On mixing rate for degenerate 2D diffusion

N. Abourashchi,* A. Yu. Veretennikov†

December 4, 2008

Abstract

An exponential uniform mixing rate is shown for a degenerate two-dimensional diffusion of Langevin type.

1 Introduction

Rate of convergence to the invariant measure is a very important issue in stochastic model theory and its applications. In a series of papers by F. Campillo et al. [2], [3], [4] the following system of SDEs in R^2 has been investigated for recurrence, invariant measure, approximation, etc.,

$$\begin{aligned} dX_t &= Y_t\,dt, \qquad X_0 = x, \\ dY_t &= b(X_t, Y_t)\,dt + dW_t, \qquad Y_0 = y, \end{aligned} \tag{1}$$

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

$$b(x, y) = (-u(x, y)y - \beta\,x - \gamma\ \mathrm{sign}(y))\,, \tag{2}$$

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 perturbations treated as a white noise. 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 appropriate assumptions, existence of a (unique) invariant measure has been proved [2]; however, the question of convergence remained

*University of Leeds, UK, e-mail: niloufar @ leeds.ac.uk

†University of Leeds, UK, and Institute of Information Transmission Problems, Moscow, Russia, e-mail: A.Veretennikov @ leeds.ac.uk

open. In this paper we show exponential bound on rate of convergence toward the stationary measure in the distance of total variation for the system (1)–(2) and a bit more general, and a similar exponential bound of beta-mixing, under suitable assumptions on the coefficients. The method of establishing *local mixing* proposed below is applicable to the equation (1), and should be suitable for a wider class of processes, in particular, not necessarily 2D.

Remind the definition of beta-mixing coefficient,

$$\beta_t^{x,y} := \inf_{s\geq 0} E_{x,y} \sup_{B\in\mathcal{B}^2} (P_{x,y}((X_{t+s}, Y_{t+s}) \in B) - P_{x,y}((X_{t+s}, Y_{t+s}) \in B \mid F_s^{X,Y})), \tag{3}$$

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, although for the first time it appeared in the joint work by his students Volkonskii and Rosanov. In his lectures in 1970s, Kolmogorov posed general problems of studying mixing coefficients for general processes. The non-stationary version of beta-coefficient for Markov processes (3) was investigated, in particular, in a series of papers by the second author.

Notice that because of the control origin of the function u in the equations, it is not desirable to assume any smoothness on it. Apart from interest for engineers, there are some mathematical issues that make this system special. In terms of recurrence properties, we apply Lyapunov's approach, using the same Lyapunov function as in [2], based on simple quadratic forms. Apparently, the use of such simple functions is limited, however, for the equation (1–2) they are quite sufficient, and possibly could serve certain even wider classes of processes. Nevertheless, for more general systems (1) possibly some other Lyapunov functions could be useful. In terms of local mixing properties, a real obstacle is a high degeneracy of the SDE system. It is comparatively not very difficult to verify a local version of the so called general Doeblin–Doob condition (see [5]), however, this kind of condition even in its global form provides only a rather reduced result about mixing (formally, about convergence in total variation) just for one particular Markov process, not for a class of processes. But clearly for the system (1) there is no global Doeblin–Doob condition available, and the question how to work with its local version is yet open. There is one more, most standard tool frequently used in similar situations, which does provide bounds uniform on some class of processes, "pétite set" condition. However, it apparently fails here completely in an even more severe fashion in compare to a non-degenerate diffusion, where it is not of any real help either. We tackle this problem by establishing some appropriate local version of *Dobrushin's ergodicity condition,* see (20) below. Notice that this kind of condition is also rather useful in the non-degenerated case, where it is provided by Harnack's inequalities, see [15], [14]. After having a good Lyapunov function and verifying a local Dobrushin condition, the remaining part of the proof is based on the method of estimating the mixing rate from above from [14].

