Uniform concentration inequality for ergodic diffusion processes observed at discrete times✩

L. Galtchouka,*, S. Pergamenshchikovb

aDepartment of Mathematics, Strasbourg University, 7, rue Rene Descartes, 67084, Strasbourg, France
bLaboratoire de Mathématiques Raphael Salem, Avenue de l’Université, BP 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex, France

Received 15 November 2011; received in revised form 9 June 2012; accepted 5 September 2012
Available online 11 September 2012

Abstract

In this paper a concentration inequality is proved for the deviation in the ergodic theorem for diffusion processes in the case of discrete time observations. The proof is based on geometric ergodicity of diffusion processes. We consider as an application the nonparametric pointwise estimation problem of the drift coefficient when the process is observed at discrete times.

© 2012 Elsevier B.V. All rights reserved.

MSC: 60F10

Keywords: Concentration inequality; Ergodic diffusion processes; Geometric ergodicity; Markov chains; Tail distribution; Upper exponential bound

1. Introduction

We consider the process \((y_t)_{t \geq 0}\) governed by the stochastic differential equation

\[ \mathrm{d}y_t = S(y_t) \, \mathrm{d}t + \sigma(y_t) \, \mathrm{d}W_t, \] (1.1)

where \((W_t, \mathcal{F}_t)_{t \geq 0}\) is a standard Wiener process and \(y_0\) is an initial condition.

✩ The second author was partially supported by the RFFI, Grant 09-01-00172-a.
* Corresponding author.
E-mail addresses: leonid.galtchouk0667@orange.fr, leonid.galtchouk@math.unistra.fr (L. Galtchouk), Serge.Pergamenchtchikov@univ-rouen.fr (S. Pergamenshchikov).
Before we state the main result of the paper on the concentration inequality we start with a nonparametric estimation problem for the process (1.1), where this kind of inequality appears. Suppose that the coefficients $S, \sigma$ are unknown and the process $(y_t)$ is observed on the interval $[0, T]$ at discrete times. We consider the pointwise estimation problem for the function $S$ at a fixed point $x_0 \in \mathbb{R}$ (i.e. $S(x_0)$) given the observations of the process (1.1)

$$(y_t)_{1 \leq j \leq N}, \quad 0 \leq t_j \leq T,$$

where $t_j = j\delta, N = \lceil T/\delta \rceil$ and $\delta$ is some positive fixed observation frequency which will be specified later. Usually, for this problem one uses kernel estimators $\hat{S}_N(x_0)$ defined as

$$\hat{S}_N(x_0) = \frac{\sum_{k=1}^{N} \psi_{h,x_0}(y_{t_k}) \Delta y_{t_k}}{\sum_{k=1}^{N} \psi_{h,x_0}(y_{t_k}) \Delta t_k}, \quad \psi_{h,x_0}(y) = \frac{1}{h} \psi \left( \frac{y - x_0}{h} \right),$$

where $\psi(y)$ is a kernel function which is equal to zero for $|y| \geq 2$ and will be specified later, $0 < h < 1$ is a bandwidth, $\Delta y_{t_k} = y_{t_k} - y_{t_k-1}$ and $\Delta t_k = \delta$.

The main difficulty in studying this estimator is that the denominator is a random variable. In particular, to obtain the convergence rate of this estimator one has to study the asymptotic behavior of the denominator; more precisely, one needs to show that

$$\sum_{k=1}^{N} \psi_{h,x_0}(y_{t_k}) \Delta t_k \approx \pi_\vartheta(\psi_{h,x_0}) h T \quad \text{as } T \to \infty,$$

where $\pi_\vartheta(\psi_{h,x_0}) = \int_{\mathbb{R}} \psi_{h,x_0}(y) q_\vartheta(y) \, dy$ (1.4) and $q_\vartheta$ is the invariant density defined in (2.2).

Unfortunately, the ergodic theorem does not permit us to obtain this kind of result because the times $t_k$ and the bandwidth $h$ depend on $T$. Usually, one obtains the desired property through concentration inequalities for the deviation in the ergodic theorem. The deviation is as follows:

$$D_T(\phi) = \sum_{k=1}^{N} \left( \phi(y_{t_k}) - \pi_\vartheta(\phi) \right) \Delta t_k,$$

where $\phi$ is some function which may depend on $T$, for example, $\phi(\cdot) = \psi_{h,x_0}(\cdot)$. The concentration inequality provides the limit behavior of tail probabilities; more precisely, it shows that, for any $\varepsilon > 0$ and for any $m > 0$, uniformly over $\vartheta$,

$$\lim_{T \to \infty} T^m P_\vartheta \left( |D_T(\psi_{h,x_0})| > \varepsilon T \right) = 0,$$

where $P_\vartheta$ is the law of the process $(y_t)_{t \geq 0}$ under the coefficients $\vartheta = (S, \sigma)$. Usually, to get properties of type (1.6) one needs to establish an exponential inequality for the deviation (1.5).

There are a number of papers devoted to concentration inequalities for functions of independent random variables (we refer the reader to [2] and references therein), and for functions of dependent random variables (see [4,5,16]). For Markov chains such inequalities were obtained in [1]. For continuous time Markov processes an exponential concentration
inequality was obtained in [3] (see also references therein). Concentration inequalities for diffusion processes are given in [8,14,18,20]. Some applications of concentration inequalities to statistics are presented in [15].

For statistical applications, we need uniform upper bounds for the tail distribution over functions $\phi$ like the exponential bounds in [8]. We cannot directly apply the method from [8], since that method is based on the continuous time version of the Ito formula. In this paper we apply this approach through simultaneous (over the functions $\vartheta = (S, \sigma)$) geometric ergodicity.

We recall (see [17]) that geometric ergodicity yields a geometric rate in the convergence

$$\lim_{t \to \infty} E_{\vartheta} (g(y_t) | y_0 = x) = \pi_{\vartheta} (g),$$

for any integrable function $g$ and an initial value $x \in \mathbb{R}$. Here $E_{\vartheta}$ means the expectation with respect to the distribution $P_{\vartheta}$. In [10] through the Lyapunov function method it is shown that the processes (1.1) are geometrically ergodic simultaneously over functions $\vartheta = (S, \sigma)$ from the functional class $\Theta$ defined in (2.1).

The main results of the paper are the concentration inequalities for the deviation (1.5) when the function $\phi$ is smooth (Theorems 2.1 and 2.2) and it is an indicator function (Theorem 2.3).

The paper is organized as follows. In Section 2 we formulate the main results. In Section 3 we introduce all necessary parameters. In Section 4 we prove a concentration inequality in the ergodic theorem for the continuous observations of the process (1.1). In Section 5 we announce the simultaneous geometric ergodic property for the processes (1.1).

In Section 6 we prove all main results. The Appendix contains the proofs of some auxiliary results.

