ayv-lvov09c.tex

On stochastic averaging and mixing

N. Abourashchi,[*] A. Yu. Veretennikov[†]

March 30, 2010

Abstract

The text contains a review and new results on stochastic averaging via mixing bounds.

1 Introduction

The advanced technique of averaging was proposed in celestial mechanics in [45]. The idea is that when the planets turn around the Sun, their motion may be approximately calculated using classical mechanics for each planet as a two bodies problem, i.e. this planet and the Sun, without taking into account the other planets. Then, the next step should be to incorporate the other planet into the analysis, by means of asymptotic series. In the first approximation, this means that the action of the other planets should be replaced by some effective change in the coefficients of the equations over the period of rotation. Further deterministic aspects were developed by many researchers, including N. N. Bogoliubov with his collaborators, see, e.g., [5]; see also [2], [44], et al. However, the goal of this review is a stochastic aspect.

Later on, a stochastic version of the theory emerged. This theory is applicable to a wide range of systems with randomness and with "slow" and "fast" components which are often called "action" and "angle" variables, due to physical reasons. The idea of the method is to replace coefficients of the slow part of the system by

[*]Nabourashchi [at] yahoo.co.uk, University of Leeds, UK

[†]A.Veretennikov [at] leeds.ac.uk, University of Leeds, UK, and Institute of Information Transmission Problems, Moscow, Russia

certain "effective" or averaged ones, which would not depend on the fast variable any more. This is a useful simplification that often plays a crucial role, especially in applications. Nowadays, stochastic systems with averaging may be found not only in mechanics, but also in many areas of physics, chemistry, mathematical biology, weather modelling, mathematical statistics, financial mathematics, et al., see [58], [7], [16], [21], [8], [51], [19], etc.; some other works will be mentioned below, but, clearly, this list is quite difficult to make complete. Hence, the authors aspirations are only perhaps to attract attention though to some less known sources.

Although it is usually impossible to find the "first word" or the first author, it seems very likely that the idea of *stochastic averaging* was proposed as a hypothesis by N. N. Bogoliubov. Technically – and perhaps a bit artificially – the area of stochastic averaging may be split into two large directions, (1°) a kind of functional law of large numbers (which may be further split into results of convergence in probability vs. weak convergence), and (2°) a kind of functional central limit theorem, the latter often being called diffusion approximation. Further directions include, in particular, large and moderate deviations, however, we do not touch them here only mentioning the pioneering works [20], [21], and some further development in [67]–[68], [26] (for discrete time systems), [18], [22], [23], [50], [70]– [72], [38], et al.

The first named direction started with a seminal paper [29] and was further advanced in [21, ch.7], [65], [26], [13], [12], et al. In [65] stochastic averaging principle has been established for a wide class of SDE systems with non-smooth drift coefficients; both types of convergence in probability and weak convergence have been treated. In [12] an answer has been given to the question in which situation convergence is in probability (strong): earlier it was noticed in several sources that if diffusion coefficient does not depend on fast component, convergence may be often established only in probability. The results from [12] are stronger: they state that if the averaged SDE has a pathwise unique strong solution, then convergence is in probability. The background for such result is the paper [28] about approximation methods for strong solutions of SDEs and theorems about pathwise uniqueness. About diffusion approximation see [20], [8], [57], [21, ch.7], [17], [7], [3], [26], [56]; close results in PDEs may be found in [4], [31], [43], [52]. Quite often (see [57], [17], [56]) such results relate to solving Poisson equations "in the whole space". About this side of the theory see also [31], [51], [52], [41], et al.

In all papers on stochastic averaging, some quantitative ergodic properties were used. Often those ergodic properties are realised as *mixing bounds*. Hence, the second part of this work concerns some new results on mixing bounds for a new class of stochastic differential equations that are genuinely highly degenerate. Simultaneously, some new weak existence and uniqueness results are established. We leave some comments to this part till the corresponding sections, in order not to overload this introduction.

Hence, the paper consists of two parts, about averaging for two-scaled processes and about (new) mixing. Notice that practically any new mixing bound for a new class of processes can be immediately used so as to get new averaging results. We do not pursue this goal here only because of an already large volume of the text. The first part of the paper – section 2 – is a review of some basic ideas of averaging for stochastic processes, with results of three sorts: of functional Law of Large Numbers (LLN) type, of diffusion approximation (= functional CLT) type, and of large deviations. In the second part of the paper – section 3 – we establish new weak existence for a degenerate SDE system with non-smooth drift, uniqueness in distribution for this system, local mixing (local Dobrushin's condition), and finally exponential mixing bounds.

2 Stochastic averaging

2.1 Averaging inequalities

Consider an SDE system with a small parameter $\epsilon > 0$,

$$\begin{aligned} dX_t^\epsilon &= b(X_t^\epsilon, Y_t^\epsilon)\,dt + \sigma(X_t^\epsilon, Y_t^\epsilon)\,dW_t, \quad X_0 = x, \\ dY_t^\epsilon &= \epsilon^{-1} B(X_t^\epsilon, Y_t^\epsilon)\,dt + \epsilon^{-1/2} C(X_t^\epsilon, Y_t^\epsilon)\,d\tilde W_t, \quad Y_0 = y. \end{aligned} \tag{1}$$

Formally, is not really very important whether the same or different Wiener processes drive the equations for both components X and Y. Often it is assumed that the two Wiener Processes are different and independent; in some other they are assumed equal; however, all this may be included in the general scheme just by choosing suitable dimensions of the matrices σ and C, although some degeneracy issues would arise. We will assume W and $\tilde W$ independent.

The idea of stochastic averaging of LLN type is based on a simple method, which has at the same time some strength and certain weakness. Firstly, the slow component is to be frozen, and the fast motion becomes an (ergodic) Markov process with a quantitative convergence rate to its invariant regime; the slow component is then replaced by a Markov process with coefficients "averaged" with respect to that invariant measure (diffusion is averaged as $\sigma\sigma^*$). Secondly, the difference has to be estimated. It should be said that there exist critical opinions that claim that the method is very rough and hardly can be optimal. However, currently there is no real alternative to this technique, although it is not impossible that such alternative might be introduced in some future.

According to [21], the following standing inequality may be used to justify the procedure: it is assumed that there exists a function $\bar b$ such that the following two inequalities hold true,

$$\sup_{t,x,y} E_{x,y} \left| \frac{1}{T} \int_t^{t+T} b(x, y_s^{x,y})\, ds - \bar b(x) \right| \le \kappa(T), \tag{2}$$

and

$$\sup_{t,x,y} E_{x,y} \left| \frac{1}{T} \int_t^{t+T} \sigma\sigma^*(x, y_s^{x,y})\, ds - \bar a(x) \right| \le \kappa(T), \tag{3}$$

with $\kappa(T) \to 0$, $T \to \infty$, possibly with a certain specified rate. Here $y_t^{x,y}$ denotes the solution of the SDE

$$dy_t = B(x, y_t)\, dt + C(x, y_t)\, d\tilde W_t, \quad Y_0 = y. \tag{4}$$

Hence, the question arises about how to verify existence of κ satisfying (2)–(3) for particular classes of processes. In fact, the assumption (2)–(3) in its original form turns out to be rather restrictive: practically, it can be checked only for SDEs with the component Y on a compact manifold, but not in R^ℓ as in (1). That is to say that, in fact, the condition (2)–(3) is, indeed, rather restrictive for the systems in $R^{d+\ell}$, which does not look compact, except for a periodic case with respect to the second component y, which makes the state space of Y equivalent to the (compact) torus. Perhaps, other compactifications may be also possible, yet, in general, the system in $R^{d+\ell}$ certainly cannot be reduced to the compact case. Does it mean that for really non-compact spaces/cases there is no averaging? The answer is that certainly there is averaging, just (2)–(3) ought to be replaced by some suitable weaker version, for example, as suggested in [65], by

$$\sup_{t,x,y} E_{x,y} \left| \frac{1}{T} \int_t^{t+T} b(x, y_s^{x,y})\, ds - \bar b(x) \right| \le \kappa(T)(1 + |x|^2 + |y|^2), \tag{5}$$