For the system (1) weak uniqueness may be established by using Girsanov's transformation with the help of a method similar to [1]. Under (2) it was claimed in [2] with a reference to [1]. Nevertheless, we should notice that apparently the direct reference does not work, neither applied to (1), nor under the restriction (2), as the paper [1] does not consider degenerate SDEs, nor does the presentation of its results in [10]. This looks a subtle matter, especially because eventually the *method* does work, as shown below, so we include the formulation of the famous result from [1] for the reader's convenience. The extension of this result to our degenerate case being done, the authors realised that the same Lyapunov function method as for the system (1)-(2) may also work for slightly more general systems of SDEs, which satisfy (1) & (6). The verification of a local Dobrushin type condition below is also based on Girsanov's transformation, although it is not a direct corollary from the section about weak solutions.

In the Section 2 we formulate our main results along with the assumptions. An extension of the approach from [1] is provided in the Section 3. In particular, in that Section we briefly discuss weak existence of solution of our system (1) with a strong Markov property. The calculus which we suggest, of course, resembles the one in [1], – and even more the one in [10], – and may be considered as a kind of complement to the latter; it also simultaneously provides Novikov's condition (see [13], [12]) for this particular case, although this observation does not lead to any further simplification. In the Section 4 a Lyapunov function is presented for this system (the same as in [2]), together with some hitting times inequalities. In the Section 5 a local Dobrushin condition is established. The proof of convergence and mixing rate is given in the last Section 6.

A local Dobrushin condition seems to be the most appropriate tool for the two-dimensional system under consideration, with severely degenerate diffusion term. The use of Girsanov's transformation in this context seems to be new, even though it looks so natural. Earlier the most general way to show local Dobrushin for *non-degenerate* SDEs was to use Harnack's inequality, see [14], although other methods were also tried, including Malliavin's calculus and some more ad hoc ideas for particular cases. A nice feature of Girsanov's approach is, of course, that it does not require any regularity.

2 Main results

Assumptions for (1) & (2)

(A1) The function b in (1) is Borel measurable, and there exists C such that

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

(A2) The function u in (2) 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, μ_t denotes the marginal distribution of (X_t, Y_t), and μ_∞ stands for its (unique) invariant distribution if the latter exists.

Theorem 1 *Let the system (1) satisfy (A1). Then the following holds true.*

1. *The equation (1) has a (weak) solution unique in distribution, which is a strong Markov process.*

2. *If additionally the drift satisfies (2) as well as (A2), then there exists a unique probability distribution μ_∞ and there exist $C, c > 0$ such that*

$$\|\mu_t - \mu_\infty\|_{TV} \le C\exp(-ct)(1+x^2+y^2), \tag{4}$$

and also

$$\beta_t^{x,y} \le C\exp(-ct)(1+x^2+y^2). \tag{5}$$

Now as we accomplished the results from [2] et al. by rate of convergence, let us describe what more general SDE systems could be tackled in a similar way. Consider the class of drifts f satisfying the following conditions. Firstly, we require (A1), which turns out to be sufficient both for the local Dobrushin's condition and weak existence and uniqueness of solution. To tackle recurrence, we require the following.

Assumptions for (1) without (2)

(A3) The function b in (1) satisfies

$$b(x,y) = b_0(x,y) - u(x,y)y - v(x,y)x, \tag{6}$$

where the term b_0 is a Borel bounded function, u satisfies the inequalities from (A2) with some $0 < u_1 \le u_2 < \infty$, and the function v *is bounded* and satisfies

$$\lim_{(x,y)\to\infty} v(x,y) = \beta > 0. \tag{7}$$

As we shall see below, the latter condition can be relaxed so as to allow for some $\beta > 0$ a "small enough" limit

$$\limsup_{(x,y)\to\infty} |v(x,y) - \beta| << 1. \tag{8}$$

For the precise formulation as to in which sense the left hand side is $<< 1$ see the calculus in the Lemma 4 and remark 3 below. Now we have a version of Theorem 1 as follows.

Theorem 2 *Let the system (1) satisfy (A1) and (A3). Then again there exists a unique invariant probability distribution μ_∞ and constants $C, c > 0$ such that (4) and (5) hold true.*

Remark 1. We do not discuss what could be a mechanical meaning of (A3) for the system (1) & (6). However, it is not difficult to imagine the same suspension in the presence of the white noise, but all terms may now be "not ideal", e.g., – most importantly, – with a non-ideal spring term.

