It follows that $P_{x_0}^{(N)}(\cdot) = P_{x_0}(\cdot)$ on σ-algebra $\sigma\left(X_{s \wedge T_N} : s \geq 0\right)$ [29, Corollary 10.1.2]. Hence,

$$\mathbb{P}\left(\sup_{\theta \in \Theta} |\ell_n(\theta) - \ell(\theta)| > A\sqrt{\Delta_n}\right) \leq$$

$$\leq \mathbb{P}\left(\{\sup_{\theta \in \Theta} |\ell_n(\theta) - \ell(\theta)| > A\sqrt{\Delta_n}, T_N \leq T\}\right) +$$

$$+ \mathbb{P}\left(\sup_{\theta \in \Theta} |\ell_n^{(N)}(\theta) - \ell^{(N)}(\theta)| > A\sqrt{\Delta_n}\right) \leq$$

$$\leq \mathbb{P}\left(T_N \leq T\right) + \frac{1}{A\sqrt{\Delta_n}}\mathbb{E}\left[|\sup_{\theta \in \Theta} |\ell_n^{(N)}(\theta) - \ell^{(N)}(\theta)|\right] \leq$$

$$\leq \mathbb{P}\left(T_N \leq T\right) + \frac{1}{A\sqrt{\Delta_n}}\left\|\sup_{\theta \in \Theta} |\ell_n^{(N)}(\theta) - \ell^{(N)}(\theta)|\right\|_{L^2} \leq$$

$$= \mathbb{P}\left(T_N \leq T\right) + \frac{1}{A}C_{X^{(N)}}.$$

Let first $n \to +\infty$, then $A \to +\infty$ and finally $N \to +\infty$. We conclude that

$$\lim_{A \to +\infty} \lim_{n \to +\infty} \mathbb{P}\left(\sup_{\theta \in \Theta} \frac{|\ell_n(\theta) - \ell(\theta)|}{\sqrt{\Delta_n}} > A\right) = 0.$$

$\square$

LEMMA 5.7. *Let Θ be a convex and relatively compact set. Assume (A1-6). Then*

$$\int_0^1 D^2\ell_n(\hat{\theta} + (\overline{\theta}_n - \hat{\theta})z)\, dz \cdot \frac{1}{\sqrt{\Delta_n}}\left(\overline{\theta}_n - \hat{\theta}\right) \overset{st}{\Rightarrow} Y(\hat{\theta}),$$

where $Y(\hat{\theta}) \sim MN(0, \Sigma(\hat{\theta}))$.

LEMMA 5.8. *Let Θ be a convex and relatively compact set. Assume (A1-6). Then*

$$\sup_{z \in [0,1]} \left\|D^2\ell(\hat{\theta}) - D^2\ell(\hat{\theta} + z(\overline{\theta}_n - \hat{\theta}))\right\|_F \overset{\mathbb{P}}{\to} 0.$$

LEMMA 5.9. *Let Θ be a convex and relatively compact set. Assume (A1-6). Then*

$$D^2\ell(\hat{\theta})\frac{1}{\sqrt{\Delta_n}}\left(\overline{\theta}_n - \hat{\theta}\right) \overset{st}{\Rightarrow} MN(0, \Sigma(\hat{\theta})).$$

The last three lemmas are the same as in the case of a one-dimensional diffusion [26] and so are their proofs.

PROOF OF THEOREM 4.7. Considering Theorem 4.3, Theorem 4.4 and Theorem 4.6 and using the equality (4.3), the conclusion of this theorem follows easily from the last three lemmas. Since $D^2\ell(\hat{\theta})$ is a negatively definite matrix (A6), it is a regular matrix, so its inverse exists. $\left(D^2\ell(\hat{\theta})\right)^{-1}$ is also a symmetric matrix. Finally, the desired result follows from Lemma 5.9. □