along with another complementary assumption,

$$\sup_{t,x,y} E_{x,y}(1 + |y_t^{x,y}|^2) \le C(1 + |x|^2 + |y|^2). \tag{6}$$

Some variations are allowed here, e.g., other increasing functions instead of squares in the right hand side of the inequality (5) would do as well, including some exponentials, however, the growth rate ought to be controlled by an appropriately changed "adjoint" condition (6). We do not go into details here. The point is that the inequalities (5)–(6) are quite realistic and may be easily verified for a wide class of SDEs in R^ℓ. For certain particular classes of processes, under assumptions on coefficients conditions of the type (5 – 6) are checked in [26], et al. In turn, those sufficient conditions are based on mixing rate bounds, as shown, e.g., in [26]. Methods of how to check mixing bounds for Markov diffusions have been developed in [61], [64], [62], [73], [39], [40], et al. In most of the sources on mixing for SDEs it is assumed that diffusion coefficient is nondegenerate. Some exception is [59], where instead some hypoellipticity condition is used; however, this is also some kind of nondegeneracy, and requires a good smoothness. In this paper we propose a new method to study mixing rates suitable for highly degenerate SDEs without smoothness. This is reasonable in all mechanical systems, because if we treat solution of an SDE as a position of some particle, then the only appropriate place where nondegenerate random noise may show up is apparently forces. This leads quite naturally to systems of the following type,

$$\dot{X} = Y, \;\; \dot{Y} = \text{“random forces”},$$

where the first component X may have no white noise term (cf. [9]) by virtue of "physical reasons".

2.2 About diffusion approximation

The problems of invariance principle kind (= functional central limit theorem type results) were studied firstly for a compact state space for the "fast" component in [57]. Even a bit earlier, similar results in terms of partial differential equations have been established in [4]. Later the theory was developed in [31], [17], [7], [53], [41], et al. In many papers on the subject starting from [57] the role of Poisson equations "in the whole space" has been emphasized, see, e.g., [17].

A general non-compact and "fully–coupled" (i.e. with all coefficients that depend on all components) case

$$\begin{aligned} dX &= b(X_t, Y_t)dt + \epsilon^{-1/2} f(X_t, Y_t)dt + \sigma(X_t, Y_t)dW_t, \quad X_0 = x, \\ dY &= \epsilon^{-1} B(X, Y)\, dt + \epsilon^{-1/2} C(X, Y)\, d\tilde{W}_t, \quad Y_0 = y. \end{aligned} \tag{7}$$

has been considered in [56] via Poisson equations. The latter were investigated in this series of papers with the help of quantitative mixing bounds developed earlier in [73] et al. A different approach for Lipschitz coefficients was earlier developed in [7] (to be precise, for discrete time, but this is not very important). Another version of fully coupled equations has been treated in [3] with a motivation from mathematical models of weather. The whole direction of diffusion approximation theory is not yet completed, although basic problems seem to have been solved. We do not go into details here, but just mention that the diffusion approximation theorem from [21] does not use directly Poisson equation. The latter is a hint that although the Poisson equation in the whole space is, indeed, a very powerful tool, nevertheless, in some interesting situations it may be avoided, or possibly it could even not applicable, while some other more probabilistic approaches may still work.

3 Degenerate case: new mixing bounds

In this section we study a 2D process that plays the role of the process $(y_t^{x,y},\, t \ge 0)$ as in (4), slightly abusing the notations accepted in the Section 2: now the first component of our 2D process is X and the second Y, but X is not averaged.

In a series of papers by F. Campillo et al. [9], [10], [11] the following system of SDEs in R^2 has been investigated for recurrence, invariant measure, approximation, etc.,

$$\begin{aligned} dX_t &= Y_t\, dt, \qquad X_0 = x, \\ dY_t &= b(X_t, Y_t)\, dt + dW_t, \qquad Y_0 = y, \end{aligned} \tag{8}$$

where W is a standard Wiener process, and drift b is a Borel measurable function satisfying a linear growth condition and having a special form,

$$b(x, y) = -u(x, y)y - \beta\, x - \gamma\, \mathrm{sign}(y), \tag{9}$$

where β and γ are some positive constants, and u satisfies (see the Assumption (A1) below) $0 < u_1 \le u(\cdot) \le u_2 < \infty$. The system describes a mechanical "semi–active" suspension device in a vehicle under external stochastic perturbation forces treated as a white noise, which, in fact, attracted much attention; we do not extend the list of references so as to cite only what is necessary for our presentation here. So, in particular, all positiveness conditions above have some clear physical nature. The term with γ corresponds to *friction,* β is a *spring* coefficient, uY corresponds to *damping* (control related to the velocity of the device), and the function u here stands for tuning of this damping control. Under certain assumptions, existence of a (unique) invariant measure has been proved [9]; as we show below, those assumptions may be relaxed. On the other hand, the question of rate of convergence to stationary regime remained completely open. We will show exponential bound on rate of convergence towards the stationary measure in the distance of total variation for the system (8)–(9), and a similar exponential bound of beta-mixing, under rather weak assumptions on the coefficients. The approach is based on *recurrence* and *local mixing*. The method of establishing local mixing proposed below is applicable to the equation (8), and should be suitable for a wider class of processes, in particular, not necessarily 2D. The method of establishing global mixing rate, as well as convergence to stationary regime, does not use "small sets" nor strong Feller property: perhaps the latter may hold true, however, due to the lack of smoothness of the drift it should not be elementary to show that.

The first question about the system (8) may be regarded as unexpected: we ought to revise existence and uniqueness questions. Why unexpected? Just because the golden period of this topic was in 60s-70s. However, motivated by the setting from [11], one may notice that it is not reasonable to assume any smoothness of the drift in the second component. The matter is that this drift may admit some control, and it is well known that optimal strategies are usually discontinues; hence, no smoothness on he drift will be assumed in the sequel. In such a case, the most traditional technique to establish (weak) existence, starting from the works [58], [46], [47] does not work. Indeed, [58] requires continuity of coefficients, while [47] requires nondegeneracy of diffusion; and we have neither assumed. Hence, we will apply another well known although less frequent approach based on Girsanov's transformation of measure [27]. Notice a similarity between the methods in the next section and in [74]: the latter is also based on Girsanov's formula and also provides weak solutions for certain class of degenerate SDE of (8) type. However, the method below is different and covers another class of equations, although there is a non-trivial intersection.

3.1 Weak existence and uniqueness

Since we are going to apply Girsanov's technique [27], we could get only weak solutions. (This does not mean, of course, that strong solution is not possible for some example or class of examples; but it should follow from some complementary analysis.) The same relates to uniqueness: we are going to check weak uniqueness, i.e. uniqueness in distribution. In fact, for non-degenerate SDE systems it was noticed by Girsanov himself in the last comment of his seminal paper (without proof), and it was realised later in [6] and [15], that linear growth of the drift suffices to establish martingale property of a stochastic exponential, and, hence, justify Girsanov's method. Remind that [27] recommended to work with bounded drifts; in case of unbounded, recommendation was to stop or truncate. Nevertheless, the results from [6] and [15] (see [33] for a more modern presentation) were very useful. However, they are not sufficient for the system (8), due to the evident degeneracy. They would be sufficient under condition

$$\sup_x |b(x,y)| \le C(1+|y|), \quad \forall \ y, \tag{10}$$

but this inequality is more restrictive than what is assumed, following [10], [11]. Remind that assumptions in [11] were based on engineering meanings of all terms of the equations. Assumption (10) is unreasonable due to physical nature of the equation.

It may be said more. Since the paper [63], it is known that any SDE in a finite-dimensional Euclidean space with a unit diffusion matrix and linear growth condition on a (Borel measurable) drift has a pathwise unique *strong* solution. Thus, weak existence result from [6] for such SDEs is redundant since about year 1980: it is fully covered by strong existence from [63]; it is actually also covered by weak existence from [46] and [47]. However, for the system (8)–(9) there is no result on strong solutions except for under rather special restrictions in [9]; we do not assume those restrictions here.

We formulate two assumptions, (A1) will be used for existence and uniqueness, and (A2) for estimating rate of mixing.

Assumptions for (8)

(A1) The function b in (8) is Borel measurable, and there exists C such that

$$|b(x,y)| \le C(1+|x|+|y|).$$

(A2) The function u in (9) is Borel measurable, and there exist constants $0 < u_1 \le u_2 < \infty$ such that $u_1 \le u \le u_2$; β and γ are strictly positive constants.

In the sequel, $\mu_t^{x,y}$ denotes the marginal distribution of (X_t, Y_t), the couple with the initial state (x,y), and μ_∞ stands for its (unique) invariant distribution if the latter exists.

