On ergodic measures for McKean–Vlasov stochastic equations

A. Yu. Veretennikov *

November 24, 2003

1 Introduction

Let us consider the McKean-Vlasov equation in $\mathbb{R}^d$,

$$dX_t = b[X_t, \mu_t]\,dt + \,dW_t,\ X_0 = x_0 \in \mathbb{R}^d, \tag{1}$$

where $b[x, \mu] := \int b(x, y)\,\mu(dy)$ for any measure μ (this is a notation convention), with locally Borel functions $b(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, and d-dimensional Wiener process W_t. Here μ_t is the marginal distribution of X_t. Strictly speaking, one should call solution of the equation (1) the couple (X_t, μ_t). However, with a slight abuse of notation we will call solution just the process X_t, having in mind that actually it is a couple.

The equation was suggested by Kac [7] as a stochastic toy model for the Vlasov kinetic equation of plasma (cf. [12]). The study of equation (1) was initiated by McKean [14]. A general introduction to the topic can be found in [19]. By ergodic measures we mean here the existence of a stationary distribution, its uniqueness, and at least weak convergence to this distribution as time goes to infinity.

The equaiton (1) relates to the following nonlinear equation for measures,

$$\partial_t \mu_t = L^*(\mu_t)\mu_t, \tag{2}$$

with

$$L(\mu) = \Delta/2 + b[x, \mu]\partial_x,$$

in the sense that the distribution of X_t solves equation (2) provided X_0 is distributed with respect to measure μ_0 and the process W is independent from X_0. Initial data in this paper will be always fixed, although generalizations to any initial measure with appropriate finite moments (see below) are straightforward.

*School of Mathematics, University of Leeds (UK) & Institute of Information Transmission Problems (Russia); email: veretenn@maths.leeds.ac.uk. This work has been done for the programme “Interaction and Growth in Complex Stochastic Systems” while visiting Isaac Newton Institute for Mathematical Sciences and Clare Hall, University of Cambridge.

An important method of solving equation (1) approximately is a use of the so called N-particle equation with weak interaction,

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^N b(X_t^{i,N}, X_t^{j,N})\,dt + \,dW_t^i,\ \ X_0^{i,N} = x_0,\ 1 \le i \le N, \tag{3}$$

with d-dimensional independent Wiener processes W_t^i. It is known that under reasonable assumptions the process $X^{i,N}$ converges weakly to the solution of the McKean-Vlasov equation with the same W^i, see [19], [1], [13], et al.,

$$d\bar X_t^i = b[\bar X_t^i, \mu_t^i]\,dt + \,dW_t^i,\ \bar X_0 = x_0 \in \mathbb{R}^d, \tag{4}$$

where μ_t^i stands for the law of $\bar X_t^i$ (given initial data). This result is called propagation of chaos for McKean-Vlasov equation. Here the law μ_t^i actually does not depend on i, if solution of the equation (1) is unique in law (e.g., see conditions for that in [6], [1], and also Theorem 2 below). The measure μ_t^i satisfies in the weak sense a non-linear PDE (2). It is also makes sense to consider different initial data for different particles. We will not use these convergence results, and hence do not impose these reasonable assumptions.

Large deviation results can be found in [5]. Approximation results are established in [4] et al.; the approach from [9] can be applied here, too. The paper [13] is based on log-Sobolev inequalities and Dirichlet forms technique. In [10], [18] one can find other propagation of chaos results.

The topic of the paper will be ergodic properties of equation (4) and some relations to the first equation. The most close works are [1] and [13] which contain condition for ergodicity and (exponential) convergence rate to equilibrium. In compare to [1], we allow any finite dimension, and do not impose assumptions of dissipation greater than linear; our approach is different. In compare to [13], we do not require gradient form of the interaction drift term, although conditions are still restrictive. Other close papers are [20], – also with gradient type drift, – [3], [6]. In [16] a different class of interacting diffusions is considered, results are also existence and uniqueness of invariant measures, however, even the setting can hardly be compared with ours. For more complete bibliography see references in [1], [15], [19] et al.

Section 2 contains main results: a new version of existence theorem, and uniqueness and ergodicity theorem under two sets of assumptions. Section 3 is a collection of proofs.

