"ver_mish_2zo (7)".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†

March 25, 2020

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. Weak existence requires a non-degeneracy of diffusion and no more than a linear growth of both coefficients in the state variable. Weak and strong uniqueness are established under the restricted assumption of diffusion, yet without any regularity of the drift; this part is based on the analysis of the total variation metric.

1 Introduction

Our subject is solutions of the stochastic Itô–McKean–Vlasov equation in $\mathbb{R}^d$

$$dX_t = B[t, X_t, \mu_t]dt + \Sigma[t, X_t, \mu_t]dW_t, \ \ t \geq 0, \qquad X_0 = x_0, \tag{1}$$

*Taras Shevchenko National University of Kyiv; myus@univ.kiev.ua

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

in a particular situation called "true McKean–Vlasov case" under the convention

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

and under certain non-degeneracy assumptions. 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 time t. The initial data x_0 may be random but independent of W; a non-random value is also allowed. Historically, Vlasov's idea, proposed originally in 1938 and contained in the reprinted paper [36], called mean field interaction in mathematical physics and stochastic analysis, assumes that for a large multi-particle ensemble with "weak interaction" between particles, this interaction for one particle with others may be effectively replaced by an averaged field. A class of equations of type (1) was proposed by M. Kac [17] as a stochastic "toy model" for the Vlasov kinetic equation of plasma. The systematic study of such equations was started by McKean [24]. The reference [30] provides an introduction to the whole area with links to the paper [8] as the most important preceding background deterministic paper.

McKean–Vlasov's equations, being clearly more involved than Itô's SDEs, arise in multi-agent systems (see [3, 4]), as well as in some other areas of high interest such as filtering (see [6]). 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, (see [13]). In what concerns "propagation of chaos" for the equation (1), we refer the reader to [30] and [5, Theorem 4.3]. In the authors' view, it may be fruitful to separate different aspects, including time discretization and "propagation of chaos" for multi-particle case, and to consider approximations differently from the basic existence and uniqueness issues; only the latter two are the main subjects of the present paper. Many control problems lead to discontinuous coefficients. This is one of motivations for looking for existence and uniqueness under minimal regularity of the coefficients.

As to earlier works in this area, one of the most important papers is [11] where the martingale problem for a similar McKean-Vlasov SDE is tackled. It is not very easy to compare our regularity assumptions with those in [11] because the latter are given not directly in terms of coefficients (please compare with (2.1) in the Assumption I from [11]). We do not assume continuity with respect to the state variable x replacing it by the non-degeneracy of the diffusion matrix. Neither our linear growth bound is comparable directly with the Lyapunov type conditions in [11]. More general growth conditions were studied in [5]; however, our regularity conditions admit just measurable coefficients in x, especially, for weak existence, and, hence, overall, our results are not covered by [5] either.

Our goal is to establish weak existence analogues to Krylov's weak existence for Itô's equations which is more general than in earlier papers. A more general equation is tackled with a possibly non-square matrix σ, which may be useful in applications and which case was not covered in [5]. Further, we propose a different method which could be of interest in some other settings. In the homogeneous case and under less general conditions, using a different technique, weak existence and weak uniqueness were established in [15] and [16]. In [34] 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, are established there under restrictive additional assumptions on the drift which include Lipschitz and some other conditions. In the present paper, weak and strong uniqueness are established for bounded and measurable drifts under additional assumptions on the (variable) diffusion coefficient.

In applications where some additional regularization by white noise is often required it may be useful to have a result for references with dimensions $d_1 \geq d$ rather than just for $d_1 = d$. This case is rarely tackled in the literature and it is not easy to find a suitable reference; this was the main reason why we included this extension. Despite the widespread intuitive belief that for weak solutions or weak uniqueness everything which may be desired only depends on the matrix $\sigma\sigma^*$, in fact, conditions in the McKean–Vlasov case usually do require certain properties of σ, not $\sigma\sigma^*$ (please compare, for example, with [11]). Hence, even if some results for $d_1 = d$ extends to $d_1 \geq d$, yet it is not automatic. Unlike the setting in the paper [5], we allow non-homogeneous coefficients depending on time; a formal reduction to a homogeneous case 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 [5]: we use explicitly Skorokhod's single probability space approach combined with Krylov's integral estimates for Itô's processes.

Strong existence for McKean–Vlasov equations in our paper is derived from strong existence for "ordinary" or "linearized" Itô's equations with a fixed flow (μ_t). The famous Yamada–Watanabe principle (see [14], [23], [37], [38]) concerning weak existence and pathwise uniqueness here has a remote analogue in terms of the equivalence of weak and strong uniqueness, yet, under additional assumptions. In all main results of the paper (but not in Propositions 1 and 2) it is assumed that the drift, and in the Theorem 1 diffusion as well, satisfies a linear growth bound condition. The linear growth is useful because of numerous applications where, at least, the drift is often unbounded; further extensions on a faster non-linear growth usually require Lyapunov type conditions, which are not considered in this paper.

The structure of the paper is as follows. In the Section 2 weak existence is established. The Theorem 1 there mimics Krylov's weak existence result for Itô's SDEs from [19] for a homogeneous case, and from [21] for a non-homogeneous case; see also [35]. No regularity of the coefficients is assumed with respect to the state variable x. The proof is split into three parts. The first two parts, given in the Proposition 1 and Proposition 2, are devoted to the case under a bit restrictive additional assumptions on the diffusion; the third part extends the consideration to the general situation, i.e. to a not necessarily quadratic and symmetric diffusion matrix. Section 3 is devoted to strong solutions and to weak and strong uniqueness. Weak uniqueness and strong uniqueness are established simultaneously under identical (for weak and for strong uniqueness) sets of conditions. The latter do involve some restriction on the diffusion coefficient which should not depend on the measure in the Theorem 3. For a completeness of the paper, three important classical Skorokhod's and Dynkin's lemmata are provided in Appendix (Section 4).

2 Weak existence

2.1 Main results

Before we turn to the main results, let us recall the definitions and a fact from functional analysis.

Definition 1. *The triple (X_t, μ_t, W_t) is called solution of the equation (1) iff (W_t) is a d_1-dimensional Wiener process with a filtration $(\mathcal{F}_t)$ such that for each t, X_t is $\mathcal{F}_t$-measurable, X_t is continuous in t, and*

$$\mathbb{P}\left(X_t - x_0 - \int_0^t B[s, X_s, \mu_s]ds - \int_0^t \Sigma[s, X_s, \mu_s]dW_s = 0,\ t \geq 0\right) = 1,$$

with all the integrals under the sign of probability being correctly defined, and with μ_t being a marginal distribution of X_t for each $t \geq 0$. This solution is called strong iff for each t the random variable X_t is measurable with respect to the sigma-algebra $\mathcal{F}^W_t$ (sigma-algebra generated by Wiener process W); all other solutions are called weak.

Note that in the case of strong solution, it exists on any probability space with a d_1-dimensional Wiener process W. Following a tradition of Itô SDE theory and slightly abusing a rigorous wording in the definition above, we will usually call solution just the first component X_t of the triple (X_t, μ_t, W_t) yet with a compulsory property that μ_t is a marginal distribution of X_t for each t. The next lemma is probably a common knowledge since it is never mentioned in the papers on the topic. Yet, it seems that this result does require some integrability conditions; so it is stated here with a brief sketch of the proof so as to make sure that the linear growth conditions suffice. Let $|\cdot|$ stand for the Euclidean norm for any vector in $\mathbb{R}^d$, and $\|\cdot\|$ for the standard matrix norm, namely, $\|\sigma\| = \left(\sum_{i,j}\sigma_{ij}^2\right)^{1/2}$.

Lemma 1. *Under the assumption of (2), let the Borel coefficients $b(t,x,y)$ and $\sigma(t,x,y)$ for each (t,x) satisfy*

$$\sup_{t,y}(|b(t,x,y)| + \|\sigma(t,x,y)\|) \le C(x)$$

with some locally bounded Borel function $C(x)$, $x \in \mathbb{R}^d$, and let $\mu_t(dy)$ be a marginal distribution of any solution X_t of the equation (1). Then the functions $\tilde b(t,x) := B[t,x,\mu_t]$ and $\tilde\sigma(t,x) := \Sigma[t,x,\mu_t]$ are Borel measurable in (t,x).

Proof. Let (X_t,μ_t,W_t) be a solution of (1) on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ with a d_1-dimensional Wiener process W, and consider another *independent* solution (ξ_t,μ_t,W'_t) *with the same marginal distribution μ_t of ξ_t*, say, on another probability space $(\Omega',\mathcal{F}',\mathbb{P}')$ with a d_1-dimensional Wiener process W':

$$d\xi_t = B[t,\xi_t,\mu_t]dt + \Sigma[t,\xi_t,\mu_t]dW'_t, \quad t \ge 0, \qquad \mathcal{L}(\xi_0) = \mathcal{L}(x_0). \tag{3}$$

Then the coefficient $B[t,x,\mu_t]$ can be written as

$$B[t,x,\mu_t] = \mathbb{E}' b(t,x,\xi_t),$$

where $\mathbb{E}'$ stands for expectation with respect to the probability measure $\mathbb{P}'$. Now, the function $b(t,x,y)$ is Borel measurable in (t,x,y) by the assumption, and the function $\xi_t(\omega')$ is $\mathcal{B}[0,\infty)\otimes\mathcal{F}$-measurable in (t,ω') due to continuity of solution ξ_t in t and its measurability in ω' (see, e.g., [22, Lemma 1.5.7]). Hence, the function $\hat b(t,x,\omega') := b(t,x,\xi_t(\omega'))$ is $\mathcal{B}[0,\infty)\otimes\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{F}$- measurable in (t,x,ω'). Further, one of the statements of Fubini theorem (see [22, Theorem 1.5.5]) claims that in this case the function

$$\mathbb{E}' b(t,x,\xi_t) = \int b(t,x,\xi_t(\omega'))\mathbb{P}'(d\omega') = \int \hat b(t,x,\omega')\mathbb{P}'(d\omega')$$

is $\mathcal{B}[0,\infty)\otimes\mathcal{B}(\mathbb{R}^d)$-measurable, as required. Here we used the condition of boundedness of b in y for each x, which implies integrability

$$\iint_D \int |b(t,x,\xi_t(\omega'))|\mathbb{P}'(d\omega')dtdx \le \iint_D C(x)dtdx < \infty,$$

over any *bounded* Borel subset $D \in \mathcal{B}[0,\infty)\otimes\mathcal{B}(\mathbb{R}^d)$. The Lemma 1 is proved. •

In this way, under the condition of the at most linear growth in x assumed in the sequel (see (4) a few lines below), the coefficients B and Σ in (1) are Borel measurable in (t, x); so, the equation (1) does make sense under this condition. The next theorem is the main result of the paper about weak existence.

Theorem 1. *Let the initial value x_0 have a finite 4th moment. For the problem (1)–(2), suppose that the following two conditions are satisfied.*

(i) *The functions b and σ admit linear growth condition in (x), i.e., there exists $C > 0$ such that for any s, x, y,*

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

(ii) *The diffusion matrix σ is uniformly non-degenerate in the following sense: there is a value $\nu > 0$ such that for any probability measure μ,*

$$\inf_{s,x} \inf_{|\lambda|=1} \lambda^* \left(\int \sigma(s, x, y)\mu(dy) \right) \left(\int \sigma^*(s, x, y)\mu(dy) \right) \lambda \ge \nu. \tag{5}$$

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 $(\mathcal{F}_t,\, t \ge 0)$.

Remark 1. *If $d_1 = d$ and σ is* symmetric and positive definite, *then the assumption (5) can be replaced by an equivalent but much easier one, which is frequently in use,*

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

The intuitive meaning of the condition (5) in the simplest 1D (that is, with $d_1 = d = 1$) situation is that the diffusion coefficient is non-degenerate and cannot change sign for any fixed (s, x) and varying y. It is plausible that any moment of order $2 + \epsilon$ for x_0 suffices for all the statements (except for the Theorem 3 under the linear growth condition on the drift where an exponential moment will be required), but this goal is not pursued here. Under the additional assumption of boundedness of b and σ, the 4th moment of the initial value x_0 is not necessary and can be further relaxed.

The structure of the proof of the Theorem 1 is such that first the case of symmetric non-degenerate σ is tackled, i.e., with $d_1 = d$ and under the assumption (6). A motivation for this approach is that under the relaxed assumption (5) and under the symmetry of σ it is easy to find a smoothing of this matrix function which would keep the non-degeneracy of the diffusion coefficient. Because of this we formulate some provisional statement under more restrictive conditions as Proposition 1. Its proof will be simultaneously the beginning of the proof of the Theorem 1.

We will need two auxiliary lemmata. The Lemma 2 is about a priori moment bounds for solutions; the Lemma 6 is a localized version of Krylov's bounds where coefficients are locally bounded and the estimate also relates to a bounded domain.

Lemma 2. *Under assumption (4) and the standard measurability, a priori estimates*

$$\mathbb{E} \sup_{0\le t\le T} |X_t|^4 \le C_T(1+\mathbb{E}|x_0|^4), \tag{7}$$

and

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|X_t - X_s|^4 \le C_T h^2, \tag{8}$$

hold true with some constants C_T (generally, different) that do not depend on n. In (7) the constant C_T, generally speaking, also depends on the value of the moment $\mathbb{E}|x_0|^4$.

Later on in the proof of the Proposition 1 similar a priori bounds will be stated for the successive approximations of solutions. Recall that $x_0 \in \mathbb{R}^d$ is the initial value of the process X and that it may be random with a certain finite moment. In fact, similar a priori bounds hold true for any power function assuming the appropriate initial moment, although, this will not be used in this paper. The proof of (7) is very standard and can be done following the lines in [12, Theorem 1.6.4], or [21, Corollary 2.5.6], or [29] (and many other places) combined with Doob's inequalities. The bounds (8) follow from a similar calculus starting from s instead of 0.

Why do we need the 4th moment, will be clarified in the proofs of the Proposition 1 and the Theorem 1: it is useful for verifying continuity for the processes with equivalent finite-dimensional distributions; for this purpose the 2nd moment is not sufficient, although $2+\epsilon$ should probably work.

A version of localised Krylov's bound required for the proof of the Theorem 1 can be found in the Appendix.

In the following intermediate simplified version of the Theorem 1, we allow both coefficients to grow, but σ is assumed symmetric and positive definite. In the first part of the proof the coefficients will be assumed bounded. We will treat now a more general structure of the coefficients B and Σ, namely, in the form

$$\bar\Sigma[t,X,\mu] = \phi(\int \sigma(t,x,y)\mu(dy)), \quad \bar B[t,X,\mu] = \psi(\int b(t,x,y)\mu(dy)), \tag{9}$$

where conditions on the matrix-functions ϕ and vector-function ψ will be specified. In fact, for the proof of the Theorem 1 we only need ϕ; however, ψ is added just by analogy since it does not bring any new difficulty. (We keep notations $\Sigma[t,X,\mu] = \int \sigma(t,x,y)\mu(dy)$, $B[t,X,\mu] = \int b(t,x,y)\mu(dy)$.) Actually, conditions on the functions ϕ and ψ might be further relaxed, but the authors do not have a motivation for that at the moment. Correspondingly, we now consider the equation (1) with coefficients $\bar{B}$ instead of B and with $\bar{\Sigma}$ instead of Σ.

Proposition 1. *Assume that $d_1 = d$, and let σ and b satisfy the linear growth condition (4), and a more general dependence,*

$$\bar{\Sigma}[t,X,\mu] = \phi(\int \sigma(t,x,y)\mu(dy)), \quad \bar{B}[t,X,\mu] = \psi(\int b(t,x,y)\mu(dy)),$$

so that there exist $m_1, m_2 > 0$ such that

$$\|\phi(\Sigma) - \phi(\tilde{\Sigma})\| \le C\|\Sigma - \tilde{\Sigma}\|(1 + (\|\Sigma\| \vee \|\tilde{\Sigma}\|)^{m_1}), \tag{10}$$

$$|\psi(b) - \psi(\tilde{b})| \le C|b - \tilde{b}|(1 + (|b| \vee |\tilde{b}|)^{m_2}) \tag{11}$$

Also, $\phi(\int \sigma(t,x,y)\mu(dy))$ is assumed to be a symmetric positive definite uniformly non-degenerate for any μ and uniformly w.r.t. t,x,μ.

Let the initial value x_0 have a finite fourth moment and suppose that the condition (4) is satisfied for the functions b and σ, and that the matrix $\sigma(t,x,y)$ is symmetric for each triple (t,x,y) and the inequality (6) holds. Then the equation (1) has a weak solution, that is, a solution on some probability space with a standard d-dimensional Wiener process with respect to some filtration $(\mathcal{F}_t,\ t \ge 0)$.

