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**[†]

October 15, 2018

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 nondegeneracy of diffusion and no more than a linear growth of both coefficients in the state variable. Weak and strong uniqueness is 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.

[*]Taras Shevchenko National University of Kiev; 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. Financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" is acknowledged.

1 Introduction

1.1 Setting, backgrounds and motivation

Our subject is solutions of the stochastic Itô–McKean–Vlasov, or, for short, McKean–Vlasov's equation in $\mathbb{R}^d$

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

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

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

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 [32], called mean field interaction in mathematical physics and stochastic analysis, assumes that for a large multiparticle ensemble with "weak interaction" between particles, this interaction for one particle with others may be effectively replaced by an averaged field. A class of equations of type (1) was proposed by M. Kac [15] as a stochastic "toy model" for the Vlasov kinetic equation of plasma. The systematic study of such equations was started by McKean [21]. The reference [26] 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 [2, 3]), as well as in some other areas of high interest such as filtering (see [5]). 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 [11]). In what concerns "propagation of chaos" for the equation (1), we refer the reader to [26] and [4, Theorem 4.3]. In the authors' view, it may be fruitful to separate different aspects, including time discretization and "propagation of chaos" for multiparticle case, and to consider approximations differently from the basic existence and uniqueness issues; the latter are the main subjects of the present paper.

Note that even existence and uniqueness – as usual in stochastic analysis, weak or strong – does require further studies. E.g., many control problems lead to discontinuous coefficients. So, establishing existence and uniqueness under minimal regularity of the coefficients is in a big demand.

As to earlier works in this area, one of the most important papers is [9] 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 [9] because the latter are given not directly in terms of coefficients (cf. with (2.1) in the Assumption I from [9]). We do not assume continuity with respect to the state variable x replcing it by the non-degeneracy of the diffusion matrix. Neither our linear growth bound is comparable directly with the Lyapunov type conditions in [9]. More general growth conditions were studied in [4]; however, here also our regularity conditions admit just measurable coefficients in x and, hence, overall, our results are not covered by [4] either. One of the first works on McKean–Vlasov equations with irregular coefficients was [23]; yet, this paper considers another class of equations, namely, with divergent operators, and so does not relate directly to the present paper.

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 [4]. 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 was established in [13] and [14]. In [30] there is a result on strong existence for the equation similar to (1) only with a unit matrix diffusion; however, strong and weak uniqueness, along with "propagation of chaos", i.e., with convergence of particle approximations, is established there under restrictive additional assumptions on the drift which include Lipschitz and some other conditions. In the present paper, weak and strong uniqueness is 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 $\sigma\sigma^*$ (cf., for example, [9]). Hence, results for $d_1 \geq d$ may not necessarily follow automatically from those for $d_1 = d$. Unlike the setting in the paper [4], 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 [4]: we use explicitly Skorokhod's single probability space approach as well as Krylov's integral estimates for Itô's processes.

Strong existence 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 [12], [20], [33], [34]) 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 results of the paper 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.

1.2 The structure of the 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 [16] for a homogeneous case, and from [18] for a non-homogeneous case; cf. also [31]. No regularity of the coefficients is assumed with respect to the state variable x. The proof is split into two unequal parts. The first is devoted to the case under a bit restrictive additional assumption on the diffusion; the second 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 2. For a completeness of the paper, a classical Skorokhod's lemma on convergence of stochastic integrals, as well as two other indispensable auxiliary lemmata also by Skorokhod are provided in the Appendix (the 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

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

in which expression under the probability all integrals are well-defined, and μ_t is 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 $\mathcal{F}_t^W$ (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.

Proposition 1 *Suppose in case of (2) for each (t, x) the Borel coefficients $b(t, x, y)$ and $\sigma(t, x, y)$ are bounded in y and integrable in x with respect to all (μ_t), $t \geq 0$, where μ_t are the marginal distributions of some (any) weak solution 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 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'. 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 $\mathcal{B}[0, \infty) \times \mathcal{F}$-measurable in (t, ω') due to continuity of solution ξ_t in t and its measurability in ω' (cf., e.g., [19, Lemma 1.5.7]). Hence, the function $\hat{b}(t, x, \omega') := b(t, x, \xi_t(\omega'))$ is $\mathcal{B}[0, \infty) \times \mathcal{B}(\mathbb{R}^d) \times \mathcal{F}$- measurable in (t, x, ω'). Further, one of the statements of Fubini Theorem (cf. [19, 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 $\mathcal{B}[0,\infty) \times \mathcal{B}(\mathbb{R}^d)$-measurable, as required. Here we used the condition of boundedness of b in y which implies integrability