6. EXAMPLE AND SIMULATIONS

The stochastic model that we used for simulations is the Heston model analyzed in [4]. The model is used for the analysis of financial data and is given by

$$(6.41) \qquad dY_t = (a - bY_t)\, dt + \sigma_1 \sqrt{Y_t}\, dW_t^1$$

$$dX_t = (\alpha - \beta Y_t)\, dt + \sigma_2 \sqrt{Y_t} \left(\rho\, dW_t^1 + \sqrt{1 - \rho^2}\, dW_t^2\right),$$

where $a > 0$, $b, \alpha, \beta \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, $\rho \in \langle -1, 1 \rangle$ and $\left(W_t^1, W_t^2\right)_{t \geq 0}$ is a two-dimensional standard Wiener process. Although this model does not satisfy the assumption of uniform ellipticity, it is shown in [4, Proposition 2.1] that there exists a pathwise unique strong solution of (6.41) for $t \geq 0$, and that the log-likelihood function can be written using results from [24, Section 7, p. 296], as was done in Section 3 of [4]. For simulation purposes, a version of Theorem 4.7 is used in which $\Sigma(\hat{\theta})$ and $D^2\ell(\hat{\theta})$ are replaced by $\Sigma_n(\overline{\theta}_n)$ and $D^2\ell_n(\overline{\theta}_n)$, respectively. Matrix $\Sigma_n(\overline{\theta}_n)$ whose elements are given by

$$\Sigma_n(\theta)_{jl} = \frac{\Delta_n}{2} \sum_{i=1}^n \sum_{p,r=1}^k S_{pr}(X_{i-1}) \langle e_r^T \nabla g_j(X_{i-1}) \nu(X_{i-1}) \,|$$

$$e_p^T \nabla g_l(X_{i-1}) \nu(X_{i-1}) \rangle$$

can be understood as the discretized version of matrix $\Sigma(\theta)$. Since the drift function in (6.41) is linear in θ, matrices $\Sigma(\theta)$ and $\Sigma_n(\theta)$ do not depend on θ. The proof of such a version of Theorem 4.7 is similar to the proofs of Lemma 5.9 and 5.11 in [26].
MLE $\hat{\theta}$ of the process in (6.41) is explicitly given by

$$(6.42) \qquad \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \frac{1}{\int_0^T Y_s\, ds \int_0^T \frac{ds}{Y_s} - T^2} \begin{bmatrix} \int_0^T Y_s\, ds \int_0^T \frac{dY_s}{Y_s} - T(Y_T - y_0) \\ T \int_0^T \frac{dY_s}{Y_s} - (Y_T - y_0) \int_0^T \frac{ds}{Y_s} \\ \int_0^T Y_s\, ds \int_0^T \frac{dX_s}{Y_s} - T(X_T - x_0) \\ T \int_0^T \frac{dX_s}{Y_s} - (X_T - x_0) \int_0^T \frac{ds}{Y_s} \end{bmatrix}$$

if $\int_0^T Y_s\, ds \int_0^T ds/Y_s > T^2$. This condition is satisfied if the parameters of the model are such that $a \in \left[\sigma_1^2/2, +\infty\right\rangle$, $b \in \mathbb{R}$, $\sigma_1 > 0$ and $y_0 > 0$. Moreover, MLE in (6.42) is unique when α, β and x_0 are real numbers, $\sigma_2 > 0$ and $\rho \in \langle -1, 1 \rangle$ [4].

We define the vector function $\mu(x, y, a, b, \alpha, \beta)$ and the matrix function $\nu(x, y)$ as

$$\mu(x, y, a, b, \alpha, \beta) = \begin{bmatrix} a - by \\ \alpha - \beta y \end{bmatrix}, \qquad \nu(x, y) = \begin{bmatrix} \sigma_1 \sqrt{y} & 0 \\ \sigma_2 \rho \sqrt{y} & \sigma_2 \sqrt{(1 - \rho^2) y} \end{bmatrix}.$$

As explained in Section 3, we simulate M realizations of discrete random sample $(Y_{t_i}, X_{t_i})_i$, for $i = 1, 2, \ldots, N$ and $N = 2^l$, with parameters $a = 2$, $b = -0.8$, $\alpha = 0.02$, $\beta = 2$, $\sigma_1 = 0.7$, $\sigma_2 = 0.6$, $\rho = -0.8$, $x_0 = \ln 100$ and $y_0 = 0.5$. The time interval is $[0, T]$ for $T = 1$ and the subdivision of points is equidistant, $t_i = \frac{T}{N} i$ for $i = 0, \ldots, N$.