Remark 2. *In the case of bounded coefficients, the functions ϕ and ψ must include the following:*

$$\psi = Id \ \ \& \ \ \phi(\langle \sigma, \mu \rangle) = (2\pi i)^{-1} \oint_{\Gamma} \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1} d\lambda$$

with $\Gamma = \Gamma(x) = \{\lambda : |\lambda| = r(x)\}$), with the radius $r(x)$ so that this contour contains all eigenvalues of $A[t,x,\mu]$, for example, $r(x) = \sup_{t,\mu} \|A[t,x,\mu] + 1\|$, with some C large enough, i.e., under our assumptions it suffices to take $r(x) = C|x|^2 + 1$. Under the boundedness condition on all coefficients which is accepted temporary for the Proposition 1, it suffices to choose a unique contour with $\bar{r} = \sup_{t,x,\mu} \|A[t,x,\mu]\| + 1$, since this value is finite. Note that in this main example we have

$$\|\phi(\langle \sigma, \mu \rangle) - \phi(\langle \sigma, \mu' \rangle)\|$$

$$= \frac{1}{2\pi} \| \oint\limits_{\Gamma} \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1} d\lambda - \oint\limits_{\Gamma} \lambda^{1/2}(\lambda - A[t,x,\mu'])^{-1} d\lambda \|$$

$$\le C\bar{r}^{3/2} \sup_{|\lambda|=\bar{r}} \sup_{\nu} \|(\lambda - A[t,x,\nu])^{-2}\| \, \|A[t,x,\mu] - A[t,x,\mu']\|$$

$$\le C\bar{r}^{2} \, \|\Sigma[t,x,\mu] - \Sigma[t,x,\mu']\|, \qquad (12)$$

where $A[t,x,\mu] = (\int \sigma(t,x,y)\mu(dy))(\int \sigma(t,x,y)\mu(dy))^* \equiv \Sigma[t,x,\mu]\Sigma^*[t,x,\mu]$*, and* Γ *is a contour in* $\mathbb{C}$*, which contains all eigenvalues of* A *for each* t,x,μ*; we aim to extend the existence result for the growing* σ *in the next Proposition. Also,* $\phi(\int \sigma(t,x,y)\mu(dy)) = \phi(\mathbb{E}^3\sigma(t,x,\xi))$ *is assumed to be a symmetric positive definite non-degenerate uniformly with respect to* t,x *and* μ*, where* $\mu = \mathcal{L}(\xi)$*.*

Under the linear growth assumptions of the Theorem 1, we will need $m_1 = 2$ *and* $m_2 = 0$*, as we need to include*

$$\psi = Id \quad \& \quad \phi(\langle \sigma, \mu\rangle) = \phi(\Sigma) = (2\pi i)^{-1} \oint_{\Gamma} \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1} d\lambda$$

with $\Gamma = \Gamma(x) = \{\lambda : |\lambda| = r(x)\})$*, with the radius* $r(x)$ *so that this contour contains all eigenvalues of* $A[t,x,\mu]$*, where* A *itself is a function of* $\langle \sigma, \mu\rangle$*; e.g.,* $r(x) = \sup_{t,\mu} \|A[t,x,\mu]\| + 1$*. Recall that* $A = \Sigma\Sigma^*$*. In this main example we have*

$$\|\phi(\langle \sigma(t,x,\cdot), \mu\rangle) - \phi(\langle \sigma(t,x,\cdot), \mu'\rangle)\| = \|\phi(\Sigma[t,x,\mu]) - \phi(\Sigma[t,x,\mu'])\|$$

$$= \frac{1}{2\pi} \| \oint\limits_{\Gamma} \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1} d\lambda - \oint\limits_{\Gamma} \lambda^{1/2}(\lambda - A[t,x,\mu'])^{-1} d\lambda \|$$

$$\le Cr(x)^{3/2} \sup_{|\lambda|=r(x)} \sup_{\nu} \|(\lambda - A[t,x,\nu])^{-2}\| \, \|A[t,x,\mu] - A[t,x,\mu']\|$$

$$= Cr(x)^{3/2} \sup_{|\lambda|=r(x)} \sup_{\nu} \|(\lambda - A[t,x,\nu])^{-2}\| \, \|\Sigma\Sigma^*[t,x,\mu] - \Sigma\Sigma^*[t,x,\mu']\|$$

$$\le Cr(x)^{2} \|\Sigma[t,x,\mu] - \Sigma[t,x,\mu']\|. \qquad (13)$$

Note that here, in turn, under the assumptions of the Theorem 1,

$$r(x) \le 1 + C|x|^2. \tag{14}$$

So, it is reasonable in the Proposition 1 to inspect the case of the conditions (10) and (11) with some $m_1, m_2 > 0$.

2.2 Proof of Proposition 1

It is important that the contour Γ may be chosen the same for each t, x, μ; hence, the assumption of the boundedness of σ is necessary. Later on in the Proposition 2 we will relax this assumption by approximations, but in these approximations will not use Cauchy – Riesz – Dunford formula. So, let us smooth both coefficients with respect to all variables by convolutions in such a way that they become globally Lipschitz in x and ξ,

$$b^n(t,x,y) = b(t,x,y) * \phi_n(x) * \phi_n(y), \tag{15}$$

and

$$\sigma^n(t,x,y) = \sigma(t,x,y) * \phi_n(x) * \phi_n(y), \tag{16}$$

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 origin of the corresponding variable as $n \to \infty$; or, in other words, that they are delta-sequences in the corresponding variables. Note that, of course, we may assume that for every n the smoothed coefficient of the drift remains to be under the linear growth condition (4) with the same constant for each n (in reality this constant may increase a little bit in comparison to the constant C from (4), but still remain uniformly bounded); also, under the assumption (6) the smoothed diffusion remains uniformly non-degenerate with ellipticity constants independent of n.

Now we shall explain why the equation with smoothed coefficients has a (strong) solution. We use successive approximations. For any fixed n, let

$$X^n(0)_t := X_0, \quad \mu^n(0)_t = \mathcal{L}(X^n(0)_t) = \mu_0, \quad t \ge 0;$$

further, if $X(m)_t)$ and $\mu(m)_t$ are already determined, let us define

$$X^n(m+1)_t := X_0 + \int_0^t B^n[s, X^n(m)_s, \mu^n(m)_s]ds - \int_0^t \Sigma^n[s, X^n(m)_s, \mu^n(m)_s]dW_s$$

where

$$B^n[t,x,\mu] = \psi\left(\int b^n(t,x,y)\mu(dy)\right), \quad \Sigma^n[t,x,\mu] = \phi\left(\int \sigma^n(t,x,y)\mu(dy)\right).$$

Recall that the integral $\int g(t,x,y)\mu^\xi(dy)$ (where $\mu^\xi(dy)$ is the distribution of the random variable ξ) can be treated as $\eth(t,x,\xi)$; in this way, we can say that $\int_0^t B^n[s, X^n(m)_s, \mu^n(m)_s] = \mathbb{E}^3 b^n(s, X^n(m)_s, \xi^n(m))_s$, where $\xi^n(m)_s$ is a random variable equivalent to $X^n(m)_s$ on some independent probability space, and, moreover, the sequence $(\xi^n(m).), m \geq 1$ can be chosen independent on $(X^n(m).), m \geq 1$, and so that the whole sequence $(\xi^n(m).), m \geq 1$ has the same distribution as the sequence $(X^n(m).), m \geq 1$. Then by induction the second moments of any $X^n(m)_t$ is finite and uniformly bounded for $t \leq T$, and by Itô's formula

$$\mathbb{E}|X^n(m+1)_t - X^n(m)_t|^2$$

$$\leq C_T \mathbb{E} \int_0^t |\psi(\mathbb{E}^3 b^n(s, X^n(m)_s, \xi^n(m)_s) - \psi(\mathbb{E}^3 b^n(s, X^n(m-1)_s, \xi^n(m-1)_s))|^2 ds$$

$$+C\mathbb{E} \int_0^t |\phi(\mathbb{E}^3 \sigma^n(s, X^n(m)_s, \xi^n(m)_s)) - \phi(\mathbb{E}^3 \sigma^n(s, X^n(m-1)_s, \xi^n(m-1)_s))|^2 ds$$

$$= C_T \mathbb{E} \int_0^t \mathbb{E}^3 |(b^n(s, X^n(m)_s, \xi^n(m)_s) - b^n(s, X^n(m-1)_s, \xi^n(m-1)_s)|^2 ds$$

$$+CE \int_0^t ds \frac{1}{2\pi} | \oint_\Gamma \lambda^{1/2}(\lambda - A[s, X^n(m)_s, \mu^n(m)_s])^{-1} d\lambda$$

$$- \oint_\Gamma \lambda^{1/2}(\lambda - A[s, X^n(m-1)_s, \mu^n(m-1)_s])^{-1} d\lambda|^2$$

$$= C_T \mathbb{E} \int_0^t |\mathbb{E}^3 (b^n(s, X^n(m)_s, \xi^n(m)_s) - b^n(s, X^n(m-1)_s, \xi^n(m-1)_s)|^2 ds$$

$$+CE\int_0^t ds\frac{1}{2\pi}|\oint_\Gamma \lambda^{1/2}(\lambda-\phi\,(\mathbb{E}\sigma^n(X^n(m)_s,\xi^n(m)_s))$$

$$\times\phi\,(\mathbb{E}\sigma^n(X^n(m)_s,\xi^n(m)_s))^*)^{-1}d\lambda$$

$$-\oint_\Gamma \lambda^{1/2}(\lambda-\phi\,(\mathbb{E}\sigma^n(X^n(m-1)_s,\xi^n(m-1)_s))$$

$$\times\phi\,(\mathbb{E}\sigma^n(X^n(m-1)_s,\xi^n(m-1)_s))^*)^{-1}d\lambda|^2$$

$$\le C_{T,n}\mathbb{E}\int_0^t |X^n(m)_s-X^n(m-1)_s|^2ds+C\mathbb{E}\int_0^t \mathbb{E}^3|\xi^n(m)_s-\xi^n(m-1)_s|^2ds$$

$$\le C_{T,n}\mathbb{E}\int_0^t |X^n(m)_s-X^n(m-1)_s|^2ds.$$

Since all terms here are finite, we obtain by induction

$$\mathbb{E}|X^n(m+1)_t-X^n(m)_t|^2\le C_{n,T}\frac{T^m}{m!},\quad t\le T.$$

Due to the Doob inequality we also get

$$\mathbb{E}\sup_{t\le T}|X^n(m+1)_t-X^n(m)_t|^2\le C_{n,T}\frac{T^m}{m!}.$$

From here by standard methods it follows easily convergence of the sequence $X^n(m)_t$ in probability uniformly with respect to $t\le T$ to a solution X^n_t, as required.

2. In a standard way (see, e.g., [21], [29]), absolutely similar to the inequalities of the Lemma 2, we get the estimates uniform in n,

$$\sup_{0\le t\le T}\mathbb{E}|X^n_t|^2\le C_T(1+\mathbb{E}|x_0|^2), \tag{17}$$

$$\sup_{0\le s\le t\le T;\,t-s\le h}\mathbb{E}|X^n_t-X^n_s|^2\le C_Th, \tag{18}$$

$$\mathbb{E}\sup_{0\le t\le T}|X^n_t|^2\le C_T(1+\mathbb{E}|x_0|^2), \tag{19}$$

$$\sup_{0\le t\le T} \mathbb{E}|X^n_t|^4 \le C_T(1+\mathbb{E}|x_0|^4), \tag{20}$$

and

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|X^n_t - X^n_s|^4 \le C_T h^2, \tag{21}$$

with some constants C_T which may be different for different inequalities but do not depend on n. In fact, similar a priori bounds hold true for any power function assuming the appropriate initial moment, although, this will not be used in this paper. The proof can be done following the lines in [12, Theorem 1.6.4]. Why do we need the fourth degree will be clear in the next step: it is useful for verifying continuity for the processes with equivalent finite-dimensional distributions; for this purpose the second degree is not sufficient, although $2+\epsilon$ should work. Note that all these bounds are valid under the linear growth assumptions in x.

3. Assume temporarily that σ and b are bounded, and instead of (10)–(11) suppose for this step that

$$\|\phi(\Sigma) - \phi(\tilde\Sigma)| \ \le C\|\Sigma - \tilde\Sigma\|, \tag{22}$$

and

$$|\psi(B) - \psi(\tilde B)| \le C|B-\tilde B|. \tag{23}$$

Let us introduce new processes ξ^n, the copy of X^n, on some other – independent – probability space (i.e., we will consider both on the direct product of the two probability spaces); it also satisfies a similar SDE. Moreover, in the sequel by $\mathbb{E}^3\sigma^n(s,X^n_s,\xi^n_s)$ or $\mathbb{E}^3\sigma(s,X_s,\xi_s)$ we denote expectation with respect to the third variable ξ^n_s, or ξ_s i.e., *conditional* expectation given the second variable X^n_s or X_s; 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 $\displaystyle\int \sigma^n(s,X^n_s,y)\mu^{\xi^n}_s(dy) - \int \sigma^n(s,X^n_s,y)\mu^{\xi}_s(dy)$, 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 $\displaystyle\int |\sigma^n(s,X^n_s,y) - \sigma^n(s,X^n_s,y')|^2\mu^{\xi^n,\xi}_s(dy,dy')$, where $\mu^{\xi^n,\xi}_s(dy,dy')$ denotes the marginal distribution of the couple (ξ^n_s,ξ_s).

Now, due to the estimates (17)–(18) and by virtue of Skorokhod's Theorem about a single probability space and convergence in probability (see the Lemma 3 in the Appendix, or [29, §6, ch. 1], or [21, 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,

$$(\tilde X^n_t, \tilde\xi^n_t, \tilde W^n_t) \stackrel{\mathbb{P}}{\to} (\tilde X_t, \tilde\xi_t, \tilde W_t), \quad n\to\infty,$$

for some **equivalent** random processes $(\tilde X^n, \tilde\xi^n, \tilde W^n)$, generally speaking, over a subsequence. Slightly abusing notations, we will denote initial values still by x_0 without tilde. Also, without loss of generality we may and will assume that each process $(\tilde\xi^n_t,\, t\ge 0)$ for any $n\ge 1$ is independent from $(\tilde X^n, \tilde W^n)$, as well as their limit $\tilde\xi_t$ may be chosen independent of the limits $(\tilde X, \tilde W)$ (this follows from the fact that on the original probability space ξ^n is independent of (X^n, W^n) and on the new probability space their joint distribution remains the same; hence, independence of $\tilde\xi^n$ is also valid and in the limit this is still true). See the details in the proof of the Theorem 2.6.1 in [21]. We may also introduce Wiener processes for ξ^n_t and $\tilde\xi^n_t$, and will do it because it will be useful at one of the steps of the proof of the Proposition 2 (for the Proposition 1 it is not necessary). Namely, on an independent probability spaces we have,

$$d\xi^n_t = B^n[t, \xi^n_t, \mu_t]dt + \Sigma^n[t, \xi^n_t, \mu_t]dW^{',n}_t, \ \ t\ge 0, \qquad \mathcal{L}(\xi^n_0) = \mathcal{L}(x_0). \tag{24}$$

and

$$d\tilde\xi^n_t = B^n[t, \tilde\xi^n_t, \mu_t]dt + \Sigma^n[t, \tilde\xi^n_t, \mu_t]d\tilde W^{',n}_t, \ \ t\ge 0, \qquad \mathcal{L}(\tilde\xi^n_0) = \mathcal{L}(x_0). \tag{25}$$

For what follows, let us fix some arbitrary $T>0$ and consider t in the interval $[0,T]$. Due to the inequality (21), the same inequality holds for $\tilde X^n, \tilde W^n$,

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|\tilde X^n_t - \tilde X^n_s|^4 \le C_T h^2. \tag{26}$$

Due to Kolmogorov's continuity theorem this means that all processes $\tilde X^n$ may be regarded as continuous, and $\tilde W^n$ can be assumed also continuous by the same reason. Further, due to the independence of the increments of W^n after time t of the sigma-algebra $\sigma(X^n_s, W^n_s, s\le t)$, the same property holds true for $\tilde W^n$ and $\sigma(\tilde X^n_s, \tilde W^n_s, s\le t)$, as well as for $\tilde W^n$ and for the completions of the sigma-algebras $\sigma(\tilde X^n_s, \tilde W^n_s, s\le t)$ which we denote by $\mathcal{F}^{(n)}_t$. Also, the processes $\tilde X^n$ are adapted to the filtration $(\mathcal{F}^{(n)}_t)$. So, all stochastic integrals which involve $\tilde X^n$ and $\tilde W^n$ are well defined. The same relates to the processes $\tilde\xi^n$.