Theorem 1 *Let the system (8) satisfy (A1). Then the equation (8) has a (weak) solution; this solution is unique in distribution and it is a strong Markov process.*

Proof. First of all let us show that there exists a weak solution of the system (8), and that it possesses a weak uniqueness property. Emphasize that (9) is not assumed in this section. Basically, there are two methods available: one based on approximations; and another based on Girsanov's transformations. In the general case, if we want to use approximations and weak convergence, then we do have a good *a priori* bound, – e.g., for the second moment, – but the function u may be discontinuous, in particular, in variable x, while the component X has no diffusion term at all. This is an obstacle while using approximations and passing to a limiting measure. So, we will work with Grisanov's transformations. We start with a couple $(X, \tilde{W})$ on *some* probability space $(\Omega, \mathcal{F}, \tilde{P})$, where $\tilde{W}$ is a Wiener process, and $X_t = x + \int_0^t \tilde{W}_s \, ds$. In other words, the process $(X, \tilde{W})$ solves the system (8) in the trivial case $b \equiv 0$. We will use Girsanov's exponential to solve a general case. Let

$$\tilde{\rho}_T := \exp\left(\int_0^T \left(b(X_t, y + \tilde{W}_t) \, d\tilde{W}_t \; - \frac{1}{2}\int_0^T \left|b(X_t, y + \tilde{W}_t)\right|^2 dt\right)\right).$$

The existence part of the Theorem will be proved if we show that this is a probability density, i.e., that $\tilde{E}\tilde{\rho}_T = 1$. It is convenient to formulate the statement as a lemma.

Lemma 1 *Under the assumption (A1), there exists $T > 0$ small enough, such that for every $R > 0$,*

$$\sup_{(x,y)\in B_R} \tilde{E}_{x,y}\tilde{\rho}_T^2 < \infty. \tag{11}$$

Moreover, for every $(x, y) \in B_R$ and every $T > 0$ (not only small),

$$\tilde{E}_{x,y}\tilde{\rho}_T = 1. \tag{12}$$

Emphasize that the value of the left hand side in (11) may depend on R, however, the value T can be chosen so that it suits all values $R > 0$.

Proof of Lemma 1. Notice that the assertion (11) guarantees uniform integrability of $\tilde{\rho}_T$ with respect to the measure $\tilde{P}$, for every $(x, y) \in B_R$, which implies (12) for small values of T. However, the latter equality is extended on any T by simple induction based on Markov property (remind that *small* T in (11) does not depend on initial data), see [6] or [33, Corollary 3.5.14]. Hence, it suffices to prove only

(11). We estimate, using Cauchy–Bouniakovsky–Schwarz' inequality (known widely as Cauchy–Schwarz' or Cauchy's),

$$
\begin{aligned}
\left(\tilde E_{x,y}\tilde\rho_T^2\right)^2 \le \Bigg(\tilde E \exp\Bigg(-4\int_0^T b(x+\int_0^t \tilde W_s\,ds, y+\tilde W_t)\,d\tilde W_t \\
-8\int_0^T \left|b(x+\int_0^t \tilde W_s\,ds, y+\tilde W_t)\right|^2\,dt\Bigg)\Bigg) \\
\times \tilde E \exp\left(+6\int_0^T \left|b(x+\int_0^t \tilde W_s\,ds, y+\tilde W_t)\right|^2\,dt\right) \\
\le \tilde E \exp\left(+6\int_0^T \left|b(x+\int_0^t \tilde W_s\,ds, y+\tilde W_t)\right|^2\,dt\right) \\
\le \tilde E \exp\left(\int_0^T C\left(1+(x+\int_0^t \tilde W_s\,ds)^2+(y+\tilde W_t)^2\right)\,dt\right) \\
\le \tilde E \exp\left(\int_0^T \left(C(1+|x|^2+|y|^2)+C(\int_0^t \tilde W_s\,ds)^2\,dt+C(W_t)^2\right)\,dt\right) \\
\le C(T,R,x,y)\,\tilde E \exp\left(C(T+T^3)\sup_{0\le t\le T}|\tilde W_t|^2\right) \\
= C(T,R,x,y)\,\tilde E \exp\left(C\,(T^2+T^4)\sup_{0\le t\le 1}|\tilde W_t|^2\right).
\end{aligned}
$$

Since, due to the André reflection principle, for any $v>0$,

$$
\tilde P(\sup_{0\le t\le 1}|\tilde W_t|>v) \le 4\tilde P(\tilde W_1>v) \le \frac{4}{v}\exp(-v^2/2),
$$

it is, indeed, easy to see that with any constant β, the latter expectation is finite if $T>0$ is chosen small enough. The Lemma 1 is proved. In particular, we have (weak) existence for the system (8–9).

Now, to show (weak) uniqueness, we suppose that the couple (X,Y) solves the system (8) under the assumption (A1). Let

$$
\rho_T := \exp\left(-\int_0^T (b(X_t,Y_t)\,dW_t\ -\frac{1}{2}\int_0^T |b(X_t,Y_t)|^2\,dt\right).
$$

In the sequel, the index T in $\rho_T = \rho$ may be dropped if T is already fixed. The statement about weak uniqueness – and strong Markov property – is convenient to formulate as a proposition.

Proposition 1 *Under the assumption (A1), weak solution of the system (8) on* $[0,\infty)$ *is unique in distribution. Any solution on any probability space is a strong Markov process. Also, for any* $T > 0$*,*

$$E\,\rho_T = 1. \tag{13}$$

Proof of Proposition 1. We already know that given x, y, for any T, weak existence follows straight away from Girsanov's transformation due to the Lemma 1. Let us show that for any T, weak uniqueness (= uniqueness in law) follows from the same Girsanov transformation. Indeed, if there is a solution of (1), we can apply the inverse Girsanov transformation and using the standard localization procedure along with Fatou's lemma, we get (13) by the Lemma 1. Hence, the distribution of (X, Y) on $[0, T]$ can be obtained from the distribution of $(\tilde X, \tilde Y)$ with $\tilde Y - y = \tilde W$ ($\tilde P$–Wiener process), by means of the Girsanov transformation $\tilde\rho_T$. So, this distribution is, indeed, unique on $[0, T]$. This kind of argumentation about using Girsanov's transformation in order to prove uniqueness in law can be found, in particular, in [25], and here we present it only for the reader's convenience. For a slightly different reasoning see [33].

Strong Markov property follows from [48], due to weak uniqueness. The proof of the Proposition 1 is completed. The Theorem 1 is also proved.

In the sequel we will use the following close assertion.

Lemma 2 *Under the assumption (A1), there exists* $T > 0$ *small enough, such that for every* $R > 0$*,*

$$\sup_{(x,y)\in B_R} E^{\rho}_{x,y}\rho_T < \infty. \tag{14}$$

Proof of Lemma 2. Notice that since $E^{\rho}_{x,y}\rho_T = E\rho_T^2$, the assertion (14) guarantees uniform integrability of ρ_T with respect to the measure P, for every $(x, y) \in B_R$, which, by the way, again implies the Proposition 1, at least, for $T > 0$ small enough. The inequality (14) can be rewritten as

$$\sup_{(x,y)\in B_R} E^{\rho}_{x,y}\rho_T = \sup_{(x,y)\in B_R} \tilde E_{x,y}(\tilde\rho_T)^{-1} < \infty.$$

In this form, it follows from the calculus quite similar to that in the proof of the Lemma 1. The Lemma 2 is proved.

Remark 3. Both weak existence and martingale property of Girsanov's exponential can be easily extended to a multidimensional case where both $X_t \in R^d$ and $Y_t \in R^d$, with just minor changes in the calculus; different dimensions for the components are also possible.

Remark 4. The result from [6] about Girsanov's transformation relates to the following SDE in R^d with a d-dimensional Wiener process (we use another notation Z_t for the process, to distinguish it from the setting (8)),

$$dZ_t = b(t, Z_t)\, dt + dW_t, \qquad Z_0 = z. \tag{15}$$

In this Remark, drift b is a d-dimensional Borel measurable vector–function, and it satisfies a linear growth condition with some constant $L > 0$,

$$|b(t, z)| \le L\,(1 + |z|), \qquad \forall\, z \in R^d. \tag{16}$$

The following Theorem is a reformulation of some combination of Lemma 0 and Theorem 1 and a discussion around them from [6], and the Lemma 7 from [27]. However, it is easier for us to cite a later presentation from [33, Corollary 3.5.16 Proposition 5.3.6]. As usual (e.g., as above in the Lemma 1), to solve (15), we consider a probability space $(\Omega, \mathcal{F}, \tilde{P})$ with a (another) Wiener process $\tilde{W}_t$, $t \ge 0$.