2. The main results

We start with the description of the functional class $\Theta$ for functions $\vartheta = (S, \sigma)$ defined in [10]. Let $x_* \geq 1$, $M > 0$ and $L > 1$ be some real numbers. We denote by $\Sigma_{L,M}$ the class of functions $S$ from $C^1(\mathbb{R})$ such that

$$\sup_{|x| \leq x_*} (|S(x)| + |\dot{S}(x)|) \leq M$$

and

$$-L \leq \inf_{|x| \geq x_*} \ddot{S}(x) \leq \sup_{|x| \geq x_*} \ddot{S}(x) \leq -L^{-1}.$$ 

Furthermore, for some fixed numbers $0 < \sigma_{\min} \leq \sigma_{\max} < \infty$, we denote by $V$ the class of functions $\sigma$ from $C^2(\mathbb{R})$ such that

$$\inf_{x \in \mathbb{R}} |\sigma(x)| \geq \sigma_{\min} \quad \text{and} \quad \sup_{x \in \mathbb{R}} \max (|\sigma(x)|, |\dot{\sigma}(x)|, |\ddot{\sigma}(x)|) \leq \sigma_{\max}.$$ 

Finally, we set

$$\Theta = \Sigma_{L,M} \times V.$$ (2.1)

It should be noted (see, for example, [11]) that, for any $\vartheta = (S, \sigma) \in \Theta$, Eq. (1.1) admits a unique strong solution which is an ergodic process with the invariant density $q_{\vartheta}$ defined as

$$q_{\vartheta} (x) = \left( \int_{\mathbb{R}} \sigma^{-2} (z) e^{\tilde{S}(z)} \, dz \right)^{-1} \sigma^{-2} (x) e^{\tilde{S}(x)}.$$ (2.2)
where \( \tilde{S}(x) = 2 \int_0^x S(v) \sigma^{-2}(v) dv \). Taking into account that, for any function \( \theta \) from \( \Theta \),

\[
-(L/\sigma_{\min}^2)[x^2 - 2|x| (x_\ast - M/L) + x_\ast^2] \leq \tilde{S}(x) \leq (\beta_1 + \beta_2 x_\ast)|x| - \beta_2 x^2
\]

and

\[
\int_\mathbb{R} \frac{e^{\tilde{S}(z)}}{\sigma^2(z)} dz \geq \frac{e^{-\beta_1 x_\ast}}{\sigma_{\max}^2} \sqrt{\frac{\pi \sigma_{\min}^2}{L}}
\]

we obtain

\[
\sup_x \sup_{\theta \in \Theta} q_\theta(x) \leq \frac{\sigma_{\max}^2}{\sigma_{\min}^2} \sqrt{\frac{L}{\pi \sigma_{\min}^2}} e^{\beta_1 x_\ast + (\beta_2/4)(x_\ast + \beta_1/2\beta_2)^2} := q^* < \infty,
\]

where \( \beta_1 = 2M/\sigma_{\min}^2 \) and \( \beta_2 = 1/(L\sigma_{\max}^2) \).

Now we describe the functional classes for the kernel functions \( \phi \). First, for any parameters \( \nu_0 > 0 \) and \( \nu_1 > 0 \), we set

\[
\mathcal{V}_{\nu_0, \nu_1} = \{ \phi \in C(\mathbb{R}) : \|\phi\|_1 \leq \nu_0, \|\phi\|_{\infty} \leq \nu_1 \},
\]

where \( \|\phi\|_1 = \int_\mathbb{R} |\phi(y)| dy \) and \( \|\phi\|_{\infty} = \sup_{y \in \mathbb{R}} |\phi(y)| \).

For any function \( \phi \) from \( C^2(\mathbb{R}) \), we denote by \( \mathcal{L}_\theta(\phi) \) the generator of the process (1.1), i.e.

\[
\mathcal{L}_\theta(\phi)(y) = S(y) \dot{\phi}(y) + \frac{\sigma^2(y)}{2} \ddot{\phi}(y).
\]

Using this notation, we set

\[
\mu(\phi) = \sup_{\theta \in \Theta} \|\mathcal{L}_\theta(\phi)\|_{\infty} \quad \text{and} \quad \tilde{\mu}(\phi) = \sup_{\theta \in \Theta} |\tilde{\pi}_\theta(\phi)|,
\]

where \( \tilde{\pi}_\theta(\phi) = \pi_0(\mathcal{L}_\theta(\phi)) \). Now, for any vector \( \nu = (\nu_0, \nu_1, \nu_2, \nu_3, \nu_4) \) from \( \mathbb{R}_+^5 \), we define the functional class

\[
\mathcal{K}_\nu = \left\{ \phi \in \mathcal{V}_{\nu_0, \nu_1} \cap C^2(\mathbb{R}) : \|\dot{\phi}\|_{\infty} \leq \nu_2, \mu(\phi) \leq \nu_3, \tilde{\mu}(\phi) \leq \nu_4 \right\}.
\]

**Theorem 2.1.** For any observation frequency \( \delta, 0 < \delta \leq 1 \), and any vector \( \nu = (\nu_0, \nu_1, \nu_2, \nu_3, \nu_4) \) from \( \mathbb{R}_+^5 \), there exist positive parameters \( z_0 = z_0(\delta, \nu), \gamma = \gamma(\delta, \nu) \) and \( \kappa = \kappa(\delta, \nu) \) such that

\[
\sup_{T \geq 1} \sup_{z \geq z_0} \sup_{\phi \in \mathcal{K}_\nu} \sup_{\theta \in \Theta} e^{\zeta_{\min}(\kappa z, \gamma)} P_{\theta} \left( |D_T(\phi)| \geq z \sqrt{N} \right) \leq 4.
\]

The parameters \( z_0, \gamma \) and \( \kappa \) are defined explicitly in (3.4)–(3.5).

**Remark 2.1.** The proof of this theorem is given in Section 6.1 and it uses the following scheme. We approximate the deviation for discrete time observations by that for continuous time. For this last deviation, the concentration inequality is proved in Proposition 4.1 like in [8], by making use of the Ito formula. Further, the approximation term is bounded thanks to the moment inequality for sums of dependent random variables given in (A.1). In order to estimate the correlations of these random variables we use geometric ergodicity for families of Markov processes (see Section 5).
Let us apply this theorem to the pointwise estimation problem, i.e. to the functions $\psi_{h,x_0}$ defined in (1.3). To this end we assume that the frequency $\delta$ in the observations (1.2) is as follows:

$$\delta = \delta_T = \frac{1}{T l_T},$$

(2.8)

where the function $l_T$ is such that, for any $m > 0$,

$$\lim_{T \to \infty} \frac{l_T}{T^m} = 0 \quad \text{and} \quad \lim_{T \to \infty} \frac{l_T}{\ln T} = +\infty.$$  

(2.9)

Further, let $\epsilon = \epsilon_T$ be a positive function satisfying the following properties:

$$\lim_{T \to \infty} \epsilon_T = 0, \quad \lim_{T \to \infty} \frac{l_T}{T \epsilon_T} = 0 \quad \text{and} \quad \lim_{T \to \infty} \frac{\epsilon_T^5 l_T}{\ln T} = +\infty.$$  

(2.10)

We can take, for example, for some $\iota > 0$,

$$l_T = \ln \left( 1 + \frac{6}{\iota} \right) (T + 1) \quad \text{and} \quad \epsilon_T = \frac{1}{\ln^\iota (T + 1)}.$$  

Theorem 2.2. Let $\Psi$ be a kernel function in (1.3) that is twice continuously differentiable. For any frequency $\delta_T$ from (2.8) with $l_T$ satisfying the properties (2.9)–(2.10), there exist real numbers $z^*_0 = z^*_0(\Psi) > 0$ and $\gamma^* = \gamma^*(\Psi) > 0$ such that

$$\limsup_{T \to \infty} \sup_{a \geq a^*_h \geq \frac{T-1}{2}} \sup_{\vartheta \in \Theta} e^{a \gamma^* l_T} P_{\vartheta} \left( |D_T(\psi_{h,x_0})| \geq a T \right) \leq 4,$$

(2.11)

where $a^*_h = z^*_0/l_T$. The parameters $z^*_0$ and $\gamma^*$ are given explicitly in Section 3.

This theorem implies immediately the following result.

Corollary 2.1. Assume the conditions of Theorem 2.2 hold true. Then, for any $m > 0$,

$$\lim_{T \to \infty} T^m \sup_{a \geq a^*_h \geq \frac{T-1}{2}} \sup_{\vartheta \in \Theta} P_{\vartheta} \left( |D_T(\psi_{h,x_0})| \geq a T \right) = 0.$$  

Now we study the deviation (1.5) for the function

$$\chi_{h,x_0}(y) = \frac{1}{h} \chi \left( \frac{y - x_0}{h} \right),$$

(2.12)

where $\chi(y) = 1_{\{|y| \leq 1\}}$.