Hence, again by using Skorokhod's lemma on convergence on a unique probability space – see the Lemma 3 in the Appendix – we may choose a subsequence $n'\to\infty$ so as to pass to the limit in the equation

$$\tilde X^{n'}_t = x_0 + \int_0^t \psi(\mathbb{E}^3 b^{n'}(s, \tilde X^{n'}_s, \tilde\xi^{n'}_s))\, ds + \int_0^t \phi(\mathbb{E}^3 \sigma^{n'}(s, \tilde X^{n'}_s, \tilde\xi^{n'}_s)) d\tilde W^{n'}_s,$$

in order to get

$$\tilde{X}_t = x_0 + \int_0^t \psi(\mathbb{E}^3 b(s, \tilde{X}_s, \tilde{\xi}_s))ds + \int_0^t \phi(\mathbb{E}^3 \sigma(s, \tilde{X}_s, \tilde{\xi}_s))\, d\tilde{W}_s,$$

or, equivalently,

$$\tilde{X}_t = x_0 + \int_0^t B[s, \tilde{X}_s, \mu_s]\, ds + \int_0^t \Sigma[s, \tilde{X}_s, \mu_s]\, d\tilde{W}_s,$$

with

$$\mu_s = \mathcal{L}(\tilde{X}_s).$$

First of all, recall that a priori bounds (17) – (26) hold true with constants not depending on n. Now, by Skorokhod's theorem on some probability space we have some equivalent processes $(\tilde{X}_t^{n'}, \tilde{\xi}_t^{n'}, \tilde{W}_t^{n'})$ and a limiting triple $(\tilde{X}_t, \tilde{\xi}_t, \tilde{W}_t)$ such that for any t,

$$(\tilde{X}_t^{n'}, \tilde{\xi}_t^{n'}, \tilde{W}_t^{n'}) \stackrel{\mathbb{P}}{\to} (\tilde{X}_t, \tilde{\xi}_t, \tilde{W}_t). \tag{27}$$

By virtue of the a priori bounds for $\tilde{W}^n$, the process $\tilde{W}_t$ is continuous and it is a Wiener process. Also, the limits are adapted to the corresponding filtration $\tilde{\mathcal{F}}_t := \bigvee_n \mathcal{F}_t^{(n)}$ and $\tilde{W}_t$ is continuous and it is a Wiener process with respect to this filtration. In particular, related the Lebesgue and stochastic integrals are all well defined. Moreover, by virtue of the uniform estimates (21), the limit $(\tilde{X}_t, \tilde{\xi}_t)$ may be also regarded as continuous due to Kolmogorov's continuity theorem because the a priori bounds (8)–(12) remain valid for the limiting processes $\tilde{X}, \tilde{\xi}$. In particular, it is useful to note for the sequel that

$$\sup_{0\le t\le T} \mathbb{E}|\tilde{X}_t|^2 \le C_T(1 + \mathbb{E}|x_0|^2). \tag{28}$$

4. Still in the case of bounded coefficients b and σ, let us now show that

$$\int_0^t \psi(\mathbb{E}^3 b^{n'}(s, \tilde{X}_s^{n'}, \tilde{\xi}_s^{n'}))ds \stackrel{\mathbb{P}}{\to} \int_0^t \psi(\mathbb{E}^3 b(s, \tilde{X}_s, \tilde{\xi}_s))ds, \tag{29}$$

and

$$\int_0^t \phi(\mathbb{E}^3 \sigma^{n'}(s, \tilde{X}_s^{n'}, \tilde{\xi}_s^{n'}))d\tilde{W}_s^{n'} \stackrel{\mathbb{P}}{\to} \int_0^t \phi(\mathbb{E}^3 \sigma(s, \tilde{X}_s, \tilde{\xi}_s))d\tilde{W}_s, \quad n' \to \infty. \tag{30}$$

We start with the draft term. Let us fix some n_0 and let $n > n_0$. Due to the assumption (23), we have for any $t \le T$,

$$
\begin{aligned}
&\mathbb{P}\left(|\int_0^t \left(\psi(\mathbb{E}^3 b^n(s, \tilde{X}_s^n, \tilde{\xi}_s^n)) - \psi(\mathbb{E}^3 b(s, \tilde{X}_s, \tilde{\xi}_s)\right))ds| > c\right) \\
&\le \mathbb{P}\left(|\int_0^t \left(\psi(\mathbb{E}^3 b^n(s, \tilde{X}_s^n, \tilde{\xi}_s^n)) - \psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}_s^n, \tilde{\xi}_s^n))\right) ds| > \frac{c}{3}\right) \\
&+ \mathbb{P}\left(|\int_0^t \left(\psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}_s^n, \tilde{\xi}_s^n)) - \psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)\right) ds| > \frac{c}{3}\right) \\
&+ \mathbb{P}\left(|\int_0^t \left(\psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)) - \psi(\mathbb{E}^3 b(s, \tilde{X}_s, \tilde{\xi}_s))\right) ds| > \frac{c}{3}\right) \\
&=: I^1 + I^2 + I^3. \qquad (31)
\end{aligned}
$$

Let

$$
\begin{aligned}
\gamma_{n,R} &:= \inf(t \ge 0 : (|\tilde{X}_t^n| \vee |\tilde{\xi}_t^n|) \ge R), \ \gamma_R := \inf(t \ge 0 : \sup_{0 \le s \le t}(|\tilde{X}_s| \vee |\tilde{\xi}_s|) \ge R), \\
\gamma_R^X &:= \inf(t \ge 0 : \sup_{0 \le s \le t} |\tilde{X}_s| \ge R), \ \ \& \ \ \gamma_R^\xi := \inf(t \ge 0 : \sup_{0 \le s \le t} |\tilde{\xi}_s| \ge R), \\
\gamma_{n,R}^X &:= \inf(t \ge 0 : \sup_{0 \le s \le t} |\tilde{X}_s^n| \ge R), \ \ \& \ \ \gamma_{n,R}^\xi := \inf(t \ge 0 : \sup_{0 \le s \le t} |\tilde{\xi}_s^n| \ge R).
\end{aligned}
$$

We have, for any $\epsilon > 0$ there exists $R > 0$ such that (to have $R-1$ instead of R will be convenient shortly)

$$
\mathbb{P}(\sup_{0 \le t \le T}(|\tilde{X}_t| \vee |\tilde{\xi}_t|) \ge R - 1) < \epsilon,
$$

or, equivalently,

$$
\mathbb{P}(\gamma_{R-1} \le T) < \epsilon, \qquad (32)
$$

and similarly,

$$
\sup_n \mathbb{P}(\sup_{0 \le t \le T}(|\tilde{X}_t^n| \vee |\tilde{\xi}_t^n|) \ge R - 1) < \epsilon,
$$

or, equivalently,

$$
\sup_n \mathbb{P}(\gamma_{n,R-1} \le T) < \epsilon. \qquad (33)
$$

Denote

$$g^{n,n_0}(s,x,\xi) := b^n(s,x,\xi) - b^{n_0}(s,x,\xi), \quad g^{n_0}(s,x,\xi) := b^{n_0}(s,x,\xi) - b(s,x,\xi).$$

Then the first summand I^1 may be estimated by Chebyshev–Markov's inequality and due to the condition (23) as

$$I^1 \le \frac{3}{c}\,\mathbb{E}\int_0^T C\,\mathbb{E}^3|b^n(s,\tilde X^n_s,\tilde\xi^n_s) - b^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds$$

$$= \frac{3C}{c}\,\mathbb{E}\mathbb{E}^3\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds = \frac{3C}{c}\,\mathbb{E}\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds$$

$$= C\mathbb{E}1(\gamma_{n,R}\le T)\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds + C\mathbb{E}1(\gamma_{n,R}> T)\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds.$$

Here the first term $\mathbb{E}1(\gamma_{n,R}\le T)\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds$ admits the bound (recall that on each line constants C can be different)

$$\mathbb{E}1(\gamma_{n,R}\le T)\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds \le CT\mathbb{P}(\gamma_{n,R}\le T).$$

which last expression is small uniformly in n if R is large enough.

The second term $\mathbb{E}1(\gamma_{n,R} > T)\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds$ admits a bound via Krylov's estimate (see the Theorems 2.4.1 or 2.3.4 in [21]) as follows: there exists a constant N depending on the dimension d and on R through the ellipticity constants of the diffusion matrix and the sup-norm of the drift on $B_R\times B_R$, so that

$$\mathbb{E}1(\gamma_{n,R}> T)\int_0^T |g^{n,n_0}(s,\tilde X^n_s,\tilde\xi^n_s)|\,ds$$

$$\le N_R\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|g^{n,n_0}(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}}$$

$$\le N_R\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^{n}(s,x,\xi)-b(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}}$$

$$+N_R\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^{n_0}(s,x,\xi)-b(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}} \to 0, \quad n,n_0\to\infty,$$

for each R, by virtue of the well-known property of mollified functions. Hence, overall, we obtain that

$$I^1 \to 0, \quad n, n_0 \to \infty.$$

Further, under the assumptions of the Proposition 1, the second term I^2 admits the estimate (for any $0 \le t \le T$),

$$I^2 = \mathbb{P}\Big(|\int_0^t \Big(\psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}^n_s, \tilde{\xi}^n_s))ds - \psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s))\Big)ds| > \frac{c}{3}\Big)$$

$$\le C\,\mathbb{E}\mathbb{E}^3 \int_0^T |b^{n_0}(s, \tilde{X}^n_s, \tilde{\xi}^n_s) - b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds,$$

which tends to zero as $n \to \infty$ under the assumptions of the Proposition 1, due to the Lebesgue bounded convergence theorem because of the convergence in probability $(\tilde{X}^n_s, \tilde{\xi}^n_s) \to (\tilde{X}_s, \tilde{\xi}_s)$. So, for each n_0

$$\lim_{n\to\infty} I^2 = 0.$$

To tackle the third term I^3, we have,

$$I^3 = \mathbb{P}\Big(|\int_0^t \Big(\psi(\mathbb{E}^3 b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)) - \psi(\mathbb{E}^3 b(s, \tilde{X}_s, \tilde{\xi}_s))\Big)ds| > \frac{c}{3}\Big)$$

$$\le C\,\mathbb{E}\mathbb{E}^3 \int_0^T |b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s) - b(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds$$

$$= C\,\mathbb{E}(1(\gamma_{n,R} \wedge \gamma_R \le T) + 1(\gamma_{n,R} \wedge \gamma_R > T)) \int_0^T |g^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds. \tag{34}$$

Firstly the term $\mathbb{E}1(\gamma_{n,R} \wedge \gamma_R \le T) \int_0^T |g^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds$ admits the bound

$$\mathbb{E}1(\gamma_{n,R} \wedge \gamma_R \le T) \int_0^T |g^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds \le C\mathbb{P}(\gamma_{n,R} \wedge \gamma_R \le T) \to 0, \quad R \to \infty,$$

uniformly in n. Hence, it follows that

$$\lim_{n_0\to\infty} \lim_{R\to\infty} \mathbb{E}1(\gamma_{n,R} \wedge \gamma_R \le T) \int_0^T |g^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds = 0.$$

In order to evaluate the term $\mathbb{E}1(\gamma_{n,R} \wedge \gamma_R > T) \int_0^T |g^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)|\,ds$, we note that the values of the function $g^{n_0}(s, x, \xi)$ outside the set $\{(x, \xi) : (|x| \vee |\xi|) \le \tilde{R}\}$ with large $\tilde{R}$ are not relevant. So, for evaluating this term, without losing of generality we may assume that $g^{n_0}(s,x,\xi)$ vanishes outside the ball $B_{R+1}\times B_{R+1}$: if not, we just truncate accepting that $g^{n_0}=0$ outside $B_{R+1}\times B_{R+1}$. Then the desired convergence follows from Krylov's bound from the Lemma 3 and 5. Indeed, denote $g^{n_0}_R(s,x,\xi) := g^{n_0}(s,x,\xi)1(|x|\le R, |\xi|\le R)$. We have,

$$\mathbb{E}1(\gamma_{n,R}\wedge\gamma_R > T)\int_0^T |g^{n_0}(s,\tilde{X}_s,\tilde{\xi}_s)|\,ds \le \mathbb{E}\int_0^T |g^{n_0}_R(s,\tilde{X}_s,\tilde{\xi}_s)|\,ds.$$

Now, for continuous g^{n_0}, in the limit as $n\to\infty$ we get,

$$\mathbb{E}1(\gamma_R > T)\int_0^T |g^{n_0}(s,\tilde{X}_s,\tilde{\xi}_s)|\,ds \le \mathbb{E}\int_0^T |g^{n_0}_R(s,\tilde{X}_s,\tilde{\xi}_s)|\,ds$$

$$= \lim_{n\to\infty}\mathbb{E}\int_0^T |g^{n_0}_R(s,\tilde{X}^n_s,\tilde{\xi}^n_s)|\,ds$$

$$\le CN_R\|g^{n_0}_R\|_{L_{2d+1}((0,\infty)\times R^d\times R^d)} = CN_R\|g^{n_0}_R\|_{L_{2d+1}((0,\infty)\times B_R\times B_R)}.$$

The latter bound extends to all Borel measurable functions g by virtue of the Lemma 5. Hence, by the properties of the mollified functions it follows that

$$\lim_{n_0\to\infty} I^3 = 0.$$

The convergence (29) is, thus, proved.

5. Now still for bounded coefficients let us consider convergence of stochastic integrals in (30). Our goal is an estimate similar to that for the drift and Lebesgue integrals above:

$$\mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^n(s,\tilde{X}^n_s,\tilde{\xi}^n_s))dW^n_s - \int_0^t \phi(\mathbb{E}^3\sigma(s,\tilde{X}_s,\tilde{\xi}_s))d\tilde{W}_s| > c) < C\epsilon, \qquad (35)$$

for any $c,\epsilon>0$ and n large enough. In principle, the task is similar to the convergence of Lebesgue integrals tackled in the previous steps. Hence, we mainly show how to tackle the additional obstacle due to different Wiener processes dW_s and dW^n_s in the stochastic integrals. We have a tool for this which is Skorokhod's Lemma 4 from the Appendix below.

By virtue of [22, Theorem 6.2.1(v)] and similarly to the calculus for the drift with Lebesgue integrals in the previous steps, yet using second moments instead of the first ones for the evident reason, we estimate

$$\begin{aligned}
&\mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s))d\tilde W^n_s - \int_0^t \phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s| > c) \\
&\le \mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s))d\tilde W^n_s - \int_0^t \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))d\tilde W^n_s| > c/3) \\
&+ \mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))d\tilde W^n_s - \int_0^t \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s))d\tilde W_s| > c/3) \\
&+ \mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s))d\tilde W_s - \int_0^t \phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s| > c/3) \\
&=: J^1 + J^2 + J^3.
\end{aligned}$$

Using the assumptions, we estimate via Itô – Skorokhod's inequality with any $\delta > 0$,

$$\begin{aligned}
&J^1 = \mathbb{P}(|\int_0^t (\phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)))d\tilde W^n_s| > c/3) \\
&\le \mathbb{P}(\int_0^t (\phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)))^2 ds > \delta) \\
&+ \frac{9}{c^2}\mathbb{E}\left(\delta \wedge \int_0^T (\phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)))^2 ds\right).
\end{aligned}$$

Here the second term is small if we choose δ small. Let us consider the first term given $\delta > 0$. We have,

$$\begin{aligned}
&\mathbb{P}(\int_0^t (\phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)))^2 ds > \delta) \\
&\le \mathbb{P}(C\int_0^t (\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s) - \mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))^2 ds > \delta)
\end{aligned}$$

Let us take $\alpha = 1/(2m)$. By Bienaymé – Chebyshev – Markov's inequality,

$$\begin{aligned}
&\mathbb{P}(C\int_0^t (\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s) - \mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))^2 ds > \delta) \\
&\le (\delta/C)^{-1}(\mathbb{E}\int_0^t (\mathbb{E}^3(\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s) - \sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))^2 ds \\
&\le (\delta/C)^{-1}(\mathbb{E}\int_0^t \mathbb{E}^3(\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s) - \sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))^2 ds
\end{aligned}$$

$$= (\delta/C)^{-1}(\mathbb{E}\mathbb{E}^3 \int_0^t (\sigma^n(s,\tilde X_s^n,\tilde\xi_s^n) - \sigma^{n_0}(s,\tilde X_s^n,\tilde\xi_s^n))^2 ds.$$

Convergence of the latter term to zero as $n, n_0 \to \infty$ follows from the same considerations as for the drift in the previous step of the proof for the analogous term I^1 via Krylov's bound: so, we have,

$$(0 \le) \lim_{n,n_0\to\infty} J^1 \le \lim_{n,n_0\to\infty} \mathbb{E}\mathbb{E}^3 \int_0^t (\sigma^n(s,\tilde X_s^n,\tilde\xi_s^n) - \sigma^{n_0}(s,\tilde X_s^n,\tilde\xi_s^n))^2 ds = 0. \quad (36)$$

The idea is to use that here S^{11} and S^{13} are small as $\tilde R \to \infty$ by virtue of Krylov's bound for diffusions with bounded coefficients like for the drift in the previous step, due to the second moment estimates above and because all $\tilde X^n, \tilde\xi^n$ are uniformly bounded in probability. We have skipped using the functions similar to g^{n,n_0} from the previous step about the drift, which works here in a totally similar way.

The term J^2 converges to zero by Skorokhod's Lemma 4 in the Appendix

$$\int_0^t \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s^n,\tilde\xi_s^n))d\tilde W_s^n \stackrel{\mathbb{P}}{\to} \int_0^t \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s))d\tilde W_s, \quad n\to\infty,$$

with $f_t^n := \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s^n,\tilde\xi_s^n))$, $\quad f_t^0 := \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s))$ in this Lemma.

Consider J^3:

$$\begin{aligned}
J^3 &= \mathbb{P}(|\int_0^t (\phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s)) - \phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s)| > c/3) \\
&\le \mathbb{P}(\int_0^t (\phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s)) - \phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s)))^2 ds > \delta) \\
&+ \frac{9}{c^2}\mathbb{E}\left(\delta \wedge \int_0^T (\phi(\mathbb{E}^3\sigma^{n}(s,\tilde X_s,\tilde\xi_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s)))^2 ds\right).
\end{aligned}$$

As for J^1, the latter term here is small for small δ. For the former we have, similarly to J^1,

$$\begin{aligned}
&\mathbb{P}(\int_0^t (\phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s)) - \phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s)))^2 ds > \delta) \\
&\le \mathbb{P}(C\int_0^t (\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s) - \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))^2 ds > \delta) \\
&\le (\delta/C)^{-1}\mathbb{E}\mathbb{E}^3 \int_0^t (\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s) - \sigma(s,\tilde X_s,\tilde\xi_s))^2 ds.
\end{aligned}$$

Convergence of this term to zero as $n_0 \to \infty$ follows from by virtue of the Lemma 3 and 5, similarly to the analogous convergence of I^3 in the previous step. So,

$$\lim_{n_0 \to \infty} J^3 = 0.$$

This finishes the proof of the desired bound (35). Thus, a weak solution of the equation (1)–(2) exists in the case of $d_1 = d$ and under the assumption (6) instead of (5). For bounded coefficients and under (22)–(23) the Proposition 1 is proved.