Proposition 2 (Benes 1971) *Under (16), for any T,*

$$\tilde{E}\zeta_T = 1, \quad \zeta_T := \exp(-\int_0^T b(s, \tilde{W}_s)\, d\tilde{W}_s - \frac{1}{2}\int_0^T |b(s, \tilde{W}_s)|^2\, ds),$$

the process $W_t := \tilde{W}_t - \int_0^t b(s, \tilde{W}_s)\, ds,\ 0 \le t \le T$, is d-dimensional Wiener under the new measure $dP \equiv d\tilde{P}^\zeta := \zeta_T\, d\tilde{P}$, and, hence, the equation (15) has a weak solution unique in the sense of distribution.

The reader may wish to check himself whether or not this Proposition is applicable directly to (8), or, at least, to (8) with the restriction (9); the authors believe that it is not.

Remark 5. The assumption (16) is essentially used in the proof of this result. It may be of interest to notice the last remark in the paper by Girsanov [27], actually related to more general processes with a variable diffusion, which (the remark) in the case of constant diffusion reduces precisely to (16). The author did not prove the claim, but promised to do it later, which apparently never occurred. Perhaps, it may be partially explained by the fact that even without that remark his method did allow applications if used with appropriate truncations.

Notice also that in [9], Girsanov's transformation was actually used to remove only the $u(X_t, Y_t)Y_t$ part of the drift, although this does not affect our comments. Despite all arguments above, the authors are still inclined to think that all results of this section are possibly just a re-discovery of something well-known, and they keep this section until a proper reference on some earlier paper(s) is advised to us by referees or readers.

Remark 6. Notice that non-Markov SDEs may be considered quite similarly, which would generalize weak existence for equations with delay from [15]; again, mention that the cited paper mainly deals with other control problems.

3.2 Local mixing via local Dobrushin's condition

In this subsection we are going to establish local mixing condition which we call local Dobrushin's. The name is because Dobrushin used global condition of this sort in his studies of central limit theorem for Markov processes. Of course, the same condition appears in the standard form of ergodic theorem for Markov chains, so – as usual – the question of who was the first to suggest this type of condition is unclear. However, remind that the expression in the left hand side of the condition (17) below in the case $B = R^d$ is called Dobrushin's ergodic coefficient in the literature. So, we are going to verify that for any R large enough and some suitable $T > 0$,

$$\inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{\mu_{T;x_0,y_0}(dx\,dy)}{\mu_{T;x_1,y_1}(dx\,dy)} \wedge 1\right) \mu_{T;x_1,y_1}(dx\,dy) =: \kappa_R > 0. \tag{17}$$

Here notation is used,

$$\mu_{T;x_0,y_0}(dx\,dy) := P_{x_0,y_0}(X_T \in dx,\, Y_T \in dy).$$

The density of one measures with respect to another is understood in the usual way, that is, as a density of the absolute continuous component. Also we notice that the notation used κ_R does not mean at all that the left hand side in (17) does not depend on anything but R. It may well depend on other parameters, however, for the time being, what is important is the choice of R; so all other parameters – which can be easily recovered – are dropped from this notation.

The assumption (17) is strictly weaker than small sets condition that is often used in such situations, and (17) provides a better constant in a final bound, and also this condition is satisfied for a wider class of processes. There is one more reason: despite a bit cumbersome outlook, the condition (17) is actually often easier to verify, and it is not an exaggeration: just imagine how to check small sets condition for the system (8). The next result is the second part of the method used in this paper and our main contribution to the technique of verification of mixing rate here. We consider any solution to the equation (8), without the restriction (9).

Lemma 3 *Let (A1) be satisfied. Then for any $R > 0$ there exists $c > 0$ such that (17) holds true.*

Proof. First of all, notice that

$$\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} > 0, \quad \text{a.s.} \tag{18}$$

Indeed, by virtue of Girsanov's transformation (cf., e.g., the Proposition 1 above), under the measure P^ρ we have a representation,

$$\rho_T = \exp\left(-\int_0^T b(x_0 + \int_0^t \tilde W_s\,ds, y_0 + \tilde W_t)\,d\tilde W_t \right.$$
$$\left. -\frac{1}{2}\int_0^T \left|b(x_0 + \int_0^t \tilde W_s\,ds, y_0 + \tilde W_t)\right|^2\,dt\right).$$

Denote

$$\mu^\rho_{T;x_0,y_0}(dx\,dy) := E^\rho_{x_0,y_0} 1(X_T \in dx, Y_T \in dy).$$

We have,

$$\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} = \frac{\mu^\rho_{T;x_0,y_0}(dx\,dy)}{dx\,dy} E_{x_0,y_0}(\rho_T^{-1} \mid X_T = x, Y_T = y),$$

where both multiples $\mu^\rho_{T;x_0,y_0}(dx\,dy)/dx\,dy$ and $E(\rho_T^{-1} \mid X_T = x, Y_T = y)$ are positive (a.s. for the second one). For the second this is because $0 < \rho^{-1} < \infty$ a.s. For the first one there is an explicit representation of a lower bound of this density, see (19) below. So, (17) can be rewritten equivalently as

$$\inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{\mu_{T;x_0,y_0}(dx\,dy)}{\mu_{T;x_1,y_1}(dx\,dy)} \wedge 1\right) \mu_{T;x_1,y_1}(dx\,dy)$$

$$= \inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} \wedge \frac{\mu_{T;x_1,y_1}(dx\,dy)}{dx\,dy}\right) dx\,dy \ge c > 0.$$

Let $L > 0$ and consider the densities,

$$\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} := \frac{E_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)}{dx\,dy},$$

$$\frac{\mu^L_{T;x_0,y_0}(dx\,dy)}{dx\,dy} := \frac{E_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)\, 1(\rho_T > L)}{dx\,dy}.$$

Clearly, the measure $\mu^L_{T;x,y}(dx\,dy)$ is absolutely continuous with respect to the Lebesgue measure $dxdy$, similarly to $\mu_{x,y}(dx\,dy)$. Moreover, ρ_T is a probability density (see the Proposition 1). So, we can use the following notations,

$$\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} \equiv \frac{E^\rho_{x_0,y_0} \rho^{-1} 1(X_T \in dx,\, Y_T \in dy)}{dx\,dy} =: p_{x_0,y_0}(x,y;T),$$

$$\frac{\mu^\rho_{T;x_0,y_0}(dx\,dy)}{dx\,dy} \equiv \frac{E^\rho_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)}{dx\,dy} =: p^\rho_{x_0,y_0}(x,y;T),$$

$$\frac{\mu^L_{T;x_0,y_0}(dx\,dy)}{dx\,dy} = \frac{E^\rho_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)\, 1(\rho_T > L)}{dx\,dy} =: p^L_{x_0,y_0}(x,y;T).$$

We estimate,

$$\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} = \frac{E^\rho_{x_0,y_0} \rho_T^{-1} 1(X_T \in dx,\, Y_T \in dy)\, 1(\rho_T \le L)}{dx\,dy}$$

$$+ \frac{E^\rho_{x_0,y_0} \rho_T^{-1} 1(X_T \in dx,\, Y_T \in dy)\, 1(\rho_T > L)}{dx\,dy}$$

$$\ge L^{-1} \frac{E^\rho_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)\, (1 - 1(\rho_T > L))}{dx\,dy}$$

$$\ge L^{-1} \left(\frac{E^\rho_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)}{dx\,dy} - \frac{E^\rho_{x_0,y_0} 1(X_T \in dx,\, Y_T \in dy)\, 1(\rho_T > L)}{dx\,dy}\right).$$

Here ρ is a probability density on Ω. So, the first term up to the multiple L^{-1} is a positive density of the two-dimensional Gaussian vector

$$\begin{pmatrix} X \\ \tilde W \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} x \\ y \end{pmatrix}, C_T\right), \qquad C_T = \begin{pmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{pmatrix},$$

under the probability measure P^ρ. In other words,

$$p^\rho_{x_0,y_0}(x,y;T) = \frac{\sqrt{12}}{2\pi T^2}\, \exp\left(-\frac{1}{2}(x-x,y-y)(C_T^{-1})(x-x,y-y)^*\right),$$

and

$$p_{x_0,y_0}(x,y;T) \geq L^{-1}\, \left(p^\rho_{x_0,y_0}(x,y;T) - p^L_{x_0,y_0}(x,y;T)\right). \tag{19}$$