Theorem 2.3. Assume the observation frequency $\delta$ has the form (2.8). Then, for any $m > 0$ and any function $\epsilon_T$ satisfying the conditions (2.9) and (2.10),

$$\lim_{T \to \infty} T^m \sup_{h \geq \frac{T-1}{2}} \sup_{\vartheta \in \Theta} P_{\vartheta} \left( |D_T(\chi_{h,x_0})| \geq \epsilon_T T \right) = 0.$$  

(2.13)

Remark 2.2. It is well known that in order to obtain the optimal convergence rate in the estimation problem for a smooth function $S$ in the process (1.1), one has to choose the bandwidth $h$ as

$$h = T^{-1/(2\alpha+1)}$$
with the smoothness parameter $\alpha \geq 1$. This means that, really for the pointwise estimation problem with absolute error risk, $h \geq T^{-1/3}$. But in the case of quadratic risk, one has to choose the parameter $h$ as $h = T^{-1/2}$ (see [6,7,9]).

3. Parameters

In this section we introduce all necessary constants and parameters. First, we set

$$\nu_1 = e^{\beta_1^2/\beta_2} \quad \text{and} \quad \nu_2 = \sqrt{\pi/\beta_2} e^{\beta_1^2/\beta_2},$$

where $\beta_1$ and $\beta_2$ are given in (2.3). Using these parameters we define the following coefficients:

$$r = r(\nu_0) = \frac{2\nu_0}{\sigma_{\min}^2} \left( 1 + \nu_1 + q^*(x_0 + \nu_1 \nu_2) \right) e^{x_0 \beta_1^2},$$

and

$$\kappa_0 = \kappa_0(\nu_0) = \frac{1}{108 \rho^2 (3 \rho^2 + \frac{y_0^2}{2} + 2 \sigma_{\max}^2)},$$

where $\nu_0$ is from (2.4) and $\rho = \max \left( |y_0|, \sigma_{\max} \sqrt{L}, 2(x_0 + ML) \right)$.

Further, for any $\delta > 0$ and any vector $v = (v_0, v_1, v_2, v_3, v_4)$ from $\mathbb{R}_5^+$, we put

$$z_0 = z_0(\delta, v) = \delta^{3/2} \max \left( 2 c_1^* v_3, 2 c_2^* v_2, T^{1/2} v_4, 8 T^{-1/2} v_1 \right),$$

$$\tau = \tau(\delta, v) = \delta^{3/2} \max \left( c_1^* v_3, c_2^* v_2 \right),$$

where

$$c_1^* = 2 e^{x_0 + 1} \sqrt{\frac{R(1 + \rho)}{\kappa}} \quad \text{and} \quad c_2^* = 2 e \sigma_{\max}.$$  

The parameters $R$ and $\kappa$ are given in Theorem 5.1. Finally, we set

$$\gamma = \frac{1}{4 \tau} \quad \text{and} \quad \nu = \nu(\delta, v) = \frac{9 \nu_0 (1 - \delta)}{64 \delta}.$$ 

Define

$$M_1 = M + L (x_0 + |x_0| + 2).$$

Further, for any integrable $\mathbb{R} \to \mathbb{R}$ function $\Psi$ that is twice continuously differentiable, we define the operator

$$k_\nu(\Psi) = \max \left( \| \dot{\Psi} \|_1, \| \ddot{\Psi} \|_1, \| \Psi \|_\infty, \| \dot{\Psi} \|_\infty, \| \dddot{\Psi} \|_\infty \right),$$

and we introduce the parameters

$$z_0^* = \lambda_1 k_\nu(\Psi), \quad \tau^* = \lambda_2 k_\nu(\Psi) \quad \text{and} \quad \gamma^* = \frac{1}{4 \tau^*},$$

where

$$\lambda_1 = \max \left( 2 c_1^* (M_1 + \sigma_{\max}^2/2), 2 c_2^* (M_1 + \sigma_{\max}^2/2) q^*, 8 \right) \quad \text{and}$$

$$\lambda_2 = \max \left( c_1^* (M_1 + \sigma_{\max}^2/2), c_2^* \right).$$
4. Continuous time observations

In this section we study the deviation in the ergodic theorem in the case of continuous time observations. We define the deviation as follows:

\[ \Delta_T(\phi) = \frac{1}{\sqrt{T}} \int_0^T (\phi(y_t) - \pi_\theta(\phi)) \, dt, \]

where \( \phi \) is any integrable function, i.e. \( \|\phi\|_1 < \infty \).

**Proposition 4.1.** For any \( \nu_0 > 0 \) and \( \nu_1 > 0 \),

\[ \sup_{z \geq 0} e^{\kappa_0 z^2} \sup_{T \geq 1} \sup_{\phi \in \mathcal{V}_{\nu_0, \nu_1}} \mathbb{P}_\theta (|\Delta_T(\phi)| \geq z) \leq 2, \]

where the parameter \( \kappa_0 \) is given in (3.3).

**Proof.** Similarly to [8] our work starts with showing that the deviation (4.1) admits an exponential moment, i.e. we show that, for the parameter \( \kappa_0 \),

\[ \mathbb{E}_\theta e^{\kappa_0 \Delta_T^2(\phi)} \leq 2. \]

To prove this inequality we have to estimate the even moments of the deviation \( \Delta_T(\phi) \). To this end we represent this deviation as the sum of a continuous martingale and a negligible term. For this, one needs to find a bounded solution of the following differential equation:

\[ \dot{v}_\theta(u) + 2 \frac{S(u)}{\sigma^2(u)} v_\theta(u) = 2 \frac{\tilde{\phi}(u)}{\sigma^2(u)}, \quad \tilde{\phi}(u) = \phi(u) - \pi_\theta(\phi). \]

One can check directly that the function

\[ v_\theta(u) = -2 \int_u^\infty \frac{\tilde{\phi}(y)}{\sigma^2(y)} \exp \left\{ 2 \int_u^y \frac{S(z)}{\sigma^2(z)} \, dz \right\} \, dy \]

yields such a solution. Due to Lemma A.1 from the Appendix, the function \( v_\theta \) is uniformly bounded. Applying the Ito formula to the function \( V(y) = \int_0^y v_\theta(u) \, du \) yields the following representation:

\[ \int_0^T \tilde{\phi}(y_s) \, ds = V(y_T) - V(y_0) - \zeta_T, \]

where \( \zeta_T = \int_0^T v_\theta(y_s) \sigma(y_s) \, dw_s \). Therefore, due to Lemma A.1, for any \( T \geq 1 \), one can estimate \( \Delta_T(\phi) \) from above as follows:

\[ |\Delta_T(\phi)| \leq r |y_T| + r |y_0| + \frac{1}{\sqrt{T}} |\zeta_T|. \]

Moreover, taking into account (see [13, Lemma 4.11]) that, for any \( m \geq 1 \),

\[ \mathbb{E}_\theta (\zeta_T)^{2m} \leq (2m - 1)!! r^{2m} \sigma_{\max}^{2m} T^m, \]
we obtain by Proposition A.2 that, for any $m \geq 1$,
\[
E_\vartheta |\Delta_T(\phi)|^{2m} \leq 3^{2m-1} \left( 2^m (E_\vartheta |y_T|^{2m} + |y_0|^{2m}) + \frac{E_\vartheta (\xi_T)^{2m}}{T^m} \right)
\]
\[
\leq (3r)^{2m} \left( 4(m + 1)(2m - 1)! \rho^{2m} + y_0^{2m} + (2m - 1)! \sigma_{\max}^{2m} \right).
\]
Therefore, taking into account the definition of $\kappa_0$, we obtain
\[
E_\vartheta e^{\kappa_0 \Delta_T^2(\phi)} \leq 1 + \sum_{m=1}^{\infty} \frac{\kappa_0^m}{m!} (3r)^{2m} \left( 4(2m + 1)! \rho^{2m} + y_0^{2m} + (2m - 1)! \sigma_{\max}^{2m} \right)
\]
\[
\leq 1 + \sum_{m=1}^{\infty} (1/2)^m = 2.
\]

From here we obtain the inequality (4.3) and by the Chebyshev inequality we come to the upper bound (4.2). Hence we have Proposition 4.1. \qed

Remark 4.1. It should be noted that the inequality (4.2) is shown in [8] for the process (1.1) with $\sigma = 1$. Thus, Proposition 4.1 extends the result from [8] to the case of any diffusion coefficient $\sigma \in \mathcal{V}$.

