ver_mish_2m-7-4.tex

Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations

In memory of A. V. Skorokhod (10.09.1930 – 03.01.2011)

Yu. S. Mishura[*] & **A. Yu. Veretennikov**[†]

June 18, 2016

Abstract

New weak and strong existence and weak and strong uniqueness results for multi-dimensional stochastic McKean–Vlasov equation are established under relaxed regularity conditions.

1 Introduction

1.1 Setting

Solutions of the stochastic Itô–McKean–Vlasov (or, for short, simply McKean–Vlasov's) equation in R^d

$$dX_t = b[t, X_t, \mu_t]dt + \sigma[t, X_t, \mu_t]dW_t, \qquad X_0 = x_0, \tag{1}$$

[*]Taras Shevchenko National University of Kiev; myus@univ.kiev.ua

[†]University of Leeds, UK & National Research University "Higher School of Economics" and Institute for Information Transmission Problems, Moscow, Russia; a.veretennikov@ leeds.ac.uk

are considered, under the convention,

$$b[t,x,\mu] = \int b(t,x,y)\mu(dy), \quad \sigma[t,x,\mu] = \int \sigma(t,x,y)\mu(dy), \tag{2}$$

except for only one result – the Theorem 4 below – where σ does *not* depend on μ; notice that, however, we consider this Theorem 4 as the most important. Here W is a standard d_1-dimensional Wiener process, b and σ are vector and matrix Borel functions of corresponding dimensions d and $d \times d_1$, μ_t is the distribution of the process X at t. The initial data x may be random, but independent of W. Such a class of equations was proposed by M. Kac [11] as a stochastic "toy model" for the Vlasov kinetic equation of plasma. By a suggestion of Kac, the study of such equations was initiated by McKean [16]. We refer to [20] as an introduction to the whole area and to [5] as an important preceding background paper. It should be noticed that, although the paper [5] relates to purely non-stochastic issues, much of the technique from it has been used in further stochastic papers on the subject; in particular, this technique allows to tackle the case of a constant unit matrix σ of the equation (1) in dimension d practically without change. In other words, Dobrushin's deterministic technique suits well certain stochastic cases, at least, for equations with a constant diffusion.

Vlasov's idea – called mean field interaction in mathematical physics and stochastic analysis – assumes that for a large multiparticle ensemble with "weak interaction" between particles, this interaction for one particle with others may be effectively replaced by an averaged field. This leads to integration of this interaction with respect to the empirical distribution of the particles approximated by the distribution of the same particle itself. In fact, Vlasov has written in [26] his version of Maxwell's equations and not equations for Newtonian particle ensemble; the former equations are more involved. Nevertheless, the term "McKean–Vlasov" is already well established in mathematical physics and relates to the equations (1)–(2), so we will stick to this understanding.

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

$$\partial_t \mu_t = L^*(\mu_t)\mu_t, \tag{3}$$

with

$$L(t,x,\mu) = \frac{1}{2}(\sigma\sigma^*)[t,x,\mu]\frac{\partial^2}{\partial x \partial x} + b[t,x,\mu]\frac{\partial}{\partial x},$$

with a given initial value μ_0 in the sense that the distribution of X_t solves the equation (3) and the distribution of X_0 is μ_0 and provided that the process W is independent from X_0. For simplicity, initial value X_0 in the paper will be fixed, although a generalization to any initial measure with (or without) appropriate finite moments is possible and straightforward.

1.2 Motivation

Why are McKean–Vlasov's equations important? Except for the authority of Kac [11] and McKean [16], such equations naturally appear as limits for multi-particle or multi-agent systems – cf. [1, 2] – and in some other areas of high interest such as filtering, cf. [4]. These processes also very closely relate to so called self-stabilizing processes (diffusions, in particular), which is, actually, another name for non-linear diffusions in the "ergodic" situation, cf. [7]. Although in this paper the authors deliberately left aside the issues of ergodicity and propagation of chaos, their hope is that new results may help in these important areas in the future.

Why earlier existence and uniqueness results do not suffice? McKean–Vlasov's equations are clearly more involved than Itô's SDEs. Hence, it is not too surprising that the theory of the former equations is yet less developed in comparison to the latter. Even existence and uniqueness (as usual, weak or strong) still requires further studies. In particular, most of control problems lead to discontinuous coefficients. Hence, establishing existence and uniqueness under minimal or no regularity at all is in a big demand, indeed.

It may be noticed that much efforts in this area related to establishing approximation results, including time discretization and "propagation of chaos" for multi-particle case. In the authors' view, it could be more natural and hopefully fruitful to separate different aspects, hence, in particular, considering approximations differently from the more basic existence and uniqueness issues. In what concerns "propagation of chaos" for the equation (1), we refer the reader to [20] and [3, Theorem 4.3]; this direction will not be considered below.

We establish weak existence more general than in earlier papers; it may be called an analogue of Krylov's weak existence for Itô's equations. Notice that although it is not easy to compare our assumptions with those in [6], because the latter are given not just in terms of coefficients, but essentially in terms of existence of Lyapunov functions, (cf. with (2.1) in the Assumption I from [6]; also note that [6] tackles the martingale problem and not an SDE). More general growth conditions were studied in [3]; however, our regularity conditions admit just measurable coefficients and, hence, overall, our results are not covered by [3] either. Only the Theorem 2 below, which is presented mainly for completeness is some variation of the result from [3]; however, firstly, we consider formally a more general equation with a possibly non-square matrix σ, which may be useful in applications and which case was not covered in [3], and secondly, we propose a different method. Note that in the homogeneous case and under less general conditions but using a different technique, weak existence and weak uniqueness was established in [9] and [10]. Notice that in [24] there is a result on strong existence for the equation similar to (1) only with a unit matrix diffusion; however, strong and weak uniqueness – along with "propagation of chaos", i.e., with convergence of particle approximations – is established under additional assumptions on the drift which include Lipschitz and some other conditions. In this paper we do not touch the problem of particle approximations; yet, weak and strong uniqueness is established for bounded and measurable drifts under additional assumptions on the (variable) diffusion coefficient.

Strong existence in this paper is derived from strong existence for "ordinary" Itô's equations. Weak uniqueness and strong uniqueness are established in the section 4 below under identical (for weak and for strong) sets of conditions. The latter do involve some restriction on the diffusion coefficient, which may not depend on the measure in the Theorem 4 below. Notice that instead of the famous Yamada–Watanabe principle [27], [28], [8], [15] – that weak existence and pathwise uniqueness together imply strong existence – here weak existence implies directly strong uniqueness, while strong existence, is a much easier claim, which, generally speaking, may not necessarily imply strong uniqueness. Notice, however, that we do not claim that weak and pathwise uniqueness are fully equivalent without additional assumptions.

In some results of the paper we assume both drift and diffusion satisfying a linear growth bound condition; in others we assume boundedness of one or both coefficients, and diffusion coefficient is always non-degenerate except for the complementary Theorem 2. The linear growth is useful because of a numerous applications where, at least, the drift is not bounded; further extensions on a faster non-linear growth usually require Lyapunov type conditions, which are not considered in this paper. The nondegeneracy is quite a standard restriction, although in certain applications degeneracy may be highly desirable; however, the authors postpone considering more general cases to future papers.

1.3 The news and the structure of the paper

Let us summarize briefly what is new in this paper: (1) new extended weak existence under non-degeneracy assumptions; (2) a new method of establishing weak existence via approximations; (3) new strong existence; (4) new strong and weak uniqueness – the Theorem 4.

The structure of the paper is as follows. The section 1 is introductory. In the section 2 a new weak existence is established. In the section 3 strong existence is discussed. The section 4 is devoted to weak and strong uniqueness. Because of these three different topics, each section is divided into two parts - results and proofs - so that the proofs are provided after the main results in each section. For a completeness of the paper, a classical Skorokhod's lemma on convergence of stochastic integrals is provided in the Appendix.

2 Weak existence

2.1 Main results

Before we turn to the main results, let us recall a simple but useful fact from functional analysis, see, for example, [17, Theorem 2.6.8].