Because of the linearity of the function μ in the parameters and the second summand in (3.2) the discrete log-likelihood is a quadratic function in a, b, α and β, thus it has a maximum. Using the abbreviations $X_{t_i} = X_i$ and $Y_{t_i} = Y_i$, the AMLE for the vector of drift parameters $\bar{\theta}_n$ is given by

$$(6.43) \qquad \begin{bmatrix} \bar{a}_n \\ \bar{b}_n \\ \bar{\alpha}_n \\ \bar{\beta}_n \end{bmatrix} = F \cdot \begin{bmatrix} \Delta_n \sum_{i=1}^n Y_{i-1} \sum_{i=1}^n \frac{Y_i - Y_{i-1}}{Y_{i-1}} - T(Y_n - y_0) \\ T \sum_{i=1}^n \frac{Y_i - Y_{i-1}}{Y_{i-1}} - \Delta_n (Y_n - y_0) \sum_{i=1}^n \frac{1}{Y_{i-1}} \\ \Delta_n \sum_{i=1}^n Y_{i-1} \sum_{i=1}^n \frac{X_i - X_{i-1}}{Y_{i-1}} - T(X_n - x_0) \\ T \sum_{i=1}^n \frac{X_i - X_{i-1}}{Y_{i-1}} - \Delta_n (X_n - x_0) \sum_{i=1}^n \frac{1}{Y_{i-1}} \end{bmatrix}$$

where $F = \left(\Delta_n^2 \sum_{i=1}^n Y_{i-1} \sum_{i=1}^n \frac{1}{Y_{i-1}} - T^2 \right)^{-1}$. After long and tedious calculation, we obtain the matrices $\Sigma_n(\bar{\theta}_n)$ and $D^2 l_n(\bar{\theta}_n)$. The formulas are

$$\Sigma_n(\bar{\theta}_n) = \frac{\Delta_n}{2} \sum_{i=1}^n \frac{1}{Y_{i-1}^2} \cdot \begin{bmatrix} \frac{1}{1 - \rho^2} & 0 & \frac{-\sigma_1 \rho}{\sigma_2(1 - \rho^2)} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{-\sigma_1 \rho}{\sigma_2(1 - \rho^2)} & 0 & \frac{\sigma_1^2}{\sigma_2^2(1 - \rho^2)} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$D^2 l_n(\bar{\theta}_n) = G \cdot \begin{bmatrix} \sigma_2^2 & -\sigma_1 \sigma_2 \rho \\ -\sigma_1 \sigma_2 \rho & \sigma_1^2 \end{bmatrix} \otimes \begin{bmatrix} -\Delta_n \sum_{i=1}^n \frac{1}{Y_{i-1}} & T \\ T & -\Delta_n \sum_{i=1}^n Y_{i-1} \end{bmatrix},$$

where $G = 1/(\sigma_1^2 \sigma_2^2 (1 - \rho^2))$ and $\otimes$ denotes the Kronecker product of two matrices [12].

Since MLE can not be calculated using (6.42), we estimate it with the formulas in (6.43). Then we compute AMLE also with (6.43), but with fewer points than for MLE. More precisely, we take a subsample of length $n = 2^k, k < l$, and $\Delta_n = \frac{T}{n}$. Then we determine the percentage of values $\left\| \frac{1}{\sqrt{\Delta_n}} \sqrt{\Sigma_n(\bar{\theta}_n)}^+ D^2 l_n(\bar{\theta}_n) \left(\bar{\theta}_n - \hat{\theta} \right) \right\|_2^2$ that are in the interval $\left[0, \chi_{1-p}^2(r) \right]$ where $\chi_{1-p}^2(r)$ is $(1 - p)$-quantile of χ^2-distribution with r degrees of freedom. The degrees of freedom r correspond to the rank of the covariance matrix of the observed expression in the norm. The matrix $\Sigma_n(\bar{\theta}_n)$ is a symmetric and singular matrix (and $\Sigma(\hat{\theta})$ as well). From the fact that $\rho \in \langle -1, 1 \rangle$, we