2 Main results

Main assumptions for existence: we assume that the function $b(x,\cdot)$ has a linear growth in the first variable,

$$|b(x,y)| \le C(1+|x|), \tag{5}$$

and continuous with respect to the second variable y for any x. Fully continuous bounded version for more general equations, – i.e. with a non-constant diffusion matrix coefficient and more general dependence of both coefficients on distribution, – can be found in [6] established by using the martingale problem method and tightness. Assumptions in [6] allow non-bounded coefficients as well, but in addition require an appropriate Lyapunov function, which may be dropped in the bounded case. Another existence and uniqueness theorem for $d = 1$ under assumptions not reduced to that from [6] (nor from this paper), can be found in [1].

Theorem 1 *Under assumption (5), there is a strong solution to equation (1).*

Various uniqueness assumptions can be found in papers [1], [6] et al. The latter (i.e. [6]) requires some kind of Lipschitz condition to this end. In the next theorem, we assume either Lipschitz condition, too (which, hence, implies uniqueness), or another set of conditions close to it although not exactly the same. Anyway, uniqueness in distribution not only for the limiting stationary measure but also for the distribution μ_t^1 for any finite t follows from the auxiliary estimate (15). Notice that in [6] uniqueness has been established for more general class of equations under Lipschitz type conditions.

Main assumptions for ergodicity and uniqueness are as follows:

- Coefficient b is decomposed into two parts,

  $$b(x, y) = b_0(x) + b_1(x, y),$$

  where the first part is responsible for the "enviromnent", while the second for the interaction itself. Next assumptions concern both whole b, and separately b_0 and b_1.

- b: recurrence–1

  $$\lim_{|x|\to\infty} \sup_y \langle b(x, y), x\rangle = -\infty, \tag{6}$$

- b: recurrence–2

  $$\lim_{|x|\to\infty} \sup_y \langle b(x, y), x\rangle \le -r < 0, \tag{7}$$

  [Essential is that the value r is fixed, and does not change with N, see the proof of theorem 2, item E.]

- b_0: attraction to zero which grows at least linearly with distance (= one-sided Lipschitz condition), for any x, x',

  $$\sup_y \langle b_0(x) - b_0(x'), x - x'\rangle \le -c_0|x - x'|^2 \qquad (c_0 > 0). \tag{8}$$

  The next two assumptions are required if c_0 is any positive. Instead, one can assume (11) which says that c_0 is large enough, along with (12) saying b_1 is Lipschitz.

- b_1: anty-symmetry of interactions,

  $$b_1(x, x') - b_1(x', x) = 0. \tag{9}$$

- b_1: "attraction" between particles, which *increases with distance*, in a certain non-rigorous sense

$$\langle (x - x') - (\bar x - \bar x'), b_1(x, x') - b_1(\bar x, \bar x')\rangle \le 0, \qquad (10)$$

  e.g., one might imagine a system of particles connected pairwise by elastic strings; the analogy, of course, is not exact, but just gives an example how interaction may not decrease with distance; needless to say that this is not a plasma. Hence, possibly the next assumptions (11-12) which replace (9-10) are more reasonable.

- b_0: *large* attraction to zero,

$$c_0 > C^{b_1}_{Lip}, \qquad (11)$$

  where $C^{b_1}_{Lip} < \infty$ is the best constant satisfying

$$\max\left(|b_1(x, y) - b_1(x', y)|, |b_1(y, x) - b_1(y, x')|\right) \le C^{b_1}_{Lip}|x - x'|, \qquad (12)$$

  for all x, x', y.

Remark. In (8) let $x' = 0$, then

$$\sup_y \langle b(x, y) - b(0, y), x\rangle \le -c|x|^2,$$

which *implies*

$$\sup_y \langle b(x, y), x\rangle \le -c|x|^2 + \sup_y \langle b(0, y), x\rangle,$$

however, this may not imply (6), in the case if $b(0, y)$ is not bounded, e.g., the rhs of the latter inequality may become even positive. Hence, we impose both conditions (6) and (8); in particular, generally speaking, the recurrence may be not strong enough to provide exponential convergence even in Markovian case (i.e., if there is no interaction), cf. [8].

Theorem 2 *Let either of the two sets of assumptions hold true:* (1°) *(5-8) with $r \ge r(d)$ large enough and (8-10) with any $c_0 > 0$; or* (2°) *(5-8) with $r \ge r(d)$ large enough and (11-12). Then, in both cases, the distribution μ_t is unique, and there is a weak limit*

$$\mu_t \Longrightarrow \mu_\infty, \quad t \to \infty,$$

and, moreover,

$$\mu^{1,N}_t \Longrightarrow \mu_\infty, \quad N, t \to \infty.$$

The measure μ_∞ is a unique invariant one for the equation (1), in particular, it does not depend on X_0.