Remark 2. Under just a measurability assumption on u, the question of strong solution even for the original version of the system (1)–(2), apparently, remains open. In [4], existence of strong solution has been proved under a certain technical assumption that the total drift may be split into a Lipschitz part and a part monotone with respect to Y.

Remark 3. The assumption of boundedness for the function b_0 in (A3) could be replaced by the property to be locally bounded and satisfy

$$\lim_{|x|+|y|\to\infty} \frac{b_0(x,y)}{|x|+|y|} = 0.$$

3 Weak solution & Girsanov's transformation

First of all let us show that there exists a weak solution of the system (1), and that it possesses a weak uniqueness property. Emphasize that neither (2) nor (6) is 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 the other words, the process $(X, \tilde W)$ solves the system (1) in the trivial case $b \equiv 0$. We will use Girsanov's exponential to solve a general case. Let

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

We ought to show that this is a probability density, i.e., that $\tilde E \tilde\rho_T = 1$.

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

$$\sup_{(x,y)\in B_R} \tilde E_{x,y}\tilde\rho_T^2 < \infty. \tag{9}$$

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

$$\tilde E_{x,y}\tilde\rho_T = 1. \tag{10}$$

Emphasize that the value of the left hand side in (9), of course, may depend on R, however, the value T may be chosen unique for all $R > 0$.

Proof. Notice that the assertion (9) guarantees uniform integrability of $\tilde\rho_T$ with respect to the measure $\tilde P$, for every $(x, y) \in B_R$, which implies (10) 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 (9) does not depend on initial data), see [1] or [10, Corollary 3.5.14]. Hence, it suffices to prove only (9). We estimate, using Cauchy–Bouniakovsky–Schwarz' inequality (known widely as Cauchy–Schwarz' or Cauchy's),

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

$$\leq C(T,R,x,y)\,\tilde E \exp\left(C(T+T^2)\sup_{0\le t\le T}|\tilde W_t|^2\right)$$

$$= C(T,R,x,y)\,\tilde E \exp\left(C\,(T^2+T^3)\sup_{0\le t\le 1}|\tilde W_t|^2\right).$$

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.

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

$$E\,\rho_T = 1. \tag{11}$$

Proof. Given x,y, for any T, weak existence follows straight away from Girsanov's transformation due to the Lemma 1 (cf. [7]).

Again 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 by the Proposition 1,

$$E\rho_T = 1. \tag{12}$$

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 to prove uniqueness in law can be found, in particular, in [7], and here we present it only for the reader's convenience. For a slightly different reasoning see [10].

Strong Markov property follows from [11], due to weak uniqueness. The proof of the Proposition 1 is completed.

In the sequel we will use one more close assertion.

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

$$\sup_{(x,y)\in B_R} E^{\rho}_{x,y}\rho_T<\infty. \tag{13}$$

Proof. Notice that since $E^\rho_{x,y}\rho_T = E\rho_T^2$, the assertion (13) 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 (13) 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 the case where both $X_t \in R^d$ and $Y_t \in R^d$, with just minor changes in the calculus.

Remark 4. The result from [1] 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 (1)),

$$dZ_t = b(t, Z_t)\, dt + dW_t, \qquad Z_0 = z. \tag{14}$$

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

$$|b(t,z)| \le L\,(1 + |z|), \qquad \forall\, z \in R^d. \tag{15}$$

The following result is a reformulation of some combination of Lemma 0 and Theorem 1 and a discussion around them from [1] and the Lemma 7 from [8]. However, it is easier for us to cite a later presentation from [10, Corollary 3.5.16 & Proposition 5.3.6]. As usual (e.g., as above in the Lemma 1), to solve (14), we consider a probability space $(\Omega, \mathcal{F}, \tilde P)$ with a (another) Wiener process $\tilde W_t$, $t \ge 0$.

Theorem 3 *[Benes 1971] Under (15), 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 $dP := d\tilde P^\zeta$, and, hence, the equation (14) has a weak solution unique in the sense of distribution.