6. Now consider the general case of unbounded coefficients satisfying the linear growth condition in x along with (10)–(11). Let

$$\sigma^n(t,x,y) = (\sigma(t,x,y)1(|x| \le n) + \sigma(0,0,0)1(|x| > n)) * \phi_n(x),$$

$$b^n(t,x,y) = b(t,x,y)1(|x| \le n).$$

Note that the function σ^n is Borel measurable, bounded, uniformly non-degenerate, and smooth (at least, Lipschitz) in x. Denote by X^n a solution of the equation

$$X^n_t = X_0 + \int_0^t \bar{B}^n[s, X^n_s, \mu^n_s]\, ds + \int_0^t \bar{\Sigma}^n[s, X^n_s, \mu^n_s]\, dW^n_s,$$

or

$$X^n_t = X_0 + \int_0^t \psi(\mathbb{E}^3 b^n[s, X^n_s, \xi^n_s])\, ds + \int_0^t \phi(\mathbb{E}^3 \sigma^n[s, X^n_s, \xi^n_s])\, dW^n_s,$$

with

$$\bar{\Sigma}^n[s,x,\mu] = \phi(\int \sigma^n(s,x,y)\,\mu(dy)) = \phi(\Sigma^n[s,x,\mu]),$$
$$\bar{B}^n[s,x,\mu] = \psi(\int b^n(s,x,y)\,\mu(dy)) = \phi(B^n[s,x,\mu]),$$

and where (ξ^n) are independent of (X^n, W^n) processes with the same distributions as X^n on some independent probability space. This weak (in fact, strong) solution exists for each n *due to the Proposition 1*. We will use again Skorokhod's technique; note that we could not apply it in one go because in the proof of the previous proposition the boundedness of both coefficients is essential.

A priori moment inequalities of the Lemma 2 hold true uniformly with respect to n. Hence, by Skorokhod's Lemma, choose a subsequence of equivalently distributed triples $(\tilde{X}^{n_k}, \tilde{\xi}^{n_k}, \tilde{W}^{n_k})$ converging in probability for any s to some limiting triple

$(\tilde X, \tilde \xi, \tilde W)$. Here $\tilde W)$ is a Wiener process of dimension d_1 (cf. with [21, Chapter 2], where, however, dimensions are equal, but it does not affect the conclusion). Convergence in the equation

$$\tilde X^{n_k}_t = \tilde X_0 + \int_0^t \psi(\mathbb{E}^3 b^{n_k}[s, \tilde X^{n_k}_s, \tilde \xi^{n_k}_s])\, ds + \int_0^t \phi(\mathbb{E}^3 \sigma^{n_k}[s, \tilde X^{n_k}_s, \tilde \xi^{n_k}_s])\, d\tilde W^{n_k}_s$$

towards the limiting equation

$$\tilde X_t = \tilde X_0 + \int_0^t \mathbb{E}^3 B[s, \tilde X_s, \tilde \xi_s]\, ds + \int_0^t \mathbb{E}^3 \sigma[s, \tilde X_s, \tilde \xi_s]\, d\tilde W_s$$

follows from the same calculus as in the proof of the Proposition 1 with the only difference that now σ may also be unbounded; however, this does require some additional care, in particular, because we want to use Krylov's bounds stated for unbounded coefficients (and do not forget about ϕ and ψ). Yet, in fact, we can use Krylov's bound from the Lemma 6.

First of all, recall that a priori bounds (17) – (26) hold true with constants not depending on n. Now, by Skorokhod's theorem on some probability space we have some equivalent processes $(\tilde X^{n'}_t, \tilde \xi^{n'}_t, \tilde W^{n'}_t)$ and a limiting triple $(\tilde X_t, \tilde \xi_t, \tilde W_t)$ such that for any t,

$$(\tilde X^{n'}_t, \tilde \xi^{n'}_t, \tilde W^{n'}_t) \stackrel{\mathbb{P}}{\to} (\tilde X_t, \tilde \xi_t, \tilde W_t). \qquad (37)$$

By virtue of the a priori bounds for $\tilde W^n$, the process $\tilde W_t$ is continuous and it is a Wiener process. Also, the limits are adapted to the corresponding filtration $\tilde{\mathcal{F}}_t := \bigvee_n \mathcal{F}^{(n)}_t$, and $\tilde W_t$ is a Wiener process with respect to this filtration. In particular, related Lebesgue and stochastic integrals are all well defined. Moreover, by virtue of the uniform estimates (21), the limit $(\tilde X_t, \tilde \xi_t)$ may be also regarded as continuous due to Kolmogorov's continuity theorem because the a priori bounds (8)–(12) remain valid for the limiting processes $\tilde X, \tilde \xi$. In particular, it is useful to note for the sequel that

$$\sup_{0\le t\le T} \mathbb{E}|\tilde X_t|^2 \le C_T(1+\mathbb{E}|x_0|^2) \quad \& \quad \mathbb{E}\sup_{0\le t\le T} |\tilde X_t|^2 \le C_T(1+\mathbb{E}|x_0|^2). \qquad (38)$$

7. Let us now show that

$$\int_0^t \psi(\mathbb{E}^3 b^{n'}(s, \tilde X^{n'}_s, \tilde \xi^{n'}_s))ds \stackrel{\mathbb{P}}{\to} \int_0^t \phi(\mathbb{E}^3 b(s, \tilde X_s, \tilde \xi_s))ds, \qquad (39)$$

and

$$\int_0^t \phi(\mathbb{E}^3 \sigma^{n'}(s, \tilde X^{n'}_s, \tilde \xi^{n'}_s))d\tilde W^{n'}_s \stackrel{\mathbb{P}}{\to} \int_0^t \phi(\mathbb{E}^3 \sigma(s, \tilde X_s, \tilde \xi_s))d\tilde W_s, \quad n' \to \infty. \qquad (40)$$

Denote $\tilde{R} := R - 1$. Then, given any $\epsilon > 0$, and slightly abusing notations by replacing n' by n, for any $t \le T$ by virtue of Chebyshev–Markov's inequality we conclude that for any $c > 0$ there exists $\tilde{R}$ such that

$$\mathbb{E}1(\gamma_{n,\tilde{R}} \wedge \gamma_{\tilde{R}} \le T) < \epsilon.$$

Further at one place we will need more precise estimates (see (17)):

$$\mathbb{P}(\gamma^X_{n,R-1} \le T) \le \frac{\mathbb{E}\sup_{0\le t\le T}|\tilde{X}^n_t|^2}{(R-1)^2} \le \frac{C(1+\mathbb{E}|x_0|^2)}{(R-1)^2},$$

and

$$\mathbb{P}(\gamma^X_{R-1} \le T) \le \frac{\mathbb{E}\sup_{0\le t\le T}|\tilde{X}_t|^2}{(R-1)^2} \le \frac{C(1+\mathbb{E}|x_0|^2)}{(R-1)^2},$$

by virtue of the Chebyshev–Markov's inequality. Hence,

$$\mathbb{P}(\gamma^X_{n,R-1} \wedge \gamma^X_{R-1} \le T) \le \frac{2C(1+\mathbb{E}|x_0|^2)}{(R-1)^2}. \tag{41}$$

Now, we estimate, with $\tilde{c} = c/C$, $m = m_1$,

$$\mathbb{P}\left(\left|\int_0^t \psi(\mathbb{E}^3 b^{n'}(s,\tilde{X}^{n'}_s,\tilde{\xi}^{n'}_s))\,ds - \int_0^t \psi(\mathbb{E}^3 b(s,\tilde{X}_s,\tilde{\xi}_s))\,ds\right| > c\right)$$

$$\le \mathbb{P}\left(C\int_0^t (1+(|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m)\left|\mathbb{E}^3\left(b^{n'}(s,\tilde{X}^{n'}_s,\tilde{\xi}^{n'}_s) - b(s,\tilde{X}_s,\tilde{\xi}_s)\right)\right|\,ds > c\right)$$

$$= \mathbb{P}\left(\int_0^t (1+(|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m)\mathbb{E}^3\Big|\big(1(\gamma_{n,\tilde{R}} \wedge \gamma_{\tilde{R}} \le T) + 1(\gamma_{n,\tilde{R}} \wedge \gamma_{\tilde{R}} > T)\big)\right.$$

$$\left.\times\left(b^{n'}(s,\tilde{X}^{n'}_s,\tilde{\xi}^{n'}_s) - b(s,\tilde{X}_s,\tilde{\xi}_s)\right)\Big|\,ds > \tilde{c}\right)$$

$$\le \mathbb{P}\left(\int_0^t (1+(|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m)\mathbb{E}^3 1(\gamma_{n,\tilde{R}} \wedge \gamma_{\tilde{R}} \le T)\left|\left(b^{n'}(s,\tilde{X}^{n'}_s,\tilde{\xi}^{n'}_s) - b(s,\tilde{X}_s,\tilde{\xi}_s)\right)\right|\,ds > \tilde{c}/2\right)$$

$$+ \mathbb{P}\left(\int_0^t (1+(|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m)\mathbb{E}^3 1(\gamma_{n,\tilde{R}} \wedge \gamma_{\tilde{R}} > T)\left|\left(b^{n'}(s,\tilde{X}^{n'}_s,\tilde{\xi}^{n'}_s) - b(s,\tilde{X}_s,\tilde{\xi}_s)\right)\right|\,ds > \tilde{c}/2\right)$$

$$\leq \mathbb{P}\left(1(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R}\leq T)\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^m)\mathbb{E}^3\left|\left(b^{n'}(s,\tilde X^{n'}_s,\tilde\xi^{n'}_s)-b(s,\tilde X_s,\tilde\xi_s)\right)\right|ds>\tilde c/2\right)$$

$$+\,\mathbb{P}\left(\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^m)\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}\leq T)\left|\left(b^{n'}(s,\tilde X^{n'}_s,\tilde\xi^{n'}_s)-b(s,\tilde X_s,\tilde\xi_s)\right)\right|ds>\tilde c/2\right)$$

$$+\,\mathbb{P}\left(1(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R}> T)\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^m)\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}> T)\right.$$

$$\left.\times\left|\left(b^{n'}(s,\tilde X^{n'}_s,\tilde\xi^{n'}_s)-b(s,\tilde X_s,\tilde\xi_s)\right)\right|ds>\tilde c/2\right)$$

$$\leq \mathbb{P}(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R}\leq T)$$

$$+\,\mathbb{P}\left(\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^m)\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}\leq T)\left|\left(b^{n'}(s,\tilde X^{n'}_s,\tilde\xi^{n'}_s)-b(s,\tilde X_s,\tilde\xi_s)\right)\right|ds>\tilde c/2\right)$$

$$+\,\mathbb{P}\left(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R}> T;\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^m)\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}> T)\right.$$

$$\left.\times\left|\left(b^{n}(s,\tilde X^{n}_s,\tilde\xi^{n}_s)-b(s,\tilde X_s,\tilde\xi_s)\right)\right|ds>\frac{\tilde c}{2}\right)=:L^1+L^2+L^3.$$

Here the first term L^1 does not exceed ϵ if R is large enough, uniformly with respect to n.

Consider the term L^2. We estimate, using the linear growth in x assumption,

$$\mathbb{P}\left(\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^m)\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}\leq T)\left|\left(b^{n'}(s,\tilde X^{n'}_s,\tilde\xi^{n'}_s)-b(s,\tilde X_s,\tilde\xi_s)\right)\right|ds>\tilde c/2\right)$$

$$\leq \mathbb{P}\left(\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}\leq T)\int_0^t (1+(|\tilde X^{n'}_s|\vee|\tilde X_s|)^{m+1})ds>\tilde c/2\right).$$

Since the processes $\tilde X^{n'}_s$ together with their limit $\tilde X_s$ are bounded in probability uniformly in n, and because

$$\mathbb{E}^3 1(\gamma^{\xi}_{n,\tilde{R}} \wedge \gamma^{\xi}_{\tilde{R}} \le T) \to 0, \quad R \to \infty,$$

we conclude that

$$\lim_{R\to\infty} \sup_n L^2 = 0.$$

Further, consider L^3. Let us fix some n_0 and let $n > n_0$. Replacing $\tilde{c}/2$ by c for simplicity, we have for any $t \le T$,

$$\begin{aligned}
&\mathbb{P}\left(\gamma^X_{n,\tilde{R}} \wedge \gamma^X_{\tilde{R}} > T;\ \int_0^t (1 + (|\tilde{X}^n_s| \vee |\tilde{X}_s|)^m) \mathbb{E}^3 1(\gamma^{\xi}_{n,\tilde{R}} \wedge \gamma^{\xi}_{\tilde{R}} > T)\right. \\
&\qquad \left. \times \left| \left(b^n(s, \tilde{X}^n_s, \tilde{\xi}^n_s) - b(s, \tilde{X}_s, \tilde{\xi}_s) \right) \right| ds > c \right) \\
&\le \mathbb{P}\left(\gamma^X_{n,\tilde{R}} \wedge \gamma^X_{\tilde{R}} > T;\ \int_0^t (1 + (|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m) \mathbb{E}^3 1(\gamma^{\xi}_{n,\tilde{R}} \wedge \gamma^{\xi}_{\tilde{R}} > T)\right. \\
&\qquad \left. \times \left| \left(b^n(s, \tilde{X}^n_s, \tilde{\xi}^n_s) - b^{n_0}(s, \tilde{X}^n_s, \tilde{\xi}^n_s) \right) \right| ds > \frac{c}{3} \right) \\
&+\mathbb{P}\left(\gamma^X_{n,\tilde{R}} \wedge \gamma^X_{\tilde{R}} > T;\ \int_0^t (1 + (|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m) \mathbb{E}^3 1(\gamma^{\xi}_{n,\tilde{R}} \wedge \gamma^{\xi}_{\tilde{R}} > T)\right. \\
&\qquad \left. \times \left| \left(b^{n_0}(s, \tilde{X}^n_s, \tilde{\xi}^n_s) - b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s) \right) \right| ds > \frac{c}{3} \right) \\
&+\mathbb{P}\left(\gamma^X_{n,\tilde{R}} \wedge \gamma^X_{\tilde{R}} > T;\ \int_0^t (1 + (|\tilde{X}^{n'}_s| \vee |\tilde{X}_s|)^m) \mathbb{E}^3 1(\gamma^{\xi}_{n,\tilde{R}} \wedge \gamma^{\xi}_{\tilde{R}} > T)\right. \\
&\qquad \left. \times \left| \left(b^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s) - b(s, \tilde{X}_s, \tilde{\xi}_s) \right) \right| ds > \frac{c}{3} \right) \\
&=: M^1 + M^2 + M^3.
\end{aligned}$$

Denote as earlier

$$g^{n,n_0}(s,x,\xi) := b^n(s,x,\xi) - b^{n_0}(s,x,\xi), \quad g^{n_0}(s,x,\xi) := b^{n_0}(s,x,\xi) - b(s,x,\xi).$$

Let $\alpha = (\le) 1/(2m+2)$. Note that

$$\mathbb{E}(1 + (\sup_s |\tilde{X}^{n'}_s| \vee \sup_s |\tilde{X}_s|)^{2m\alpha} \le C < \infty,$$

and that this constant C is uniform in n' and does not depend on R. By CBS we decode Cauchy – Buniakovsky – Schwarz inequality. The first summand M^1 may be estimated by Bienaymé – Chebyshev – Markov's inequality as

$$
\begin{aligned}
M^1 \le C\,\mathbb{E}\Bigg(1(\gamma^X_{n,\tilde R} > T, \gamma^X_{\tilde R} > T)\int_0^T (1 + (|\tilde X^{n'}_s| \vee |\tilde X_s|)^m)
\\
\times \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)|b^n(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s)|\, ds\Bigg)^\alpha
\\
\le C\,\mathbb{E}1(\gamma^X_{n,\tilde R} > T, \gamma^X_{\tilde R} > T)(1 + (\sup_s |\tilde X^{n'}_s| \vee \sup_s |\tilde X_s|)^{m\alpha})
\\
\times \left(\int_0^T \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)|b^n(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s)|\, ds\right)^\alpha
\\
\stackrel{\text{CBS}}{\le} C\left(\mathbb{E}1(\gamma^X_{n,\tilde R} > T, \gamma^X_{\tilde R} > T)(1 + (\sup_s |\tilde X^{n'}_s| \vee \sup_s |\tilde X_s|)^{2m\alpha})\right)^{1/2}
\\
\times \left(\mathbb{E}\left(\int_0^T \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)|b^n(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s)|\, ds\right)^{2\alpha}\right)^{1/2}
\\
\stackrel{\text{Hölder}}{\le} C\left(\int_0^T \mathbb{E}1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)|b^n(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s)|\, ds\right)^\alpha
\\
= C\left(\int_0^T \mathbb{E}1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)|g^{n,n_0}(s, \tilde X^n_{s\wedge\gamma_{n,\tilde R}}, \tilde \xi^n_{s\wedge\gamma_{n,\tilde R}})|\, ds\right)^\alpha.
\end{aligned}
\tag{42}
$$

Here the couple $(\tilde X^n_{s\wedge\gamma_{n,\tilde R}}, \tilde \xi^n_{s\wedge\gamma_{n,\tilde R}})$ is a stopped diffusion with coefficients bounded by norm in state variables (x,ξ) by the value $C\tilde R$ uniformly with respect to n, and with the diffusion coefficient uniformly non-degenerate.