Proposition 1 *For any Borel function $f(z,y)$ and any probability measure $\mu(dy)$, the function $f[z,\mu] := \int f(z,y)\,\mu(dy)$ is a Borel function in z.*

Note that in most of textbooks on function analysis a similar statement is usually presented as a part of the Fubini theorem and only in the Lebesgue measurability format, which is not sufficient for the aims of this paper.

The next corollary shows why this Borel measurability is important here: it guarantees that the equation (1) is well-posed without any additional regularity assumptions on the coefficients.

Corollary 1 *Given the marginal distribution (μ_t), $t \geq 0$, the functions $\tilde{b}(t,x) := b[t,x,\mu_t]$ and $\tilde{\sigma}(t,x) := \sigma[t,x,\mu_t]$ are Borel in (t,x).*

Proof. In the sequel, measurability is always understood as a Borel one. We have, $\beta(t,x,\omega) := b(t,x,X_t)$ is a measurable function of (t,x,ω) as a composite function of two measurable ones: $b(t,x,y)$ and $X_t(\omega)$; recall that $X_t(\omega)$ is measurable with respect to ω and continuous with respect to t, hence, is measurable with respect to (t,ω). Thus, $b[t,x,\mu_t]$ may be understood as a result of integration of a measurable function of (t,x) and ω (i.e., $\beta(t,x,\omega)$) with respect to the measure $P(d\omega)$. In other words, we have

$$b[t,x,\mu_t] = \mathbb{E}\beta(t,x,\omega) \equiv \int \beta(t,x,\omega)\,\mathbb{P}(d\omega),$$

where the right hand side is measurable w.r.t. (t, x) due to the Proposition 1 with $z = (t, x)$, $y = \omega$, $\mu = P$, $f = \beta$, as required. The Corollary 1 is proved.

Recall what was said earlier: due to this Corollary, the equation (1) is well-posed since a substitution of any (weak) solution X_t in both coefficients leads to adapted and integrable in the sense of the standard Itô's integral processes.

In the next two theorems, the first one is new; the second one is just a new version of some earlier result presented here for completeness. The Theorem 1, in fact, mimics Krylov's weak existence result for Itô's SDEs – see [12] for a homogeneous case, and [14] and [25] for a non-homogeneous case – which does not assume any regularity of the coefficients. To the best of the authors' knowledge, the only previous result about strong existence for the Mckean–Vlasov stochastic equation with a (bounded) Borel measurable drift without any further regularity restriction has been established in [24] for the case of a *unit matrix diffusion coefficient*. In this respect, notice that the hypotheses in [6] and [3] do require certain continuity assumptions for weak existence. Also note that Krylov's weak existence theorems, in turn, use Skorokhod's single probability space method, cf. [12].

The second theorem is an existence result in the style of Skorokhod for Itô's SDEs with both continuous coefficients. It is also a variation of the Theorem 2.1 from [6] and the Theorem 3.6 from [3] on weak existence and is provided here only for completeness. Yet, there are some little news here related to dimensions and there is also some difference in the assumptions. On one hand, our conditions are a little bit more restrictive than those in [3] because our intention here was to give conditions explicitly in terms of properties of the original coefficients $b(t, x, y)$ and $\sigma(t, x, y)$, which is not the case of the result from [3]. On the other hand, unlike in the paper [3], we allow non-homogeneous coefficients, i.e., they may depend on time; a formal reduction of this case to a homogeneous one by considering a couple (t, X_t) would require unnecessary additional conditions due to the degeneracy. Our *method of proof* is also different from that used in [3]: we use explicitly Skorokhod's single probability space approach as well as Krylov's integral estimates for Itô's processes. It is a bit less evident how the Theorem 2 below may be compared to analogous results in [6] because both the conditions and the statement in the [6, Theorem 2.1] are formulated in different terms; in any case, apparently, they do not allow linear growth in y.

Theorem 1 *Suppose the following two conditions are both satisfied.*

1. *The functions b and σ admit linear growth condition in (x, y), i.e., there exists*

$C > 0$ such that

$$|b(s,x,y)| + \|\sigma(s,x,y)\| \le C(1 + |x| + |y|), \quad \forall\ s,x,y, \tag{4}$$

where $|\cdot|$ stands for the Euclidean norm in R^d for b and $\|\cdot\|$ for the $\|\sigma\| = \sqrt{\sum_{i,j}\sigma_{ij}^2}$.

2. Diffusion matrix σ is uniformly nondegenerate in the following sense:

$$\inf_{s,x,y}\inf_{|\lambda|=1} \lambda^*\sigma(s,x,y)\sigma(s,x,y)^*\lambda =: \sigma_0^2 > 0. \tag{5}$$

Then the equation (1) has a weak solution, that is, a solution on some probability space with a standard d_1-dimensional Wiener process with respect to some filtration $(\mathcal{F}_t,\ t \ge 0)$.

Recall that in the Theorem 1 no regularity of any coefficient is assumed.

Theorem 2 *Let the following two conditions be satisfied together.*

1. The functions b and σ of dimensions $d \times 1$ and $d \times d_1$ respectively satisfy linear growth condition (4).

2. Both coefficients are continuous in (x,y) for any t.

Then the equation (1) has a weak solution.

2.2 Proof of Theorem 1.

1. Firstly we establish the Theorem under a more restrictive assumption that $d_1 = d$, and that the matrix σ is symmetric and satisfies

$$\inf_{s,x,y}\inf_{|\lambda|=1} \lambda^*\sigma(s,x,y)\lambda = \sigma_0 > 0. \tag{6}$$

Exactly this assumption was assumed in [14]. In the end of the proof, all these additional restrictions will be dropped.

2. The proof is based on Krylov's integral estimate for any non-degenerate Itô process (not necessarily a solution of an SDE) with *bounded coefficients*,

$$\mathbb{E}\int_0^T f(t,X_t)dt \le N\|f\|_{L_{d+1}},$$

see [14]. Here the constant N may depend on d, T and the bounds for sup–norm of coefficients and inverse $\sigma\sigma^*$. This estimate will be applied to a *couple* of processes,

$$\mathbb{E}\int_0^T f(t,X_t,\hat{X}_t)dt \le N\|f\|_{L_{2d+1}},$$

where $\hat{X}_t$ is an independent copy of X_t and, hence, has exactly the same distribution; the constant N in the latter inequality depends on the same norms as earlier but now the dimension is $2d$ and, respectively, all norms for the new process may be different, however, still finite. Of course, in order to apply this estimate, we will need first to truncate all coefficients if they are not bounded from the beginning.

For bounded coefficients the hint is to smooth them so as to use existence theorems which are known in the literature (in particular, for continuous coefficients, see [6], and then to pass to the limit by using Skorokhod's single space method.

3. Now, let us firstly truncate coefficients – of course, with diffusion remaining non-degenerate – and only after this truncation let us smooth them. Eventually, we will have to check tightness of corresponding measures, which tightness should let us pass to the limit in the equation as $n \to \infty$. For each coordinate b^j of b we use the following truncation,

$$b^j_K(t,x,y) = b^j(t,x,y)1(|x|+|y| \le K),$$

while for σ,

$$\sigma_K(t,x,y) = \sigma(t,x,y)1(|x|+|y| \le K) + I\,1(|x|+|y| > K),$$

where I is a $d \times d$ unit matrix. Below, in this part of the proof we consider the case of b and σ **bounded** *and will not change notations* because of that. However, because for tightness we wish to have bounds uniform with respect to this truncation, recall that the assumptions (4) are still valid with a constant independent on the truncation parameter K.

Now, let us smooth both coefficients with respect to all variables, i.e., let

$$b^n(t,x,y) = b(t,x,y) * \psi_n(t) * \phi_n(x) * \phi_n(y),$$

where $\psi_n(t), \phi_n(x), \phi_n(y)$ are defined in a standard way, i.e., as non-negative C^∞ functions with a compact support, integrated to one, and so that this compact support squeezes to the one point set $\{O\}$ (the origin for the corresponding variable); or, in other words, that they are delta-sequences in the corresponding variables. Note that, of course, for every n the coefficients remain bounded due to the truncation; at the same time, without losing a generality, we may and will assume that the linear growth assumption (4) holds true with the same constant for each n; in reality this constant may increase a little bit still remaining uniformly bounded.