5. Simultaneous geometric ergodicity for a class of diffusion processes

Here we announce a result on geometric ergodicity obtained in [10].

Theorem 5.1. Let coefficients $(S, \sigma) = \vartheta$ belong to the space $\Theta$. There exist some constants $R \geq 1$ and $\kappa > 0$ such that
\[
\sup_{t \geq 0} e^{\kappa t} \sup_{\|g\|_\infty \leq 1} \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Theta} \left| E_\vartheta (g(y_t) | y_0 = x) - \pi_\vartheta(g) \right| \leq R.
\]

The explicit parameters $R$ and $\kappa$ are given in [10].

6. Proofs

6.1. Proof of Theorem 2.1

We start with remark that, due to Corollary A.1, one has, for any $\alpha \in [1, \infty[$,
\[
\sup_{t \geq 0} \sup_{\vartheta \in \Theta} E_\vartheta (|y_t|^\alpha | y_0 = x) \leq (2 (\alpha + 1)^{1/2} \rho)^\alpha.
\]

Now we represent the deviation $D_T(\phi)$ as follows:
\[
D_T(\phi) = \int_0^T (\phi(y_t) - \pi_\vartheta(\phi)) dt + A_{1,T} - A_{2,T}
\]
\[
= \sqrt{T} \Delta_T(\phi) + A_{1,T} - A_{2,T},
\]

From here we obtain the inequality (4.3) and by the Chebyshev inequality we come to the upper bound (4.2). Hence we have Proposition 4.1. \qed

Remark 4.1. It should be noted that the inequality (4.2) is shown in [8] for the process (1.1) with $\sigma = 1$. Thus, Proposition 4.1 extends the result from [8] to the case of any diffusion coefficient $\sigma \in \mathcal{V}$.

5. Simultaneous geometric ergodicity for a class of diffusion processes

Here we announce a result on geometric ergodicity obtained in [10].

Theorem 5.1. Let coefficients $(S, \sigma) = \vartheta$ belong to the space $\Theta$. There exist some constants $R \geq 1$ and $\kappa > 0$ such that
\[
\sup_{t \geq 0} e^{\kappa t} \sup_{\|g\|_\infty \leq 1} \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Theta} \left| E_\vartheta (g(y_t) | y_0 = x) - \pi_\vartheta(g) \right| \leq R.
\]

The explicit parameters $R$ and $\kappa$ are given in [10].

6. Proofs

6.1. Proof of Theorem 2.1

We start with remark that, due to Corollary A.1, one has, for any $\alpha \in [1, \infty[$,
\[
\sup_{t \geq 0} \sup_{\vartheta \in \Theta} E_\vartheta (|y_t|^\alpha | y_0 = x) \leq (2 (\alpha + 1)^{1/2} \rho)^\alpha.
\]

Now we represent the deviation $D_T(\phi)$ as follows:
\[
D_T(\phi) = \int_0^T (\phi(y_t) - \pi_\vartheta(\phi)) dt + A_{1,T} - A_{2,T}
\]
\[
= \sqrt{T} \Delta_T(\phi) + A_{1,T} - A_{2,T},
\]
where

$$A_{1,T} = \sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \left( \phi(y_t) - \phi(y_{t_j}) \right) dt$$

and

$$A_{2,T} = \int_{\delta N}^{T} \left( \phi(y_t) - \pi_\theta(\phi) \right) dt.$$

In order to estimate the term $A_{1,T}$ we write the difference $\phi(y_t) - \phi(y_{t_j})$ through the Ito formula as

$$\phi(y_t) - \phi(y_{t_j}) = \int_{t}^{t_j} \mathcal{L}_\theta(\phi)(y_s) ds + \int_{t}^{t_j} \dot{\phi}(y_s) \sigma(y_s) dW_s$$

$$= \tilde{\pi}_\theta(\phi)(t_j - t) + \Psi_j(t) + \int_{t}^{t_j} \dot{\phi}(y_s) \sigma(y_s) dW_s,$$

where $\Psi_j(t) = \int_{t}^{t_j} \psi(y_s) ds$, and $\psi(y) = \mathcal{L}_\theta(\phi)(y) - \tilde{\pi}_\theta(\phi)$. Now setting $X_j = \int_{t_{j-1}}^{t_j} \Psi_j(t) dt$ and $\eta_j = \int_{t_{j-1}}^{t_j} \omega_j(t) dt$, $\omega_j(t) = \int_{t}^{t_j} \dot{\phi}(y_s) \sigma(y_s) dW_s$, one has

$$A_{1,T} = \tilde{\pi}_\theta(\phi) \frac{N \delta^2}{2} + \sum_{j=1}^{N} X_j + \sum_{j=1}^{N} \eta_j. \quad (6.3)$$

To estimate the second term in the right-hand part of (6.3), we make use of Proposition A.1. We start with verifying its conditions. In view of Theorem 5.1, setting $\mathcal{F}_s = \sigma\{y_u, 0 \leq u \leq s\}$ yields, for any $t \geq s$ and for any $\phi$ from the functional class (2.6),

$$|E_{\theta} (\psi(y_t)|\mathcal{F}_s)| \leq \mu(\phi) R (1 + |y_s|) e^{-\kappa(t-s)} \leq v_3 R (1 + |y_s|) e^{-\kappa(t-s)}.$$

Therefore, for any $k > j$,

$$|E_{\theta} (X_k|\mathcal{F}_j)| \leq R e^\kappa (1 + |y_{t_j}|) v_3 \delta^2 e^{-\kappa \delta (k - j)}. \quad (6.4)$$

It should be noted also that the random variables $X_j$ are bounded, i.e. $|X_j| \leq v_3 \delta^2$. To estimate the tail probability for the sum $\sum_{j=1}^{n} X_j$ we will use the inequality (A.1). For this we need to estimate the coefficients $b_{j,N}(p)$ for any $p \geq 1$. From here, taking into account that

$$1 - e^{-\kappa \delta} \geq \kappa \delta e^{-\kappa},$$

we can estimate the coefficient $b_{j,N}(p)$ as

$$b_{j,N}(p) \leq \frac{1}{\kappa} R e^{2\kappa} \delta^2 \left( 1 + (E|y_{t_j}|^{p/2})^{2/p} \right),$$

where $\delta^2 = v_3^2 \delta^3$. Now the inequality (6.1) yields

$$b_{j,N}(p) \leq R_1 \delta^2 \sqrt{2 + p} \leq R_1 \delta^2 \sqrt{2} p,$$

where $R_1 = R e^{2\kappa}(1 + \rho)/\kappa$. Using this in (A.1) we obtain that, for any $p \in [2, \infty[$,

$$E_{\theta} \left[ \right] \sum_{k=1}^{N} X_k \right|^p \leq (2p)^{p/2} N_p^{p/2} R_1^{p/2} \delta^p (2p)^{p/4} \leq \left( 2\sqrt{R_1} \delta \right)^p N^{p/2} p^p.$$
Therefore, by Chebyshev’s inequality one has
\[
P_{\delta} \left( \left| \sum_{k=1}^{N} X_k \right| \geq z \sqrt{N} \right) \leq e^{p \ln(a) + p \ln p}
\]
with \( a = 2 \sqrt{R_1 \varsigma / \varsigma} \). Minimizing now the right-hand part over \( p \in [2, \infty[ \), we obtain, for \( z \geq 2 \varsigma_1 \),
\[
P_{\delta} \left( \left| \sum_{k=1}^{N} X_k \right| \geq z \sqrt{N} \right) \leq \exp \left\{ -z / \varsigma_1 \right\}, \text{ (6.5) }
\]
where \( \varsigma_1 = 2 e \sqrt{R_1} \nu_3 \delta^{3/2} \).