conclude that its submatrix of nonzero elements is a strictly positive definite matrix of rank 2. Since symmetric matrices are orthogonally diagonalizable, the eigenvalue decomposition of the submatrix is the same as its singular value decomposition (SVD) [12]. The SVD of the matrix $\Sigma_n(\overline{\theta}_n)$ is then obtained by placing zeros on the diagonal matrix of SVD of the submatrix and adding vectors to the set of its eigenvectors to complete the basis of $\mathbb{R}^4$. Let us denote the resulting SVD by UDU^T. It is now easy to see that the generalized inverse of the square root of $\Sigma_n(\overline{\theta}_n)$ is given by $U\tilde{D}U^T$ where $\tilde{D}$ is such a diagonal matrix that for all $i \leq 2$, $\tilde{d}_{ii} = 1/\sqrt{d_{ii}}$ and for $i > 2$, $\tilde{d}_{ii} = 0$. For this reason, after a short calculation, we conclude that in our case the covariance matrix of $\frac{1}{\sqrt{\Delta_n}}\sqrt{\Sigma_n(\overline{\theta}_n)}^{+} D^2 l_n(\overline{\theta}_n)\left(\overline{\theta}_n - \hat{\theta}\right)$ is an identical matrix of rank $r = 2$.

TABLE 1. $M = 1000$, $p = 0.025$, $l = 12$

k	3	4	5	6	7	8
%	0.675	0.812	0.867	0.942	0.952	0.972

TABLE 2. $M = 1000$, $p = 0.05$, $l = 14$

k	3	4	5	6	7	8	9	10
%	0.589	0.740	0.826	0.884	0.911	0.935	0.944	0.95

TABLE 3. $M = 1000$, $p = 0.05$, $l = 16$

k	3	4	5	6	7	8	9	10	11
%	0.596	0.726	0.831	0.904	0.921	0.931	0.933	0.939	0.955

Tables 1, 2 and 3 show that an increase in k causes an increase in the percentage so that the value, for a given p, approaches $1 - p$.

7. Appendix

Proof of Lemma 4.2. Let $c_1, \ldots, c_7$ be positive constants occurring in the proof. For function $g_j(x) \equiv g_j(x, \theta)$ $((x, \theta) \in E \times \Theta)$ as in (4.8), under (A2) and (A4) we use Lemma 2.1 in (4.5) and (4.7).

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \langle g_j(X_s) - g_j(X_{i-1}) \,|\, \mu(X_s, \theta_0)\rangle \, ds =$$

$$= \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \left(\nabla g_j(X_u)\mu(X_u, \theta_0) + \right.\right.$$

$$(7.44) \qquad + \frac{1}{2}\nabla_2 g_j(X_u)\Big) du \,\Big|\, \mu(X_s,\theta_0)\Big\rangle ds+$$

$$(7.45) \qquad + \frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left\langle\int_{t_{i-1}}^{s}\nabla g_j(X_u)\nu(X_u)\,dW_u \,\Big|\, \mu(X_s,\theta_0)\right\rangle ds.$$

$$\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left\langle g_j(X_s)-g_j(X_{i-1})\,|\,\nu(X_s)\,dW_s\right\rangle =$$

$$= \frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left\langle\int_{t_{i-1}}^{s}\Big(\nabla g_j(X_u)\mu(X_u,\theta_0)+\right.$$

$$(7.46) \qquad + \frac{1}{2}\nabla_2 g_j(X_u)\Big) du \,\Big|\, \nu(X_s)\,dW_s\Big\rangle +$$

$$+ \frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left\langle\int_{t_{i-1}}^{s}\nabla g_j(X_u)\nu(X_u)\,dW_u \,\Big|\, \nu(X_s)\,dW_s\right\rangle$$

Also, under (A3) and (A4) we use Lemma 2.2 in (4.6) for function:

$$(7.47) \qquad f_j(x) \equiv f_j(x,\theta) := \left\langle S^{-1}(x)\mu(x,\theta)\,|\,\partial_j\mu(x,\theta)\right\rangle, \quad (x,\theta)\in E\times\Theta.$$

$$\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left(f_j(X_{i-1})-f_j(X_s)\right) ds =$$

$$= -\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left(f_j(X_s)-f_j(X_{i-1})\right) ds =$$

$$= -\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\left(\langle\nabla f_j(X_u)\,|\,\mu(X_u,\theta_0)\rangle+\right.$$

$$(7.48) \qquad + \frac{1}{2}\operatorname{Tr}\left(S(X_u)\nabla\left(\nabla f_j(X_u)\right)\right)\Big) du\,ds-$$

$$(7.49) \qquad - \frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{s}\left\langle\nabla f_j(X_u)\,|\,\nu(X_u)\,dW_u\right\rangle ds$$

For each $j=1,2,\ldots,d$, we define a component of vector V_n as $V_n^j(\theta) = (7.44)+ +(7.45)+(7.46)+(7.48)+(7.49)$. We will prove that for all $j=1,2,\ldots,d$ V_n^j converges in probability to zero, when $n\to+\infty$. Obviously, then vector $V_n(\theta)$ converges in probability to $\mathbf{0}_d$, when $n\to+\infty$.