4. By standard estimates weak compactness of the triple $(X^n_t, \xi^n_t, W^n_t,\ t \ge 0)$ in $C(0,\infty; R^{3d})$ can be verified; the next natural step will be a return to non-regularized coefficients. In a standard way (see, e.g., [19], [14]) we get,

$$\sup_{0\le t\le T} \mathbb{E}|X^n_t|^2 \le C_T(1+|x_0|^2), \tag{7}$$

and also

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|X^n_t - X^n_s|^2 \le C_{T,x_0} h, \tag{8}$$

with constants C_t, C_{T,x_0} that does not depend on n. Recall that $x_0 \in R^d$ is the initial (non-random) value of the process X. Bounds similar to (7) and (8) hold true also for the component ξ^n and naturally for W^n. So, the sequence (X^n, ξ^n, W^n) is weakly compact (= tight) in $C([0,T; R^d])$.

Indeed, the inequality (8) follows straightforward from the bound

$$|X^n_t - X^n_s| \le |\int_s^t \sigma[r, X^n_r, \mu^n_r]\, dW^n_r| + |\int_s^t b[r, X^n_r, \mu^n_r]\, dr|,$$

because of the isometry property of stochastic integrals and by virtue of the inequality (7). Hence, for the convenience of the reader let us recall how to establish the inequality (7). In the sequel by $\mathbb{E}^3\sigma^n(s, X^n_s, \xi^n_s)$ we denote expectation with respect to the third variable ξ^n_s, i.e., expectation conditional on X^n_s, or, in other words, $\mathbb{E}^3\sigma^n(s, X^n_s, \xi^n_s) = \int \sigma^n(s, X^n_s, y)\mu^{\xi^n}_s(dy)$, where $\mu^{\xi^n}_s$ stands for the marginal distribution of ξ^n_s; likewise, $\mathbb{E}^3(\sigma^n(s, X^n_s, \xi^n_s) - \sigma^n(s, X_s, \xi_s))$ means simply $\int \sigma^n(s, X^n_s, y)\mu^{\xi^n}_s(dy) - \int \sigma^n(s, X^n_s, y)\mu^{\xi}_s(dy)$, where μ^{ξ}_s is the marginal distribution of ξ_s, and, finally, $\mathbb{E}^3|\sigma^n(s, X^n_s, \xi^n_s) - \sigma(s, X_s, \xi_s))|^2$ is understood as $\int |\sigma^n(s, X^n_s, y) - \sigma^n(s, X^n_s, y)|^2 \mu^{\xi^n,\xi}_s(dy, dy')$, where $\mu^{\xi^n,\xi}_s(dy, dy')$ denotes the marginal distribution of the couple (ξ^n_s, ξ_s). From

$$|X^n_t| \le |x_0| + |\int_0^t \sigma^n[r, X^n_r, \mu^n_r]\, dW^n_r| + |\int_0^t b^n[r, X^n_r, \mu^n_r]\, dr|,$$

we get,

$$\begin{aligned}
\frac{1}{3}\mathbb{E}|X^n_t|^2 &\le |x_0|^2 + \mathbb{E}(\int_0^t |b^n[s, X^n_s, \mu^n_s]|ds)^2 + (\int_0^t \mathbb{E}|\sigma^n[s, X^n_s, \mu^n_s]|^2 ds) \\
&\le |x_0|^2 + t(\mathbb{E}\int_0^t |b^n[s, X^n_s, \mu^n_s]|^2 ds) + (\int_0^t \mathbb{E}|\sigma^n[s, X^n_s, \mu^n_s]|^2 ds)
\end{aligned}$$

$$
\begin{aligned}
&= |x_0|^2 + t(\int_0^t \mathbb{E}|\mathbb{E}^3 b^n(s, X_s^n, \xi_s^n)|ds)^2 + (\int_0^t \mathbb{E}|\mathbb{E}^3 \sigma^n(s, X_s^n, \xi_s^n)|^2 ds) \\
&\le |x_0|^2 + t(\int_0^t \mathbb{E}\mathbb{E}^3 |b^n(s, X_s^n, \xi_s^n)|ds)^2 + (\int_0^t \mathbb{E}\mathbb{E}^3 |\sigma^n(s, X_s^n, \xi_s^n)|^2 ds) \\
&\le |x_0|^2 + t(\int_0^t \mathbb{E}\mathbb{E}^3 |b^n(s, X_s^n, \xi_s^n)|^2 ds) + (\int_0^t \mathbb{E}\mathbb{E}^3 |\sigma^n(s, X_s^n, \xi_s^n)|^2 ds) \\
&\le |x_0|^2 + Ct(\int_0^t \mathbb{E}\mathbb{E}^3 (1 + |X_s^n| + |\xi_s^n|)^2 ds) + C(\int_0^t \mathbb{E}\mathbb{E}^3 (1 + |X_s^n| + |\xi_s^n|)^2 ds)
\end{aligned}
$$

Since $\mathbb{E}^3|\xi_s^n|^k = \mathbb{E}|\xi_s^n|^k$ and $\mathbb{E}^3|\xi_s^n|^k = \mathbb{E}|X_s^n|^k$ for any k, we obtain,

$$
\frac{1}{3}\mathbb{E}|\xi_t^n|^2 \le |x_0|^2 + C(t+1)\int_0^t \mathbb{E}(1 + |\xi_s^n|)^2 ds,
$$

or, equivalently,

$$
\frac{1}{3}\mathbb{E}|X_t^n|^2 \le |x_0|^2 + C(t+1)\int_0^t \mathbb{E}(1 + |X_s^n|)^2 ds,
$$

Here all expectations are finite and C does not depend neither on truncations, nor on n. So, we can use Gronwall's lemma with constants that are the same for all truncations to obtain the following:

$$
\sup_{0\le t\le T} \mathbb{E}|X_t^n|^2 \le 3(|x_0|^2 + CT + CT^2)\exp(CT(T+1)),
$$

which justifies (7).

Now, due to this weak compactness implied by (7)–(8) and by virtue of Skorokhod's Theorem about single probability space and convergence in probability (see [19, §6, ch. 1], or [14, Lemma 2.6.2]), without loss of generality we may and will assume that not only $\mu^n \Longrightarrow \mu$, but also on some probability space for any t,

$$
(X_t^n, \xi_t^n, W_t^n) \stackrel{\mathbb{P}}{\to} (X_t, \xi_t, W_t), \quad n \to \infty,
$$

generally speaking, over a sub-sequence. Now, to be fully rigorous with notations we should have redenoted all our processes; however, with a slight abuse of notations we will not do it. Also, notice – this will be important in the sequel – that without loss of generality we may and will assume that each process $(\xi_t^n,\ t \ge 0)$ for any $n \ge 1$

is independent from (X^n, W^n), as well as their limit ξ_t may be chosen independent from the limits $(X{\cdot}W)$; the processes (W_t^n) are the Wiener ones, and Itô's stochastic integrals are all well-defined. See the details in the proof of the Theorem 2.6.1 in [14]. We could have also introduced Wiener processes for ξ_t^n, but they will not show up in this proof. For what follows, let us fix an arbitrary $T > 0$ and consider $t \in [0, T]$.