Moreover, from the inequality for stochastic integrals (see, for example, [13]), for any \( \alpha \geq 1 \), it follows that
\[
E_{\theta} |\omega_j(t)|^\alpha \leq (2 \alpha)^{\alpha/2} \nu_2^\alpha \sigma_{\text{max}}^\alpha (t_j - t)^{\alpha/2}.
\]

By the Hölder inequality, from here we get
\[
E_{\theta} |\eta_j|^\alpha \leq \delta^{\alpha-1} E_{\theta} \int_{t_{j-1}}^{t_j} |\omega_j(t)|^\alpha dt \leq (2 \alpha)^{\alpha/2} \delta^{3\alpha/2} \nu_2^\alpha \sigma_{\text{max}}^\alpha.
\]

Note that, in this case, in the right-hand part of the inequality (A.1),
\[
b_{j,c}(p) = \left( E_{\theta} |\eta_j|^p \right)^{2/p}.
\]

Therefore, we find, similarly to the inequality (6.5), that, for any \( z \geq 2 \varsigma_2 \),
\[
P_{\delta} \left( \left| \sum_{k=1}^{N} \eta_k \right| \geq z \sqrt{N} \right) \leq \exp \left\{ -z / \varsigma_2 \right\}, \text{ (6.6) }
\]
where \( \varsigma_2 = 2 e \delta^{3/2} \nu_2 \sigma_{\text{max}} \). Now from (6.3), (6.5)–(6.6) it follows that, for \( z \geq z_0 \),
\[
P_{\delta} \left( |\mathbf{A}_{1,T}| \geq z \sqrt{N} \right) \leq P_{\delta} \left( \left| \sum_{k=1}^{N} X_k \right| \geq z \sqrt{N} / 4 \right)
+ P_{\delta} \left( \left| \sum_{k=1}^{N} \eta_k \right| \geq z \sqrt{N} / 4 \right) \leq 2 \exp \left\{ -z / 4 \tau \right\}, \text{ (6.7) }
\]
where the parameters \( z_0 \) and \( \tau \) are given in (3.4). Moreover, note that, due to (2.6), the last term in (6.2) is bounded, i.e.
\[
|\mathbf{A}_{2,T}| \leq 2 \delta \|\phi\|_{\infty} \leq 2 \delta \nu_1 \leq z_0 \sqrt{N} / 4.
\]

Finally, from (6.2) for \( z \geq z_0 \), one has
\[
P_{\delta} \left( |D_T(\phi)| \geq z \sqrt{N} \right) \leq P_{\delta} \left( \sqrt{T} |\Delta_T(\phi)| + |\mathbf{A}_{1,T}| \geq 3z \sqrt{N} / 8 \right)
\leq P_{\delta} \left( \sqrt{T} |\Delta_T(\phi)| \geq 3z \sqrt{N} / 8 \right) + P_{\delta} \left( |\mathbf{A}_{1,T}| \geq 3z \sqrt{N} / 8 \right).
\]
Taking into account here that \( N/T \geq (1-\delta)/\delta \), for any \( 0 < \delta < 1 \) and \( T \geq 1 \), we obtain
\[
P(\Delta_T(\phi) | \geq \frac{3\sqrt{T-\delta}}{8\sqrt{\delta}}) + P(\Delta_{1,T} | \geq \frac{3\sqrt{N}}{8}).
\]

Therefore, applying here the inequalities (4.2) and (6.7) we come to the upper bound (2.7) with the parameter \( \kappa \) given in (3.5). Hence we have Theorem 2.1.

6.2. Proof of Theorem 2.2

Firstly, note that in this case,
\[
\| \psi_{h,x_0} \|_1 = \| \psi \|_1, \quad \| \psi_{h,x_0} \|_{\infty} = \frac{1}{h} \| \psi \|_{\infty} \quad \text{and} \quad \| \dot{\psi}_{h,x_0} \|_{\infty} = \frac{1}{h^2} \| \dot{\psi} \|_{\infty}.
\]

Moreover, taking into account that \( |S(y)| \leq M + Lx + L|y| \), we find that
\[
\sup_{|y| \leq |x_0| + 2} |S(y)| \leq M_1,
\]
where \( M_1 \) is given in (3.6).

In view of the facts that \( 0 < h < 1 \) and that \( \psi \) is a finite function of the support \([-2, 2]\), we can estimate from above the parameters (2.5) as
\[
\mu(\psi_{h,x_0}) \leq \mu_* h^{-3} \quad \text{and} \quad \tilde{\mu}(\psi_{h,x_0}) \leq \tilde{\mu}_* h^{-2},
\]
where \( \mu_* = \max(\| \dot{\psi} \|_{\infty}, \| \ddot{\psi} \|_{\infty})(M_1 + \sigma^2_{\text{max}}/2) \) and
\[
\tilde{\mu}_* = \max \left( |\dot{\psi}|_1, |\ddot{\psi}|_1 \right) \left( M_1 + \frac{\sigma^2_{\text{max}}}{2} \right) q^*.
\]

Therefore, the function \( \psi_{h,x_0} \) belongs to the class (2.6) with the parameters
\[
\nu_0 = \| \psi \|_1, \quad \nu_1 = \frac{\| \psi \|_{\infty}}{h}, \quad \nu_2 = \frac{\| \dot{\psi} \|_{\infty}}{h^2}, \quad \nu_3 = \frac{\mu_*}{h^3}, \quad \nu_4 = \frac{\tilde{\mu}_*}{h^2}.
\]

In this case, the coefficient (3.3) is equal to \( \kappa_0(\| \psi \|_1) \) and the parameters (3.4) can be represented as
\[
z_0 = \frac{\delta^{3/2}}{h^3} \max \left( 2c_1^* \mu_*, 2c_2^* \| \dot{\psi} \|_{\infty} h, \tilde{\mu}_* h T^{1/2}, 8\| \dot{\psi} \|_{\infty} h^2 T^{-1/2} \right)
\]
\[
\tau = \frac{\delta^{3/2}}{h^3} \max \left( c_1^* \mu_*, c_2^* \| \dot{\psi} \|_{\infty} h \right).
\]

Thanks to the condition (2.9), for any \( T^{-1/2} \leq h \leq 1 \),
\[
z_0 \leq l_T^{-3/2} z_0^* \quad \text{and} \quad \tau \leq l_T^{-3/2} \tau^*,
\]
where the parameters \( z_0^* \) and \( \tau^* \) are given in (3.8). Note now that, by the condition (2.8), one has
\[
P(\| D_T(\psi_{h,x_0}) \| \geq a) \leq P(\| D_T(\psi_{h,x_0}) \| \geq z_1 \sqrt{N})
\]
where $z_1 = a/\sqrt{lT}$. The first inequality in (6.11) implies that $z_1 \geq z_0$ for all $a \geq a^* = z_0^*/lT$. Moreover, from the last inequality in (6.11) it follows that, for $a \geq a^*$,

$$\min (\kappa z_1, \gamma) = \min \left( \kappa z_1, \frac{1}{4\tau} \right) \geq \min \left( \kappa \frac{z_0^*}{\sqrt{lT}}, \frac{lT\sqrt{lT}}{4\tau^*} \right).$$

Taking into account here the definition of $\kappa$ in (3.5) and the form for $\delta$ given by (2.8) we obtain that, for sufficiently large $T$,

$$\min \left( \kappa \frac{z_0^*}{\sqrt{lT}}, \frac{lT\sqrt{lT}}{4\tau^*} \right) = \frac{lT\sqrt{lT}}{4\tau^*}.$$ 

Thus, through Theorem 2.1 we come to the inequality (2.11). Hence we have Theorem 2.2. 