3 Proofs

3.1 Auxuliary result

Lemma 1 *Let (5) and (6) hold true. Then for any $m > 0$,*

$$\sup_N \sup_t E|X_t^{i,N}|^m < \infty.$$

If (5) and (7) hold true instead, the same bound is valid for a fixed m if r is large enounh, that is, $r \geq r_m$.

This Lemma will be only used in subsection 3.3. Even more, it could be avoided in that subsection, too. However, for a possible future progress we would like to keep all technical estimates to be as good as possible.

Proof follows directly from comparison theorem for the $|X^{1,N}|$ and a corresponding one-dimensional markovian diffusion with reflection, and bounds for the invariant measure for this reflected markovian diffusion, cf. [22] concerning a comparison of similar type.

3.2 Proof of Theorem 1

follows from Krylov's successive approximations due to tightness and Krylov's bounds for stochastic integrals, cf. [11] for the ordinary Itô SDE. The advantage is that Krylov's technique does not require continuity of coefficients. The approximations read,

$$dX_t^n = b^n[X_t^n, \mu_t^n]\, dt + dW_t^{(n)}, \quad X_0 = x_0,$$

where $b^n(\cdot, y)$ is a *smooth* approximation of the function $b(\cdot, y)$ in the function space $L_{p,loc}(R^d)$ which is also bounded uniformly with respect to y for any fixed n, and, moreover, satisfies a uniform linear growth condition (5) for all n's. Due to [6], we have a solution X^n with a corresponding measure μ^n and a corresponding d-dimensional Wiener process denoted by $W^{(n)}$ (just notice that it is not at all W^n).

Next, due to tightness of the couple $(X_\cdot^n, W_\cdot^{(n)})$ in $C([0,t]; R^d)$ for any t, – which follows from standard stochastic integral inequalities, – one can find a sub-sequence $n' \to \infty$ such that $\mu^{n'}$ has a weak limit in $C([0,t]; R^d)$ for any t. Due to Skorokhod embedding theorem (see [17] or [11]), one can change probability space and find another sub-sequence $n'' \to \infty$ such that, moreover, $(X_\cdot^{n''}, W_\cdot^{(n'')}) \to (X_\cdot^0, W_\cdot^{(0)})$ in $C([0,t]; R^d)$ almost surely. Here all $W_\cdot^{n''}$ are d-dimensional Wiener processes, and all $X_t^{n''}$ are $\mathcal{F}_t^{W^{n''}}$ measurable (since all $X_\cdot^{n''}$ are strong solutions). In the limit, from

$$X_t^{n''} - x_0 = \int_0^t \left(\int b^{n''}(X_s^{n''}, y) \mu_s^{n''}(dy) \right) dt + W_t^{(n'')} \tag{13}$$

we get

$$X_t^0 - x_0 = \int_0^t \left(\int b(X_s^0, y) \mu_s^0(dy) \right) dt + W_t^{(0)}, \tag{14}$$

by the Lebesgue dominated convergence theorem. Details of this convergence based on Krylov's bounds for Itô processes and on the approximation $\|b^n(\cdot,y) - b(\cdot,y)\|_{L_{p,loc}} \to 0$, $n \to \infty$, may be found in [11] and [17] for ordinary SDEs: they include freezing of $b^{n''}$. For the equations (13) one has, in addition, to take into account the continuity with respect to the variable y and weak convergence of marginal distributions $\mu_s^{n''} \Longrightarrow \mu_s^0$.
Concerning strong solutions for all these equations see [21].

3.3 Proof of Theorem 2

A. We are going to show that

$$\sup_t E|X_t^{i,N} - \bar{X}_t^i|^2 \le C/N. \tag{15}$$

This part in case (1°) follows closely [1] and [13]. Since we do not use directly any uniqueness result for the equation (1), we shall say precisely what is meant by $\bar{X}_t^i$: this is *any* solution of the equation (4). However, we take a solution with the same distribution $\mu_t = \mu_t^i$ for any i. The latter is certainly possible due to the theorem 1 which asserts, in particular, that any solution X_t for (1) is strong. Hence, for any W^i we can take the *same* functional of the corresponding Wiener path (W^i) for all i's, which implies the same distribution, too. We add that since *any* solution of the equation (1) is strong, then the estimate (15) relates to any such solution, even if it not unique.