5. Now the next task is to pass to the limit in the integral equality,

$$X_t^n = x_0 + \int_0^t b^n[s, X_s^n, \mu_t^n]\, ds + \int_0^t \sigma^n[s, X_s^n, \mu_t^n]\, dW_s^n, \quad 0 \le t \le T. \tag{9}$$

Note that due to the earlier truncation – see the item 3 of the proof – and because of the compact support of the functions ϕ_n and ψ_n and the condition of squeezing of this support,

$$b^n(t, x, y) \equiv b(t, x, y), \qquad |x| + |y| > C + 1,$$

and

$$\sigma^n(t, x, y) \equiv \sigma(t, x, y) \equiv I, \qquad |x| + |y| > C + 1,$$

at least, for n large enough. Now, we will be using Krylov's approach as in [14] so as to establish the required convergence. We have, by Minkowskii's inequality,

$$\begin{aligned}
&\mathbb{E}|\int_0^t (\mathbb{E}^3 b^n(s, X_s^n, \xi_s^n) - \mathbb{E}^3 b(s, X_s, \xi_s))ds| \\
&\le \mathbb{E}\int_0^t \mathbb{E}^3 |b^n(s, X_s^n, \xi_s^n) - b(s, X_s, \xi_s))|ds \\
&= \mathbb{E}\mathbb{E}^3 \int_0^t |b^n(s, X_s^n, \xi_s^n) - b(s, X_s, \xi_s))|ds,
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\mathbb{E}|\int_0^t (\mathbb{E}^3 \sigma^n(s, X_s^n, \xi_s^n) - \mathbb{E}^3 \sigma(s, X_s, \xi_s))dW_s|^2 \\
&\le \mathbb{E}\int_0^t \mathbb{E}^3 |\sigma^n(s, X_s^n, \xi_s^n) - \sigma(s, X_s, \xi_s))|^2 ds.
\end{aligned}$$

Here are the details. Let $\epsilon > 0$. Let us find $R > 0$ and n_0 such that

$$\sup_{n \ge n_0} \mathbb{P}(\sup_{0 \le s \le t} |(X_s^n, \xi_s^n)| > R) \le \epsilon, \tag{10}$$

and for $n \geq n_0$,

$$\|b^n(\cdot) - b(\cdot)\|_{L_{2d+1}([0,T]\times B_R^{(2d)})} + \|\sigma^n(\cdot) - \sigma(\cdot)\|^2_{L_{2(2d+1)}([0,T]\times B_R^{(2d)})} < \varepsilon, \qquad (11)$$

where $B_R^{(2d)} = \{(x,y) : x, y \in R^d, |(x,y)| \leq R\}$. Indeed, it is well-known that smoothed functions converge to their originals in any L_p space on any bounded domain. Since b and σ are bounded, they belong to every $L_p([0,T]\times B_R^{(2d)})$ for any $T, R > 0$. Hence, (11) holds true as required.

Denote $T_R := \inf(t \geq 0 : \ |(X_t, \xi_t)| \geq R)$ and $T_R^n := \inf(t \geq 0 : \ |(X_t^n, \xi_t^n)| \geq R)$. We have,

$$\begin{aligned}
&\mathbb{E}\int_0^t |b^n[s, X_s^n, \mu_s^n] - b[s, X_s, \mu_s]|ds \\
&= \mathbb{E}\int_0^t \left|\mathbb{E}^3 b^n(s, X_s^n, \xi_s^n) - \mathbb{E}^3 b(s, X_s, \xi_s)\right| \, ds \\
&\leq \mathbb{E}\mathbb{E}^3 \int_0^t |b^n(s, X_s^n, \xi_s^n) - b(s, X_s, \xi_s)| \, ds \\
&\leq \mathbb{E}\mathbb{E}^3 \int_0^t |b^n(s, X_s^n, \xi_s^n) - b^{n_0}(s, X_s^n, \xi_s^n)| \, ds \\
&+ \mathbb{E}\mathbb{E}^3 \int_0^t |b^{n_0}(s, X_s^n, \xi_s^n) - b^{n_0}(s, X_s, \xi_s)| \, ds \\
&+ \mathbb{E}\mathbb{E}^3 \int_0^t |b^{n_0}(s, X_s, \xi_s) - b(s, X_s, \xi_s)| \, ds. \\
&\equiv J_1 + J_2 + J_3.
\end{aligned}$$

Due to Krylov's estimate applied to (X^n, ξ^n) and by virtue of (10) with (11) we have,

$$\begin{aligned}
J_1 &= \mathbb{E}\mathbb{E}^3 \int_0^t |b^n(s, X_s^n, \xi_s^n) - b^{n_0}(s, X_s^n, \xi_s^n)|ds \\
&= \mathbb{E}\int_0^t |b^n(s, X_s^n, \xi_s^n) - b^{n_0}(s, X_s^n, \xi_s^n)|ds \\
&= \mathbb{E}\int_0^{t\wedge T_R^n} |b^n(s, X_s^n, \xi_s^n) - b^{n_0}(s, X_s^n, \xi_s^n)|ds
\end{aligned}$$

$$+\mathbb{E}\int_{t\wedge T^n_R}^{t} |b^n(s, X^n_s, \xi^n_s) - b^{n_0}(s, X^n_s, \xi^n_s)| ds$$

$$\le K_{2d+1}\|b^n - b^{n_0}\|_{L_{2d+1}([0,T]\times B^{(2d)}_R)} + 2C\epsilon \le \epsilon(2C + K_{2d+1}).$$

Similarly, for J_3 by Krylov's estimate applied to (X, ξ),

$$J_3 = \mathbb{E}\mathbb{E}^3\int_0^t |b^{n_0}(s, X_s, \xi_s) - b(s, X_s, \xi_s)| ds$$

$$= \mathbb{E}\int_0^t |b^n(s, X_s, \xi_s) - b^{n_0}(s, X_s, \xi_s)| ds$$

$$= \mathbb{E}\int_0^{t\wedge T_R} |b^n(s, X_s, \xi_s) - b^{n_0}(s, X_s, \xi_s)| ds$$

$$+\mathbb{E}\int_{t\wedge T_R}^{t} |b^n(s, X_s, \xi_s) - b^{n_0}(s, X_s, \xi_s)| ds$$

$$\le K_{2d+1}\|b^n - b^{n_0}\|_{L_{2d+1}([0,T]\times B^{(2d)}_R)} + 2C\epsilon \le \epsilon(2C + K_{2d+1}).$$

Finally, $J_2 \to 0$ as $n \to \infty$ by Lebesgue's dominated convergence theorem.

6. For stochastic integrals the reasoning is similar, although, instead of Lebesgue's dominated convergence we have to use a special tool about convergence of stochastic integrals established in [19, ch.2, §3] (see also [14, Lemma 2.6.3] without proof); we formulate it in the Appendix for the reader's convenience.

Further, we estimate,

$$\int_0^T \sigma^n[s, X^n_s, \mu^n_s]\, dW^n_s - \int_0^T \sigma[s, X_s, \mu_s]\, dW_s$$

$$= \int_0^T \mathbb{E}^3\sigma^n(s, X^n_s, \xi^n_s)\, dW^n_s - \int_0^T \mathbb{E}^3\sigma^{n_0}(s, X^n_s, \xi^n_s)\, dW^n_s$$

$$+ \int_0^T \mathbb{E}^3\sigma^{n_0}(s, X^n_s, \xi^n_s)\, dW^n_s - \int_0^T \mathbb{E}^3\sigma^{n_0}(s, X_s, \xi_s)\, dW_s$$

$$+ \int_0^T (\mathbb{E}^3\sigma^{n_0}(s, X_s, \xi_s) - \mathbb{E}^3\sigma(s, X_s, \xi_s))\, dW_s.$$

$$\equiv \tilde J_1 + \tilde J_2 + \tilde J_3.$$

By virtue of Skorokhod's result (see the Lemma 1 in the Appendix), we have $\tilde{J}_2 \to 0,\ n \to \infty$. Indeed, in our case $f^n_s = \sigma^{n_0}(s, X^n_s, \xi^n_s)$ (cf. notations in the Lemma 1), these processes are all uniformly bounded due to smoothing, and finally the condition (26) of the Lemma holds true because

$$\lim_{h\to 0} \sup_n \sup_{|s-t|\le h} \mathbb{P}\{|\sigma^{n_0}(s, X^n_s, \xi^n_s) - \sigma^{n_0}(t, X^n_t, \xi^n_t)| > \varepsilon\} = 0, \tag{12}$$

due to (7) and uniform continuity of σ^{n_0} on any compact. Both other terms are estimated with the help of Krylov's bound as follows.