For simplicity, we propose some new notation: $G_j(x,\theta) := \nabla g_j(x)\mu(x,\theta)+ \frac{1}{2}\nabla_2 g_j(x)$. Using Cauchy-Schwartz inequality for vectors and boundedness of functions we conclude that for (7.44) there exist constant c_1 such that

$$\left|\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i}\left\langle\int_{t_{i-1}}^{s}G_j(X_u,\theta_0)\,du \,\Big|\, \mu(X_s,\theta_0)\right\rangle ds\right| \le c_1\sqrt{\Delta_n}T.$$

In the same manner, we conclude that for (7.48) there exists constant c_2 such that

$$\left| -\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left(\langle \nabla f_j(X_u) \,|\, \mu(X_u, \theta_0) \rangle + \right.\right.$$

$$\left.\left. + \frac{1}{2} \operatorname{Tr} \left(S(X_u) \nabla \left(\nabla f_j(X_u) \right) \right) \right) du\, ds \right| \le c_2 \sqrt{\Delta_n} T.$$

In the sequel, we analyze (7.45) using Lemma 2.1 for function $x \mapsto \mu(x, \theta_0)$.

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle|\, \mu(X_s, \theta_0) \right\rangle ds =$$

$$(7.50) \qquad = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle|\, \mu(X_{i-1}, \theta_0) \right\rangle ds +$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle| \right.$$

$$(7.51) \qquad \left. \int_{t_{i-1}}^{s} \nabla \mu(X_u, \theta_0) \mu(X_u, \theta_0)\, du \right\rangle ds +$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle| \right.$$

$$(7.52) \qquad \frac{1}{2} \int_{t_{i-1}}^{s} \nabla_2 \mu(X_u, \theta_0)\, du \right\rangle ds +$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle| \right.$$

$$(7.53) \qquad \left. \int_{t_{i-1}}^{s} \nabla \mu(X_u, \theta_0) \nu(X_u)\, dW_u \right\rangle ds$$

For (7.50) and (7.51) + (7.52) it is sufficient to prove that it converges to zero in L^2. Then it converges in probability to zero, too. The L^2 convergence of (7.50) is proved in the sequel.

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle|\, \mu(X_{i-1}, \theta_0) \right\rangle ds \right)^2 \right] =$$

$$= \frac{1}{\Delta_n} \sum_{i=1}^{n} \mathbb{E}\left(\int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle|\, \mu(X_{i-1}, \theta_0) \right\rangle ds \right)^2 +$$

$$+ \frac{2}{\sqrt{\Delta_n}} \sum_{1 \le i < l \le n} \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u)\, dW_u \,\middle|\, \mu(X_{i-1}, \theta_0) \right\rangle ds \right) \cdot \right.$$

$$\cdot \left(\int_{t_{l-1}}^{t_l} \left\langle \int_{t_{l-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \,\middle|\, \mu(X_{l-1}, \theta_0) \right\rangle ds \right) \Bigg]$$

$$= \frac{1}{\Delta_n} \sum_{i=1}^{n} \mathbb{E} \left(\int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \,\middle|\, \mu(X_{i-1}, \theta_0) \right\rangle ds \right)^2 \leq$$

$$\leq \frac{1}{\Delta_n} \sum_{i=1}^{n} \mathbb{E} \left(\int_{t_{i-1}}^{t_i} \left(\left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \,\middle|\, \mu(X_{i-1}, \theta_0) \right\rangle \right)^2 ds \right) \leq$$

$$\leq \frac{1}{\Delta_n} \sum_{i=1}^{n} \mathbb{E} \left(\int_{t_{i-1}}^{t_i} \left\| \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \right\|_2^2 \cdot \| \mu(X_{i-1}, \theta_0) \|_2^2 \, ds \right) \leq$$

$$\leq \frac{c_3}{\Delta_n} \sum_{i=1}^{n} \Delta_n \int_{t_{i-1}}^{t_i} \mathbb{E} \left(\left\| \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \right\|_2^2 \right) =$$

$$= \frac{c_3}{\Delta_n} \sum_{i=1}^{n} \Delta_n \int_{t_{i-1}}^{t_i} \mathbb{E} \left(\int_{t_{i-1}}^{s} \| \nabla g_j(X_u) \nu(X_u) \|_F^2 \, du \right) \leq$$

$$\leq c_4 \Delta_n T$$

Using Doob's maximal inequality for vector martingale [28, Theorem 1.7] and Lemma 2.3 (i) we conclude that there exists constant c_5 such that (7.51) + (7.52) is bounded in the following way

$$\left| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \,\middle|\, \int_{t_{i-1}}^{s} \nabla \mu(X_u, \theta_0) \mu(X_u, \theta_0) \, du \right\rangle ds \right.$$