Denote by $\hat B^{n,\tilde R}[s,x,\mu]$ and $\hat \Sigma^{n,\tilde R}[s,x,\mu]$ bounded vector and matrix functions in x respectively, with $\hat \Sigma^{n,\tilde R}[s,x,\mu]$ uniformly non-degenerate and smooth (e.g., C^1), such that

$$
\hat B^{n,\tilde R}[s,x,\mu] = B^n[s,x,\mu], \;\; \hat \Sigma^{n,\tilde R}[s,x,\mu] = \Sigma^n[s,x,\mu], \quad |x| \le \tilde R.
$$

Let $(\hat X^n_s) = (\hat X^{n,\tilde R}_s)$ be a (strong) solution of the Itô equation,

$$
d\hat X^n_t = \hat B^{n,\tilde R}[t, \hat X^n_t, \mu^n_t]\, dt + \phi(\hat \Sigma^{n,\tilde R}[t, \hat X^n_t, \mu^n_t])\, d\tilde W^n_t, \quad \hat X^n_0 = x_0, \tag{43}
$$

where μ_t^n is still the marginal distribution of X_t^n and $\tilde X_t^n$. Let also $\hat\xi^n$ be an equivalent independent copy of the process $\hat X_t^n$ satisfying the equation

$$d\hat\xi_t^n = \hat B^{n,\tilde R}[t,\hat\xi_t^n,\mu_t]dt + \hat\Sigma^{n,\tilde R}[t,\hat\xi_t^n,\mu_t]d\tilde W_t^{\prime,n}, \quad t\ge 0, \qquad \mathcal{L}(\xi_0^n)=\mathcal{L}(x_0). \tag{44}$$

We may assume that the Wiener processes $\tilde W^{\prime,n}$ here are the same as in the equation (24) for $\tilde\xi^n$. (Emphasize that solutions $\hat\xi^n$ are strong ones; this is why we have mollified σ in the variable x). Then it follows that on $[0,t\wedge\gamma_{n,\tilde R}]$ the processes $\tilde X^n$ and $\hat X^n$ coincide (see [22, Theorem 6.2.1(v)]), as well as $\tilde\xi^n$ coincide with $\hat\xi^n$. (We highlight that the same stopping times $\gamma_{n,\tilde R}$ can be used!) Then the bound for I^1 in the second line of (42) may be rewritten as

$$\begin{aligned} M^1 &\le C\left(\mathbb{E}1(\gamma_{n,\tilde R}>T,\gamma_{\tilde R}>T)\int_0^T |g^{n,n_0}(s,\hat X^n_{s\wedge\gamma_{n,\tilde R}},\hat\xi^n_{s\wedge\gamma_{n,\tilde R}})|\,ds\right)^\alpha \\ &\le C\left(\mathbb{E}\int_0^T |g^{n,n_0}(s,\hat X^n_{s\wedge\gamma_{n,\tilde R}},\hat\xi^n_{s\wedge\gamma_{n,\tilde R}})|\,ds\right)^\alpha. \end{aligned}\tag{45}$$

The values of the function $g^{n,n_0}(s,x,\xi)$ outside the set $\{(x,\xi):\,(|x|\vee|\xi|)\le\tilde R\}$ are not relevant to the evaluation of the expression in the second line of (45). So, without losing of generality we may assume for our goal that $g^{n,n_0}(s,x,\xi)$ vanishes outside of this ball. Then, by Krylov's estimate from the Lemma 6 we obtain with some constant N depending on the dimension d and on R through the ellipticity constants of the diffusion matrix and the sup-norm of the drift with some N_R,

$$M^1 \le C\left(\mathbb{E}\int_0^T |g^{n,n_0}(s,\hat X^n_{s\wedge\gamma_{n,\tilde R}},\hat\xi^n_{s\wedge\gamma_{n,\tilde R}})|\,ds\right)^\alpha$$

$$= N_R\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^n(s,x,\xi)-b^{n_0}(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{\alpha}{2d+1}}$$

$$\le N_R\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^n(s,x,\xi)-b(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{\alpha}{2d+1}}$$

$$+N_R\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^{n_0}(s,x,\xi)-b(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{\alpha}{2d+1}} \to 0, \quad n,n_0\to\infty,$$

for each R, by virtue of the properties of mollified functions. Hence, for any R,

$$\lim_{n,n_0\to\infty} M^1 = 0.$$

Further, the second term M^2 admits the estimate (for any $0 \le t \le T$),

$$M^2 \le \mathbb{P}\Big(\gamma^X_{n,\tilde R} \wedge \gamma^X_{\tilde R} > T; \int_0^t (1 + (|\hat X^n_{s\wedge\gamma_{n,\tilde R}}| \vee |\hat X_{s\wedge\gamma_{n,\tilde R}}|)^m)$$

$$\times \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)\Big|\Big(b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s)ds - b^{n_0}(s, \tilde X_s, \tilde \xi_s)\Big)\Big| ds > \tfrac{c}{3}\Big)$$

$$\le \frac{3}{c}(1 + \tilde R^m)\mathbb{E}\mathbb{E}^3 1(\gamma_{n,\tilde R} > T, \gamma_{\tilde R} > T) \int_0^T |b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X_s, \tilde \xi_s)|\, ds,$$

which tends to zero as $n \to \infty$ due to the Lebesgue bounded convergence theorem. Indeed, on the set $(\gamma_{n,\tilde R} > T, \gamma_{\tilde R} > T)$, the random variable $\int_0^T |b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X_s, \tilde \xi_s)|\, ds$ is bounded uniformly in n, and

$$\int_0^T |b^{n_0}(s, \tilde X^n_s, \tilde \xi^n_s) - b^{n_0}(s, \tilde X_s, \tilde \xi_s)|\, ds \to 0, \quad n \to \infty,$$

in probability. Therefore,

$$\lim_{n,n_0\to\infty} M^2 = 0.$$

To tackle the third term M^3, the indicators $1(\gamma_{\tilde R} > T)$ are not enough and we need some new auxiliary function. Let $R > 1$ and let $0 \le w(x,\xi) \le 1$ be any continuous function which equals 1 for every $|x| \vee |\xi| \le R - 1\,(= \tilde R)$ and zero for every $|x| \vee |\xi| > R$. Then we have,

$$M^3 = \mathbb{P}\Big(\gamma^X_{n,\tilde R} \wedge \gamma^X_{\tilde R} > T; \int_0^T (1 + (|\tilde X^n_s| \vee |\tilde X_s|)^m)\mathbb{E}^3 1(\gamma^\xi_{n,\tilde R} \wedge \gamma^\xi_{\tilde R} > T)$$

$$\times \Big|\Big(b^{n_0}(s, \tilde X_s, \tilde \xi_s)ds - b(s, \tilde X_s, \tilde \xi_s)\Big)\Big| ds > \frac{c}{3}\Big)$$

$$\le C(1 + \tilde R^m)\, \mathbb{E} 1(\gamma_{\tilde R} \wedge \gamma_{n,\tilde R} > T) \int_0^T |g^{n_0}(s, \tilde X_s, \tilde \xi_s)|\, ds$$

$$\le C(1 + \tilde R^m)\, \mathbb{E} \int_0^T |g^{n_0}_R(s, \tilde X_s, \tilde \xi_s)|\, ds \tag{46}$$

For any Borel function g_R Krylov's bound (Lemma 6) implies the inequality

$$\sup_n \mathbb{E} \int_0^T |g_R(s, \tilde X^n_s, \tilde \xi^n_s)|\, ds \le N_R \|g_R\|_{L_{2d+1}([0,T]\times B_R \times B_R)}$$

Assume g_R continuous in (x, y) and vanishing outside $B_R \times B_R$. Then in the limit we obtain,

$$\mathbb{E} \int_0^T |g_R(s, \tilde{X}_s, \tilde{\xi}_s)| \, ds \le N_R \|g_R\|_{L_{2d+1}([0,T]\times B_R \times B_R)}.$$

By the Lemma 5, this bound is valid for any $g \in L_{2d+1}([0, T] \times B_R \times B_R)$. Now,

$$M^3 \le C(1 + \tilde{R}^m) \, \mathbb{E}1(\gamma_{\tilde{R}} \wedge \gamma_{n,\tilde{R}} > T) \int_0^T |g^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)| \, ds$$

$$\le C(1 + \tilde{R}^m) \, \mathbb{E} \int_0^T |g_R^{n_0}(s, \tilde{X}_s, \tilde{\xi}_s)| \, ds \le C(1 + \tilde{R}^m) \, N_R \|g_R^{n_0}\|_{L_{2d+1}([0,T]\times B_R \times B_R)}. \quad (47)$$

Hence,

$$\lim_{n_0 \to \infty} M^3 = 0.$$

The convergence (39) is, thus, proved.

8. Now let us consider convergence of stochastic integrals in (40). Our goal is an estimate similar to that for the drift and Lebesgue integrals above:

$$\mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^n(s, \tilde{X}_s^n, \tilde{\xi}_s^n))dW_s^n - \int_0^t \phi(\mathbb{E}^3\sigma(s, \tilde{X}_s, \tilde{\xi}_s))d\tilde{W}_s| > c) < C\epsilon, \quad (48)$$

for any $c, \epsilon > 0$ and n large enough. In principle, the task is similar to the convergence of Lebesgue integrals tackled in the previous steps. Hence, we mainly show how to tackle the additional obstacle due to different Wiener processes dW_s and dW_s^n in the stochastic integrals. Fortunately, we have a tool for this which is Skorokhod's Lemma 4 from the Appendix below. However, it is not applicable directly because our processes may be unbounded, so we should overcome this with the help of a truncation which reduces the problem to bounded processes.

By virtue of [22, Theorem 6.2.1(v)] and similarly to the calculus for Lebesgue integrals in the previous steps, yet using second moments instead of the first ones by the evident reason we estimate,

$$\mathbb{P}(|\int_0^t \phi(\mathbb{E}^3\sigma^n(s, \tilde{X}_s^n, \tilde{\xi}_s^n))d\tilde{W}_s^n - \int_0^t \phi(\mathbb{E}^3\sigma(s, \tilde{X}_s, \tilde{\xi}_s))d\tilde{W}_s| > c)$$

$$= \mathbb{P}(\gamma_{\tilde{R}}^X \wedge \gamma_{n,\tilde{R}}^X \le T; |\int_0^t \phi(\mathbb{E}^3\sigma^n(s, \tilde{X}_s^n, \tilde{\xi}_s^n))d\tilde{W}_s^n - \int_0^t \phi(\mathbb{E}^3\sigma(s, \tilde{X}_s, \tilde{\xi}_s))d\tilde{W}_s| > c)$$

$$+\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^t\phi(\mathbb{E}^3\sigma^n(s,\tilde X^n_s,\tilde\xi^n_s))d\tilde W^n_s-\int_0^t\phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s|>c)$$

$$\leq\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}\leq T)$$

$$+\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^t\phi(\mathbb{E}^3\sigma^n(s,\hat X^n_s,\hat\xi^n_s))d\tilde W^n_s-\int_0^t\phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s|>c).$$

The first term here $\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}\leq T)$ is small if R is large enough, as we have seen in the earlier steps. It remains to evaluate the second term. We estimate,

$$\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^t\phi(\mathbb{E}^3\sigma^n(s,\hat X^n_s,\hat\xi^n_s))d\tilde W^n_s-\int_0^t\phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s|>c)$$

$$\leq\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^T\phi(\mathbb{E}^3\sigma^n(s,\hat X^n_s,\hat\xi^n_s))d\tilde W^n_s-\int_0^t\phi(\mathbb{E}^3\sigma^{n_0}(s,\hat X^n_s,\hat\xi^n_s))d\tilde W^n_s|>c/3)$$

$$+\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^T\phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s))d\tilde W^n_s-\int_0^t\phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s))d\tilde W_s|>c/3)$$

$$+\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^T\phi(\mathbb{E}^3\sigma^{n_0}(s,\tilde X_s,\tilde\xi_s))d\tilde W_s-\int_0^t\phi(\mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s))d\tilde W_s|>c/3)$$

$$=:J^1+J^2+J^3.$$

Now for all three terms the evaluation is similar to that for the drift term, except for one difference. We start with J^1:

$$J^1=\mathbb{P}(\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}>T;|\int_0^T(\phi(\mathbb{E}^3\sigma^n(s,\hat X^n_s,\tilde\xi^n_s))d\tilde W^n_s-\phi(\mathbb{E}^3\sigma^{n_0}(s,\hat X^n_s,\tilde\xi^n_s)))d\tilde W^n_s|>c/3)$$

$$\leq\mathbb{P}(|\int_0^{T\wedge\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R}}(\phi(\mathbb{E}^3\sigma^n(s,\hat X^n_s,\tilde\xi^n_s))-\phi(\mathbb{E}^3\sigma^{n_0}(s,\hat X^n_s,\tilde\xi^n_s)))d\tilde W^n_s|>c/3)$$

$$= \mathbb{P}(|\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(\phi(\mathbb{E}^3\sigma^n(s, \hat X^n_s, \tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s)))d\tilde W^n_s| > c/3)$$

By virtue of the (simplified one, without sup) Itô – Skorokhod's inequality (see [22, Theorem 6.3.5]: for any $\delta > 0$,

$$\mathbb{P}(|\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(\phi(\mathbb{E}^3\sigma^n(s, \hat X^n_s, \tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s)))d\tilde W^n_s| > c/3)$$

$$\le P(\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(\phi(\mathbb{E}^3\sigma^n(s, \hat X^n_s, \tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s)))^2 ds > \delta)$$

$$+ \frac{9}{c^2}\mathbb{E}\left(\delta \wedge \int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(\phi(\mathbb{E}^3\sigma^n(s, \hat X^n_s, \tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s)))^2 ds\right)$$

$$=: F^1 + F^2.$$

The term F^2 is small if we choose δ small enough. Let us consider F^1. We estimate,

$$P(\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(\phi(\mathbb{E}^3\sigma^n(s, \hat X^n_s, \tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s)))^2 ds > \delta)$$

$$\le P(\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(\phi(\mathbb{E}^3\sigma^n(s, \hat X^n_s, \tilde\xi^n_s)) - \phi(\mathbb{E}^3\sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s)))^2 ds > \delta)$$

$$\le P(\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(1 + |\hat X^n_s|^m)^2 |\mathbb{E}^3(\sigma^n(s, \hat X^n_s, \tilde\xi^n_s) - \sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s))|^2 ds > \delta)$$

$$= P(\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})(1 + |\hat X^n_s|^m)^2 |\mathbb{E}^3(1(T \ge \gamma^\xi_{\tilde R} \wedge \gamma^\xi_{n,\tilde R}) + 1(T < \gamma^\xi_{\tilde R} \wedge \gamma^\xi_{n,\tilde R}))$$

$$\times(\sigma^n(s, \hat X^n_s, \tilde\xi^n_s) - \sigma^{n_0}(s, \hat X^n_s, \tilde\xi^n_s))|^2 ds > \delta)$$

$$\le P(\int_0^T 1(s < \gamma^X_{\tilde R} \wedge \gamma^X_{n,\tilde R})\mathbb{E}^3(1 + |\hat X^n_s|^m)^2 1(T \ge \gamma^\xi_{\tilde R} \wedge \gamma^\xi_{n,\tilde R})$$

$$\times \|\sigma^n(s,\hat X^n_s,\tilde\xi^n_s)-\sigma^{n_0}(s,\hat X^n_s,\tilde\xi^n_s)||^2 ds > \delta)$$

$$+P(\int_0^T 1(s<\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R})(1+|\hat X^n_s|^m)^2\mathbb{E}^3 1(T<\gamma^\xi_{\tilde R}\wedge\gamma^\xi_{n,\tilde R})$$

$$\times\|\sigma^n(s,\hat X^n_s,\tilde\xi^n_s)-\sigma^{n_0}(s,\hat X^n_s,\tilde\xi^n_s)\|\|^2 ds>\delta)$$

$$=: S^1+S^2.$$

For S^2 we can replace $\tilde\xi$ by $\hat\xi$ and then use Bienaymé – Chebychev – Markov's inequality as in the drift evaluation, to obtain an L_{4d+1}-bound (not L_{2d+1} because of $\|\sigma^n-\sigma^{n_0}\|^2$) by virtue of Krylov's bound (see the Lemma 6):

$$S^2 = P(\int_0^T 1(s<\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R})(1+|\hat X^n_s|^m)^2\mathbb{E}^3 1(T<\gamma^\xi_{\tilde R}\wedge\gamma^\xi_{n,\tilde R})$$

$$\times\|\sigma^n(s,\hat X^n_s,\hat\xi^n_s)-\sigma^{n_0}(s,\hat X^n_s,\hat\xi^n_s)\|^2 ds>\delta)$$

$$\le C_R\|g^{n,n_0}\|^2_{L_{4d+1}([0,T]\times B_R\times B_R)}\to 0,\quad n,n_0\to\infty.$$

For S^1 let $\beta=1/(4m+4)$ and also use Bienaymé – Chebychev – Markov:

$$P(\int_0^T 1(s<\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R})\mathbb{E}^3(1+|\hat X^n_s|^m)^2 1(T\ge\gamma^\xi_{\tilde R}\wedge\gamma^\xi_{n,\tilde R})$$

$$\times\|\sigma^n(s,\hat X^n_s,\tilde\xi^n_s)-\sigma^{n_0}(s,\hat X^n_s,\tilde\xi^n_s)||^2 ds>\delta)$$

$$\le C\mathbb{E}(\int_0^T 1(s<\gamma^X_{\tilde R}\wedge\gamma^X_{n,\tilde R})\mathbb{E}^3(1+|\hat X^n_s|^m)^2 1(T\ge\gamma^\xi_{\tilde R}\wedge\gamma^\xi_{n,\tilde R})$$

$$\times\|\sigma^n(s,\hat X^n_s,\tilde\xi^n_s)-\sigma^{n_0}(s,\hat X^n_s,\tilde\xi^n_s)||^2 ds)^\beta$$

$$\overset{\text{CBS}}{\leq} C(T\mathbb{E}\sup_s(1+|\hat{X}^n_s|))^{1/2}(\mathbb{P}(T \geq \gamma^\xi_{\tilde{R}} \wedge \gamma^\xi_{n,\tilde{R}}))^{\beta/2}.$$

Due to the a priori bounds from the Lemma 2, the first multiplier here is bounded, while the second is small if R is large enough, all uniformly in n. Hence, we have,

$$\lim_{R\to\infty} S^1 = 0.$$

This proves the desired bound (48). So, (39) and (40) hold true, and thus, the Proposition is proved.