$$\begin{aligned}
\mathbb{E}|\tilde{J}_1|^2 &= \mathbb{E}\int_0^T \|\sigma^n(s, X^n_s, \mu^n_s) - \sigma^{n_0}(s, X^n_s, \mu^n_s)\|^2\, ds \\
&= \mathbb{E}\int_0^T \|\mathbb{E}^3\sigma^n(s, X^n_s, \xi^n_s) - \mathbb{E}^3\sigma^{n_0}(s, X^n_s, \xi^n_s)\|^2\, ds \\
&\le \mathbb{E}\mathbb{E}^3\int_0^T \|\sigma^n(s, X^n_s, \xi^n_s) - \sigma^{n_0}(s, X^n_s, \xi^n_s)\|^2\, ds \\
&\le K_{2d+1}\|\sigma^n - \sigma^{n_0}\|^2_{L_{2(2d+1)}([0,T]\times B_R)} + 4C^2\epsilon \le \epsilon(2K_{2(2d+1)} + 4C^2).
\end{aligned}$$

Similarly, for the last term we have

$$\begin{aligned}
\mathbb{E}|\tilde{J}_3|^2 &= \mathbb{E}\int_0^T \|\sigma^{n_0}(s, X_s, \mu_s) - \sigma(s, X_s, \mu_s)\|^2\, ds \\
&= \mathbb{E}\int_0^T \|E^3\sigma^{n_0}(s, X_s, \xi_s) - \mathbb{E}^3\sigma(s, X_s, \xi_s)\|^2\, ds \\
&\le \mathbb{E}\mathbb{E}^3\int_0^T \|\sigma^{n_0}(s, X_s, \xi_s) - \sigma(s, X_s, \xi_s)\|^2\, ds \\
&\le K_{2d+1}\|\sigma^{n_0} - \sigma\|^2_{L_{2(2d+1)}([0,T]\times B_R)} + 4C^2\epsilon \le \epsilon(K_{2(2d+1)} + 4C^2).
\end{aligned}$$

Hence, it now follows that the triple (X_t, ξ_t, W_t) is a solution of the corresponding limiting SDE

$$X_t = x_0 + \int_0^t \mathbb{E}^3 b(s, X_s, \xi_s)\, ds + \int_0^t \mathbb{E}^3\sigma(s, X_s, \xi_s) dW_s,$$

that is,

$$X_t = x_0 + \int_0^t b[s, X_s, \mu_s]\, ds + \int_0^t \sigma[s, X_s, \mu_s]\, dW_s,$$

as required.

7. Now let us return to the non-truncated coefficients. To this end, we need a priori estimates of some moments for solutions with truncated ones – denote them temporarily by X^N, μ^N and ξ^N – preferably uniform with respect to N. We shall see that such estimates do exist. Indeed, partially repeating the calculus in the step 4, we have,

$$\frac{1}{3}\mathbb{E}|X_t^N|^2 \le |x_0|^2 + (\int_0^t \mathbb{E}|b[s, X_s^N, \mu_s^N]|ds)^2 + (\int_0^t \mathbb{E}|\sigma[s, X_s^N, \mu_s^N]|^2 ds)$$
$$= |x_0|^2 + (\int_0^t \mathbb{E}|\mathbb{E}^3 b(s, X_s^N, \xi_s^N)|ds)^2 + (\int_0^t \mathbb{E}|\mathbb{E}^3 \sigma(s, X_s^N, \xi_s^N)|^2 ds)$$
$$\le |x_0|^2 + (\int_0^t \mathbb{E}\mathbb{E}^3|b(s, X_s^N, \xi_s^N)|ds)^2 + (\int_0^t \mathbb{E}\mathbb{E}^3|\sigma(s, X_s^N, \xi_s^N)|^2 ds)$$
$$\le |x_0|^2 + t(\int_0^t \mathbb{E}\mathbb{E}^3|b(s, X_s^N, \xi_s^N)|^2 ds) + (\int_0^t \mathbb{E}\mathbb{E}^3|\sigma(s, X_s^N, \xi_s^N)|^2 ds)$$
$$\le |x_0|^2 + Ct(\int_0^t \mathbb{E}\mathbb{E}^3(1 + |X_s^N| + |\xi_s^N|)^2 ds) + C(\int_0^t \mathbb{E}\mathbb{E}^3(1 + |X_s^N| + |\xi_s^N|)^2 ds)$$

Since $\mathbb{E}^3|\xi_s^N|^k = \mathbb{E}|\xi_s^N|^k$ and $\mathbb{E}^3|\xi_s^N|^k = \mathbb{E}|X_s^N|^k$ for any k, we obtain,

$$\frac{1}{3}\mathbb{E}|\xi_t^N|^2 \le |x_0|^2 + C(t+1)\int_0^t \mathbb{E}(1 + |\xi_s^N|)^2 ds).$$

Here all expectations are finite. So, we can use Gronwall's lemma with constants that do not involve N.

$$\sup_{0\le s\le t} \mathbb{E}|X_s^N|^2 \le 3|x_0|^2 + 3C(t+1)\int_0^t \mathbb{E}(1 + |X_s^N|)^2 ds).$$

From here,

$$\sup_{0\le t\le T} \mathbb{E}|X_t^N|^2 \le 3(|x_0|^2 + CT + CT^2)\exp(CT(T+1)), \tag{13}$$

where C does not depend on N, as required.

8. For $t \le T$, we similarly obtain

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|X_t^N - X_s^N|^2 \le C_T(1 + |x_0|^2)h.$$

This signifies that (μ^N) is weakly compact (tight). Hence, again by using Skorokhod's lemma, we may choose a subsequence so as to pass to the limit as $N\to\infty$ from

$$X_t^N = x_0 + \int_0^t \mathbb{E}^3 b^N(s, X_s^N, \xi_s^N)ds + \int_0^t \mathbb{E}^3\sigma^N(s, X_s^N, \xi_s^N)dW_s^N,$$

to

$$X_t = x_0 + \int_0^t \mathbb{E}^3 b(s, X_s, \xi_s)ds + \int_0^t \mathbb{E}^3\sigma(s, X_s, \xi_s)dW_s,$$

i.e., to

$$X_t = x_0 + \int_0^t b[s, X_s, \mu_s]ds + \int_0^t \sigma[s, X_s, \mu_s]dW_s.$$

Thus, weak solution of the equation (1) exists in the case of $d_1 = d$ and under a more restrictive assumption (6) instead of (5). Recall that once μ_t is the distribution of ξ_t, and distributions of ξ_t and X_t coincide, then μ_t is also the distribution of X_t.

9. Now we will show how to drop the assumption (6) and, in particular, the condition $d_1 = d$. We will use a hint from [25, section 4]; however, due to a more involved structure of the equation and its coefficients in this paper, we repeat the details here. Denote $\tilde\sigma(t,x,y) := \sqrt{a(t,x,y)}$, where $a(t,x,y) := \sigma(t,x,y)\sigma^*(t,x,y)$, $a[t,x,\mu] := \sigma[t,x,\mu]\sigma^*[t,x,\mu]$, and let $\tilde X_t$ be a solution of the equation,

$$\tilde X_t = x + \int_0^t b[s, \tilde X_s, \mu_s]ds + \int_0^t \tilde\sigma[s, \tilde X_s, \mu_s]d\tilde W_s,$$

with some d-dimensional Wiener process $(\tilde W_t,\, t\ge 0)$ on some probability space and where μ_s stands for the distribution of $\tilde X_s$. Without losing a generality we may and will assume that on the same probability space there exists another *independent* d_1-dimensional Wiener process $(\bar W_t,\, t\ge 0)$. Let I denote a $d_1\times d_1$-dimensional unit matrix and let

$$p[s,x,\mu] = \tilde\sigma[s,x,\mu]^{-1}\,\sigma[s,x,\mu].$$

Note that (dropping the arguments for the brevity of presentation in some cases)

$$p^*p[s,x,\mu] = \sigma^*(\tilde\sigma[s,x,\mu]^*)^{-1}\tilde\sigma[s,x,\mu]^{-1}\,\sigma[s,x,\mu] = \sigma^*(a)^{-1}\sigma,$$

$$p^*pp^*p = \sigma^*(a)^{-1}\sigma\sigma^*(a)^{-1}\sigma = \sigma^*(a)^{-1}(a)(a)^{-1}\sigma = \sigma^*(a)^{-1}\sigma,$$

and let

$$W_t^0 := \int_0^t p^*[s, \tilde X_s, \mu_s]\, d\tilde W_s + \int_0^t (I - p^*[s, \tilde X_s, \mu_s]p[s, \tilde X_s, \mu_s])\, d\bar W_s.$$