$$\iint\limits_D |b(t,x,\xi_t(\omega'))|\mathbb{P}'(d\omega')dtdx < \infty,$$

over any *bounded* Borel subset $D \in \mathcal{B}[0,\infty) \times \mathcal{B}(\mathbb{R}^d)$ which integrability is the assumption of Fubini Theorem ([19, Theorem 1.5.5]).

Theorem 1 *Let the initial value x_0 have a finite fourth moment. For the problem (1)– (2), suppose that the following two conditions are both satisfied. Firstly, the functions b and σ admit linear growth condition in (x), i.e., there exists $C > 0$ such that for any s, x, y,*

$$|b(s,x,y)| + \|\sigma(s,x,y)\| \le C(1+|x|), \tag{3}$$

where $|\cdot|$ stands for the Euclidean norm in R^d for b and $\|\cdot\|$ for the $\|\sigma\| = \sqrt{\sum_{i,j}\sigma_{ij}^2}$. Secondly, the diffusion matrix σ is uniformly nondegenerate in the following sense: there is a value $\nu > 0$ such that for any probability measure μ,

$$\inf_{s,x,y}\inf_{|\lambda|=1} \lambda^* \left(\int \sigma(s,x,y)\mu(dy)\right)\left(\int \sigma^*(s,x,y)\mu(dy)\right)\lambda \ge \nu. \tag{4}$$

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 *Note that if $d_1 = d$ and if the matrix σ is quadratic and* symmetric, *then the assumption (4) can be replaced by an easier and more frequently in use one,*

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

Note that the intuitive meaning of condition (4) in the simplest 1D (i.e., $d_1 = d = 1$) 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 the Theorem 2 under the linear growth condition on the drift where an exponential moment will be required), but we do not pursue this goal here. Under the additional assumption of boundedness of b and σ, the fourth moment of the initial value x_0 is not necessary and can be further relaxed. On weak uniqueness *there will be a remark in the last section.*

2.2 Proof of Theorem 1.

1. Firstly we establish the Theorem under a more restrictive assumption that $d_1 = d$, and that the matrix σ is symmetric and satisfies the condition (5). Exactly this was assumed in [18] for Itô's equations where, in addition, the coefficients were assumed bounded. None of these two restrictions is actually necessary, which was, of course, very well-known to the author of [18] and which extensions were covered in other publications; yet, some efforts are required to relax them here for the McKean-Vlasov equation setting.

In the end of the present proof, the restriction (5) will be dropped. To explain the motivation of this approach note that under the relaxed assumption (4) of the Theorem, a smoothing which would keep the non-degeneracy of the diffusion coefficient is to be found.

The proof is based on Krylov's integral estimate for non-degenerate Itô processes (i.e. for those possessing a stochastic differential but not necessarily a solution of any SDE) with *bounded coefficients*,

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

see [18, Chapter 2]. Here the constant N may depend on d, T and the bounds for sup–norm of coefficients and of the inverse $\sigma\sigma^*$. This estimate will be applied to a *couple* (X_t, ξ_t),

$$\mathbb{E}\int_0^T f(t,X_t,\xi_t)dt \le N\|f\|_{L_{2d+1}}, \tag{6}$$

where the process ξ_t is an independent copy of the process X_t and, hence, has exactly the same distribution (not only the same marginals); the constant N in the latter inequality depends on the same norms as earlier but now the dimension is $2d$.

Similarly to the Itô case the standard hint is to smooth the coefficients so as to use existence theorems which are known in the literature (in particular, for continuous coefficients, see [9]), and then to pass to the limit by using Skorokhod's convergence on a single probability space method. So, let us smooth both coefficients with respect to all variables, i.e., let

$$b^n(t,x,y) = b(t,x,y) * \psi_n(t) * \phi_n(x) * \phi_n(y), \tag{7}$$

and

$$\sigma^n(t,x,y) = \sigma(t,x,y) * \psi_n(t) * \phi_n(x) * \phi_n(y), \tag{8}$$

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. While smoothing, assume for definiteness that the coefficients b and σ are defined for any $t < 0$, e.g., as zero vector function and the constant unity matrix $I_{d\times d}$, respectively. Note that, of course, we may assume that for every n the smoothed coefficients remain to be under the linear growth condition (3) with the same constant for each n (in reality this constant may increase a little bit in comparison with C from (3), but still remain uniformly bounded); also, under the assumption (5) the smoothed diffusion remains uniformly nondegenerate with ellipticity constants independent of n.