$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u \,\middle|\, \frac{1}{2} \int_{t_{i-1}}^{s} \nabla_2 \mu(X_u, \theta_0) \, du \right\rangle ds \Bigg| \leq$$

$$\leq c_5 \sqrt{\Delta_n} T.$$

To prove convergence of (7.53) we use Itô formula for the function $F : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$, $F(y, z) = \langle y \,|\, z \rangle$ on $[t_{i-1}, s]$ and vector martingales

$$Y_s := \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) \, dW_u, \quad Z_s := \int_{t_{i-1}}^{s} \nabla \mu(X_u, \theta_0) \nu(X_u) \, dW_u.$$

When we calculate quadratic variation $\langle Y^l, Z^l \rangle_s$, it is crucial to use independence of components of Brownian motion. Then, (7.53) equals

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} \nabla g_j(X_u) \nu(X_u) dW_u \,\middle|\, \int_{t_{i-1}}^{s} \nabla \mu(X_u, \theta_0) \nu(X_u) dW_u \right\rangle ds =$$

(7.54)
$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left\langle \int_{t_{i-1}}^{u} \nabla\mu(X_v, \theta_0)\nu(X_v)\, dW_v \,\middle|\, \nabla g_j(X_u)\nu(X_u)\, dW_u \right\rangle ds +$$

(7.55)
$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \left\langle \int_{t_{i-1}}^{u} \nabla g_j(X_v)\nu(X_v)\, dW_v \,\middle|\, \nabla\mu(X_u, \theta_0)\nu(X_u)\, dW_u \right\rangle ds +$$

(7.56)
$$+ \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{s} \sum_{p,r=1}^{k} ((\nabla g_j(X_u)\nu(X_u)) \circ (\nabla\mu(X_u, \theta_0)\nu(X_u)))_{pr}\, du\, ds.$$

Expressions (7.49), (7.54) and (7.55) we treat in the same manner as (7.50). Because of boundedness of function in (7.56), it is bounded by $c_6\sqrt{\Delta_n}T$. By a similar reasoning as for (7.50) there exists constant c_7 such that

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left\langle \int_{t_{i-1}}^{s} G_j(X_u, \theta_0)\, du \,\middle|\, \nu(X_s)\, dW_s \right\rangle \right)^2 \right] \leq c_7 \Delta_n T.$$

Hence, we can conclude that (7.46) converges in L^2 to zero so it converges also in probability to zero.

In the end, every component of $V_n(\theta)$ converges in probability to zero so we can conclude that the whole vector $V_n(\theta) \xrightarrow{\mathbb{P}} \mathbf{0}_d$, $n \to +\infty$. □

PROOF OF LEMMA 5.2.

$$\mathbb{E}\left[\|X_{s_2} - X_{s_1}\|_2^2 \right] \leq 2\mathbb{E}\left[\left\| \int_{s_1}^{s_2} \mu(X_t, \theta_0)\, dt \right\|_2^2 \right] + 2\mathbb{E}\left[\left\| \int_{s_1}^{s_2} \nu(X_t)\, dW_t \right\|_2^2 \right]$$

For the first integral we use Cauchy-Schwarz inequality for integrals, and for the second one we use Lemma 2.3 (i). Finally, we have

$$\mathbb{E}\left[\|X_{s_2} - X_{s_1}\|_2^2 \right] \leq 2(s_2 - s_1)\mathbb{E}\left[\int_{s_1}^{s_2} \|\mu(X_t, \theta_0)\|_2^2\, dt \right] +$$
$$+ 2\mathbb{E}\left[\int_{s_1}^{s_2} \|\nu(X_t)\|_F^2\, dt \right] \leq$$
$$\leq 2K\left((s_2 - s_1)^2 + (s_2 - s_1) \right).$$

□

PROOF OF LEMMA 5.3. Without loss of generality (Remark 5.1) we can assume that assumption (P2) is fulfilled for function f. We prove (5.11).