6.3. Proof of Theorem 2.3

First we represent the tail probability as

$$P_{\vartheta} \left( |D_T(\chi_{h,x_0})| \geq \epsilon_T T \right) = I_1 + I_2,$$

where

$$I_1 = P_{\vartheta} \left( \sum_{j=1}^{N} \chi_{h,x_0}(y_{t_j}) \Delta t_j \leq (\pi_{\vartheta}(\chi_{h,x_0}) - \epsilon_T) T \right)$$

and

$$I_2 = P_{\vartheta} \left( \sum_{j=1}^{N} \chi_{h,x_0}(y_{t_j}) \Delta t_j \geq (\pi_{\vartheta}(\chi_{h,x_0}) + \epsilon_T) T \right).$$

Let us define now the following smoothing indicator functions:

$$\Psi_{1,\eta}(u) = \frac{1}{\eta} \int_{-\infty}^{+\infty} \mathbf{1}_{\{|z| \leq 1-\eta\}} V \left( \frac{z - u}{\eta} \right) dz$$

and

$$\Psi_{2,\eta}(u) = \frac{1}{\eta} \int_{-\infty}^{+\infty} \mathbf{1}_{\{|z| \leq 1+\eta\}} V \left( \frac{z - u}{\eta} \right) dz,$$

where $\eta$ is a positive smoothing parameter which will be specified later, $V$ is an even $\mathbb{R} \to \mathbb{R}$ function that is twice continuously differentiable such that $V(z) = 0$, for $|z| \geq 1$, and

$$\int_{-1}^{1} V(z) dz = 1.$$

It is easy to see that, for any $y \in \mathbb{R}$ and $0 < \eta \leq 1/2$,

$$\Psi_{1,\eta}(y) \leq \chi(y) \leq \Psi_{2,\eta}(y)$$

and $\Psi_{2,\eta}(y) = 0$ for $|y| \geq 2$. Moreover, for the functions

$$\psi_{i,h}(y) = \frac{1}{h} \Psi_{i,\eta} \left( \frac{y - z_0}{h} \right).$$
using the inequality (2.3), we can estimate the difference between the corresponding ergodic integrals (1.4) as
\[ |\pi_\theta (\chi_{h,x_0}) - \pi_\theta (\psi_{i,h})| \leq 4\eta^*. \]
Choosing here \( \eta = \epsilon_T^2 \) we obtain, for sufficiently large \( T \),
\[ I_i \leq P_\theta \left( |D_T(\psi_{i,h})| \geq \epsilon_T T/2 \right). \]
One can check directly that, in this case, the operator (3.7) has the following asymptotic (as \( T \to \infty \)) form:
\[ k_*(\Psi_{i,\eta}) = O \left( \eta^{-2} \right). \]
Therefore, from (3.8) and (6.11) it follows that, for sufficiently large \( T \),
\[ z_0(\psi_{i,h}) = O \left( \eta^{-2}l_T^{-3/2} \right) \quad \text{and} \quad \tau(\psi_{i,h}) = O \left( \eta^{-2}l_T^{-3/2} \right), \]
i.e.
\[ z_0(\psi_{i,h}) = O \left( \frac{1}{\epsilon_T^4 l_T^3/2} \right) \quad \text{and} \quad \tau(\psi_{i,h}) = O \left( \frac{1}{\epsilon_T^4 l_T^3/2} \right). \]
Now we have
\[ P_\theta \left( |D_T(\psi_{i,h})| \geq \epsilon_T T \right) \leq P_\theta \left( |D_T(\psi_{i,h})| \geq z_1 \sqrt{N} \right), \]
where \( z_1 = \epsilon_T / \sqrt{l_T} \). The last equality in (2.10) implies \( z_1 \geq z_0 \), for sufficiently large \( T \). Now, take into account that there exists a constant \( c_* > 0 \) such that, for sufficiently large \( T \),
\[ \kappa z_1 \geq c_* T \sqrt{l_T} \epsilon_T \quad \text{and} \quad \gamma \geq c_* l_T \sqrt{l_T} \epsilon_T^4, \]
i.e.,
\[ \min (\kappa z_1, \gamma) \geq c_* l_T \sqrt{l_T} \epsilon_T^4. \]
Therefore, by Theorem 2.1, for sufficiently large \( T \),
\[ P_\theta \left( |D_T(\psi_{i,h})| \geq \epsilon_T T \right) \leq 4e^{-c_* l_T \epsilon_T^5}. \]
Now the last condition in (2.10) yields the equality (2.13). Hence we have Theorem 2.3. \( \square \)

**Appendix**

**A.1. Correlation inequality**

In this subsection we give the following inequality from [4,19].

**Proposition A.1.** Let \( (\Omega, \mathcal{F}, (\mathcal{F}_j)_{1 \leq j \leq n}, \mathbf{P}) \) be a filtered probability space and \( (X_j, \mathcal{F}_j)_{1 \leq j \leq n} \) be a sequence of random variables such that, for some \( p \geq 2 \),
\[ \max_{1 \leq j \leq n} \mathbf{E} |X_j|^p < \infty. \]
Then

\[
E \left| \sum_{j=1}^{n} X_j \right|^p \leq (2p)^{p/2} \left( \sum_{j=1}^{n} b_{j,n}(p) \right)^{p/2}.
\]  

(A.1)

**Proof.** We set \( h_n(t) = E|S_{n-1} + tX_n|^p \) and \( S_n = \sum_{j=1}^{n} X_j \). Using the induction method we assume that, for any \( 1 \leq k \leq n-1 \) and \( 0 \leq t \leq 1 \),

\[
h_k(t) \leq (2p)^{p/2} B_k^{p/2}(t),
\]

where \( B_k(t) = \sum_{j=1}^{k-1} b_{j,k}(p) + tb_{k,k}(p) \). Note that as is shown in [19, Theorem 2.3]

\[
E|S_n|^p = p(p-1) \sum_{j=1}^{n} \int_{0}^{1} E|S_{j-1} + vX_j|^p - vX_j^2 + T(j,n))dv \tag{A.3}
\]

and \( T(j,n) = X_j \sum_{k=j}^{n} E(X_k|F_j) \). Therefore,

\[
h_n(t) = p(p-1) \sum_{j=1}^{n-1} \int_{0}^{1} E|S_{j-1} + vX_j|^p - vX_j^2 + G(i,n,t))dv \\
+ p(p-1) \int_{0}^{1} E|S_{n-1} + vX_n|^p - vX_n^2 (1-v)X_n^2 dv,
\]

where \( G(j,n,t) = T(j,n-1) + tX_j E(X_n|F_j) \).