2. In a standard way (see, e.g., [18], [25]) we get the estimates,

$$\sup_{0\le t\le T} \mathbb{E}|X^n_t|^2 \le C_T(1 + \mathbb{E}|x_0|^2), \tag{9}$$

and also

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|X^n_t - X^n_s|^2 \le C_T h, \tag{10}$$

with constants C_t, C_T that do not depend on n. Recall that $x_0 \in R^d$ is the initial value of the process X and that it may be random with a certain moment. Bounds similar to (9) and (10) hold true also for the component ξ^n and naturally for W^n. These bounds suffice for the applicability of Skorokhod's single probability space theorem (see the Appendix 1, Lemma 2).

Note that by Doob's inequality for stochastic integrals the bound (9) immediately extends to the further bound useful in the sequel,

$$\mathbb{E} \sup_{0\le t\le T} |X^n_t|^2 \le C(\mathbb{E}|x_0|^2 + CT + CT^2)\exp(CT(T+1)), \tag{11}$$

where C does not depend on n. Further, similarly, the following higher moment bounds can be established,

$$\sup_{0\le t\le T} \mathbb{E}|X^n_t|^4 \le C_T(1 + \mathbb{E}|x_0|^4), \tag{12}$$

and

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|X^n_t - X^n_s|^4 \le C_T h^2, \tag{13}$$

also with a (new) constant C_T that 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 [10, Theorem 1.6.4].

3. Let us introduce new processes ξ^n equivalent to X^n on some other – independent – probability space (i.e., we will consider both on the direct product of the two probability spaces). 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 (9)–(10) and by virtue of Skorokhod's Theorem about a single probability space and convergence in probability (see the Lemma 1 in the Appendix, or [25, §6, ch. 1], or [18, 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 $(\tilde{\xi}^n_t,\ t \geq 0)$ for any $n \geq 1$ 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 [18]. We could have also introduced Wiener processes for ξ^n_t and $\tilde{\xi}^n_t$, but they will not show up in this proof. For what follows, let us fix some arbitrary $T > 0$ and consider t in the interval $[0, T]$.

Due to the inequality (13), the same inequality holds for $\tilde{X}^n, \tilde{W}^n$,

$$\sup_{0\le s\le t\le T;\, t-s\le h} \mathbb{E}|\tilde{X}^n_t - \tilde{X}^n_s|^4 \le C_T h^2. \tag{14}$$

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 1 in the Appendix – we may choose a subsequence $n'\to\infty$ so as to pass to the limit in the equation

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

in order to get

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

or, equivalently,

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

This requires some additional explanation because we want to use Krylov's bounds stated for uniformly bounded coefficients while in our setting they may grow infinitely. However, in fact, we can use these bounds because the processes we deal with are uniformly bounded in probability with suitable a priori bounded moments. The details are provided below.

First of all, recall that a priori bounds (9) – (14) 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,

$$(\tilde{X}^{n'}_t, \tilde{\xi}^{n'}_t, \tilde{W}^{n'}_t) \xrightarrow{\mathbb{P}} (\tilde{X}_t, \tilde{\xi}_t, \tilde{W}_t).$$

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 Lebesgue and stochastic integrals are all well defined. Moreover, by virtue of the uniform estimates (13), 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

$$\sup_{0\le t\le T} \mathbb E|\tilde X_t|^2 \le C_T(1+\mathbb E|x_0|^2). \tag{15}$$

4. Let us now show that

$$\int_0^t \mathbb E^3 b^{n'}(s, \tilde X_s^{n'}, \tilde\xi_s^{n'})ds \stackrel{\mathbb P}{\to} \int_0^t \mathbb E^3 b(s,\tilde X_s,\tilde\xi_s)ds, \tag{16}$$

and

$$\int_0^t \mathbb E^3 \sigma^{n'}(s, \tilde X_s^{n'}, \tilde\xi_s^{n'})d\tilde W_s^{n'} \stackrel{\mathbb P}{\to} \int_0^t \mathbb E^3 \sigma(s,\tilde X_s,\tilde\xi_s)d\tilde W_s, \quad n'\to\infty. \tag{17}$$

We are to explain how to use Krylov's estimate for Itô processes with bounded coefficients while in our case they may be unbounded.