After having established this precaution, we have,

$$d(X_t^{i,N} - \bar{X}_t^i)^2 = 2(X_t^{i,N} - \bar{X}_t^i)(b[X_t^{i,N}, \hat{\mu}_t^N] - b[\bar{X}_t^i, \mu_t^1])\,dt$$

$$= 2(X_t^{i,N} - \bar{X}_t^i)(b[X_t^{i,N}, \hat{\mu}_t^N] - b[\bar{X}_t^i, \mu_t^1])\,dt$$

$$+2(X_t^{i,N} - \bar{X}_t^i)(b_0(X_t^{i,N}) - b_0(\bar{X}_t^i))\,dt$$

$$\le \left[2(X_t^{i,N} - \bar{X}_t^i)(b[X_t^{i,N}, \hat{\mu}_t^N] - b[\bar{X}_t^i, \mu_t^1]) - 2c_0|X_t^{i,N} - \bar{X}_t^i|^2\right]\,dt$$

Hence,

$$d\sum_{i=1}^N (X_t^{i,N} - \bar{X}_t^i)^2 = 2\sum_{i=1}^N (X_t^{i,N} - \bar{X}_t^i)(b[X_t^{i,N}, \hat{\mu}_t^N] - b[\bar{X}_t^i, \mu_t^1])\,dt$$

$$\le -2c_0 \sum_{i=1}^N |X_t^{i,N} - \bar{X}_t^i|^2\,dt + 2\sum_{i=1}^N (X_t^{i,N} - \bar{X}_t^i)(b[X_t^{i,N}, \hat{\mu}_t^N] - b[X_t^{i,N}, \mu_t^1])\,dt$$

Therefore,

$$E\sum_{i=1}^N (X_t^{i,N} - \bar{X}_t^i)^2 - E\sum_{i=1}^N (X_s^{i,N} - \bar{X}_s^i)^2$$

$$= 2E\int_s^t \sum_{i=1}^N (X_r^{i,N} - \bar{X}_r^i)(\hat{b}[X_r^{i,N}, \hat{\mu}_r^N] - \hat{b}[\bar{X}_r^i, \mu_r^1])\,dr$$

$$\le -2c_0 E\int_s^t \sum_{i=1}^N |X^{i,N}_r - \bar X^i_r|^2\, dr$$

$$+2E\int_s^t \sum_{i=1}^N (X^{i,N}_r - \bar X^i_r)(b[X^{i,N}_r, \hat\mu^N_r] - b[\bar X^i_r, \mu^1_r])\, dr$$

We have,

$$A := E\sum_{i=1}^N (X^{i,N}_r - \bar X^i_r)(b[X^{i,N}_r, \hat\mu^N_r] - b[\bar X^i_r, \mu^1_r]) \qquad (16)$$

$$= E\sum_{i=1}^N (X^{i,N}_r - \bar X^i_r)\left(\frac{1}{N}\sum_{j=1}^N (b(X^{i,N}_r, X^{j,N}_r) - b[\bar X^i_r, \mu^1_r])\right)$$

$$= E\sum_{i=1}^N (X^{i,N}_r - \bar X^i_r)\left(\frac{1}{N}\sum_{j=1}^N (b(X^{i,N}_r, X^{j,N}_r) - b(\bar X^i_r, \bar X^j_r))\right)$$

$$+E\sum_{i=1}^N (X^{i,N}_r - \bar X^i_r)\left(\frac{1}{N}\sum_{j=1}^N (b(\bar X^i_r, \bar X^j_r) - b[\bar X^i_r, \mu^1_r])\right) =: A_1 + A_2.$$

Case (1°). Using anty-symmetry and increase of interaction conditions on b_1, we get,

$$(X^{i,N}_r - \bar X^i_r)\left((b(X^{i,N}_r, X^{j,N}_r) - b(\bar X^i_r, \bar X^j_r))\right)$$

$$+(X^{j,N}_r - \bar X^j_r)\left((b(X^{j,N}_r, X^{i,N}_r) - b(\bar X^j_r, \bar X^i_r))\right)$$

$$= \left((X^{i,N}_r - X^{j,N}_r) - (\bar X^i_r - \bar X^j_r)\right)\left([b(X^{i,N}_r, X^{j,N}_r) - b(\bar X^i_r, \bar X^j_r)]\right) \le 0.$$