In particular, the main term in the lower bound of the density p_T is uniformly bounded from below on any compact. We estimate,

$$\inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{\mu_{T;x_0,y_0}(dx\,dy)}{\mu_{T;x_1,y_1}(dx\,dy)} \wedge 1\right) \mu_{T;x_1,y_1}(dx\,dy)$$

$$= \inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} \wedge \frac{\mu_{T;x_1,y_1}(dx\,dy)}{dx\,dy}\right) dx\,dy$$

$$\geq \inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} L^{-1}\left(p^\rho_{x_0,y_0}(x,y;T) \wedge p^\rho_{x_1,y_1}(x,y;T) \right.$$

$$\left. - p^L_{x_0,y_0}(x,y;T) - p^L_{x_1,y_1}(x,y;T)\right) dx\,dy$$

$$\geq L^{-1}\left(\inf_{(x,y),(x',y')\in B_R} p^\rho_{x,y}(x',y';T)\,|B_R| - 2 \sup_{(x,y)\in B_R} P^\rho_{x,y}(\rho_T > L)\right).$$

We used the elementary inequality, $(a-b)\wedge(c-d) \geq (a\wedge c) - b - d$. Next, clearly,

$$\inf_{(x,y),(x',y')\in B_R} p^\rho_{x,y}(x',y';T)\,|B_R| = \pi R^2 \inf_{(x,y),(x',y')\in B_R} p^\rho_{x,y}(x',y';T) > 0,$$

and this value does not depend on L. The second term admits the following bound due to Bienaimé–Chebyshev (it would do with any power),

$$\sup_{(x_0,y_0)\in B_R} P^\rho_{x_0,y_0}(\rho_T \geq L) \leq L^{-1} \sup_{(x_0,y_0)\in B_R} E^\rho_{x_0,y_0}\rho_T.$$

Hence, in order to complete the proof of the Lemma, it *suffices* to notice that

$$\sup_{(x_0,y_0)\in B_R} E^\rho_{x_0,y_0}\rho_T < \infty, \tag{20}$$

at least, for $T > 0$ small enough. Indeed, the inequality (20) has been established in the Lemma 2 above. The Lemma 3 is proved.

Remark 9. Let us emphasize once more again that the calculus above does not guarantee local boundedness of the transition density p (unlike p^ρ). Hence, an applicability of small sets condition remains an open question, not speaking of its optimality.

The Lemma 3 could be very helpful on its own. However, in the sequel it will be more convenient to use some modification of its statement. Let us start our process (X, Y) at $(x_0, y_0) \in B_R$, and consider the exit measure $\nu^{R'}_{x_0,y_0}(\cdot)$ of this process from the cylinder

$$Q^T_{R'} := \{(t, x, y) : \ t \le T, \ |(x, y)| \le R'\}, \quad R' \ge R.$$

This exit measure is concentrated on the parabolic boundary of $Q^T_{R'}$. Clearly, if let $R' \to \infty$ – with R fixed – the mass of this function on the surface where $t < T$ will tend to zero. Moreover, on the part of the boundary where $t = T$, for any $A \in \mathcal{B}(R^2)$,

$$\nu^{R'}_{x_0,y_0}((T, x, y) : \ (x, y) \in A) \uparrow \mu_{T;x_0,y_0}(A), \quad R' \uparrow \infty.$$

Due to the domination $\nu^{R'}_{x_0,y_0}(\{T\} \times A) \le \mu_{T;x_0,y_0}(A)$, the measure $\nu^{R'}_{x_0,y_0}(\{T\} \times \cdot)$ has a density with respect to $\mu_{T;x_0,y_0}$ and, hence, also with respect to the Lebesgue measure on R^2. Let us denote the latter

$$q^{R'}_{x_0,y_0}(dxdy) := \frac{\nu^{R'}_{x_0,y_0}(dxdy)}{dxdy}.$$

Consider a modified local Dobrushin's condition, with any $R' \ge R$ and for simplicity written in a truncated style,

$$\inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{d\nu^{R'}_{T;x_0,y_0}}{d\nu^{R'}_{T;x_1,y_1}} \wedge 1\right) d\nu^{R'}_{T;x_1,y_1} > 0. \tag{21}$$

Because of existence of densities with respect to the Lebesgue measure, it is possible to rewrite the latter equivalently as

$$\inf_{(x_0,y_0),(x_1,y_1)\in B_R} \int_{B_R} \left(\frac{d\nu^{R'}_{T;x_0,y_0}}{d\Lambda} \wedge \frac{d\nu^{R'}_{T;x_1,y_1}}{d\Lambda}\right) d\Lambda =: \kappa_{R,R'} > 0, \tag{22}$$

where Λ denotes the Lebesgue measure on the whole parabolic boundary of the cylinder $Q_T^{R'}$, i.e. on the set $\{(T,x,y) : |(x,y)| \le R'\}$.

Lemma 4 *Let (A1) be satisfied. Then for any $R > 0$ there exist $R' \ge R$ and $c > 0$ such that (22) holds true.*

Proof follows from the monotone convergence theorem. Indeed, as $R' \uparrow \infty$, the densities $\dfrac{d\nu^{R'}}{dxdy}$ on the boundary where $t = T$ increase, and they almost everywhere converge to the density $\dfrac{d\mu}{dxdy}$. Hence, the sequence of minimums of two sub-probability densities also monotonically converges almost everywhere on the boundary $t = T$,

$$\left(\frac{d\nu^{R'}_{T;x_0,y_0}}{d\Lambda} \wedge \frac{d\nu^{R'}_{T;x_1,y_1}}{d\Lambda}\right)_{t=T} \uparrow \left(\frac{\mu_{T;x_0,y_0}(dx\,dy)}{dx\,dy} \wedge \frac{\mu_{T;x_1,y_1}(dx\,dy)}{dx\,dy}\right), \quad R' \uparrow \infty.$$

So, the condition (22) follows from (17). The Lemma 4 is proved.

3.3 Lyapunov functions and hitting time bounds

The main ideas of this section are due to [9]-[11]; however, our presentation contains some further news adjusted so as to serve establishing mixing bounds.

Lemma 5 *Let the assumptions (A1)–(A2) be satisfied. Then for the system (8–9) there exists a constant $C > 0$ such that*

$$\sup_{t\ge 0} E_{x,y}(|X_t|^2 + |Y_t|^2) \le C(1 + |x|^2 + |y|^2).$$

Proof follows from [9], with the Lyapunov function suggested there,

$$f(x,y) = \beta x^2 + \epsilon xy + y^2,$$

where $\epsilon > 0$ is small enough. Let us remind the main line, entirely for the reader's convenience. Let $g_{x,y}(t) \equiv g(t) := E_{x,y} f(X_t, Y_t)$; from general martingale inequalities it easily follows that this function is locally bounded. There exists $\epsilon_0 > 0$ such that for any $\epsilon_0 > \epsilon > 0$

$$f(x,y) \ge \frac{1}{2}(\beta x^2 + y^2),$$

and, of course,
$$f(x,y) \le C\,(x^2+y^2).$$
Hence it suffices to show that $g(t) \le C(1+|x|^2+|y|^2)$ for any $t \ge 0$ with some $C>0$. Naturally, $g(0) = f(x,y) \le C(1+|x|^2+|y|^2)$. Applying Itô's formula, we find that there exist positive constants ϵ and δ such that
$$\frac{d}{dt} g(t) \le -C(\epsilon,\delta) g(t) + \frac{\epsilon}{2\delta} + \sigma^2, \tag{23}$$
where $C(\epsilon,\delta)>0$. From here it follows that
$$g(t) \le g(0)\exp(-C(\epsilon,\delta)t) + \,(\frac{\epsilon}{2\delta}+\sigma^2)\,C(\epsilon,\delta)^{-1}. \tag{24}$$
Clearly, the arguments above may require some localization procedure which is quite standard. The Lemma 5 is proved.