The assumption (15) is essentially used in the proof of this result. It may be of interest to notice the last remark in the paper by Girsanov [8], – which actually relates to more general processes with a variable diffusion, – which (the remark) in the case of constant diffusion reduces precisely to (15). The author does not prove the claim, but promises to do it later, which apparently never occurred. In respect to the *weak* solution existence, we ought to mention that the existence and pathwise uniqueness of *strong* solution is also known for the equation (14) under the same condition for quite a while [17]. This is one more point to the question of applicability of Beneš' result to the equation (1): unlike for (14), for this equation existence of strong solution is yet unknown without additional assumptions. The reader may now wish to check himself whether or not the Theorem 3 is applicable directly to (1), or, at least, to (1) with the restriction (2). Notice also that in [2] Girsanov's transformation is actually used to remove only the $u(X_t, Y_t)Y_t$ part of the drift. Overall, the authors still think that all results of this section are just a re-discovery of something well-known, and we keep this section until a proper reference on some earlier paper(s) is advised to us by readers.

Remark 5. One of the conclusions of this section is, of course, that the use of Beneš' result in [2] et al. was, in fact, correct, if one takes into account the Proposition 1 above; the latter may be called complementary to the Beneš Theorem.

4 Lyapunov functions and hitting time bounds

Lemma 3 *Let (A1)–(A2) be satisfied. Then for the system (1–2) there exists a constant C such that*

$$\sup_{t\geq 0} E(|X_t|^2 + |Y_t|^2) \leq C.$$

Proof follows from [2], with the Lyapunov function suggested there,

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

with $\epsilon > 0$ small enough. Below we remind the main points, entirely for the reader's convenience. Let $g(t) := Ef(X_t, Y_t)$. There exists $\epsilon_0 > 0$ such that for any $\epsilon_0 > \epsilon > 0$

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

Hence it suffices to show that $g(t) \leq C$ for any $t \geq 0$. Applying Itô's formula, we find that there exist positive constants ϵ and δ such that

$$\frac{d}{dt}g(t) \leq -C(\epsilon, \delta)g(t) + \frac{\epsilon}{2\delta} + \sigma^2, \qquad (16)$$

where $C(\epsilon, \delta) > 0$. From here it follows,

$$g(t) \leq (\frac{\epsilon}{2\delta} + \sigma^2)exp(-C(\epsilon, \delta)t). \qquad (17)$$

Clearly, the arguments above may require some localization procedure which is quite standard. The Lemma 3 is proved.

Lemma 4 *Let (A1) and (A3) be satisfied. Then for the system (1) & (6) there exists a constant C such that*

$$\sup_{t\ge 0} E(|X_t|^2 + |Y_t|^2) \le C.$$

Proof. We will use the same Lyapunov function

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

where ϵ is to be chosen. The calculus is similar to the one in the previous Lemma. We apply Itô's formula to $f(X_t, Y_t)$:

$$df(X_t,Y_t) = 2\beta X_t\, dX_t + 2Y_t\, dY_t + (dY_t)^2 + \epsilon X_t\, dY_t + \epsilon Y_t\, dX_t$$

$$= 2\beta X_t Y_t\, dt + 2Y_t\, dW_t + 2Y_t(b_0(X_t,Y_t) - u(X_t,Y_t)Y_t - v(X_t,Y_t)X_t)\, dt$$

$$+ dt + \epsilon Y_t^2\, dt + \epsilon X_t(b_0(X_t,Y_t) - u(X_t,Y_t)Y_t - v(X_t,Y_t)X_t)\, dt$$

$$\le 2Y_t\, dW_t + 2(Y_t + \epsilon X_t) b_0(X_t,Y_t)\, dt$$

$$-((u_1 - \epsilon)Y_t^2 + v(X_t,Y_t)\epsilon X_t^2 + (2v(X_t,Y_t) - 2\beta + \epsilon u(X_t,Y_t))X_tY_t)\, dt.$$