Due to the inequality (11), for any $\epsilon>0$ there exists R such that for any n,

$$\mathbb P(\sup_{0\le t\le T} |\tilde X_t^n| \ge R) < \epsilon.$$

The same holds true for $\tilde\xi^n$ (since they are equivalent). Hence, for any $\epsilon>0$ there exists $R>0$ such that for any n,

$$\mathbb P(\sup_{0\le t\le T} (|\tilde X_t^n| \vee |\tilde\xi_t^n|) \ge R-1) < \epsilon$$

(the reason for using $R-1$ instead of R will be clear in the proof), or, equivalently,

$$\mathbb P(\gamma_{n,R-1}\le T) < \epsilon, \tag{18}$$

where

$$\gamma_{n,R} := \inf(t\ge 0: \sup_{0\le t\le T} (|\tilde X_t^n| \vee |\tilde\xi_t^n|) \ge R).$$

The same holds true for the limiting process $(\tilde{X}_t, \tilde{\xi}_t)$ by virtue of its continuity due to Kolmogorov's continuity theorem, that is, 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,

$$\mathbb{P}(\gamma_{R-1} \le T) < \epsilon, \tag{19}$$

where

$$\gamma_R := \inf(t \ge 0: \sup_{0\le s\le t}(|\tilde{X}_s| \vee |\tilde{\xi}_s|) \ge R).$$

In the sequel it will be convenient to define

$$\gamma_R^X := \inf(t \ge 0: \sup_{0\le s\le t}|\tilde{X}_s| \ge R), \ \ \& \ \ \gamma_R^\xi := \inf(t \ge 0: \sup_{0\le s\le t}|\tilde{\xi}_s| \ge R),$$

and similarly

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

Note that

$$1(\gamma_R > T) = 1(\gamma_R^X > T)1(\gamma_R^\xi > T),$$

and similarly

$$1(\gamma_R \wedge \gamma_{n,R} > T) = 1(\gamma_R^X \wedge \gamma_{n,R}^X > T)1(\gamma_R^\xi \wedge \gamma_{n,R}^\xi > T).$$

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 (9)):

$$\mathbb{P}(\gamma_{n,R-1} \le T) \le \frac{\mathbb{E}\sup_{0\le t\le T}(|\tilde{X}_t|^2 \vee |\tilde{\xi}_t|^2)}{(R-1)^2} \le \frac{C(1+\mathbb{E}|x_0|^2)}{(R-1)^2}, \tag{20}$$

by virtue of the Chebyshev–Markov inequality. Now, we estimate,

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

$$\le \mathbb{P}(\gamma_{n,\tilde R}\wedge\gamma_{\tilde R}\le T)$$

$$+\mathbb{P}\left(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T;\, |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R} > T)\left(b^n(s,\tilde X^n_s,\tilde\xi^n_s) - b(s,\tilde X_s,\tilde\xi_s)\right)ds| > c\right).$$

Here the first term does not exceed ϵ if R is large enough, uniformly with respect to n. Further, let us fix some n_0 and let $n > n_0$. We have for any $t\le T$,

$$\mathbb{P}\left(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T;\, |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R} > T)\left(b^n(s,\tilde X^n_s,\tilde\xi^n_s) - b(s,\tilde X_s,\tilde\xi_s)\right)ds| > c\right)$$

$$\le \mathbb{P}\left(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T;\, |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R} > T)\left(b^n(s,\tilde X^n_s,\tilde\xi^n_s) - b^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)\right)ds| > \frac{c}{3}\right)$$

$$+\mathbb{P}\left(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T;\, |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R} > T)\left(b^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s) - b^{n_0}(s,\tilde X_s,\tilde\xi_s)\right)ds| > \frac{c}{3}\right)$$

$$+\mathbb{P}\left(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T;\, |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R} > T)\left(b^{n_0}(s,\tilde X_s,\tilde\xi_s) - b(s,\tilde X_s,\tilde\xi_s)\right)ds| > \frac{c}{3}\right)$$

$$=: I^1 + I^2 + I^3. \tag{21}$$

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 as

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

$$= \frac{3}{c}\,\mathbb{E}\mathbb{E}^3 1(\gamma_{n,\tilde R} > T, \gamma_{\tilde R} > T)\int_0^T |g^{n,n_0}(s,\tilde X^n_{s\wedge\gamma_{n,\tilde R}},\tilde\xi^n_{s\wedge\gamma_{n,\tilde R}})|\,ds. \tag{22}$$