Corollary 1 *There exists a stationary distribution μ_∞ with the property*
$$\int (x^2+y^2)\,\mu_\infty(dx\,dy) < \infty.$$
The proof follows from the Lemma 5 as in [9], for example.

Lemma 6 *Let (A1)–(A2) be satisfied, and R be large enough. Then for the system (8–9) there exist $C,\alpha>0$ such that*
$$E_{x,y}\exp(\alpha\tau) \le C(1+f(x,y)),$$
The proof of Lemma 6 follows easily from the standing inequality above (23), similarly to the calculus in [73] or [61].

We will need a similar technical inequality for a process in a double–dimension state space. Namely, we consider another independent copy $(\bar X_t, \bar Y_t,\, t\ge 0)$ of the process $(X_t, Y_t,\, t \ge 0)$, possibly with another initial condition. Let $Z_t = (X_t, Y_t)$, $\bar Z_t = (\bar X_T, \bar Y_t)$.

Lemma 7 *Let (A1)–(A2) be satisfied, and R be large enough. Then for the system (8–9) there exist $C,\alpha>0$ such that*
$$E_{z,z'}\exp(\alpha\gamma) \le C\,(1+f(z)+f(z')),$$
where γ is defined as follows,
$$\gamma := \inf(t\ge 0:\ |Z_t| \vee |\bar Z_t| \le R).$$

The proof follows similarly from the Lyapunov inequality above (23), cf. [73] or [61].

Remark 8. Such inequalities are frequently needed in techniques which use coupling. They usually do follow *from a similar analysis* as without space-doubling. However, it is a bit unclear whether such assertions may follow, say, from the Lemma 6 automatically, i.e. without a new calculus.

3.4 Mixing rate bounds

Next step is mixing and convergennce rate to the stationary regime. Remind the definition of beta-mixing coefficient,

$$\beta_t^{x,y} := \sup_{s\ge 0} E_{x,y} \sup_{B\in\mathcal{B}^2} \left(P_{x,y}((X_{t+s}, Y_{t+s}) \in B) - P_{x,y}((X_{t+s}, Y_{t+s}) \in B \mid F_s^{X,Y})\right), \quad (25)$$

where (x, y) is the initial condition for the equation. The coefficient $\beta_t^{x,y}$ dominates the (non-stationary) *alpha-mixing* coefficient introduced (in the stationary form) by Rosenblatt, and the latter is widely used for establishing all kinds of limit theorems. Hence, naturally, $\beta_t^{x,y}$ is also suitable for this goal. The stationary version of the coefficient β_t is widely known as Kolmogorov's coefficient (for the first time it appeared in the joint work by his students Volkonskii and Rosanov). The non-stationary version of beta-coefficient for Markov processes (25) was investigated, in particular, in a series of papers by the second author. The approach consists of two parts, recurrency – e.g., via Lyapunov functions – and "local mixing condition". Both issues have been studied in the previous sections, and now we can turn to our second main goal, i.e. beta-mixing bounds.

In the sequel, $\mu_t^{x,y}$ denotes the marginal distribution of (X_t, Y_t), the couple with the initial state (x, y), and μ_∞ stands for its (unique) invariant distribution if the latter exists.

Theorem 2 *Let the system (8) satisfy (A1) and (A2). Then there exists a unique probability distribution μ_∞ which does not depend on initial data (x, y), and there exist $C, c > 0$ such that*

$$\|\mu_t^{x,y} - \mu_\infty\|_{TV} \le C \exp(-ct)(1 + x^2 + y^2), \qquad t \ge 0, \quad (26)$$

and also

$$\beta_t^{x,y} \le C \exp(-ct)(1 + x^2 + y^2), \qquad t \ge 0. \quad (27)$$

Proof. The plan is to use the Lemmas 7 and 4 and the calculus from [73], with a natural replacement of polynomial inequalities by exponential ones.

1. Consider a couple of independent processes $Z_t = (X_t, Y_t),\ t \geq 0$, and $\tilde{Z}_t = (\tilde{X}_t, \tilde{Y}_t),\ t \geq 0$, where (X_t, Y_t) is a solution of the (8), while $\tilde{Z}_t$ now is a stationary version of the Markov process with the same generator and with a finite seond moment (see the Corollary 1). On the direct product of those two probability spaces, construct a sequence of stopping time, following [29],

$$\hat{\tau}_1 = \inf(t \geq 0 : |Z_t| \vee |\tilde{Z}_t| \leq R),$$

and for all $n \geq 1$,

$$T_n = \inf(t \geq \hat{\tau}_n : |Z_t| \geq R',\ \text{or}\ |\tilde{Z}_t| \geq R') \wedge (\hat{\tau}_n + 1),$$

$$\hat{\tau}_{n+1} = \inf(t \geq T_n : |Z_t| \vee |\tilde{Z}_t| \leq R).$$

2. Using the coupling method as in [62], due to the Lemma 4, we can construct a new process $\bar{Z}_.$ (a copy of $Z_.$) and a stopping time $L \geq 0$ on some extended probability space, so that

$$P_{z,\tilde{Z}_0}(\bar{Z}_t = Z_t\ , t \leq L-1) = P_{z,\tilde{Z}_0}(\bar{Z}_t = \tilde{Z}_t,\ t \geq L) = 1. \tag{28}$$

Here $z = (x, y)$, while $\tilde{Z}_0$ has a stationary distribution, $P_{z,\tilde{Z}_0}$ means conditional probability *given* $\tilde{Z}_0$ *and* $Z_0 = z$; P_z stands for probability given only $Z_0 = z$. It follows from the implementation of the coupling method as in [60] that there exists $q \leq 1 - \kappa_{R,R'}$ such that

$$P_{z,\tilde{Z}_0}(L > \hat{\tau}_n) \leq q^n, \quad \forall n. \tag{29}$$

Naturally, we choose here R, R' so that $\kappa_{R,R'} > 0$, which is possible due to the Lemma 4. We have, $\forall C \in \mathcal{B}(R^2)$,

$$\begin{aligned} |P_z(Z_t \in C) - P(\tilde{Z}_t \in C)| &= |P_z(\bar{Z}_t \in C) - \int P_{\tilde{z}}(\tilde{Z}_t \in C)\,\mu_\infty(d\tilde{z})| \\ &= |\int \Big(P_z(\bar{Z}_t \in C) - P_{\tilde{z}}(\tilde{Z}_t \in C)\Big)\ \mu_\infty(d\tilde{z})| \leq E_z\, P_{z,\tilde{Z}_0}(L \geq t). \end{aligned} \tag{30}$$

So,

$$||\mu_t^{x,y} - \mu_\infty||_{TV} := 2\sup_C\,(\mu_t^{x,y}(C) - \mu_\infty(C)) \leq 2E_z P_{z,\tilde{Z}_0}(L \geq t).$$

3. Now, with $a^{-1}+b^{-1}=1$, $a,b>1$, by *Rogers – Hölder's* inequality (known usually as Hölder's), the following holds:

$$P_{z,\tilde Z_0}(L>t)=\sum_{n=0}^{\infty}E_{z,\tilde Z_0}1(L>t)1(\hat\tau_n\le t<\hat\tau_{n+1})$$

$$\le\sum_{n\ge 0}P_{z,\tilde Z_0}(L>\hat\tau_n)^{1/a}P_{z,\tilde Z_0}(\hat\tau_{n+1}>t)^{1/b}\le\sum_{n\ge 0}q^{n/a}P_{z,\tilde Z_0}(\hat\tau_{n+1}>t)^{1/b}.$$

By Bienaimé–Chebyshev, the Lemmae 5 and 7, and by induction,

$$P_{z,\tilde Z_0}(\hat\tau_{n+1}>t)\le e^{-\alpha t}E_{z,\tilde Z_0}e^{\alpha\hat\tau_{n+1}}$$

$$=e^{-\alpha t}\,E_{z,\tilde Z_0}\,e^{\alpha(\hat\tau_1+\sum_{k=1}^{n}(\hat\tau_{k+1}-\hat\tau_k))}\le e^{-\alpha t}C_R^n\,C(1+|x|^2+|y|^2+|\tilde X_0|^2+|\tilde Y_0|^2).$$

Hence, given the initial values X_0 and Y_0 for the process Z_t, we get

$$P_{z,\tilde Z_0}(L>t)\le(1+|x|^2+|y|^2+|\tilde X_0|^2+|\tilde Y_0|^2)\,\exp(-\alpha b^{-1}t)$$

$$\times\sum_{n\ge 0}\exp(-n(a^{-1}\ln q^{-1}-b^{-1}\ln C_R)).$$

By choosing a,b, so that $a^{-1}\ln q^{-1}-b^{-1}\ln(C_R)>0$, which is possible due to

$$\lim_{b\to\infty}b^{-1}\ln(C_R)=0,\quad\text{and}\quad\lim_{a\to 1}a^{-1}\ln q^{-1}=\ln q^{-1}>0,$$

we get here in the right hand side a convergent series and, hence, due to the Corollary 1, the required bound (26) follows after integration over $\tilde X_0$ and $\tilde Y_0$.
4. Beta-mixing is established similarly (see, for example, [73]), and we drop the details. The Theorem 2 is proved.