Notice that

$$\sigma p^* = a(a)^{-1/2} = (a)^{1/2},$$

$$\sigma p^* p = (a)^{1/2} p = (a)^{1/2}(a)^{-1/2}\sigma = \sigma.$$

Due to the multivariate Lévy characterization theorem this implies that W^0 is a d_1-dimensional Wiener process, since its matrix angle characteristic (also known as a matrix angle bracket) equals

$$\langle W^0, W^0\rangle_t = \int_0^t p^* p\, ds + \int (I - p^* p)^* (I - p^* p)\, ds$$

$$= \int (p^* p + I - 2p^* p + p^* p p^* p)\, ds = \int (I - p^* p + p^* p p^* p)\, ds$$

$$= \int (I - \sigma^* (a)^{-1}\sigma + \sigma^* (a)^{-1}(a)(a)^{-1}\sigma)\, ds = \int_0^t I\, ds = t\, I.$$

Next, due to the stochastic integration rules (see [8]),

$$\int_0^t \sigma[s, \tilde X_s, \mu_s]\, dW_s^0 = \int \sigma p^*[s, \tilde X_s, \mu_s]\, d\tilde W + \int \sigma(I - p^* p)[s, \tilde X_s, \mu_s]\, d\bar W$$

$$= \int (a)^{1/2}[s, \tilde X_s, \mu_s]\, d\tilde W = \int \tilde\sigma[s, \tilde X_s, \mu_s]\, d\tilde W = \tilde X_t - x - \int_0^t b[s, \tilde X_s, \mu_s]\, ds.$$

In other words, $(\tilde X, W^0)$ is a (weak) solution of the equation (1):

$$\tilde X_t - x - \int_0^t b[s, \tilde X_s, \mu_s]\, ds - \int_0^t \sigma[s, \tilde X_s, \mu_s]\, dW_s^0 = 0.$$

It remains to notice that since we did not change measures, μ_s is still the distribution of $\tilde X_s$ by the assumption. So, the proof of the Theorem 1 is completed.

Remark. Notice that due to the Fatou lemma we have from (13),

$$\mathbb{E}_x \sup_{0\le t\le T} |X_t|^2 \le 3(|x|^2 + CT + CT^2)\exp(CT(T+1)). \qquad (14)$$

2.3 Proof of Theorem 2

1. In the case $d_1 = d$ it is possible to truncate and smooth the coefficients like in the previous proof and then to pass to the limit in the equation (9) as in the Skorokhod's Lemma 1 from the Appendix, using only continuity of the coefficients.

2. For $d_1 > d$ the assertion follows in the same way as in the step 9 of the proof of the previous result. The Theorem 2 is proved.

3 Strong existence

3.1 Main result

In this section it is shown that strong solution exists under appropriate conditions. The first version mimics strong existence for Itô's equations by Itô himself (see, e.g., [8]); notice that it requires Lipschitz condition in x, but not in y; on the other hand, we do not claim strong *uniqueness* here, unlike in the classical Itô result. The second version mimics strong existence for Itô's equations from [22] and [23], but not in y; on the other hand, emphasize that we do not claim strong *uniqueness* in this theorem, but only strong existence. We also notice for interested readers that in [23] the assumption of continuity in time was dropped in comparison to [22].

Theorem 3 *Let either of the following two conditions hold true.*

1. *Let the coefficients b and σ satisfy a linear growth condition (4) in (x, y) uniformly with respect to s and let them also satisfy the following Lipschitz condition in x uniformly with respect to s and locally with respect to y,*

$$|b(t, x, y) - b(t, x', y)| + \|\sigma(t, x, y) - \sigma(t, x', y)\| \le C(1 + |y|^2)|x - x'|. \quad (15)$$

2. *Let the coefficients b and σ satisfy a linear growth estimate in x uniformly with respect to s, y and let σ be Lipschitz in x in the sense of (15) and uniformly nondegenerate.*

Then the equation (1) has a strong solution and, moreover, every solution is strong and, in particular, solution may be constructed on any probability space equipped with a d_1-dimensional Wiener process.

Remark. Under such assumptions, strong (= pathwise) uniqueness still remains an open problem.

Proof. **1.** The proof is based on the result from [22] about strong solutions for SDEs and Borel drift, bounded or with a linear growth condition (see [22] and [23]; in the latter reference the assumption on continuity of the diffusion matrix with respect to t is dropped). Some (weak) solution exists, and whatever is its distribution μ, the process X may be considered as an ordinary SDE with coefficients depending on time,

$$\tilde{b}(t,x) = b[t,x,\mu_t], \qquad \tilde{\sigma}(t,x) = \sigma[t,x,\mu_t],$$

and, hence,

$$dX_t = \tilde{b}(t,X_t)dt + \tilde{\sigma}(t,X_t)dW_t, \qquad X_0 = x. \tag{16}$$

2a. Now we ought to verify that new coefficients $\tilde{b}$ and $\tilde{\sigma}$ satisfy (local) linear growth and Lipschitz conditions. We have, for any $T > 0$ and $0 \le t \le T$,

$$|\tilde{b}(t,x)| = |b[t,x,\mu_t]| = |\int b(t,x,y)\,\mu_t(dy))|$$
$$\le C|\int (1+|x|+|y|)\,\mu_t(dy))| \le C_T\,(1+|x|),$$

due to the moment estimate (14) above. Similarly, we show that also

$$\|\tilde{\sigma}(t,x)\| \le C\int (1+|x|+|y|)\,\mu_t(dy)) \le C_T\,(1+|x|).$$

2b. Further, we estimate, by virtue of the same moment estimate (14),

$$|\tilde{\sigma}(t,x) - \tilde{\sigma}(t,x')| = |\sigma[t,x,\mu_t] - \sigma[t,x',\mu_t]|$$
$$= |\int \sigma(t,x,y)\,\mu_t(dy)) - \int \sigma(t,x',y)\,\mu_t(dy))|$$
$$\le C\,|x-x'|\int (1+|y|^2)\,\mu_t(dy)) \le C_T\,|x-x'|.$$

Under the assumption 1, we similarly establish Lipschitz condition for the drift coefficient.

3. Now, under either set of conditions 1 or 2 of the Theorem above, the latter equation has a strong solution due to Itô's or "Zvonkin-Veretennikov's theorems" on $[0,T]$, with any $T > 0$. The Theorem 3 is proved.

Remark. Notice that as a solution of the equation (16), X is pathwise unique, but so far it is not known if this implies the same property for X as a solution of (1), unless *weak uniqueness* for the equation (1) has been established. In a restricted framework this will be done in the next Theorem.

Remark. In the case of dimension one, Lipschitz condition may be relaxed to Hölder of order 1/2 and, actually, a little bit further by using techniques from [27] and [21].

4 Strong and weak uniqueness

4.1 Main result

In this section it will be shown that in certain cases weak uniqueness implies strong uniqueness for the equation (1) and also both properties will be established under appropriate conditions. This result – the Theorem 4 below – requires only a measurability of the drift with respect to the state variable x, but assumes that diffusion σ does not depend on y along with Lipschitz in x and nondegeneracy.

The next result is the main theorem of the paper. The linear growth condition is a bit different from (4).

Theorem 4 *Assume the functions b and σ are Borel measurable,*

$$\sigma(s,x,y) \equiv \sigma(s,x),$$

that is, σ does not depend on the variable y; let σ satisfy the non-degeneracy assumption (5); let $d_1 = d$ and let there exist $C > 0$ such that the function $\tilde{b}(s,x,y) := \sigma^{-1}(s,x)\, b(s,x,y)$ is bounded. Assume that the equation

$$dX^0_t = \sigma(t, X^0_t)\, dW_t, \quad X^0_0 = x, \tag{17}$$

has a unique strong solution for any x. Then solution of the equation (1) is weakly and strongly unique.

Recall that no regularity on b is needed here; however, the price is a special form of σ which should not depend on the measure. Denote by X^0_t the solution of the Itô equation (17).

Remark. About strong solutions of Itô's equation (17) see [22, 23].