Moreover, we can estimate \( h_n(t) \) as

\[
h_n(t) \leq \sum_{j=1}^{n-1} \int_{0}^{1} E|S_{j-1} + vX_j|^p - |G(i,n,t)|dv + \int_{0}^{t} E|S_{n-1} + sX_n|^p - X_n^2 X_n^2 ds.
\]

Taking into account that \( \max_{0 \leq t \leq 1} \left( E|G(j,n,t)|^{p/2} \right)^{2/p} \leq b_{j,n}(p) \), we obtain, by the Hölder inequality,

\[
\int_{0}^{1} E|S_{j-1} + vX_j|^p - |G(i,n,t)|dv \leq \int_{0}^{1} h_j^\alpha(v) b_{j,n}(p) dv,
\]

where \( \alpha = 1 - 2/p \). Therefore,

\[
h_n(t) \leq \sum_{j=1}^{n-1} b_{j,n}(p) \int_{0}^{1} h_j^\alpha(v) dv + b_{n,n}(p) \int_{0}^{t} h_n^\alpha(s) ds.
\]

By the induction assumption, for any \( 1 \leq j \leq n-1 \), one has

\[
b_{j,n}(p) \int_{0}^{1} h_j^\alpha(v) dv \leq (2p)^{(p-2)/2} \int_{0}^{1} B_j^{(p-2)/2} (v) dv b_{j,n}(p).
\]
Due to the inequality $B_j(v) \leq \sum_{i=1}^{j-1} b_i, n + vb_{j,n}(p)$, we obtain that
\[
\int_0^1 B_j((p-2)/2)(v) \, dv \, b_{j,n}(p) \leq \frac{2}{p} \left( \left( \sum_{i=1}^{j} b_{i,n} \right)^{p/2} - \left( \sum_{i=1}^{j-1} b_{i,n} \right)^{p/2} \right).
\]
This implies, for any $0 \leq t \leq 1$,
\[
h_n(t) \leq k_n \int_0^t h_n^\alpha(v) \, dv + f_n, \tag{A.4}
\]
where $k_n = p^2 b_{n,n}(p)$ and $f_n = \left( 2p \sum_{j=1}^{n-1} b_{j,n}(p) \right)^{p/2}$. Setting
\[
Z(t) = \int_0^t h_n^\alpha(s) \, ds + \frac{f_n}{k_n},
\]
we obtain from (A.4) that $\dot{Z}(t) \leq k_n^\alpha Z(t)$. Therefore, we can write the differential equation
\[
\dot{Z}(t) = k_n^\alpha Z(t) + g(t)
\]
where $g(t) \leq 0$. From here one has
\[
Z^{2/p}(t) = Z^{2/p}(0) + \frac{2}{p} k_n^\alpha t + \int_0^t \frac{g(u)}{Z^\alpha(u)} \, du \leq Z^{2/p}(0) + \frac{2}{p} k_n^\alpha t,
\]
i.e.
\[
Z(t) \leq \left( Z^{2/p}(0) + \frac{2}{p} k_n^\alpha t \right)^{p/2}.
\]
Substituting this bound in (A.4) implies
\[
h_n(t) \leq k_n Z(t) \leq k_n \left( Z^{2/p}(0) + \frac{2}{p} k_n^\alpha t \right)^{p/2} = \left( 2p \sum_{j=1}^{n-1} b_{j,n}(p) + 2pt b_{n,n}(p) \right)^{p/2}.
\]

Hence we have Proposition A.1. \qed

A.2. The moment bound for the process $y_t$.

**Proposition A.2.** For any integer $m \geq 1$,
\[
\sup_{t \geq 0} \sup_{\vartheta \in \Theta} \mathbb{E}_{\vartheta} |y_t|^{2m} \leq 4(m + 1)(2m - 1)!! \rho^{2m} \leq 4(2m)^m \rho^{2m},
\]
where $\rho$ is given in (3.3).

**Proof.** The proof is based on the moment bounds for the solution of a linear stochastic differential equation from the book [12]. In the case of identity diffusion coefficient, the moment bound was proved in [8].

Through the Ito formula, we can write for the function $z_t(m) = \mathbb{E}_{\vartheta} y_t^{2m}$ the following integral equality:
\[
z_t(m) = z_0(m) + 2m \int_0^t \mathbb{E}_{\vartheta} y_\sigma^{2m-1} S(y_\sigma) \, ds + m(2m - 1) \int_0^t \mathbb{E}_{\vartheta} y_\sigma^{2m-2} \sigma^2(y_\sigma) \, ds,
\]
which can be rewritten as the differential equality
\[
\dot{z_t}(m) = 2m \mathbb{E}_{\vartheta} y_t^{2m-1} S(y_t) + m(2m - 1) \mathbb{E}_{\vartheta} y_t^{2m-2} \sigma^2(y_t).
\]
Taking into account that $\sup_{x \in \mathbb{R}} \sigma^2(x) \leq \sigma_{\max}^2$ we obtain that, for any $m \geq 1$ and $t \geq 0$,
\[
\dot{z}_t(m) \leq 2mE_\theta y_t^{2m-1}S(y_t) + m(2m - 1)\sigma_{\max}^2 z_t(m - 1).
\]
Now we need to estimate from above the function $x^{2m-1}S(x)$. Obviously we have that, for any $K > x_*$,
\[
x^{2m-1}S(x) \leq K^{2m-1}\sup_{|x| \leq K} |S(x)|1_{(|x| \leq K)} + x^{2m} \frac{S(x)}{x}1_{(|x| > K)}.
\]
Taking into account that $\sup_{|x| > K} |\dot{S}(x)| \leq L$, we obtain, for any $x \in [x_*, K]$, $|S(x)| \leq |S(x_*)| + L|x - x_*| \leq M + L(K - x_*)$.

Similarly, we obtain the same upper bound for $x \in [-K, -x_*]$. Therefore,
\[
\sup_{|x| \leq K} |S(x)| \leq M + L(K - x_*).
\]
Consider now the case $|x| > K$. We recall that $\sup_{|x| \geq x_*} \dot{S}(x) \leq -L^{-1}$. This implies
\[
\frac{S(x)}{x} \leq \frac{M}{K} - \frac{K - x_*}{LK}.
\]
Choosing $K = 2(x_* + ML)$ yields
\[
\frac{S(x)}{x} \leq -\frac{1}{2L}.
\]
One has
\[
x^{2m-1}S(x) \leq K^{2m-1}(M + L(K - x_*)) - \frac{1}{2L}x^{2m}1_{(|x| > K)}
\]
\[
= K^{2m-1}(M + L(K - x_*)) + \frac{\beta}{2}x^{2m}1_{(|x| \leq K)} - \frac{1}{2L}x^{2m}
\]
\[
\leq A_m - \frac{\beta}{2}x^{2m},
\]
where
\[
A_m = (2(x_* + ML))^{2m-1} \left(2M + x_* \left(L + L^{-1}\right) + 2L^2M\right).
\]
From here it follows that
\[
\dot{z}_t(m) \leq 2m A_m - L^{-1} m z_t(m) + m(2m - 1)\sigma_{\max}^2 z_t(m - 1).
\]
We can rewrite this inequality as follows:
\[
\dot{z}_t(m) = -L^{-1} m z_t(m) + m(2m - 1)\sigma_{\max}^2 z_t(m - 1) + \psi_t,
\]
where $\sup_{t \geq 0} \psi_t \leq 2m A_m$. This equality provides
\[
z_t(m) = z_0(m)e^{-mL^{-1}t} + m(2m - 1)\sigma_{\max}^2 \int_0^t e^{-mL^{-1}(t-s)}z_s(m - 1)ds
\]
\[
+ \int_0^t e^{-mL^{-1}(t-s)}\psi_s ds
\]
\[
\leq m(2m - 1)\sigma_{\max}^2 \int_0^t e^{-mL^{-1}(t-s)}z_s(m - 1)ds + B_m,
\]
where $B_m = y_0^{2m} + 2A_m L$. Setting $B_0 = 1$ and resolving this inequality by recurrence yields

$$z_t(m) \leq 4 (2m - 1)!! \sum_{j=0}^{m} \left( \frac{\sigma_{\max}^2 L}{m-j} \right)^{m-j} B_j.$$

It is easy to see that

$$B_m \leq 4 \left( \max \left( |y_0|^2, 4(x_* + ML)^2 \right) \right)^m.$$

Therefore,

$$\sup_{t \geq 0} z_t(m) \leq 4 (m+1)(2m-1)!! \rho^{2m} \leq 4(2m)^m \rho^{2m},$$

where $\rho$ is defined in (3.3) and the last inequality follows from the inequality $(m+1)(2m-1)!! \leq 4(2m)^m$. Hence we have Proposition A.2.