When $m = k_1 + \cdots + k_d$ and $\varepsilon_j = \text{sign}(k_j)$, $j = 1, \ldots, d$, we have

$$(1 + |k_1| + |k_2| + \cdots + |k_d|)^m \, \|C_{\mathbf{k}}(x) - C_{\mathbf{k}}(y)\|_2 =$$
$$= (1 + \varepsilon_1 k_1 + \cdots + \varepsilon_d k_d)^m \, \|C_{\mathbf{k}}(x) - C_{\mathbf{k}}(y)\|_2 =$$
$$= \sum_{j_0 + \cdots + j_d = m} \binom{m}{j_0, j_1, \ldots, j_d} \varepsilon_1^{j_1} \cdots \varepsilon_d^{j_d} \mathbf{k}^{\mathbf{j}} \, \|C_{\mathbf{k}}(x) - C_{\mathbf{k}}(y)\|_2 \leq$$
$$(7.57) \qquad \leq (d+1)^m \left\| C_{\mathbf{k}}^{(\mathbf{j})}(x) - C_{\mathbf{k}}^{(\mathbf{j})}(y) \right\|_2.$$

Here we use the multinomial theorem and the property of Fourier coefficients, namely $C_{\mathbf{k}}^{(\mathbf{j})}(x) = i^m \mathbf{k}^{\mathbf{j}} C_{\mathbf{k}}(x)$. Using the definition of Fourier coefficient and Lemma 2.4, inequality (7.57) yields

$$\|C_{\mathbf{k}}(x) - C_{\mathbf{k}}(y)\|_2 \leq$$
$$\leq K_m(\mathbf{k}) \left\| C_{\mathbf{k}}^{(\mathbf{j})}(x) - C_{\mathbf{k}}^{(\mathbf{j})}(y) \right\|_2 \leq$$
$$\leq K_m(\mathbf{k}) \frac{1}{(2\pi)^d} \int_{Cl(\mathcal{K}_0)} \left\| D_{\mathbf{j}}^m f(x, \theta) - D_{\mathbf{j}}^m f(y, \theta) \right\|_2 d\theta \leq$$
$$\leq K_m(\mathbf{k}) \frac{1}{(2\pi)^d} \int_{Cl(\mathcal{K}_0)} \int_0^1 \left\| \nabla_x D_{\mathbf{j}}^m f\left(x + s(y-x), \theta\right) \right\|_F ds \cdot \|x - y\|_2 \, d\theta \leq$$
$$\leq k_1 K_m(\mathbf{k}) \|x - y\|_2.$$

The beginning of proving (5.12) is the same as in (7.57). Afterwards we analyze it in the following way.

$$\|C_{\mathbf{k}}(x)\|_2 \leq \left(\frac{d+1}{1 + |k_1| + \cdots + |k_d|} \right)^m \frac{1}{(2\pi)^d} \int_{Cl(\mathcal{K}_0)} \left\| D_{\mathbf{j}}^m f(x, \theta) \right\|_2 d\theta \leq$$
$$\leq k_2 K_m(\mathbf{k})$$

$\square$

PROOF OF LEMMA 5.4. The proof is the same as one part of proof of Lemma 6.7 in [17]. $\square$

REFERENCES

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

[2] Y. Aït-Sahalia, *Closed-form likelihood expansions for multivariate diffusions,* Ann. Statist. **36(2)**(2008), 906–937.

[3] A. Alfonsi and B. Jourdain and A. Kohatsu-Higa, *Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme,* Electron. J. Probab. **20**(2015), 1–31.

[4] M. Barczy and G. Pap, *Asymptotic properties of maximum-likelihood estimators for Heston models based on continuous time observations,* Statistics **50(2)**(2016), 389–417.

[5] I. V. Basawa and B. L. S. Prakasa Rao, Stochastic Inference for Stochastic Processes, Academic Press, London, 1980.

[6] P. Billingsley, Convergence of probability measures, John Wiley & Sons, New York, 1999.

[7] Y. G. Borisovich and N. M. Bliznyakov and T. N. Fomenko and Y. A. Izrailevich, Introduction to differential and algebraic topology, Springer Science+Business Media, Dordrecht, 1995.