Here the inequality sign, of course, relates to the dt terms, while the term dW_t remains the same. Clearly, to establish the Lyapunov condition, the terms of the first order are not important if $|(X_t,Y_t)| > R$ and if R is chosen large enough. Next, since the difference $2v(X_t,Y_t) - 2\beta$ is *small enough* by modulus for $|(X_t,Y_t)| > R$ due to (7), clearly we can choose $\epsilon > 0$ small enough, so that the expression

$$\ell_t := ((u_1 - \epsilon)Y_t^2 + v(X_t,Y_t)\epsilon X_t^2 + (2v(X_t,Y_t) - 2\beta + \epsilon u(X_t,Y_t))X_tY_t) \quad (18)$$

is no less than some positive definite quadratic form, say,

$$\frac{u_1}{2} Y_t^2 + c\epsilon X_t^2 \qquad (c > 0),$$

if $|(X_t,Y_t)| > R$. Also, of course,

$$\frac{u_1}{2} Y_t^2 + c\epsilon X_t^2 \ge C^{-1} f(X_t,Y_t) \qquad (C > 0).$$

On the other hand, if $|(X_t, Y_t)| \le R$, then the whole expression in (18) is bounded. Hence, with the same notation $g(t) := Ef(X_t, Y_t) < \infty$ as above, taking expectations, we get,

$$g'(t) \le -C^{-1} E\ell_t 1(|(X_t, Y_t)| > R) - E\ell_t\, 1(|(X_t, Y_t)| \le R)$$

$$\le -E\ell_s - 2\, E\ell_s\, 1(|(X_s, Y_s)| \le R)$$

$$\le -C^{-1} g(s)\, ds + 2C_0,$$

with

$$C_0 = \sup_{|(x,y)| \le R} |\ell(x, y)|,$$

and

$$\ell(x, y) = ((u_1 - \epsilon)y^2 + v(x, y)\epsilon x^2 + (2v(x, y) - 2\beta + \epsilon u(x, y))xy).$$

This shows

$$\frac{d}{dt} g(t) \le -Cg(t) + C, \tag{19}$$

from which, in turn, it follows,

$$(0 \le) \quad g(t) \le C(1 + \exp(-Ct)).$$

The Lemma 4 is proved.

Lemma 5 *Let (A1)–(A2) be satisfied, and R be large enough. Then for the system (1–2) there exist $C, \alpha > 0$ such that*

$$E_{x,y} \exp(\alpha\tau) \le C(1 + f(x, y)),$$

Lemma 6 *Let (A1) and (A3) be satisfied, and R be large enough. Then for the system (1) & (6) there exist $C, \alpha > 0$ such that*

$$E_{x,y} \exp(\alpha\tau) \le C(1 + f(x, y)),$$

The proofs of both Lemmas 5 and 6 follow easily from the standing inequality above (16), similarly to the calculus in [15] or [14].

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 $(X_t, Y_t,\, t \ge 0)$. 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 (1–2) 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,

$$\gamma := \inf(t \ge 0 :\ |Z_t| \vee |\bar{Z}_t| \le R).$$

Lemma 8 *Let (A1) and (A3) be satisfied, and R be large enough. Then for the system (1) & (6) 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 proofs of the Lemmas 7 and 8 follow similarly from the Lyapunov inequality above (16) or (19), cf. [15] or [14].

5 Dobrushin's local mixing condition

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 (1), without restriction (2) or (6).

Lemma 9 *Let (A1) be satisfied. Then for any $R > 0$ there exists $c > 0$ such that*

$$\inf_{(x_0,y_0)\in B_R} \int_{B_R} \left(\frac{\mu_{x_0,y_0}(dx\,dy)}{dx\,dy} \wedge 1\right) dx\,dy \ge c > 0. \tag{20}$$

Proof. Due to the Proposition 1, under the measure P^ρ we have a representation,

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

Let $L > 0$ and consider the densities,

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

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

Since $\rho_T < \infty$ a.s., it is clear that both measures $\mu_{x,y}(dx\,dy)$ and $\mu^L_{x,y}(dx\,dy)$ are absolutely continuous with respect to the Lebesgue measure $dxdy$. Moreover, ρ_T is a probability density (see the Proposition 1). So, we can use the following representations,

$$\frac{\mu_{x_0,y_0}(dx\,dy)}{dx\,dy} = \frac{E^{\rho}_{x_0,y_0}\rho^{-1}1(X_T \in dx,\, Y_T \in dy)}{dx\,dy},$$

$$\frac{\mu^L_{x_0,y_0}(dx\,dy)}{dx\,dy} = \frac{E^{\rho}_{x_0,y_0}\rho^{-1}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T > L)}{dx\,dy}.$$