Remark 8. Hence, some part of analysis in [9] et al. concerning invariant measures for systems (8)–(9) can be accomplished by the exponential rate of convergence. More than that, clearly, this conclusion may be extended on a wider class of equations, but we will not pursue it here. It is interesting to notice that we have achieved even a bit more than promised: in the right hand side of the bound (26) we may actually have a multiple $(1+|x|^2+|y|^2)^{1/b}$ with some $b>1$, rather than $(1+|x|^2+|y|^2)$.

The last assertion is a trivial consequence of the Theorem 2 – by integration – and it returns the reader to the standing assumption needed for averaging from the Section 2.

Corollary 2 *Let function* $(f(x,y),\ x,y \in R^1)$ *be Borel and bounded and let the process* $Z_t = (X_t, Y_t)$ *satisfy (8)–(9). Then there exists* $C > 0$ *such that for any* x, y *and any* $T > 0$,

$$\sup_{t,x,y} E_{x,y} \left| \frac{1}{T} \int_t^{t+T} f(X_s, Y_s)\, ds - \int f(x', y') \mu_\infty(dx'dy') \right| \\ \le C \|f\|_B \times \min \left(\frac{1 + |x|^2 + |y|^2}{T}, 1 \right). \tag{31}$$

This is a way to establish (5) with $\kappa(T) \le C\, \|f\|_B T^{-1}$, because (31) is a special form of the former, while the Lemma 5 provides (6). Remind that the couple (X, Y) here plays the role of the component $(y_t^{x,y},\ t \ge 0)$ in (5)–(6).

Acknowledgements

The section 3 of the paper is based on the results from the PhD thesis by the first author [1], except for the Corollary 2 which is an analogue of the Lemma 6.1 from [26]. The second author thanks the RFBR grant 08-01-00105a for support.

References

[1] N. Abourashchi, Stability of Stochastic Differential Equations, PhD Thesis, University of Leeds, 2009.

[2] V. I. Arnold, V. V. Kozlov, A. I. Neistadt, Mathematical Aspects of Classical and Celestial Mechanics, Encyclopaedia of Mathematical Sciences, vol.3, Dynamical Systems-3, Springer-Verlag, 1988, 1998, 2006

[3] V. Bakhtin, Yu. Kifer, Diffusion approximation for slow motion in fully coupled averaging, Probability Theory and Related Fields, 129(2) (2004), 157-181.

[4] N. S. Bakhvalov, Averaging of partial differential equations with rapidly oscillating coefficients, Sov. Math., Dokl. 16 (1975), 351-355 (1975); transl. from Dokl. Akad. Nauk SSSR 221 (1975), 516-519.

[5] N. N. Bogoliubov, Y. A. Mitropolsky, Asymptotic Methods in the Theory of Nonlinear Oscillations, Gordon and Breach Inc., New York, 1961.

[6] V. E. Beneš, Existence of optimal stochastic control laws, SIAM J. Control. 9(3) (1971), 446-472.

[7] A. Benveniste, M. Metivier, P. Priovret, Adaptive Algorithems And Stochastic Approximation, Springer-Verlag, Berlin, Heidelberg, New York, 1990.

[8] A. N. Borodin, A limit theorem for solutions of differential equations with a random right-hand side. Theory Probab. Appl. 22 (1977), 482-497; transl. from Teor. Veroyatn. Primen. 22 (1977), 498-512.

[9] F. Campillo, Optimal ergodic control for a class of nonlinear stochastic systems. In *Proceedings of the 28th Conference on Decision and Control,* 1190-1195, Tampa, December 1989, IEEE.

[10] F. Campillo, F. Le Gland, E. Pardoux, Approxmation of a stochastic ergodic control problem. In: J. Descusse, M. Fliess, A. Isidori and D. Leborgne, eds., *New trends in Nonlinear Control Theory*, pages 379-395, CNRS, Nants 1988, Lecture notes in Control and information Sciences 122, Springer-Verlag, 1989.

[11] F. Campillo, E. Pardoux, Numerical methods in ergodic optimal stochastic control and application. In: I. Karatzas and D. Ocone, eds., *Applied Stochastic Analysis*, Proceedings of a US-French Workshop, Rutgers University,New Brunswick, NJ, April 29-May 2, 1991, volume 177 of lecture notes in Control and Information Sciences, pages 59-73. Berlin, 1992, Springer-Verlag.

[12] S. Cerrai, A Khasminskii type averaging principle for stochastic reaction-diffusion equations, Ann. Appl. Probab. 19(3) (2009) 899-948.

[13] S. Cerrai, M. Freidlin, Averaging Principle For A Class Of Stochastic Reaction Diffusion Equations, Probability Theory and Related Fields, 144(1-2) (2009), 137-177.

[14] J. L. Doob, Stochastic processes, J. Wiley and Sons, New York, 1953.

[15] T. Duncan, P. Varayja, On solutions of stochastic control systems, SIAM Journal of Control, 9(3) (1971), 354-371.

[16] M. Duflo, Recursive Stochastic Methods, Springer-Verlag, Berlin, 1993.

[17] S. N. Ethier, T. G. Kurtz, Markov processes: Characterization and Convergence, John Wiley & Sons, New York et al., 1986.

[18] J. Feng, T. G. Kurtz, Large deviations for stochastic processes. Providence, RI: American Mathematical Society (AMS), 2006.

[19] J.-P. Fouque, G. Papanicolaou, K. R. Sircar, Derivatives in Financial Markets with Stochastic Volatility, CUP, 2000.

[20] M. I. Freidlin, The averaging principle and theorems on large deviations, Russian Math. Surveys, 33(5) (1978), 117-176.

[21] M. I. Freidlin, A. D. Wentzell, Random Perturbations of Dynamical Systems. Springer-Verlag, Berlin, 1984.

[22] M. I. Freidlin, A. D. Wentzell, Random perturbations of Hamiltonian systems. Mem. Am. Math. Soc. 523 (1994), 82 p.

[23] M. I. Freidlin, R. B. Sowers, A comparison of homogenization and large deviations, with applications to wavefront propagation. Stochastic Processes Appl. 82(1) (1999), 23-52.

[24] I. I. Gikhman, On a theorem of N. N. Bogoliubov, Ukr. Matem. Zhurnal, 4, 1952, 215-218 (Russian).

[25] I. I. Gihman, A. V. Skorohod, Stochastic differential equations. Springer-Verlag, Berlin, 1972.

[26] O. V. Gulinsky, A. Yu. Veretennikov, Large deviation for discrete time processes with averaging, VSP, Utrecht, The Netherlands, 1993.

[27] I. V. Girsanov, On transforming a class of stochastic processes by absolutely continuous substitution of measures. (Russian) Teor. Verojatnost. i Primenen. 5 (1960), 314–330; Engl. transl. Theory Probab. Appl. 3(5) (1960), 285-301.

[28] I. Gyongy, N. V. Krylov, Existence of strong solutions of Ito's stochastic equations via approximations. Probability Theory and Related Fields, 105(2), 143–158.

[29] R. Z. Has'minski, On averaging principle for stochastic differential Itô's equations (Russian), Kibernetika, 4(3) (1968), 260-279.

[30] R. Z. Has'minski, Stochastic stability of differential equations. Sijthoff Noordhoff, Alphen aan den Rijn, The Netherlands, and Rockville, Maryland, USA, 1980.

[31] V. V. Jikov, S. M. Kozlov, O. A. Oleinik, Homogenization of Differential Operators and Integral Functionals, Springer-Verlag, New York, 1994.