Hence, the first term is not positive, while the second possesses the bound,

$$|A_2| = |E\sum_{i=1}^N (X^{i,N}_r - \bar X^i_r)\left(\frac{1}{N}\sum_{j=1}^N (b(\bar X^i_r, \bar X^j_r) - b[\bar X^i_r, \mu^1_r])\right)|$$

$$\le \frac{1}{N}\sum_{i=1}^N \left(E\left(X^{i,N}_r - \bar X^i_r\right)^2\right)^{1/2}\left(E\left(\sum_{j=1}^N b(\bar X^i_r, \bar X^j_r) - b[\bar X^i_r, \mu^1_r]\right)^2\right)^{1/2}$$

$$\le CN^{1/2}\left(E|X^{1,N}_r - \bar X^1_r|^2\right)^{1/2}, \quad (17)$$

because random variables $b(\bar X^i_r, \bar X^j_r) - b[\bar X^i_r, \mu^1_t]$ are non-correlated for different j's. In the other words, if $\alpha(t) := E(X^{1,N}_t - \bar X^1_t)^2$, then $(t > s)$

$$N\alpha(t) - N\alpha(s) \le -2c_0\int_s^t N\alpha(r)\, dr + CN^{1/2}\int_s^t \alpha^{1/2}(r)\, dr. \qquad (18)$$

This implies

$$\alpha(t) \le \frac{C^2}{4c_0^2 N}.$$

Indeed, the function $\alpha(t)$ is differentiable, $\alpha(t) \ge 0$, and $\alpha(0) = 0$. Its derivative due to (18) satisfies

$$\alpha'(t) \le -2c_0\alpha(t) + (C/\sqrt{N})\alpha^{1/2}(r).$$

If $\alpha(t) = 0$, there is nothing to show. Suppose $\alpha(t) > 0$, and let $t_0 := \sup(s < t : \alpha(s) = 0)$. Notice that $0 \le t_0 \le t$, since $\alpha(0) = 0$. Denote $\beta(t) = \alpha(t)\exp(2c_0(t-t_0))$, then on (t_0, t) we have,

$$\begin{aligned}
\alpha'(s) &\le -2c_0\alpha(s) + (C/\sqrt{N})\alpha^{1/2}(s) \\
\text{implies} & \\
\beta'(s) &\le (C/\sqrt{N})\beta^{1/2}(s)\exp(c_0(s-t_0)), \\
\text{whence} & \\
\frac{\beta'(s)}{\beta^{1/2}(s)} &\le (C/\sqrt{N})\exp(c_0(s-t_0)), \\
\text{so,} & \\
2\beta^{1/2}(t) &\le \frac{C}{c_0\sqrt{N}}(\exp(c_0(t-t_0)) - 1), \\
\text{which implies} & \\
\beta(t) &\le \left(\frac{C}{2c_0\sqrt{N}}\right)^2 (\exp(c_0(t-t_0)) - 1)^2, \\
\text{or, equivalently,} & \\
\alpha(t) &\le \left(\frac{C}{2c_0\sqrt{N}}\right)^2 (1 - \exp(-c_0(t-t_0)))^2 \le \left(\frac{C}{2c_0\sqrt{N}}\right)^2.
\end{aligned}$$

Sorry for the boring details.

Case (2°). The value (A) from (16) can be considered as follows. The term $A_2 = S$ possesses the bound (17). Let us estimate A_1:

$$\begin{aligned}
|A_1| &\le |E\sum_{i=1}^N (X_r^{i,N} - \bar{X}_r^i)\left(\frac{1}{N}\sum_{j=1}^N (b(X_r^{i,N}, X_r^{j,N}) - b(X_r^{i,N}, \bar{X}_r^j))\right)| \\
&+ |E\sum_{i=1}^N (X_r^{i,N} - \bar{X}_r^i)\left(\frac{1}{N}\sum_{j=1}^N (b(X_r^{i,N}, \bar{X}_r^j)) - b(\bar{X}_r^i, \bar{X}_r^j))\right)| \\
&\le C_{Lip}^{b_1}\frac{1}{N}\sum_{i,j=1}^N E\,|(X_r^{i,N} - \bar{X}_r^i)(X_r^{j,N} - \bar{X}_r^j)| \\
&+ C_{Lip}^{b_1}\frac{1}{N}\sum_{i,j=1}^N E\,|(X_r^{i,N} - \bar{X}_r^i)|^2 \\
&\le 2C_{Lip}^{b_1}N\alpha(r).
\end{aligned}$$