Corollary 2 *If the assumptions of the Theorem 4 holds, then there is a strong solution, which is unique in distribution and pathwise unique.*

4.2 Proof of Theorem 4

1. We start with the case where both b and σ are bounded; later this restriction will be dropped. Note that under the assumptions of the theorem, any solution is strong by virtue of the Theorem 3. Hence, it suffices to show weak uniqueness, after which strong uniqueness will follow from strong uniqueness for the equation (16). We will do it by contradiction. Suppose there are two solutions X^1 and X^2 with distributions μ^1 and μ^2 respectively in the space of trajectories $C[0,\infty; R^d]$. Without a loss of generality, we may and will assume that both processes X^1 and X^2 are realized on the same probability space and with the same Wiener process. It will be shown that $\mu^1 = \mu^2$ and $X^1 = X^2$ a.s. Note that both solutions are Markov processes (see [13]).

Both solutions (X^i, μ^i) may be obtained from the same Wiener process W via Girsanov's transformations using the following stochastic exponents:

$$\gamma^i = \exp(-\int_0^T \tilde{b}[s, X^0_s, \mu^i_s]\, dW_s - \frac{1}{2}\int_0^T |\tilde{b}[s, X^0_s, \mu^i_s]|^2\, ds), \quad i = 1, 2,$$

where $\tilde{b}(t,x,y) := \sigma^{-1}(t,x)\, b(t,x,y)$, $|\tilde{b}|$ stands for the module of the vector $\tilde{b}$, and $\tilde{b}[s, X^0_s, \mu^i_s]\, dW_s$ is understood as a scalar product.

Denote

$$\tilde{W}_t := W_t + \int_0^t \tilde{b}(s, X^0_s, \mu^1_s)\, ds, \quad 0 \le t \le T.$$

This is a new Wiener process on $[0,T]$ under the probability measure P^{γ^1} defined by its density as $(dP^{\gamma^1}/dP)(\omega) = \gamma^1$. Note that

$$d\tilde{W}_t = dW_t + \tilde{b}(t, X^0_t, \mu^1_s)\, dt, \ \text{ or, } \ dW_t = d\tilde{W}_t - \tilde{b}(t, X^0_t, \mu^1_s)\, dt.$$

Recall that $\tilde{b} = b\sigma^{-1}$ is bounded. Then the density of μ^2 with respect to μ^1 equals

$$\rho = \exp(-\int_0^T (\tilde{b}[s, X_s^0, \mu_s^2] - \tilde{b}[s, X_s^0, \mu_s^1]) d\tilde{W}_s$$
$$-\frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^2] - \tilde{b}[s, X_s^0, \mu_s^1]|^2 ds).$$

For the convenience of the reader we show the details below. Indeed, since for $i = 1, 2$

$$\frac{dP^{\gamma^i}}{dP}(\omega) = \gamma^i = \exp(-\int_0^T \tilde{b}[s, X_s^0, \mu_s^i]\, dW_s - \frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^i]|^2\, ds),$$

we conclude that

$$\frac{dP^{\gamma^2}}{dP^{\gamma^1}}(\omega) = \frac{\gamma^2}{\gamma^1}$$
$$= \exp(-\int_0^T \tilde{b}[s, X_s^0, \mu_s^2]\, dW_s - \frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^2]|^2\, ds)$$
$$\times \exp(+\int_0^T \tilde{b}[s, X_s^0, \mu_s^1]\, dW_s + \frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^1]|^2\, ds)$$
$$= \exp(-\int_0^T (\tilde{b}[s, X_s^0, \mu_s^2] - \tilde{b}[s, X_s^0, \mu_s^1])\, dW_s$$
$$\times \exp(-\frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^2]|^2\, ds + \frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^1]|^2\, ds)$$
$$= \exp(-\int_0^T (\tilde{b}[s, X_s^0, \mu_s^2] - \tilde{b}[s, X_s^0, \mu_s^1])\, (d\tilde{W}_s - \tilde{b}(s, X_s^0, \mu_s^1)\, ds)$$
$$\times \exp(-\frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^2]|^2\, ds + \frac{1}{2}\int_0^T |\tilde{b}[s, X_s^0, \mu_s^1]|^2\, ds)$$
$$= \exp(-\int_0^T (\tilde{b}[s, X_s^0, \mu_s^2] - \tilde{b}[s, X_s^0, \mu_s^1])\, d\tilde{W}_s$$
$$\times \exp(\int_0^T (-\frac{1}{2}|\tilde{b}[s, X_s^0, \mu_s^2]|^2 + \tilde{b}[s, X_s^0, \mu_s^1]\tilde{b}[s, X_s^0, \mu_s^2] - \frac{1}{2}|\tilde{b}[s, X_s^0, \mu_s^1]|^2)\, ds)$$

$$= \exp(-\int_0^T (\tilde b[s, X^0_s, \mu^2_s] - \tilde b[s, X^0_s, \mu^1_s])\, d\tilde W_s$$

$$-\frac{1}{2}\int_0^T |\tilde b[s, X^0_s, \mu^2_s] - \tilde b[s, X^0_s, \mu^1_s]|^2 ds),$$

as required.

Further, recall that the total variation norm between two probability measures is defined as $\|\mu - \nu\|_{TV} = 2\sup_A(\mu - \nu)(A)$, and it is known that

$$v(t) := \|\mu^1_{[0,t]} - \mu^2_{[0,t]}\|_{TV} = 1 - \mathbb{E}^{\gamma^1}\rho \wedge 1 \tag{18}$$

$$\le \sqrt{E^{\gamma^1}\rho^2 - 1}.$$

Let us show the calculus for completeness:

$$1 - \mathbb{E}^{\gamma^1}(\rho \wedge 1) = \mathbb{E}^{\gamma^1}(1 - \rho \wedge 1)$$

$$\le \sqrt{\mathbb{E}^{\gamma^1}(1 - \rho \wedge 1)^2} = \sqrt{E^{\gamma^1}(1 - \rho 1(\rho \le 1) - 1(\rho > 1))^2}$$

$$= \sqrt{\mathbb{E}^{\gamma^1}(1(\rho \le 1) - \rho 1(\rho \le 1))^2} = \sqrt{\mathbb{E}^{\gamma^1} 1(\rho \le 1)(\rho - 1)^2}$$

$$\le \sqrt{\mathbb{E}^{\gamma^1}(\rho - 1)^2} = \sqrt{\mathbb{E}^{\gamma^1}\rho^2 - 1},$$

as required. We used the Cauchy–Bouniakovsky–Schwarz inequality. Now, again by virtue of the Cauchy–Bouniakovsky–Schwarz inequality,

$$\mathbb{E}^{\gamma^1}\rho^2 = \mathbb{E}^{\gamma^1}\exp(-2\int_0^T (\tilde b[s, X^0_s, \mu^2_s] - \tilde b[s, X^0_s, \mu^1_s]) d\tilde W_s$$

$$-\int_0^T |\tilde b[s, X^0_s, \mu^2_s] - \tilde b[s, X^0_s, \mu^1_s]|^2 ds)$$

$$= \mathbb{E}^{\gamma^1}\exp(-2\int_0^T (\tilde b[s, X^0_s, \mu^2_s] - \tilde b[s, X^0_s, \mu^1_s]) d\tilde W_s$$

$$-4\int_0^T |\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1]|^2 ds)$$

$$\times \exp(+3\int_0^T |\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1]|^2 ds)$$

$$\le \Bigg(\mathbb{E}^{\gamma^1}\exp(-4\int_0^T (\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1])d\tilde W_s$$

$$-8\int_0^T |\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1]|^2 ds)\Bigg)^{1/2}$$

$$\times \left(\mathbb{E}^{\gamma^1}\exp(6\int_0^T |\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1]|^2 ds)\right)^{1/2}$$

$$\le (=)\sqrt{\mathbb{E}^{\gamma^1}\exp\left(6\int_0^T |\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1]|^2 ds\right)}. \tag{19}$$

We estimate,

$$\begin{aligned}&\mathbb{E}^{\gamma^1}\exp\left(6\int_0^T |\tilde b[s,X_s^0,\mu_s^2]-\tilde b[s,X_s^0,\mu_s^1]|^2 ds\right)\\ &\le \mathbb{E}^{\gamma^1}\exp\left(6\|\tilde b\|_B^2\int_0^T \|\mu_s^1-\mu_s^2\|_{TV}^2\, ds\right).\end{aligned} \tag{20}$$