We estimate,

$$\begin{aligned}
\frac{\mu_{x_0,y_0}(dx\,dy)}{dx\,dy} &= \frac{E^{\rho}_{x_0,y_0}\rho^{-1}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T \le L)}{dx\,dy} \\
&+ \frac{E^{\rho}_{x_0,y_0}\rho^{-1}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T > L)}{dx\,dy} \\
&\ge L^{-1}\,\frac{E^{\rho}_{x_0,y_0}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T \le L)}{dx\,dy} \\
&+ \frac{E^{\rho}_{x_0,y_0}\rho^{-1}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T > L)}{dx\,dy} \\
&= L^{-1}\,\frac{E^{\rho}_{x_0,y_0}1(X_T \in dx,\, Y_T \in dy)\,(1 - 1(\rho_T > L))}{dx\,dy} \\
&+ \frac{E^{\rho}_{x_0,y_0}\rho^{-1}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T > L)}{dx\,dy} \\
&= L^{-1}\,\frac{E^{\rho}_{x_0,y_0}1(X_T \in dx,\, Y_T \in dy)}{dx\,dy} \\
&+ \frac{E^{\rho}_{x_0,y_0}(\rho^{-1} - L^{-1})1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T > L)}{dx\,dy} \\
&\ge L^{-1}\,\frac{E^{\rho}_{x_0,y_0}1(X_T \in dx,\, Y_T \in dy)}{dx\,dy} \\
&- L^{-1}\,\frac{E^{\rho}_{x_0,y_0}1(X_T \in dx,\, Y_T \in dy)\,1(\rho_T > L)}{dx\,dy}.
\end{aligned}$$

Since here ρ is a probability density on Ω, the first term up to the multiple L^{-1} is a positive Gaussian density on R^2 for the 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 the other words,

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

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

In particular, the density p is uniformly bounded by the value $\sqrt{3}/T^2$. Next, with any L, the second term is also a density, $p^L_{x_0,y_0}(x,y;T)$, which is dominated by p. Let us choose the constant L so large that

$$L^{-1} \sup p_T \le 1.$$

Then, the lower bound for our density does not exceed one, so that the operation "minimum with one" disappears under the integral, and we may estimate,

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

$$\ge \inf_{(x_0,y_0)\in B_R} \int_{B_R} \left(L^{-1}(p_{x_0,y_0}(x,y;T) - p^L_{x_0,y_0}(x,y;T)) \wedge 1\right) dx\,dy$$

$$= L^{-1} \inf_{(x_0,y_0)\in B_R} P^\rho_{x_0,y_0}((X_T,Y_T)\in B_R; \rho_T \le L)$$

$$\ge L^{-1}\left(\inf_{(x,y)\in B_R} P^\rho_{x,y}((X_T,Y_T)\in B_R) - \sup_{(x,y)\in B_R} P^\rho_{x,y}(\rho_T > L)\right).$$

Here, clearly,

$$\inf_{(x_0,y_0)\in B_R} P^\rho_{x_0,y_0}((X_T,Y_T)\in B_R) > 0, \tag{21}$$

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),

$$\sup_{(x_0,y_0)\in B_R} P^\rho_{x_0,y_0}(\rho_T(x,y) \ge L) \le L^{-1} \sup_{(x_0,y_))\in B_R} E^\rho_{x_0,y_0}\rho_T(x,y).$$

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

$$E\rho_T = 1, \quad \& \quad \sup_{(x_0,y_0)\in B_R} E^\rho_{x_0,y_0}\rho_T < \infty, \tag{22}$$

at least, for $T > 0$ small enough. Both inequalities have been established in the Lemma 2 above. The Lemma 9 is proved.