**Corollary A.1.** For any $\alpha \in [1, \infty[$,

$$\sup_{t \geq 0} \sup_{\vartheta \in \Theta} \mathbb{E}_\vartheta (|y_t|^\alpha |y_0 = x) \leq 2^\alpha (\alpha + 1)^{\alpha/2} \rho^\alpha.$$

**Proof.** Let $\beta$ be a smallest integer, $\beta \geq \alpha$. By the Hölder inequality,

$$\mathbb{E}_\vartheta |y_t|^\alpha \leq (\mathbb{E}_\vartheta |y_t|^\beta)^{\alpha/\beta}.$$

Further, by the Cauchy–Schwarz inequality,

$$\mathbb{E}_\vartheta |y_t|^\beta \leq \sqrt{\mathbb{E}_\vartheta |y_t|^2} \leq \sqrt{4(2\beta)\beta \rho^{2\beta}} = 2(2\beta)^{\beta/2} \rho^\beta.$$

Finally,

$$\mathbb{E}_\vartheta |y_t|^\alpha \leq (2(2\beta)^{\beta/2} \rho^\beta)^{\alpha/\beta} \leq 2^\alpha (\beta)^{\alpha/2} \rho^\alpha \leq 2^\alpha (\alpha + 1)^{\alpha/2} \rho^\alpha.$$

**A.3. Properties of the function (4.5)**

**Lemma A.1.** For any integrable function $\phi$, the solution (4.5) is uniformly bounded, i.e.

$$\sup_{\vartheta \in \Theta} \sup_{y \in \mathbb{R}} |v_\vartheta(y)| \leq r,$$

where the upper bound $r$ is introduced in (3.2).

**Proof.** Note that, for any $\vartheta$ from $\Theta$ and any integrable $\mathbb{R} \to \mathbb{R}$ function $\phi$,

$$|\pi_\vartheta(\phi)| \leq q^* \|\phi\|_1.$$

Due to the definition of the functional class in (2.1), for any $0 \leq u \leq y$, we get

$$2 \int_u^y \frac{S(v)}{\sigma^2(v)} dv \leq \beta_1 (y-u) - \beta_2 (y-u)^2,$$

where the coefficients $\beta_1$ and $\beta_2$ are given in (2.3). Therefore, for $u \geq 0$, one can estimate the function $v_\vartheta$ as

$$|v_\vartheta(u)| \leq \int_u^\infty |\phi(y)| e^{2\beta_1(y-u) - \beta_2 (y-u)^2} dy + q^* \|\phi\|_1 \int_0^\infty e^{2\beta_1 z - \beta_2 z^2} dz \leq \|\phi\|_1 v_1 \left(1 + q^* v_2\right),$$
where the parameters $\upsilon_1$ and $\upsilon_2$ are introduced in (3.1). Taking into account the definition (3.2), one has from the last inequality
\[
\sup_{\vartheta \in \Theta} \sup_{u \geq 0} |v_{\vartheta}(u)| \leq r. \tag{A.6}
\]

Let now $u \leq 0$. Taking into account that
\[
\int_{R} \frac{\tilde{\phi}(y)}{\sigma^2(y)} \exp \left\{ 2 \int_{0}^{y} \frac{S(z)}{\sigma^2(z)} \, dz \right\} \, dy = 0,
\]
we can represent the function $v_{\vartheta}$ as
\[
v_{\vartheta}(u) = 2 \int_{|u|}^{\infty} \frac{\tilde{\phi}(-y)}{\sigma^2(-y)} \, e^{-2 \int_{|u|}^{-y} S(-z)\sigma^{-2}(-z) \, dz} \, dy.
\]
Like for (A.5), one can check directly that, for any $y \geq |u|$,\[
-2 \int_{y}^{0} \frac{S(-z)}{\sigma^2(-z)} \, dz \leq \beta_1 (y - |u|) - \beta_2 (y - |u|)^2.
\]
Therefore, in the same way as in the proof of (A.6) we can estimate the function $v_{\vartheta}(u)$ as
\[
\sup_{\vartheta \in \Theta} \sup_{u \leq 0} |v_{\vartheta}(u)| \leq r.
\]
Hence we have Lemma A.1. \qed

References

[1] P. Bertail, S. Clémençon, Sharp bounds for the tails of functionals of Markov chains, Teor. Veroyatn. Primen. 54 (3) (2009) 609–619.
[2] S. Boucheron, G. Lugosi, P. Massart, Concentration inequalities using the entropy method, Ann. Probab. 31 (2003) 1583–1614.
[3] D.A. Cattiaux, A. Guillin, Deviation bounds for additive functionals of Markov process, ESAIM Probab. Stat. 12 (2008) 12–29.
[4] J. Dedecker, P. Doukhan, A new covariance inequality and applications, Stochastic Process. Appl. 106 (2003) 63–80.
[5] J. Dedecker, C. Prieur, New dependence coefficients. examples and applications to statistics, Probab. Theory Related Fields 132 (2005) 203–236.
[6] L. Galtchouk, S. Pergamenshchikov, Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes, Math. Methods Statist. 10 (2001) 316–330.
[7] L. Galtchouk, S. Pergamenshchikov, Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes, Stat. Inferences Stoch. Process. 9 (2006) 1–16.
[8] L. Galtchouk, S. Pergamenshchikov, Uniform concentration inequality for ergodic diffusion processes, Stochastic Process. Appl. 117 (2007) 830–839.
[9] L. Galtchouk, S. Pergamenshchikov, Adaptive sequential estimation for ergodic diffusion processes in quadratic metric, J. Nonparametr. Stat. 23 (2) (2011) 255–285.
[10] L. Galtchouk, S. Pergamenshchikov, Geometric ergodicity for families of homogeneous Markov chains, version 2, 2012. http://hal.archives-ouvertes.fr/hal-00455976/fr.
[11] I.I. Gihman, A.V. Skorohod, Stochastic Differential Equations, Springer, New York, 1972.
[12] Yu.M. Kabanov, S.M. Pergamenchtchikov, Two-Scale Stochastic Systems. Asymptotic Analysis and Control, in: Applications of Mathematics, Stochastic Modelling and Applied Probability, Springer, 2003.
[13] R.Sh. Liptser, A.N. Shiryaev, Statistics of Random Processes, I, Springer, New York, 1977.
[14] E. Löcherbach, D. Loukianova, O. Loukianov, Polynomial bounds in the ergodic theorem for one-dimensional diffusions and integrability of hitting times, Ann. Inst. Henri Poincaré Probab. Stat. 47 (2) (2011) 425–449.
[15] P. Massart, Some applications of concentration inequalities to statistics, Ann. Fac. Sci. Toulouse Math. 9 (2000) 245–303.
[16] V. Maume-Deschamps, Concentration inequalities and estimation of conditional probabilities. Université de Bourgogne, Décembre 2004, Prépublication n. 396.
[17] S. Meyn, R. Tweedie, Markov Chains and Stochastic Stability, Springer-Verlag, 1993.
[18] S.M. Pergamenshchikov, On large deviation probabilities in ergodic theorem for singularly perturbed stochastic systems. Weierstrass Institut für Angewandte Analysis und Stochastik, Berlin, 1998, Preprint n. 414.
[19] E. Rio, Théorie asymptotique des processus faiblement dépendants, in: Collection : Mathématiques & Applications, 31, Springer, Berlin, 2000.
[20] A.Yu. Veretennikov, On large deviations for diffusion processes with measurable coefficients, Uspekhi Mat. Nauk 50 (1995) 135–146.