Hence, we get

$$N\alpha(t) - N\alpha(s) \le -(2c_0 - 2C_{Lip}^{b_1})\int_s^t N\alpha(r)\,dr + CN^{1/2}\int_s^t \alpha^{1/2}(r)\,dr, \quad (19)$$

which implies, as we know,

$$\alpha(t) \le \frac{C^2}{4(c_0 - C^{b_1}_{Lip})^2 N}.$$

B. Uniqueness of distribution μ_t now follows directly from the bound (15). Indeed, since distribution $\hat\mu^N_t \equiv \mathrm{Law}(X^1_t, \ldots, X^N_t)$ is unique, and μ^i_t has been chosen the same for any i, that is, it does not depend on N at all, then we have, $\mu^{i,N}_t \Longrightarrow \mu^i_t$ as $N \to \infty$. But clearly $\mu^{i,N}_t = \mu^{1,N}_t$ for any i, due to uniqueness of $\hat\mu^N_t$ and symmetry (remind that the initial data X^i_0 is the same for each i). Hence, the limit μ^i_t is indeed unique.

C. The following statement also follows directly from the bound (15), although it will not be used in the sequel.

Corollary 1 *Under assumptions of the Theorem 2, for any finite number of indices $i_1 < i_2 \ldots < i_k$,*

$$\left(X^{i_1,N}_t, \ldots, X^{i_k,N}_t\right) \Longrightarrow \left(\bar X^{i_1}_t, \ldots, \bar X^{i_k}_t\right), \quad N \to \infty, \tag{20}$$

uniformly with respect to $t \ge 0$, where the random variables in the right hand side are independent.

Indeed, all processes $\bar X^i$ are independent because W^i are, while finite dimensional convergence (20) follows from (15). This is a propagation of chaos type result, – the term suggested by M. Kac, – saying that different particles behave nearly independently if their total number is large.

D. Show that $\hat X^N_t = (X^{1,N}, \ldots, X^{N,N})$ is ergodic, possesses mixing, and hence, the law of $\hat X^N_t$ tends in TV topology to some limiting measure, μ^N_∞. Then this measure is stationary for the equation system (3). Naturally, this implies the convergence for projections, too, $\|\mu^{1,N}_t - \mu^{1,N}_\infty\|_{\mathrm{TV}} \to 0$ as $t \to \infty$.

We firstly show the ergodicity and mixing under (6), which is easier and shows clearly the idea. For this end it suffices to notice that the mixing condition (see [23]) for the large (Markov diffusion) process $\hat X^N \in \mathbb{R}^{dN}$,

$$\lim_{|\hat x^N| \to \infty} \langle \hat x^N, \hat b^N(\hat x^N)\rangle = -\infty, \tag{21}$$

with an obvious drift notation $\hat b^N(\hat x)$, follows directly from (6).

E. Show the same under assumption (7). This can be done by the method used in [22, 23] for couples of independent recurrent processes. Firstly, we need r to be large enough so that for some m greater than 2 the assertion of the Lemma 1 holds true. Although in this lemma the values r_m are not explicit, however, certain bounds which are linear in m are actually available for them, cf. [23].

The foundation of the approach to establishing beta-mixing as well as convergence rate to a (unique) equilibrium measure in total variation in [23] consists of two estimates,

$$\sup_{t \ge 0} E_{\hat X_0} |\hat X_t|^m 1(t < \tau) \le C(1 + |\hat X_0|^m),$$

and

$$E_{\hat{X}_0}\tau^{k+1} \le C(1+|\hat{X}_0|^m),$$

with $\tau := \inf(s: |\hat{X}_t| \le R)$, for R large enough, with the appropriate k.