Here the value under the expectation is non-random; hence, the symbol of this expectation may be dropped. Therefore, we have with $C=6\|b\|_B^2$,

$$v(T)\le \sqrt{\exp\left(C\int_0^T v(s)^2 ds\right)-1}. \tag{21}$$

Recall that $v(t)\le 2$, and the function v increases in t. Let us choose $\alpha>0$ small so that $\exp(4\alpha)-1\le 8\alpha$, and take $T\le \alpha/C$. Then, the inequality

$$v(T)\le \sqrt{\exp\left(C\int_0^T v(s)^2\, ds\right)-1}\le \sqrt{\exp\left(CTv(T)^2\right)-1},$$

implies

$$v(T)\le \sqrt{2CTv(T)^2}=\sqrt{2CT}v(T). \tag{22}$$

If we choose T so small that $\sqrt{2CT} < 1$, that is, $T < 1/(2C)$, then it follows that $v(T) = 0$. Since $v(t)$ is continuous in t – which may be easily seen, for example, from (18) – we conclude that $v(T) = 0$ for $T = \min(1/(2C), 1/(\alpha C))$.

2. We conclude by induction that

$$v(2T) = v(3T) = \ldots = 0 \qquad (T = \min(1/(2C), 1/(\alpha C))). \tag{23}$$

Indeed, assume that $v(kT) = 0$ is already established for some integer $k > 0$. Redefine the stochastic exponents:

$$\gamma^i = \exp(-\int_{kT}^{(k+1)T} \tilde{b}[s, X_s^0, \mu_s^i]\, dW_s - \frac{1}{2}\int_{kT}^{(k+1)T} |\tilde{b}[s, X_s^0, \mu_s^i]|^2\, ds), \quad i = 1, 2,$$

and re-denote

$$\tilde{W}_t := W_t + \int_{kT}^{kT+t} \tilde{b}(s, X_s^0, \mu_s^1)\, ds, \quad 0 \le t \le T.$$

This is a new Wiener process on $[kT, (k+1)T]$ starting at W_{kT} under the probability measure P^{γ^1} defined by its density as $(dP^{\gamma^1}/dP)(\omega) = \gamma^1$. Repeating the calculus leading to (19), (20), and (21), and having in mind the assumption $v(kT) = 0$, we obtain **with the same constant** C,

$$v((k+1)T) \le \sqrt{\exp\left(C\int_{kT}^{(k+1)T} v(s)^2 ds\right) - 1}, \tag{24}$$

which straightforward implies

$$v((k+1)T) \le \sqrt{2CTv((k+1)T)^2} = \sqrt{2CT}v((k+1)T). \tag{25}$$

As earlier, the condition $T = \min(1/(2C), 1/(\alpha C))$ guarantees that

$$v((k+1)T) = 0,$$

as required. This completes the induction (23).

Hence, solution is weakly unique on R_+. As noticed above, strong uniqueness also follows. The Theorem 4 is proved.

5 Appendix

Lemma 1 (Skorokhod) *Let $f^n : R \times \Omega \to R$, $(n \geq 0)$ be uniformly bounded random processes on some probability space; let $(W^n\,(n \geq 0))$ be a sequence of Wiener processes on the same probability space, and let all Itô's stochastic integrals $\int_0^T f_s^n\,dW_s^n, n \geq 0$ be well-defined. Assume that for any $\varepsilon > 0$,*

$$\lim_{h\to 0} \sup_n \sup_{|s-t|\leq h} \mathbb{E}\{|f_s^n - f_t^n| > \varepsilon\} = 0, \tag{26}$$

and let for each $s \in [0,T]$

$$(f_s^n, W_s^n) \stackrel{\mathbb{P}}{\to} (f_s^0, W_s^0).$$

Then

$$\int_0^T f_s^n\,dW_s^n \stackrel{\mathbb{P}}{\to} \int_0^T f_s^0\,dW_s^0.$$

Acknowledgements

For the second author the paper was prepared within the framework of a subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

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(2): 173-201, 1998.

[2] M. Bossy, D. Talay, A stochastic particle method for the McKean-Vlasov and the Burgers equation. *Math. Comp.* 66: 157-192, 1997.

[3] T. S. Chiang. McKean-Vlasov equations with discontinuous coefficients. *Soochow J. Math.*, 20(4):507–526, 1994. Dedicated to the memory of Professor Tsing-Houa Teng.

[4] D. Crisan, J. Xiong, Approximate McKeanVlasov representations for a class of SPDEs. *Stochastics: An International Journal of Probability and Stochastic Processes*, 82(1): 53-68, 2010.

[5] R. Dobrushin. Vlasov equations. *Funct. Anal. Appl.*, 13:115–123, 1979.

[6] T. Funaki. A certain class of diffusion processes associated with nonlinear parabolic equations. *Z. Wahrsch. Verw. Gebiete*, 67(3): 331–348, 1984.

[7] S. Herrmann, J. Tugaut, Nonuniqueness of stationary measures for self-stabilizing processes. *Stochastic Processes and Their Applications.* 120(7): 1215-1246, 2010.

[8] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*. Elsevier, 2014.

[9] B. Jourdain. Diffusions with a nonlinear irregular drift coefficient and probabilistic interpretation of generalized Burgers' equations. *ESAIM Probab. Statist.*, 1:339–355 (electronic), 1995/97.

[10] B. Jourdain and S. Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. *Ann. Inst. H. Poincaré Probab. Statist.*, 34(6):727–766, 1998.

[11] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pages 171–197. University of California Press, Berkeley and Los Angeles, 1956.

[12] N. V. Krylov. On Ito's stochastic integral equations. *Theory Probab. Appl.*, 14:330–336, 1969.

[13] N. V. Krylov. On the selection of a markov process from a system of processes and the construction of quasi-diffusion processes. *Izvestiya: Mathematics*, 7(3):691–709, 1973.

[14] N. V. Krylov. *Controlled diffusion processes*, volume 14 of *Applications of Mathematics*. Springer-Verlag, New York-Berlin, 1980. Translated from the Russian by A. B. Aries.

[15] T. G. Kurtz. Weak and strong solutions of general stochastic models. *Electron. Commun. Probab.*, 19(58):1–16, 2014.

[16] H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Nat. Acad. Sci. U.S.A.*, 56:1907–1911, 1966.

[17] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.

[18] A. V. Skorohod. *Stochastic equations for complex systems*, volume 13 of *Mathematics and its Applications (Soviet Series)*. D. Reidel Publishing Co., Dordrecht, 1988. Translated from the Russian by L. F. Boron.

[19] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[20] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 165–251. Springer, Berlin, 1991.

[21] A. Yu. Veretennikov. On the strong solutions of stochastic differential equations. *Theory of Probability & Its Applications*, 24(2):354–366, 1980.

[22] A. Yu. Veretennikov. On strong solutions and explicit formulas for solutions of stochastic integral equations. *Math. USSR, Sb.*, 39:387–403, 1981.

[23] A. Yu Veretennikov. On the criteria for existence of a strong solution of a stochastic equation. *Theory Probab. Appl.*, 27:441–449, 1982.

[24] A. Yu. Veretennikov. On ergodic measures for McKean-Vlasov stochastic equations. In *Monte Carlo and quasi-Monte Carlo methods 2004*, 471–486. Springer, Berlin, 2006.

[25] A. J. (A. Yu.) Veretennikov and N. V. Krylov. On explicit formulas for solutions of stochastic equations. *Mathematics of the USSR-Sbornik*, 29(2):239–256, 1976.

[26] A. A. Vlasov. The vibrational properties of an electron gas. *Physics-Uspekhi*, 10(6):721–733, 1968.

[27] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.

[28] A. K. Zvonkin and N. V. Krylov. Strong solutions of stochastic differential equations. In *Proceedings of the School and Seminar on the Theory of Random Processes (Druskininkai, 1974), Part II (Russian)*, pages 9–88. Inst. Fiz. i Mat. Akad. Nauk Litovsk. SSR, Vilnius, 1975.