[8] P. J. Brockwell and R. A. Davis, Time series: theory and methods, Springer Science+Business Media, New York, 2006.

[9] J. Chang and S. X. Chen, *On the approximate maximum likelihood estimation for diffusion processes,* Ann. Statist. **39(6)**(2011), 2820–2851.

[10] Q. Clairon and A. Samson, *Optimal control for estimation in partially observed elliptic and hypoelliptic linear stochastic differential equations,* Stat. Inference Stoch. Process. **23(1)**(2020), 105–127.

[11] R. Durrett, Probability: Theory and Examples, Cambridge University Press, New York, 2010.

[12] J. E. Gentle, Matrix algebra, Springer, New York, 2007.

[13] E. Gobet, *Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach,* Bernoulli (2001), 899–912

[14] E. Gobet, *LAN property for ergodic diffusions with discrete observations,* Ann. Inst. Henri Poincaré Probab. Stat. **38(5)**(2002), 711–737.

[15] E. Gobet and R. Munos, *Sensitivity analysis using Itô-Malliavin calculus and martingales, and application to stochastic optimal control,* SIAM J. Control Optim. **43(5)**(2005), 1676–1713.

[16] M. Huzak, *A general theorem on approximate maximum likelihood estimation,* Glas. Mat. Ser. III **36(1)**(2001), 139–153.

[17] M. Huzak, *Estimating a class of diffusions from discrete observations via approximate maximum likelihood method,* Statistics **52(2)**(2018), 239–272.

[18] J. Jacod, *On continuous conditional Gaussian martingales and stable convergence in law,* Sémin. Probab. (1997), 232-246.

[19] J. Jacod and P. Protter, Discretization of processes, Springer Science+Business Media, Berlin-Heidelberg, 2012.

[20] R. Khasminskii, Stochastic stability of differential equations, Springer-Verlag, Berlin-Heidelberg, 2012.

[21] P. E. Kloeden and E. Platen, Numerical Solution of Stochastic Differential Equations, Springer-Verlag, Berlin-Heidelberg, 1999.

[22] H. Lee and G. Trutnau, *Existence and uniqueness of (infinitesimally) invariant measures for second order partial differential operators on Euclidean space,* J. Math. Anal. Appl. **507(1)**(2022), 125778.

[23] C. Li, *Maximum-likelihood estimation for diffusion processes via closed-form density expansions,* Ann. Statist. **41(3)**(2013), 1350–1380.

[24] R. S. Liptser and A. N. Shiryaev, Statistics of random processes: General theory, Springer-Verlag, Berlin-Heidelberg, 2001.

[25] S. Lubura Strunjak, *Local asymptotic properties of approximate maximum likelihood estimator of drift parameters in diffusion model,* Ph.D. thesis, University of Zagreb (in Croatian), 2015.

[26] S. Lubura Strunjak and M. Huzak, *Local asymptotic mixed normality of approximate maximum likelihood estimator of drift parameters in diffusion model,* Glas. Mat. Ser. III **52(2)**(2017), 377–410.

[27] B. Oksendal, Stochastic Differential Equations, Springer-Verlag, Berlin, 2003.

[28] D. Revuz and M. Yor, Continuous Martingales and Brownian Motion, Springer-Verlag, Berlin-Heidelberg, 1999.
[29] D. W. Stroock and S. R. S. Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, Berlin-Heidelberg, 2006.
[30] M. E. Taylor, Partial differential equations I: Basic Theory, Springer Science+Business Media, New York, 2011.
[31] A. W. Van der Vaart, Asymptotic statistics, Cambridge University Press, New York, 1998.
[32] N. Yang and N. Chen and X. Wan, *A new delta expansion for multivariate diffusions via the Itô-Taylor expansion,* J. Econometrics, **209(2)**(2019), 256–288.

M. Huzak
Department of Mathematics
Faculty of Science
University of Zagreb
10 000 Zagreb
Croatia
E-mail: `miljenko.huzak@math.hr`

S. Lubura Strunjak
Department of Mathematics
Faculty of Science
University of Zagreb
10 000 Zagreb
Croatia
E-mail: `snjezana.lubura.strunjak@math.hr`

A. Vlahek Štrok
Faculty of Chemical Engineering and Technology
University of Zagreb
10 000 Zagreb
Croatia
E-mail: `avlahek@fkit.hr`