In turn, for both bounds the main technical tool is the inequality $\langle \hat{x}, \hat{b}(\hat{x})\rangle \le -(r-\varepsilon) < 0$ for any fixed $\varepsilon > 0$ and for $|\hat{x}| > R$, where one can choose R to be arbitrary large. This inequality is inappropriate, however, in our case; instead we can apply the method used also in [22, 23] for *couples* of independent recurrent processes. The basic estimate for this (in the present setting) is

$$\lim_{|\hat{x}|\to\infty} \sum_{i=1}^{N} \langle \hat{x}^i, \hat{b}^i(\hat{x})\rangle\, |\hat{x}^i|^{m-2} = -\infty. \tag{22}$$

The latter follows from two remarks. (1) We notice that for each d-tuple of the form $(\hat{x}^{id+1}, \dots, \hat{x}^{(i+1)d})$, the value

$$\sum_{j=1}^{d} \hat{x}^{id+j}\hat{b}^{id+j}(\hat{x})$$

is bounded from above; moreover, it is negative once $|\hat{x}^i| = |(\hat{x}^{id+1}, \dots, \hat{x}^{(i+1)d})|$ is greater than some constant, R_0; finally, it approaches the value $-r$ or less if $|\hat{x}^i|$ is large enough. (2) Since we actually compare the values

$$\langle \hat{x}^i, \hat{b}^i(\hat{x})\rangle\, |\hat{x}^i|^{m-2},$$

and m is greater than two (see above), it remains to notice that as $|\hat{x}|$ tends to infinity, for each $1 \le i \le N$ the value $\langle \hat{x}^i, \hat{b}^i(\hat{x})\rangle\, |\hat{x}^i|^{m-2}$ either remains bounded, or tends to $-\infty$; and at least one of them does tend to $-\infty$. Hence, (22) holds true. The rest of the proof follows the calculus and arguments from [23], as we mentioned above, for the couples of independent processes and stopping time denoted in [23] by γ_n.

F. Now we use the double limit theorem for $\mu_t^{1,N}$, and the weakest among the two topologies, both being stronger than weak one. One limit is uniform due to the bound (15). All assertions of the theorem follow from the double limit theorem. Indeed, both convergence assertions are straightforward, and uniqueness of μ_∞ follows from uniqueness of $\hat{\mu}_t^N$. The limiting measure cannot depend on X_0 because this property has the measure $\hat{\mu}_\infty^N$. As long as the measure μ_∞ is invariant for the equation (1), it is unique as well, because we have convergence to this measure, while prelimiting distributions $\mu_\infty^{1,N}$, $N = 2, 3, \dots$, *are* unique. Finally, to show that μ_∞ is indeed invariant, it suffices to pass to the limit as $t \to \infty$ in the integral equality $(t, s \ge 0)$

$$X_{t+s} = X_t + \int_t^{t+s} b[X_r, \mu_r]\, dr + W_s^{\{t\}}, \quad W_s^{\{t\}} := W_{t+s} - W_t, \tag{23}$$

using Skorokhod's technique (see [17] or [11]), quite similarly to the passage from (13) to (14) indeed. We denote $Y_s^t = X_{s+t}$, then (23) reads,

$$Y_s^t = Y_0^t + \int_0^s b[Y_r^t, \nu_r^t]\, dr + W_s^{\{t\}}, \tag{24}$$

with notation $\nu_r^t = \mu_{t+r}$. Due to tightness, exactly as in the proof of the Theorem 1, we can choose a subsequence $t' \to \infty$ such that the couple $(Y_\cdot^{t'}, W_\cdot^{\{t'\}})$ weakly converges. Changing the probability space by the Skorokhod method, we can assume that the couple $(Y_\cdot^{t''}, W_\cdot^{\{t''\}})$ converges along some new subsequence $t'' \to \infty$ almost surely, $(Y_\cdot^{t''}, W_\cdot^{\{t''\}}) \to (Y_\cdot^0, W_\cdot^{\{0\}})$ in the weak topology in $C([0,\infty); R^d)$. In particular, $\nu_r^{t''} \Longrightarrow \nu_r^0$. But we already know that $\nu_r^{t''} = \mu_{t''+r} \Longrightarrow \mu_\infty$. Hence, from (24) rewritten on the new probability space, we get in the limit

$$Y_s^0 = Y_0^0 + \int_0^s b[Y_r^0, \nu_r^0]\, dr + W_s^{\{0\}},$$

where $W_\cdot^{\{0\}}$ is a new Wiener process in R^d, and $\nu_r^0 \equiv \mu_\infty$ is the distribution of Y_r^0 for any r. Thus, μ_∞ is an invariant measure indeed. $\square$

Remark Any of the two sets of assumptions in the Theorem 2 is rather strong indeed, as any implies a uniform bound (15). For the purposes of establishing the assertions of the theorem, this bound seems to be far too strong than needed. The author can hardly believe that there is no other way of establishing ergodic properties for the equation (1) which would not use this bound. However, this seems to be a fair description of the current state of the question.