2.3 Proof of Theorem 1

1. We will use a hint from [35, section 4] in order to reduce the statement to the case considered earlier in the Proposition 1. Due to a more involved structure of the equation in this paper, it is desirable to repeat the details here.

Denote $\tilde{\Sigma}[t,x,\mu] := \sqrt{A[t,x,\mu]}$, where $A[t,x,\mu] := \Sigma[t,x,\mu]\Sigma^*[t,x,\mu]$, and *suppose* that there exists a (weak) solution $\tilde{X}$ of the equation,

$$\tilde{X}_t = x_0 + \int_0^t B[s,\tilde{X}_s,\mu_s]ds + \int_0^t \tilde{\Sigma}[s,\tilde{X}_s,\mu_s]d\tilde{W}_s, \tag{49}$$

with some d-dimensional Wiener process $(\tilde{W}_t,\, t\geq 0)$ on some probability space and where μ_s stands for the distribution of $\tilde{X}_s$.

Existence of this weak solution is already justified in the Proposition 1. Here the matrix-function $\sqrt{A[t,x,\mu]}$ is defined via the Cauchy – Riesz – Dunford formula for a function of a positive self-adjoint square root of the matrix A (see, e.g., [9, VII.3.9]),

$$\sqrt{A[t,x,\mu]} = \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1}\,d\lambda \tag{50}$$

with a unique contour Γ for bounded sigma, or via its analogue

$$\sqrt{A[t,x,\mu]} = \frac{1}{2\pi i}\sum_{i=1}^\infty 1(i-1\leq |x| < i)\oint_{\Gamma_i} \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1}\,d\lambda, \tag{51}$$

for unbounded, where

$$\Gamma_i = \{\lambda\in\mathbb{C}:\, |\lambda| = \sup_{t,x,\mu:\,|x|\leq i}\|A[t,x,\mu]\| + 1\},$$

where the contour $\Gamma_i \subset \mathbb{C}$ in the complex plane is chosen in a way so that its interior contains all the eigenvalues of the (elliptic) matrix $A[s,x,\cdot]$ for $|x| \le i$.

2. Further, without loss of 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

$$p[s,x,\mu] := \tilde\Sigma[s,x,\mu]^{-1}\,\Sigma[s,x,\mu]. \tag{52}$$

Note that the matrix $\tilde\Sigma[s,x,\mu]$ is symmetric and that

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

$$= \Sigma^*[s,x,\mu](A)^{-1}[s,x,\mu]\Sigma[s,x,\mu],$$

$$p^*[s,x,\mu]p[s,x,\mu]p^*[s,x,\mu]p[s,x,\mu]$$

$$= \Sigma^*[s,x,\mu](A)^{-1}[s,x,\mu]\Sigma[s,x,\mu]\Sigma^*[s,x,\mu](A)^{-1}[s,x,\mu]\Sigma[s,x,\mu]$$

$$= \Sigma^*(A)^{-1}(A)A)^{-1}\Sigma[s,x,\mu] = \Sigma^*(A)^{-1}\Sigma[s,x,\mu],$$

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

$$\Sigma[s,x,\mu]p^*[s,x,\mu] = a[s,x,\mu](a[s,x,\mu])^{-1/2} = (a[s,x,\mu])^{1/2},$$

$$\Sigma[s,x,\mu]p^*[s,x,\mu]p[s,x,\mu] = (a[s,x,\mu])^{1/2}p[s,x,\mu]$$

$$= (a[s,x,\mu])^{1/2}(a[s,x,\mu])^{-1/2}\Sigma[s,x,\mu] = \Sigma[s,x,\mu].$$

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

$$\langle W^0, W^0\rangle_t = \int_0^t p^*p[s,\tilde X_s,\mu_s]\,ds + \int (I - p^*p[s,\tilde X_s,\mu_s])^*(I - p^*p[s,\tilde X_s,\mu_s])\,ds$$

$$= \int (p^* p[s, \tilde X_s, \mu_s] + I - 2p^* p[s, \tilde X_s, \mu_s] + p^* p p^* p[s, \tilde X_s, \mu_s])\, ds$$

$$= \int (I - p^* p[s, \tilde X_s, \mu_s] + p^* p p^* p[s, \tilde X_s, \mu_s])\, ds$$

$$= \int (I - \Sigma^* (A)^{-1} \Sigma[s, \tilde X_s, \mu_s] + \Sigma^* (A)^{-1} (A) (A)^{-1} \Sigma[s, \tilde X_s, \mu_s])\, ds = \int_0^t I\, ds = t\, I.$$

Next, due to the stochastic integration rules (see [14]),

$$\int_0^t \Sigma[s, \tilde X_s, \mu_s]\, dW^0_s = \int \Sigma p^*[s, \tilde X_s, \mu_s]\, d\tilde W + \int \Sigma (I - p^* p)[s, \tilde X_s, \mu_s]\, d\bar W$$
$$= \int (A)^{1/2}[s, \tilde X_s, \mu_s]\, d\tilde W = \int \tilde\Sigma[s, \tilde X_s, \mu_s]\, d\tilde W = \tilde X_t - x_0 - \int_0^t B[s, \tilde X_s, \mu_s]\, ds. \tag{53}$$

In other words, $(\tilde X, W^0)$ is a (weak) solution of the equation (1). It remains to notice that since we did not change measures, μ_s is still the distribution of $\tilde X_s$ by the assumption. The proof of the Theorem 1 is thus completed.

Remark 3. *There is a non-rigorous view that for SDE solutions everything related to weak solutions and weak uniqueness depends only on the matrix $\sigma^*\sigma$ and not on σ itself. This is not precise. Firstly, for strong solutions this is not true because regularity such as Lipschitz condition or even a simple continuity may fail for a badly chosen square root, let us forget about non-Borel square roots. Secondly, even for weak solutions in the absence of non-degeneracy and if the square root is not continuous, there is no guarantee that weak solution exists for* any *square root. Also, existing results about weak solutions and weak uniqueness – see [5, 11] – impose conditions on σ and not on $\sigma\sigma^*$. Hence, a vague "common knowledge" is not sufficient and had to show the calculus.*

3 Strong solutions; strong and weak uniqueness

3.1 On strong existence

In this section it is shown that strong solution of the equation (1)–(2) exists under appropriate conditions. Emphasize that we do not claim strong *uniqueness* in this section, but only strong existence in the sense of the Definition 1. We also notice for interested readers that in [33] the assumption of continuity in time was dropped in comparison to [32]; so, just a certain (local) Lipschitz condition suffices for our aim.

Proposition 2. *Let $\mathbb{E}|x_0|^4 < \infty$. Let the coefficients b and σ satisfy all conditions of the Theorem 1 and the non-degeneracy assumption (6), and let just σ be Lipschitz in x uniformly with respect to s and locally with respect to y,*

$$\|\sigma(t,x,y) - \sigma(t,x',y)\| \le C(1+|y|^2)|x-x'|. \tag{54}$$

Then the equation (1)–(2) 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.

This result is likely to be a common knowledge. A brief sketch of the proof is presented below for completeness and because the authors were unable to find an exact reference

1. First of all, note that weak solutions exist and a priori bounds (17)–(21) are valid. Considerations are based on the results from [32] and [33] about strong solutions for SDEs for a Borel measurable drift which is assumed bounded or with a linear growth in both papers. Since weak solution does exist, 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,

$$dX_t = \tilde b(t,X_t)dt + \tilde\sigma(t,X_t)dW_t, \qquad X_0 = x. \tag{55}$$

Recall that according to the Corollary 1, the new coefficients $\tilde b(t,x)$ and $\tilde\sigma(t,x)$ are Borel measurable.

2. Now in order to establish strong existence it suffices to verify that the new coefficient and $\tilde\sigma$ satisfies linear growth in x condition uniform in time, and Lipschitz condition in x, and is uniformly non-degenerate.

(a) In the case 1 we have, for any $T>0$ and $0\le t\le T$,

$$|\tilde b(t,x)| = |\int b(t,x,y)\,\mu_t(dy)| \le C|\int (1+|x|)\,\mu_t(dy)| = C\,(1+|x|).$$

Similarly, it also follows that

$$\|\tilde\sigma(t,x)\| \le C\int (1+|x|)\,\mu_t(dy) = C\,(1+|x|).$$

Further, we estimate, by virtue of the moment estimate (17),

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

The uniform non-degeneracy of σ – and, hence, also of $\sigma\sigma^*$ – follows from the inequality (6) by integration with respect to μ_t.

These properties suffice for the local strong uniqueness of solution of the equation (2) by virtue of the results from [32]. However, because weak solution is well-defined for all values of time, strong uniqueness is global. According to the Yamada–Watanabe principle ([37]), any solution of the equation (2) is strong. So, any solution of the original equation (1) is also strong.

(b) In the case (2), Lipschitz conditions on both diffusion and drift are checked similarly. Now, under the set of conditions 2 of the Proposition, the equation (2) has a strong solution X_t due to Itô's Theorem. Hence, X_t is also a strong solution of the equation (1). This completes the proof of the Proposition 2.

Remark 4. *Notice that as a solution of the "linearized" equation (55), 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 Theorem 3 below.*

Remark 5. *In the case of $d = 1$, Lipschitz condition may be relaxed to Hölder of order $1/2$ and, actually, a little bit further by using techniques from [37] and [31].*

3.2 Weak uniqueness

In this section weak uniqueness will be shown for the equation (1) – (2) under appropriate conditions. This result requires only a Borel measurability of the drift with respect to the state variable x, although, it assumes that diffusion σ does not depend on y along with some additional continuity condition in x and non-degeneracy. The drift may be unbounded in the state variable x.

Theorem 2. *Let $\mathbb{E}\exp(r|x_0|^2) < \infty$ for some $r > 0$, and let the functions b and σ be Borel measurable, and*

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

that is, σ does not depend on the variable y; let σ satisfy the non-degeneracy assumption (6); let $\mathrm{d}_1 = d$, the matrix σ be bounded, symmetric and invertible, and let there exist $C > 0$ such that the function

$$\tilde{B}[s, x, \mu] := \sigma^{-1}(s, x)\, B[s, x, \mu]$$

satisfies the linear growth condition: there is $C > 0$ such that for all $x \in \mathbb{R}^d$,

$$\sup_{s,\mu} |\tilde{B}[s, x, \mu]| \le C(1 + |x)). \tag{56}$$

Also assume that the matrix-function $\sigma(t, x)$ satisfies the uniform continuity condition in x which guarantees that the equation

$$dX_t^0 = \sigma(t, X_t^0)\, dW_t, \quad X_0^0 = \xi, \tag{57}$$

with an $\mathcal{F}_0$-measurable initial data ξ possessing a given distribution μ_0, has a unique weak solution (see [32, 33]). Then (under all the assumptions of the Theorem 1) solution of the equation (1)–(2) is weakly unique.

Remark 6. *In case of $d = 1$ continuity of sigma in x is not needed. Under the additional assumption of boundedness of $\tilde{b}$ exponential moment of the initial value x_0 is not necessary and can be replaced by the fourth moment, or even weaker.*

3.3 Proof of Theorem 2

Denote by X_t^0 any (weakly unique) weak solution of the Itô equation (57). Note that

$$dW_t = \sigma^{-1}(t, X_t^0) dX_t^0.$$

1. Recall that under the assumptions of the Theorem, any solution of the equation (1)–(2) is strong by virtue of the Proposition 2. Hence, it suffices to show weak uniqueness, after which strong uniqueness for this equation will follow from strong uniqueness for the "linearized" equation (55). We will show this weak uniqueness by contradiction. Suppose there are two solutions (X^1, W^1) and (X^2, W^2) of the equation (1) with distributions μ^1 and μ^2 respectively in the space of trajectories $C[0, \infty; \mathbb{R}^d]$:

$$dX_t^1 = \sigma(t, X_t^1)\, dW_t^1 + B[t, X_t^1, \mu_t^1]\, dt, \quad X_0^1 = \xi^1, \tag{58}$$

and

$$dX_t^2 = \sigma(t, X_t^2)\, dW_t^2 + B[t, X_t^2, \mu_t^2]\, dt, \quad X_0^2 = \xi^2, \tag{59}$$

respectively, with $\mathcal{L}(\xi^1) = \mathcal{L}(\xi^2)$. Yet, under the present setting it will be shown that firstly $\mu^1 = \mu^2$ and secondly $X^1 = X^2$ a.s. Note that both X^1 and X^2 are Markov processes ([20]).

Both solutions (X^i, μ^i) *in the weak sense* may be obtained from the Wiener process W and solution X^0 of the equation without the drift (57) via Girsanov's transformations using the following stochastic exponents:

$$\gamma_T^i = \exp(\int_0^T \tilde{B}[s, X_s^0, \mu_s^i]\, dW_s - \frac{1}{2}\int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2\, ds), \quad i = 1, 2,$$

where $\tilde{b}(t,x,y) := \sigma^{-1}(t,x)\, b(t,x,y)$, $\tilde{B}[t,x,\mu] := \sigma^{-1}(t,x)\, B[t,x,\mu]$, $|\tilde{B}|$ stands for the modulus of the vector $\tilde{B}$, and $\tilde{B}[s, X_s^0, \mu_s^i]\, dW_s$ is understood as a scalar product, $\sum_{j=1}^d \tilde{B}^j[s, X_s^0, \mu_s^i] d\tilde{W}_s^j$.

It is well-known that in the case of bounded $\tilde{B}$ the random variables γ_T^i, $i = 1, 2$, are probability densities due to Girsanov's theorem (see, e.g., [22, Theorem 6.8.8]). So, till the step 4 we assume $\tilde{B}$ bounded; note that in this case we have,

$$|B[s,x,\mu] - B[s,x,\nu]| \le C\|\mu - \nu\|_{TV}. \tag{60}$$

The calculus with a bounded B is needed so as to explain the idea which will be further expanded to the case without this restriction. Also this will justify the statement in the Remark 6.

Denote

$$\tilde{W}_t^1 := W_t - \int_0^t \tilde{B}[s, X_s^0, \mu_s^1]\, ds, \quad 0 \le t \le T.$$

This is a new Wiener process on $[0,T]$ under the probability measure P^{γ^1} defined by its density as $(dP^{\gamma^1}/dP)(\omega) = \gamma_T^1$. Then, on the same interval $[0,T]$, on the probability space with a Wiener process $(\Omega, \mathcal{F}, (\tilde{W}_t^1, F_t), \mathbb{P}^{\gamma^1})$, the process $(X_t^0,\, 0 \le t \le T)$ satisfies the equation,

$$\begin{aligned} dX_t^0 &= \sigma(t, X_t^0) d\tilde{W}_t^1 + \sigma(t, X_t^0)\tilde{B}[t, X_t^0, \mu_t^1] dt \\ &= \sigma(t, X_t^0) d\tilde{W}_t^1 + B[t, X_t^0, \mu_t^1]\, dt, \end{aligned} \tag{61}$$

with the initial condition $X_0^0 = x_0$. In other words, the process X^0 on $[0,T]$ satisfies the equation (58), just with another Wiener process and under another probability measure. However, given μ_t^1, $0 \le t \le T$, this solution considered as a solution of Itô's

– or "linearized" – equation is weakly unique. This is a well-known fact for bounded coefficients due to the results on uniqueness for solutions of parabolic equations, see [27]. For unbounded coefficients under the linear growth conditions this follows by truncation and via stopping times in a standard way. Further, this uniqueness for X^0 implies weak uniqueness for the pair (X^0, W), see [2] et al. So, the pair $(X^0_t, \tilde W^1_t,\, 0\le t\le T)$ has the same distribution under the measure $\mathbb{P}^{\gamma^1}$ as the pair $(X^1_t, W_t,\, 0\le t\le T)$ under the measure $\mathbb{P}$. Therefore, the marginal distribution of X^0_t under the measure $\mathbb{P}^{\gamma^1}$ equals μ^1_t, i.e., the couple (X^0_t, μ^1_t) under $\mathbb{P}^{\gamma^1}$ solves the McKean–Vlasov equation (1), that is, it is equivalent to the pair $(X^1_t, \mu^1_t,\, 0\le t\le T)$ under the measure $\mathbb{P}$.