6 Proof of Theorems 1 and 2

Proof. We remind the main steps of the proof for the reader's convenience, because formally the case is, of course, new. The plan is to use the Lemmas 7 and 9 and the calculus from [14], with a natural replacement of polynomial inequalities by exponential ones. Both Theorems require the same calculus.

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 (1), while $\tilde{Z}_t$ now is a stationary version of this Markov process. Notice that the existence of invariant measure follows form lemmas 3 and 5, as in [2]. On the direct product of those two probability spaces, construct a sequence of stopping time, following [9],

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

and for all $n \geq 1$,

$$T_n = \inf(t \geq \hat{\tau}_n : \, |Z_t| \geq R+1, \text{ or } |\tilde{Z}_t| \geq R+1) \wedge (\hat{\tau}+1),$$

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

2. Using the coupling method as in [16], due to the lemma (6), 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

$$P(\bar{Z}_t = Z_t \ , t \leq L-1) = P(\bar{Z}_t = \tilde{Z}_t, \ t \geq L) = 1. \tag{23}$$

Moreover, there exists $q \in (0,1)$ such that

$$P(L > \hat{\tau}_n) \leq q^n, \quad \forall n. \tag{24}$$

Hence, $\forall C \in \mathcal{B}(R^2)$,

$$|P(Z_t \in C) - P(\tilde{Z}_t \in C)| = |P(\bar{Z}_t \in C) - P(\tilde{Z}_t \in C)| \leq P(L \geq t). \tag{25}$$

So,

$$||\mu_t - \mu_\infty||_{TV} := 2 \sup_C (\mu_t(C) - \mu_\infty(C)) \leq 2P(L \geq t).$$

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

$$P(L > t) = \sum_{n=0}^{\infty} E1(L > t)1(\hat{\tau}_n \leq t < \hat{\tau}_{n+1})$$

$$\leq \sum_{n \geq 0} P(L > \hat{\tau}_n)^{1/a} P(\hat{\tau}_{n+1} > t)^{1/b}$$

$$\leq \sum_{n \geq 0} q^{n/a} P(\hat{\tau}_{n+1} > t)^{1/b}.$$

By Bienaimé–Chebyshev, the Lemmas 3 and 7, and by induction,

$$P(\hat\tau_{n+1} > t) \le e^{-\alpha t} E e^{\alpha \hat\tau_{n+1}}$$

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

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

$$P(L > t) \le (1 + |X_0|^2 + |Y_0|^2)\, \exp(-\alpha b^{-1} t) \sum_{n \ge 0} \exp(-n(a^{-1} \ln q^{-1} - b^{-1} \ln C_R)).$$

By choosing a, b, so that $a^{-1} \ln q^{-1} - b^{-1} \ln(C_R) > 0$, – which is possible as $\lim_{b\to\infty} b^{-1} \ln(C_R) = 0$ and $\lim_{a\to 1} a^{-1} \ln q^{-1} = \ln q^{-1} > 0$, – we get here in the right hand side a convergent series and, hence, the required bound (4). The Theorems 1 and 2 are both proved.

Remark 6. Hence, some part of analysis in [2] et al. concerning invariant measures for systems (1)–(2) can be accomplished by the (exponential) rate of convergence; more than that, this conclusion holds true also for a slightly wider class of equations (1) & (6). It is interesting to notice that we have achieved even a bit more than promised: in the right hand side of the bound (4) we may have a multiple $(1 + |x|^2 + |y|^2)^{1/b}$ with some $b > 1$, rather than $(1 + |x|^2 + |y|^2)$.

Acknowledgements

The second author thanks the grant RFBR 08-01-00105a for support.

References

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

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

[3] 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.

[4] 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.

[5] J. L. Doob, Stochastic processes, J. Wiley and Sons, 1953.

[6] T. Duncan, P. Varaiya, On the solutions of a stochastic control system, SIAM J. Control, 9(3) (August 1971), 354-371.

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

[8] 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.

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

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

[11] 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.

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

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

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

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

[16] A. Yu. Veretennikov, On polynomial mixing bounds for stochastic differential equations. Stochastic process and their applications 70 (1997) 115-127.

[17] 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.