Acknowledgements

This work was supported by Isaac Newton Institute for Mathematical Sciences, University of Cambridge, and grant INTAS-99-0590.

References

[1] Benachour, S.; Roynette, B.; Talay, D.; Vallois, P. Nonlinear self-stabilizing processes. I: Existence, invariant probability, propagation of chaos. Stochastic Processes Appl. 75, No.2, 173-201 (1998).

[2] Benachour, S.; Roynette, B.; Vallois, P. Nonlinear self-stabilizing processes. II: Convergence to invariant probability. Stochastic Processes Appl. 75, No.2, 203-224 (1998).

[3] Benedetto, D.; Caglioti, E.; Carrillo, J.A.; Pulvirenti, M. A non-Maxwellian steady distribution for one-dimensional granular media. J. Stat. Phys. 91, No.5-6, 979-990 (1998).

[4] Bossy, M., Talay, D. A stochastic particle method for the McKean-Vlasov and the Burgers equation. Math. Comput. 66, No.217, 157-192 (1997).

[5] Dawson, D. A.; Gärtner, J. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics 20, 247-308 (1987).

[6] Funaki, T. A certain class of diffusion processes associated with nonlinear parabolic equations. Z. Wahrscheinlichkeitstheor. Verw. Geb. 67, 331-348 (1984).

[7] Kac, M. Foundations of kinetic theory. Proc. 3rd Berkeley Sympos. Math. Statist. Probability 3, 171-197 (1956).

[8] Klokov, S. A. On lower bounds for mixing rates for a class of Markov Processes, submitted; preprint version at http://www.mathpreprints.com/math/Preprint/klokov/20020920/1/

[9] Klokov, S. A., Veretennikov, A. Yu. Polynomial mixing and convergence rate for a class of stochastic difference equations, submitted

[10] Kotelenez, P. A class of quasilinear stochastic partial differential equations of McKean-Vlasov type with mass conservation. Probab. Theory Relat. Fields 102, No.2, 159-188 (1995).

[11] Krylov, N. V. Controlled diffusion processes. NY et al., Springer (1980).

[12] Landau, L. D., Lifshitz, E. M. Course of theoretical physics, v. 10, Physical kinetics / by E. M. Lifshitz and L. P. Pitaevskii, Oxford, Pergamon (1981).

[13] Malrieu, F. Logarithmic Sobolev inequalities for some nonlinear PDE's http://www.mathpreprints.com/statistics/0106017/

[14] McKean, H.P. A class of Markov processes associated with nonlinear parabolic equations Proc. Natl. Acad. Sci. USA 56, 1907-1911 (1966).

[15] Nagasawa, M. Schrödinger equations and diffusion theory. Basel: Birkhäuser (1993).

[16] Shiga, T. Stationary distribution problem for interacting diffusion systems. Dawson, D. A. (ed.), Measure-valued processes, stochastic partial differential equations, and interacting systems. Providence, RI: American Mathematical Society. CRM Proc. Lect. Notes. 5, 199-211 (1994).

[17] Skorokhod, A. V. Studies in the theory of random processes. Reading, Mass.: Addison-Wesley (1965).

[18] Skorokhod, A. V. Stochastic equations for complex systems. Dordrecht (Netherlands) etc.: D. Reidel Publishing Company (1988).

[19] Sznitman, A.-S. Topics in propagation of chaos. Calcul des probabilités, Ec. d'été, Saint-Flour/Fr. 1989, Lect. Notes Math. 1464, 165-251 (1991).

[20] Tamura, Y. On asymptotic behavior of the solution of a non-linear diffusion equation. J. Fac. Sci., Univ. Tokyo, Sect. I A 31, 195-221 (1984).

[21] Veretennikov, A.Yu. On strong solutions and explicit formulas for solutions of stochastic integral equations. Math. USSR Sb. 39, 387-403 (1981)

[22] Veretennikov, A.Yu. On polynomial mixing bounds for stochastic differential equations. Stochastic Processes Appl. 70, No.1, 115–127 (1997).

[23] Veretennikov, A. Yu. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. (Russian) Teor. Veroyatnost. i Primenen. 44, 1999, no. 2, 312–327; English translation: Theory Probab. Appl. 44, no. 2, 361–374 (1999).