Note for the sequel that $d\tilde W^1_t$ admits a representation

$$d\tilde W^1_t = \sigma^{-1}(t,X^0_t)\, dX^0_t - \sigma^{-1}(t,X^0_t) B[t,X^0_t,\mu^1_t]\, dt = \sigma^{-1}(t,X^0_t) dX^0_t - \tilde B[t,X^0_t,\mu^1_t]\, dt,$$

or, equivalently,

$$\sigma^{-1}(t,X^0_t) dX^0_t = d\tilde W^1_t + \tilde B[t,X^0_t,\mu^1_t]\, dt.$$

Similarly, let

$$\tilde W^2_t := W_t - \int_0^t \tilde B[s,X^0_s,\mu^2_s]\, ds, \quad 0\le t\le T.$$

This is a new Wiener process on $[0,T]$ under the probability measure P^{γ^2} defined by its density as $(dP^{\gamma^2}/dP)(\omega) = \gamma^2$. Then, on the interval $[0,T]$, on the probability space with a Wiener process $(\Omega, \mathcal{F}, (\tilde W^2_t, F_t), \mathbb{P}^{\gamma^2})$, the process $(X^0_t,\, 0\le t\le T)$ satisfies the equation,

$$dX^0_t = \sigma(t,X^0_t) d\tilde W^2_t + B[t,X^0_t,\mu^2_t]\, dt,$$

with the initial condition $X^0_0 = x_0$. In other words, the process X^0 on $[0,T]$ satisfies the equation (59), just with another Wiener process and under another measure. However, given μ^2_t, $0\le t\le T$, this solution considered as a solution of Itô's equation is weakly unique. Therefore, the couple (X^0_t, μ^2_t) under the probability measure $\mathbb{P}^{\gamma^2}$ solves the McKean–Vlasov equation (1), that is, it is equivalent to the pair $(X^2_t, \mu^2_t,\, 0\le t\le T)$ under the measure $\mathbb{P}$.

2. This provides us a way to write down the density of the distribution of X^1 on $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to the distribution of X^2 on $(\Omega, \mathcal{F}, \mathbb{P})$ on the interval of time $[0,T]$. We have, for any measurable $A\subset C[0,T;\mathbb{R}^d]$,

$$\mu^1_{0,T}(A) := \mathbb{P}(X^1\in A) = \mathbb{P}^{\gamma^1}(X^0\in A) = \mathbb{E}^{\gamma^1} 1(X^0\in A) = \mathbb{E}\gamma^1_T 1(X^0\in A), \quad (62)$$

and

$$\mu^2_{0,T}(A) := \mathbb{P}(X^2 \in A) = \mathbb{P}^{\gamma^2}(X^0 \in A) = \mathbb{E}^{\gamma^2} 1(X^0 \in A) = \mathbb{E}\gamma^2_T 1(X^0 \in A). \quad (63)$$

So, on the sigma-algebra $\mathcal{F}^W_T$ we obtain,

$$\frac{\mu^2_{[0,T]}(dX)}{\mu^1_{[0,T]}(dX)}(X^0) = \frac{\gamma^2_T}{\gamma^1_T}(X^0) = \exp(\int_0^T \tilde{B}[s, X^0_s, \mu^2_s]dW_s - \frac{1}{2}\int_0^T |\tilde{B}[s, X^0_s, \mu^2_s]|^2 ds)$$

$$\times \exp(-\int_0^T \tilde{B}[s, X^0_s, \mu^1_s]dW_s + \frac{1}{2}\int_0^T |\tilde{B}[s, X^0_s, \mu^1_s]|^2 ds)$$

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

$$= \exp(\int_0^T (\tilde{B}[s, X^0_s, \mu^2_s] - \tilde{B}[s, X^0_s, \mu^1_s])\sigma^{-1}(s, X^0_s)dX^0_s)$$

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

$$= \exp(\int_0^T (\tilde{B}[s, X^0_s, \mu^2_s] - \tilde{B}[s, X^0_s, \mu^1_s])(d\tilde{W}^1_s + \tilde{B}[s, X^0_s, \mu^1_s]\, ds$$

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

$$= \exp(\int_0^T (\tilde{B}[s, X^0_s, \mu^2_s] - \tilde{B}[s, X^0_s, \mu^1_s])\, d\tilde{W}^1_s - \frac{1}{2}\int_0^T |\tilde{B}[s, X^0_s, \mu^2_s] - \tilde{B}[s, X^0_s, \mu^1_s]|^2 ds).$$

Further, due to (62) and (63) the measure μ^i is an image of $\mathbb{P}^{\gamma^i}$ under the mapping X^0 for $i = 1, 2$. So,

$$v(t) := \|\mu^1_{[0,t]} - \mu^2_{[0,t]}\|_{TV} \le \|P^{\gamma^1}|_{\mathcal{F}^W_t} - P^{\gamma^2}|_{\mathcal{F}^W_t}\|_{TV}. \quad (64)$$

Since the two measures P^{γ^1} and P^{γ^1} on $\mathcal{F}^W_t$ are equivalent with the density

$$\frac{dP^{\gamma^2}}{dP^{\gamma^1}}\Big|_{\mathcal{F}^W_t}(\omega) = \frac{\gamma^2_t}{\gamma^1_t}(\omega),$$

the total variation distance between them equals (denoting $\rho_t = \gamma^2_t/\gamma^1_t$),

$$\frac{1}{2}\|P^{\gamma^2}|_{\mathcal{F}^W_t} - P^{\gamma^1}|_{\mathcal{F}^W_t}\|_{TV} = \int_\Omega \left(1 - \frac{\gamma^2_t}{\gamma^1_t}(\omega) \wedge 1\right) \mathbb{P}^{\gamma^1}(d\omega) = 1 - \mathbb{E}^{\gamma^1}\rho_t \wedge 1 \le \sqrt{E^{\gamma^1}\rho^2_t - 1}.$$

Let us justify the last inequality for completeness, dropping the sub-index t:

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

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

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

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

as required. We used the Cauchy–Bunyakovsky–Schwarz inequality. So, due to (64),

$$v(t)\le 2\sqrt{\mathbb{E}^{\gamma^1}\rho_t^2-1}. \tag{65}$$

Now, again by virtue of the Cauchy–Bunyakovsky–Schwarz inequality,

$$\mathbb{E}^{\gamma^1}\rho_T^2=\mathbb{E}^{\gamma^1}\exp(-2\int_0^T(\tilde B[s,X_s^0,\mu_s^2]-\tilde B[s,X_s^0,\mu_s^1])d\tilde W_s^1$$

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

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

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

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

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

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

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

$$\leq (=) \sqrt{\mathbb{E}^{\gamma^1} \exp\left(6\int_0^T |\tilde{B}[s, X_s^0, \mu_s^2] - \tilde{B}[s, X_s^0, \mu_s^1]|^2 ds\right)}. \tag{66}$$

(NB: The last *inequality* is always true; for a *bounded* $\tilde{B}$ it is, apparently, an equality.)

3. We estimate, $\tilde{B}$ being *bounded*,

$$\begin{aligned} &\mathbb{E}^{\gamma^1} \exp\left(6\int_0^T |\tilde{B}[s, X_s^0, \mu_s^2] - \tilde{B}[s, X_s^0, \mu_s^1]|^2 ds\right) \\ &\leq \mathbb{E}^{\gamma^1} \exp\left(6\|\tilde{B}\|_B^2 \int_0^T \|\mu_s^1 - \mu_s^2\|_{TV}^2 \, ds\right). \end{aligned} \tag{67}$$

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

$$v(T) \leq 2\sqrt{\exp\left(C\int_0^T v(s)^2 ds\right) - 1}. \tag{68}$$

Recall that $v(t) \leq 2$, and the function v increases in t. Let us choose $\alpha_0 > 0$ small so that for any $0 \leq \alpha \leq \alpha_0$,

$$\exp(\alpha) - 1 \leq 2\alpha, \tag{69}$$

and take $T \leq \alpha_0/(4C)$. Then $C\displaystyle\int_0^T v(s)^2 \, ds \leq CTv(T)^2 \leq 4CT \leq \alpha_0$. So,

$$v(T) \leq 2\sqrt{\exp\left(C\int_0^T v(s)^2 \, ds\right) - 1} \leq 2\sqrt{2CTv(T)^2} = 2\sqrt{2CT}v(T). \tag{70}$$

If we choose T so small that $2\sqrt{2CT} < 1$, that is, $T < 1/(8C)$, then it follows that $v(T) = 0$. Hence, $v(T) = 0$ for any $T < \min(1/(8C), \alpha_0/(4C))$. Let us fix some $T > 0$ satisfying this inequality.

Further, we conclude by induction that

$$v(2T) = v(3T) = \ldots = 0. \tag{71}$$

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

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

and re-denote

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

Then $\tilde{W}_t^1$ 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 $(dP^{\gamma^1}/dP)(\omega) = \gamma^1_{kT,(k+1)T}$. Repeating the calculus leading to (66), (67), and (68), and having in mind the induction assumption $v(kT) = 0$, we obtain with the same constant C,

$$v((k+1)T) \le \sqrt{\exp\left(C\int_{kT}^{(k+1)T} v(s)^2 ds\right) - 1}, \tag{72}$$

which straightforward implies

$$v((k+1)T) \le \sqrt{2CTv((k+1)T)^2} = \sqrt{2CT}v((k+1)T). \tag{73}$$

As earlier, the condition $T < \min(1/(2C), 1/(\alpha C))$ (see (69)) guarantees that

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

as required. This completes the induction (71).

Hence, solution is weakly unique on the whole $\mathbb{R}_+$. As noticed above, strong uniqueness also follows. For bounded $\tilde{b}$ the statements of the Theorem 3 as well as of the Remark 6 are justified.

4. Now let us return to the inequality (66) and explain how to drop the additional assumption of boundedness of $\tilde{B}$, and also how to deal with a localised version of (60). First of all, prior to (66) we have to show that γ^i, $i = 1, 2$, are, indeed, probability densities for which it suffices to show uniform integrability for $T > 0$ small enough: for example, it suffices to check that

$$\mathbb{E}(\gamma^i)_T^2 < \infty, \quad i = 1, 2.$$

Via the estimates similar to (66) by virtue of Cauchy–Bunyakovsky–Schwarz, this problem is reduced to the question whether or not the following expression is finite:

$$\mathbb{E}(\gamma^i)^2 \le \left(\mathbb{E}\exp(4\int_0^T \tilde{B}[s, X_s^0, \mu_s^i]\, dW_s - 8\int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2\, ds)\right)^{1/2}$$

$$\times \left(\mathbb{E} \exp(6 \int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2 \, ds) \right)^{1/2}$$
$$\leq \left(\mathbb{E} \exp(6 \int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2 \, ds) \right)^{1/2}$$
$$\leq \left(\mathbb{E} \exp(C \int_0^T (1 + |X_s^0|^2) \, ds) \right)^{1/2}. \tag{74}$$

In the last inequality the assumption on the linear growth of $\tilde{B}$ was used.

Suppose for instant that the finiteness of the last expectation in the last line of (74) has been shown; then, by standard induction arguments with conditional expectations it follows that both γ_T^i are, indeed, probability densities for *any* $T > 0$. Hence, the calculus leading to (65) and (66) is valid and we have,

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

and

$$\mathbb{E}^{\gamma^1} \rho^2 \leq \sqrt{\mathbb{E}^{\gamma^1} \exp \left(6 \int_0^T |\tilde{B}[s, X_s^0, \mu_s^2] - \tilde{B}[s, X_s^0, \mu_s^1]|^2 ds \right)}.$$

It is a general fact which does not use any boundedness of b in any variable but only in the last variable is,

$$|\tilde{B}[s, X_s^0, \mu_s^2] - \tilde{B}[s, X_s^0, \mu_s^1]| \leq \sup_y |\tilde{b}(s, X_s^0, y)| \|\mu_s^2 - \mu_s^1\|_{TV}. \tag{75}$$

Due to the linear growth assumption (56), the inequality (75) implies

$$|\tilde{B}[s, X_s^0, \mu_s^2] - \tilde{B}[s, X_s^0, \mu_s^1]| \leq C(1 + |X_s^0|) \|\mu_s^1 - \mu_s^2\|_{TV}. \tag{76}$$

Hence, by virtue of (75) we obtain

$$\mathbb{E}^{\gamma^1} \rho^2 \leq \mathbb{E}^{\gamma^1} \exp \left(6 \int_0^T [C(1 + |X_s^0|) \|\mu_s^1 - \mu_s^2\|_{TV}]^2 ds \right)$$
$$\leq \mathbb{E}^{\gamma^1} \exp \left(6C^2 v(T)^2 \int_0^T (1 + |X_s^0|^2 \, ds) \right). \tag{77}$$

Recall that the process X^0 satisfies the equation (61) on $[0, T]$ with respect to the measure $\mathbb{P}^{\gamma^1}$. We want to show that given C, the right hand side in (77) is finite for any T small enough. For this end, denote $6C^2 v(T)^2 := r \geq 0$. We would like to show that for any fixed constant $K > 0$ ($K = 24C^2$ suffices), the value

$$\mathbb{E}^{\gamma^1} \exp\left(r \int_0^T (1 + |X_s^0|^2 \, ds\right)$$

is finite for $0 \leq r < K$, and differentiable with respect to r, and that this derivative is non-negative and small uniformly in $r \in [0, K)$ if $T > 0$ is small enough.

It suffices to show the same properties – still for small enough T – for the function

$$\psi(r) = \mathbb{E} \exp\left(r \int_0^T (1 + |X_s^1|^2) \, ds\right), \tag{78}$$

where X^1 solves the equation (61) on $[0, T]$ with respect to the original measure $\mathbb{P}$, because X^1 solves the same equation with respect to the measure $\mathbb{P}$ as the process X^0 with respect to the measure $\mathbb{P}^{\gamma^1}$ on $[0, T]$.

First of all, note that this claim is true for the function (see, for example, [1])

$$\tilde{\psi}(r) = \mathbb{E} \exp\left(r \int_0^T (1 + |W_s|^2) \, ds\right).$$

Denote $\bar{X}_t := \sup_{0 \leq s \leq t} |X_s|$. From the equation () we have,

$$|X_t| \leq C(1 + |X_0| + \int_0^t (1 + |X_s|) \, ds + \int_0^t \sigma_s(?) \, dW_s|).$$

From here by virtue of Gronwall's inequality – since both sides in this inequality are finite - we obtain with some $C > 0$,

$$\bar{X}_t \leq C \exp(Ct)(1 + |X_0| + \sup_{0 \leq s \leq t} |\int_0^s \sigma_s(?) \, dW_s|).$$

Therefore, for each $r > 0$

$$\exp(r \sup_{0 \leq s \leq t} |X_s^1|^2) \leq C_t \exp(r(1 + |X_0|^2)) \exp(r \sup_{0 \leq s \leq t} |\int_0^s \sigma_r(...) dW_r|^2),$$

with some $C_t < \infty$. Note that here the function σ iz bounded. Due to the exponential martingale inequalities (see, e.g., [25]) we have for any $a > 0$,

$$\mathbb{P}(\sup_{0 \leq s \leq t} |\int_0^s \sigma_r(...) dW_r| > a) \leq C \exp(-a^2/(Ct))$$

It follows that for $r > 0$ small enough

$$E\exp(r\sup_{0\le s\le t}|\int_0^s \sigma_r(...)dW_r|^2) < \infty,$$

and, hence, also

$$E\exp(r\sup_{0\le s\le t}|X_s^1|^2) < \infty, \tag{79}$$

as required.

Thus, the function ψ (see (78)) is finite for r from some finite range $0 \le r < K$. Hence, it is easy to see that it is differentiable in r with a bounded derivative within this range. In particular, since $\psi(0) = 1$, for $r > 0$ close to zero we obtain,

$$\psi(r) \le 1 + Cr(1 + \mathbb{E}|X_0|^2).$$

Also, it follows that all expressions in (74) for small enough $T > 0$ are finite. So, in particular, both γ_T^i are, indeed, probability densities for small $T > 0$ under the linear growth condition (56), too. Hence, we can return to the inequalities (65) earlier established for bounded $\tilde{b}$, and by virtue of (77) we get,

$$v(T) \le \sqrt{E^{\gamma^1}\rho_T^2 - 1} \le \sqrt{CTv(T)^2},$$

with some constant C which constant may depend on the initial distribution (or value). Therefore, $v(T) = 0$ for $T > 0$ small enough.

7. Denote

$$\mathcal{N} := \{t \ge 0 : v(t) = 0\}.$$

The previous steps show that $\sup(\mathcal{N}) > 0$ and that $0 \in \mathcal{N}$. Note that $t \in \mathcal{N} \Longrightarrow s \in \mathcal{N}$, $0 \le s \le t$. Recall that $v(t) \le \sqrt{\mathbb{E}^{\gamma^1}\rho_t^2 - 1}$ (see (65)) where the right hand side is clearly continuous in t. Moreover, as it follows from (77),

$$v(t)^2 \le \mathbb{E}^{\gamma^1}\rho_t^2 - 1 \le \mathbb{E}^{\gamma^1}\exp\left(6\int_0^t [C(1+|X_s^0|)\|\mu_s^1 - \mu_s^2\|_{TV}]^2 ds\right) - 1,$$

which implies that the set $\mathcal{N}$ is closed.