[32] V. V. Jikov, S. M. Kozlov, Multiscaled homogenization. In: Berdichevsky, V. (ed.) et al., Homogenization. In memory of Serguei Kozlov. Singapore: World Scientific. Ser. Adv. Math. Appl. Sci. 50, 35-64 (1999).

[33] I. Karatzas, S. E. Shreve, Brownian motion and stochastic calculus. Springer-Verlag, 1991.

[34] R. Z. Khasminskii, Principle of Averaging for Parabolic and Elliptic Differential Equations and for Markov Processes with Small Diffusion, Theory Probab. Appl., 8(1) (1963), 1-21.

[35] R. Z. Khasminskii, On the averaging principle for Itô stochastic differential equations, Kibernetika (Prague), 4 (1968), 260-279 (in Russian).

[36] Yu. Kifer, Some recent advances in averaging. In: Brin, M. (ed.) et al., Modern dynamical systems and applications. Dedicated to Anatole Katok on his 60th birthday. Cambridge: Cambridge University Press. 385-403 (2004).

[37] Yu. Kifer, Averaging principle for fully coupled dynamical systems and large deviations, Ergodic Theory and Dynamical Systems, 24(3) (2004), 847-871.

[38] Yu. Kifer, Large deviations and adiabatic translations for dynamical systems and Markov processes in fully coupled averaging, Memoirs of AMS, Providence, Rhode Island, 2009.

[39] S. A. Klokov, A. Yu. Veretennikov, On subexponential mixing rate for Markov processes. (English. Russian original) Theory Probab. Appl. 49(1) (2005), 110-122; translation from Teor. Veroyatn. Primen. 49(1) (2004), 21-35.

[40] S. A. Klokov, A. Yu. Veretennikov, On mixing and convergence rates for a family of Markov processes approximating SDEs. Random Oper. Stoch. Equ. 14(2) (2006), 103-126.

[41] Yu. V. Kolomiets, Averaging of randomly perturbed evolutionary equations, Ukrainian Mathematical Journal, 45(7) (1993), 1066–1076.

[42] Yu. V. Kolomiets, Fractional Stability of Diffusion Approximation for Random Differential Equations, In: S. N. Ethier, J. Feng and R. H. Stockbridge, eds.,

Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz (Beachwood, Ohio, USA: IMS, 2008), 41–61.

[43] S. M. Kozlov, Geometric aspects of averaging. Russ. Math. Surv. 44, No.2, 91-144 (1989); translation from Usp. Mat. Nauk 44, No.2(266), 79-120 (1989).

[44] M. A. Krasnosel'skii, S. G. Krein, On the averaging principle in nonlinear mechanics, Uspekhi Mat. Nauk, 1955.

[45] N. Krylov, N. Bogoliubov, Introduction to non-linear mechanics. Princeton, N.J.: Princeton University Press, 1943.

[46] N. V. Krylov, On Ito's stochastic integral equations, Theor. Probab. Appl. 14 (1969), 330-336.

[47] N. V. Krylov, Controlled diffusion processes. Springer-Verlag, New York et al., 1980.

[48] N. V. Krylov, The selection of a Markov process from a Markov system of processes, and the construction of quasidiffusion processes. (Russian) Izv. Akad. Nauk SSSR Ser. Mat. 37 (1973), 691–708; Engl. transl.: Math. USSR Izv. 7 (1973) 691-709.

[49] N. V. Krylov, A simple proof of a result of A. Novikov, arXiv:math/0207013v1

[50] R. Liptser, V. Spokoiny, A. Yu. Veretennikov, Freidlin-Wentzell type large deviations for smooth processes. Markov Process. Relat. Fields 8(4) (2002), 611-636.

[51] R. Liptser, J. Stoyanov, On the averaging principle for diffusion type processes, Stochastic and Stochastic Reports 32, 145-163.

[52] S. Ya. Mahno, The averaging principle for stochastic partial differential equations. (Russian) Teor. Sluchajnykh Protsessov 8 (1980), 113-117.

[53] S. Ya. Mahno, Sufficient conditions for convergence of solutions of stochastic equations, J. Sov. Math. 67(4) (1993), 3216-3222; transl. from Teor. Sluchajnykh Protsessov 16 (1988), 66-73.

[54] G. N. Milstein, A. Yu. Veretennikov, On deterministic and stochastic sliding modes via small diffusion approximation. Markov Process. Related Fields 6(3) (2000), 371-395.

[55] A. A. Novikov, On moment inequalities and identities for stochastic integrals, Proc. Second Japan-USSR Symp. Probab. Theory, Lecture Notes in Math., 1973, Vol. 330, 333-339.

[56] E. Pardoux, A. Yu. Veretennikov, On the Possion equation and diffusion approximation 1, 2, 3, Ann.Probab. 29(3), 2001, 1061-1085; ibid., 31(3), 2003, 1166-1192; 33(3), 2005, 1111-1133.

[57] G. C. Papanicolaou, D. W. Stroock, S. R. S. Varadhan, Martingale approach to some limit theorems. 1976 Duke Turbul. Conf., Durham 1976, (1977) VI.1-VI.120.

[58] A. V. Skorokhod, Asymptotic methods in the theory of stochastic differential equations. Providence, RI: American Mathematical Society (AMS), 1989.

[59] A. Yu. Veretennikov, On rate of mixing and the averaging principle for hypoelliptic stochastic differential equations. Math. USSR, Izv. 33(2) (1989), 221-231; transl. from Izv. Akad. Nauk SSSR, Ser. Mat. 52(5) (1988), 899-908.

[60] A. Yu. Veretennikov, An essay on the theme of a theorem of Boris Vladimirovich Gnedenko, Teor. Imovir. Mat. Stat. 69 (2003), 16-25.

[61] A. Yu. Veretennikov, Bounds for the mixing rate in the theory of stochastic equations. Theory Probab. Appl. 32(2), 1987, 273-281.

[62] A. Yu. Veretennikov, On polynomial mixing bounds for stochastic differential equations. Stoch. Processes Appl., 70 (1997) 115-127.

[63] A. Yu. Veretennikov, Strong solutions and explicit formulas for solutions of stochastic integral equations. (Russian) Mat. Sb. (N.S.) 3, vol. 111(153) (1980), 434–452; Engl. transl.: Math. USSR Sb. 39 (1981), 387-403.

[64] A. Yu. Veretennikov, Estimates for the mixing rate for Markov processes. Lithuanian Math. J. 31(1) (1991), 27-34.

[65] A. Yu. Veretennikov, On the averaging principle for systems of stochastic differential equations. (Russian) Mat. Sb. 181(2) (1990), 256-268.

[66] A. Yu. Veretennikov, On hypoellipticity conditions and estimates of the mixing rate for stochastic differential equations. Soviet Math. Dokl. 40(1) (1990), 94-97.

[67] A. Veretennikov, On large deviations in averaging principle for stochastic differential equations with periodic coefficients. 1. - Probab. Theory and Math. Statistics. Proc. Fifth Vilnius Conf. (1989). Ed. by B.Grigelionis et. al.Vilnius, Lithuania: Mokslas and Utrecht, The Netherlands: VSP, Vol. 2, 542-551

[68] A. Yu. Veretennikov, On large deviations in the averaging principle for stochastic differential equations with periodic coefficients. II. Math. USSR, Izv. 39(1) (1992), 677-701; transl. from Izv. Akad. Nauk SSSR, Ser. Mat. 55(4) (1991), 691-715.

[69] A. Yu. Veretennikov, Large deviations in averaging principle for stochastic differential equation systems (noncompact case). Stochastics and Stochastics Reports, 48 (1994), 83-96.

[70] A. Veretennikov, On large deviations for SDEs with small diffusion and averaging. Stochastic Process. Appl. 89(1) (2000), 69-79.

[71] A. Yu. Veretennikov, On large deviations in the averaging principle for SDEs with a "full dependence". Ann. Probab. 27(1) (1999), 284-296.

[72] A. Yu. Veretennikov, On large deviations in the averaging principle for SDE's with a "full dependence", correction. Arxiv, math.PR/0502098, 4 Feb 2005.

[73] A. Yu. Veretennikov, On polynomial mixing and the rate of convergence for stochastic differential and difference equations. Teor. Veroyatn. i Primenen. 44 (1999), 2, 312–327; Engl. transl. in Theory Probab. Appl. 44 (2000), 2, 361–374.

[74] L. M. Wu, Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems, Stochastic Processes Appl., 91 (2001), 205-238.