On the other hand, consider any $N \in (0, \sup(\mathcal{N}))$. Recall that $\mathbb{E}\exp(c\sup_{s\le N}|X_s^0|^2) < \infty$ with some positive c. Hence, the same calculus as above shows that $v(t) = 0$ in some small right neighbourhood of N. In other words, the set on the positive half-line $\mathbb{R}_+$ where $v(t) = 0$ is non-empty, closed and open in $\mathbb{R}_+$. Thus, it coincides with $\mathbb{R}_+$ itself. In other words, for all $t \ge 0$,

$$v(t) = 0,$$

which finishes the proof of the Theorem 2.

3.4 Strong uniqueness

In this section it will be shown that in certain cases weak uniqueness implies strong uniqueness for the equation (1) – (2), and both properties will be established under appropriate conditions. This result – the Theorem 3 below – requires only a Borel measurability of the drift with respect to the state variable x, although, it assumes that diffusion σ does not depend on y along with Lipschitz condition in x and non-degeneracy. The drift may be unbounded in the state variable x.

Theorem 3. *Let $\mathbb{E}\exp(r|x_0|^2) < \infty$ for some $r > 0$, and let the functions b and σ be Borel measurable, and*

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

that is, σ does not depend on the variable y; let σ satisfy the non-degeneracy assumption (6); let $d_1 = d$, the matrix σ be bounded, symmetric and invertible, and let there exist $C > 0$ such that the function

$$\tilde{B}[s, x, \mu] := \sigma^{-1}(s, x)\, B[s, x, \mu]$$

satisfies the linear growth condition: there is $C > 0$ such that for all $x \in R^d$,

$$\sup_{s,\mu} |\tilde{B}[s, x, \mu]| \le C(1 + |x|). \tag{80}$$

Also assume that the matrix-function $\sigma(t, x)$ satisfies the following Lipschitz condition (for simplicity) which guarantees that the equation

$$dX^0_t = \sigma(t, X^0_t)\, dW_t, \quad X^0_0 = x_0, \tag{81}$$

has a unique strong solution for any x (see [32, 33]):

$$\sup_{t\ge 0} \sup_{x,x' :\, x'\neq x} \frac{\|\sigma(t, x) - \sigma(t, x')\|}{|x - x'|} < \infty. \tag{82}$$

Then solution of the equation (1)–(2) is weakly and strongly unique; this solution is strong.

Proof follows straightforwardly from the Theorem 2 and from the fact that with a given μ_t any solution is strong [32, 33] (note that linear growth of the drift is allowed in both [32, 33]).

Remark 7. *Note that under the condition (82), not only the equation (57) but any equation with the same diffusion and a Borel measurable drift with a linear growth* *assumption in x will have a strong solution. It concerns both solutions of the equation (1) and its "linearized" version (55).*

Emphasize that no regularity on the function b is needed in either variable. Also, a linear growth condition on the drift in x is equivalent to the condition (56); the latter was assumed in order to make the presentation more explicit. The price for the no regularity and linear growth is a special form of σ which may not depend on the "measure variable" y; in particular, in such a case $\Sigma(t,x) = \sigma(t,x)$, and we will use the lower case to denote the diffusion coefficient in the remaining sections.

Remark 8. *Instead of Lipschitz condition (82), it suffices if diffusion coefficient σ belongs to the Sobolev class $\sigma(t,x) \in W^{0,1}_{2d+2,loc}$. More general conditions on Sobolev derivatives for σ can be found in [32, Theorem 1] and [33], and any of them can be used in our Theorem 3 above. Note that in the latter reference σ is assumed Lipschitz but it is shown that continuity is necessary only with respect to the state variable x, which is also applied to the conditions from [32]. As usual, a more relaxed conditions on σ can be stated in the case of dimension one as in [37], or in [32, Theorem 2], the simplest version of both being just Hölder 1/2.*

4 Appendix

Lemma 3 (Skorokhod (on unique probability space and convergence))**.** *Let ξ^n_t ($t \geq 0$, $n = 0, 1, \ldots$) be some d-dimensional stochastic processes defined on some probability space and let for any $T > 0$, $\epsilon > 0$ the following hold true:*

$$\lim_{c\to\infty} \sup_n \sup_{t\le T} \mathbb{P}(|\xi^n_t| > c) = 0,$$

and

$$\lim_{h\downarrow\infty} \sup_n \sup_{t,s\le T;\,|t-s|\le h} \mathbb{P}(|\xi^n_t - \xi^n_s| > \epsilon) = 0,$$

Then there exists a subsequence $n' \to \infty$ and a new probability can be constructed with processes $\tilde\xi^{n'}_t$, $t \geq 0$ and $\tilde\xi_t$, $t \geq 0$, such that all finite-dimensional distributions of $\tilde\xi^{n'}_\cdot$ coincide with those of $\xi^{n'}_\cdot$ and such that for any $\epsilon > 0$ and $t \geq 0$,

$$\mathbb{P}(|\tilde\xi^{n'}_t - \tilde\xi_t| > \epsilon) \to 0, \quad n' \to \infty.$$

See [29, Ch.1, §6].

Lemma 4 (Skorokhod)**.** *Let* $f^n : \mathbb{R} \times \Omega \to \mathbb{R}\, (n \geq 0)$ *be uniformly bounded random processes on some probability space; let* $(W^n\, (n \geq 0))$ *be a sequence of (one-dimensional) Wiener processes on the same probability space, and let all Itô's stochastic integrals* $\displaystyle\int_0^T f_s^n\, dW_s^n, n \geq 0$ *be well-defined. Assume that for any* $\varepsilon > 0$*,*

$$\lim_{h \to 0} \sup_n \sup_{|s-t| \leq h} \mathbb{P}\{|f_s^n - f_t^n| > \varepsilon\} = 0, \tag{83}$$

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

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

See [29, Ch.2, §3, Theorem], where W^n are allowed to be more general martingales with brackets converging to that of a Wiener process.

In the proof of the Propositions 1 and 2 (and Thm 1?) the following lemma was needed. We use verbatim notations and terminology from [10] to save some space.

Definition 2. *A system of subsets* $\mathcal{F}$ *of some set* E *is called* π*-system iff* $A, B \in \mathcal{F}$ *implies* $AB \in \mathcal{F}$*. Let* $\mathcal{L}$ *be a set of functions on* E *which contains* f_+ *and* f_- *along with any* f*. A system of functions* $\mathcal{H}$ *is called* $\mathcal{L}$*-system iff the three conditions hold:*

(a) $1 \in \mathcal{H}$*;*

(b) if $f_1, f_2 \in \mathcal{H}$ *and* $c_1, c_2 \in \mathbb{R}$*, then* $c_1 f_1 + c_2 f_2 \in \mathcal{H}$*;*

(c) if $f_n \in \mathcal{H}$*,* $0 \leq f_n \uparrow f$*, and either* f *is bounded, or* $f \in \mathcal{L}$*, then* $f_n \in \mathcal{H}$*.*

Lemma 5 (Dynkin)**.** *If some* $\mathcal{L}$*-system* $\mathcal{H}$ *contains indicators of all subsets from a* π*-system* $\mathcal{C}$*, then* $\mathcal{H}$ *contains all* $\sigma(\mathcal{C})$*-measurable functions from* $\mathcal{L}$*.*

See [10, Lemmae 0.2 - 0.3]. In our case, $\mathcal{C}$ consists of all open sets in $[0, T] \times B_R \times B_R$, $\sigma(\mathcal{C})$ is a Borel sigma-algebra in $[0, T] \times B_R \times B_R$ (which is a π-system), $\mathcal{L}$ is a family of all Borel measurable functions satisfying Krylov's bound (86) for the couple $(\tilde{X}, \tilde{\xi})$, and $\mathcal{H}$ is a family of all continuous functions satisfying the same Krylov's bound and monotonically increasing limits of such continuous functions, for which limits this Krylov's bound with a finite or infinite right hand side also holds true (by the monotonic convergence theorem). The point is that the indicator of any compact may be monotonically approximated from above by bounded continuous functions with compact supports. As a consequence, the closure of the linear hull of continuous functions on a compact domain coincides with the closure of the linear hull of indicators of all open subsets. As a result of the Lemma 5, $\mathcal{H}$ contains all functions from $\sigma(\mathcal{C})$, that is, all Borel measurable functions. So, Krylov's bound (86) holds for them all (possibly with an infinite right hand side which does not contradict to the bound).

As promised, the last lemma is a version of Krylov's bounds. Let Z_t be a strong Markov process in $\mathbb{R}^d$ satisfying an SDE

$$dZ_t = b_t(Z_t)dt + \sigma_t(Z_t)dW_t, \quad Z_0 = z_0,$$

where (non-random) functions $b_t(z)$ and $\sigma_t(z)$ are d-vector and matrix $d \times d$ respectively, σ_t is uniformly non-degenerate, and locally in z bounded uniformly in (t, ω), that is,

$$d(R) := \sup_{|z|\le R} \sup_t (|b_t(z)| + \|\sigma_t(z)\|) < \infty, \quad \forall \ R > 0, \tag{84}$$

and the random variable z_0 is independent of the filtration of the Wiener process $W = (W_t, \mathcal{F}_t)$. Let D be a bounded domain in $B_R = (z \in \mathbb{R}^d : \ |z| \le R)$, $g : \mathbb{R}^d \to \mathbb{R}^1$, and $f : \mathbb{R}^1 \times \mathbb{R}^d \to \mathbb{R}^1$ be Borel measurable functions, and

$$\tau_D := \inf(t \ge 0 : Z_t \not\in D).$$

Then the following versions of Krylov's bounds hold true: we assume only local boundedness of the coefficients, and the statements are also local,

Lemma 6. *Under assumption (84), for any $p \ge d$ there exists a constant $N = N_R$ which also depends on d, on the constants of ellipticity of $\sigma\sigma^*$ and on the upper bounds for the norms of b and σ for $|x|, |y| \le R$, such that for any g satisfying $g(z) \equiv 0$, $z \not\in D$,*

$$\mathbb{E} \int_0^T g(Z_t)dt \le N_R \|g\|_{L_p(D)}, \tag{85}$$

where $R = diam(D)$, and for any f satisfying $f(t, z) \equiv 0$, $z \not\in D$,

$$\mathbb{E} \int_0^T f(t, Z_t)dt \le N_R \|f\|_{L_{p+1}([0,T]\times D)}. \tag{86}$$

Recall that in the proof of the Theorem 1 the role of Z is played by the pairs $(\tilde{X}^n, \tilde{\xi}^n)$. These are yet not full Krylov's estimates for general Itô processes, however, they suffice for our goals here.

Proof is based on Krylov's bounds for Itô processes with bounded coefficients and from the hint similar to the one in the proof of the Lemma 4.3.1 [18]. We only show the second inequality, since the first one follows similarly. Let D' be another bounded domain containing the closure of D: $\bar D \subset\subset D'$. Without loss of generality, we may assume that $\mathrm{diam}(D') \le \mathrm{diam}(D') + 1$, and $D' \subset B_{R+1}$. Denote

$$\tau^0 = 0, \quad T^1 := \inf(t \ge \tau_0 : Z_t \not\in \bar D'),$$
$$\tau^k := \inf(t \ge T^k : Z_t \in D),\ T^{k+1} := \inf(t \ge \tau^k : Z_t \not\in \bar D'),\ k \ge 1,$$

and let

$$\hat Z^0_t := Z_0 + \int_0^t 1(s<T^1) b_s(Z_s)ds + \int_0^t (1(s<T^1)\sigma_s(Z_s) + 1(s\ge T^1))dW_s,\ t\ge 0, \quad \ldots,$$

$$\hat Z^k_t := Z_{\tau^k} + \int_{\tau^k}^t 1(s<T^{k+1}) b_s ds + \int_{\tau^k}^t (1(s<T^{k+1})\sigma_t + 1(s\ge T^{k+1}))dW_s,\ t\ge \tau^k, \quad \ldots$$

Note that $Z_t = \hat Z^0_t$ on the set $(t<T^1)$, and $\hat Z^k_t = Z_t$ on $(\tau^k \le t < T^{k+1})$ (see [22, Theorem 6.3.7(iii)]. Since $f(s,Z_s) = 0$ on any interval $T^k \le s \le \tau^k$, we have with $p \ge d$,

$$\mathbb{E}\int_0^T |f(s,Z_s)|\,ds = \sum_{k=0}^\infty \mathbb{E}\int_{\tau^k\wedge T}^{T^{k+1}\wedge T} |f(s,Z_s)|\,ds = \sum_{k=0}^\infty \mathbb{E}1(\tau^k\le T)\int_{\tau^k\wedge T}^{T^{k+1}\wedge T}|f(s,Z_s)|\,ds$$
$$= \sum_{k=0}^\infty \mathbb{E}1(\tau^k\le T)\int_{\tau^k}^{T^{k+1}\wedge T}|f(s,Z_s)|\,ds \le \sum_{k=0}^\infty \mathbb{E}1(\tau^k\le T)\int_{\tau^k}^{T^{k+1}}|f(s,Z_s)|\,ds$$
$$= \sum_{k=0}^\infty \mathbb{E}1(\tau^k\le T)\mathbb{E}\left(\int_{\tau^k}^{T^{k+1}}|f(s,\hat Z^k_s)|\,ds|\mathcal{F}_{\tau^k}\right) \le N_R\|f\|_{L_{p+1}([0,T]\times R^d)}\sum_{k=0}^\infty P(\tau^k\le T),$$

according to Krylov's bound [21, Theorems 2.2.2, 2.2.4]. Recall that $P(\tau^k \le T) \le P(T^k \le T)$. For a strong Markov process Z_s with a positive probability to exit from $\bar D'$ on any finite interval of time due to the non-degeneracy of its diffusion coefficient and boundedness of both coefficients in $\bar D'$, the probabilities $P(T^k\le T)$ admit exponential bounds

$$P(T^k \le T) \le Cq^{k-1},\ k\ge 1,$$

with some $C<\infty, q<1$. So, with a new constant C and since $(1-1_D)f\equiv 0$,

$$\mathbb{E}\int_0^T |f(s,Z_s)|\,ds \le N_R\|f\|_{L_{p+1}([0,T]\times\mathbb{R}^d)}\sum_{k=0}^\infty \mathbb{P}(T^k\le T) \le CN_R\|f\|_{L_{p+1}([0,T]\times D)}. \tag{87}$$

The bound (85) follows similarly. The Lemma 6 is proved.

Acknowledgements

For the second author this study has been funded by the Russian Academic Excellence Project '5-100' and by the Russian Science Foundation project 17-11-01098. Certain stages of this work have been fulfilled while the second author was visiting Bielefeld university to which programme SFB1283 this author is grateful. Both authors are grateful to Denis Talay, Mireille Bossy and Sima Mehri who posed valuable questions and thus drew attention of the authors to various gaps in the earlier proofs mainly in the Theorem 1 and in this way stimulated us to improve our presentation. Their help was indispensable. Also, the authors are really grateful to the anonymous referee for many useful comments and remarks which helped a lot.

References

[1] N. Abourashchi, A.Yu. Veretennikov, On stochastic averaging and mixing, Theory of Stochastic Processes, 16(32) (1), 111-130, 2010.

[2] R. Bass, Diffusions and Elliptic Operators, Springer, New York, 1998.

[3] S. Benachour, B. Roynette, D. Talay, P. Vallois, Nonlinear self-stabilizing processes. I: Existence, invariant probability, propagation of chaos. *Stochastic Processes Appl.* 75(2): 173-201, 1998.

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

[5] 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.

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

[7] P.-E. Chaudru de Raynal, Strong well-posedness of McKean-Vlasov stochastic differential equation with Hölder drift, arXiv:1512.08096.

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

[9] N. Dunford, J.T. Schwartz, Linear operators, Part 1, Interscience, New York, 1958.

[10] E.B. Dynkin, Markov processes, Academic Press, New York, 1965. vv. 1 - 2.

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

[12] I.I. Gihman, A.V. Skorohod, Stochastic differential equations, Springer, Berlin, 1972.

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

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

[15] 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.

[16] 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.

[17] 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.

[18] R.Z. Khasminskii, Stochastic Stability of Differential Equations. 2nd ed., with contributions by G.N. Milstein and M.B. Nevelson, Springer, Berlin, 2012.

[19] N. V. Krylov. On Ito's stochastic integral equations. *Theory Probab. Appl.*, 14:330–336, 1969; Addendum: On Ito's stochastic integral equations, *ibid.,* 17:2, 373-374, 1973.

[20] 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.

[21] N. V. Krylov. *Controlled diffusion processes*, Springer, Berlin, 2nd edition, 2009.

[22] N. V. Krylov, Introduction to the Theory of Random Processes, AMS, Providence, R.I., 2002.

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

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

[25] H. P. McKean. Stochastic integrals. New York, NY, Acad. Press, 1969.

[26] P.-A. Meyer, Probability and Potentials, (or Dellacherie and Meyer, same title, later edition.)

[27] O.A. Ladyzhenskaya, V.A. Solonnikov, N.N. Uraltseva, Linear and Quasi-linear Equations of Parabolic Type, AMS, R.I., 1968.

[28] A. V. Skorohod. *Stochastic equations for complex systems*, volume 13 of *Mathematics and its Applications (Soviet Series)*. D. Reidel Publishing Co., Dordrecht, 1988.

[29] 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.

[30] 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.

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

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

[33] A. Yu. Veretennikov. Parabolic equations and Ito's stochastic equations with coefficients discontinuous in the time variable. Math. notes, 1982, 31, 278-283.

[34] 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.

[35] 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.

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

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

[38] 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.