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 $\hat \sigma^{n,\tilde R}[s,x,\mu]$ smooth (e.g., C^1) bounded vector and matrix functions in x respectively, with $\hat \sigma^{n,\tilde R}[s,x,\mu]$ uniformly nondegenerate, such that

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

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

$$d\hat X^n_t = \hat b^{n,\tilde R}[t,\hat X^n_t,\mu^n_t]\,dt + \hat\sigma^{n,\tilde R}[t,\hat X^n_t,\mu^n_t]\,d\tilde W^n_t, \quad \hat X^n_0 = x_0, \tag{23}$$

where μ^n_t is still the marginal distribution of X^n_t and $\tilde X^n_t$. Let also $\hat \xi^n$ be an equivalent independent copy of the process $\hat X^n_t$. Note that on $[0, t\wedge\gamma_{n,\tilde R}]$ the processes $\tilde X^n$ and $\hat X^n$ coincide (see [19, Theorem 6.2.1(v)]). Then the bound for I^1 in the second line of (22) may be rewritten as

$$\begin{aligned} I^1 &\le \frac{3}{c}\,\mathbb{E}\mathbb{E}^3 1(\gamma_{n,\tilde R}>T, \gamma_{\tilde R}>T)\int_0^T |g^{n,n_0}(s,\hat X^n_{s\wedge\gamma_{n,\tilde R}}, \hat \xi^n_{s\wedge\gamma_{n,\tilde R}})\,ds \\ &\le \frac{3}{c}\,\mathbb{E}1(\gamma_{n,\tilde R}>T)\int_0^T |g^{n,n_0}(s,\hat X^n_s,\hat\xi^n_s)|\,ds. \end{aligned} \tag{24}$$

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 (24). 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 (see the Theorem 2.4.1 or the Theorem 2.3.4 in [18]) we obtain,

$$\begin{aligned} I^1 &\le \frac{3}{c}\,\mathbb{E}\int_0^T |g^{n,n_0}(s,\hat X^n_s,\hat\xi^n_s)|\,ds \le N\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|g^{n,n_0}(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}} \\ &= \frac{3N}{c}\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^n(s,x,\xi)-b^{n_0}(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}} \\ &\le \frac{3N}{c}\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^n(s,x,\xi)-b(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}} \\ &+\frac{3N}{c}\left(\int_0^T\int_{|x|\le\tilde R}\int_{|\xi|\le\tilde R}|b^{n_0}(s,x,\xi)-b(s,x,\xi)|^{2d+1}\,dxd\xi ds\right)^{\frac{1}{2d+1}} \to 0, \quad n,n_0\to\infty, \end{aligned}$$

by virtue of the well-known property of mollified functions. Hence, the term I^1 goes to zero as $n\to\infty$ since this term does not depend on n_0.

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

$$I^2 \le \mathbb{P}\Big(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T; |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}>T)\Big(b^{n_0}(s,\tilde X^n_s,\tilde\xi^n_s)ds - b^{n_0}(s,\tilde X_s,\tilde\xi_s)\Big)ds| > \frac{c}{3}\Big)$$

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

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, 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.

To tackle the third term I^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,

$$I^3 = \mathbb{P}\Big(\gamma^X_{n,\tilde R}\wedge\gamma^X_{\tilde R} > T; |\int_0^t \mathbb{E}^3 1(\gamma^\xi_{n,\tilde R}\wedge\gamma^\xi_{\tilde R}>T)\Big(b^{n_0}(s,\tilde X_s,\tilde\xi_s)ds - b(s,\tilde X_s,\tilde\xi_s)\Big)ds| > \frac{c}{3}\Big)$$

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

$$\le \frac{3}{c}\,\mathbb{E}\int_0^T w(\tilde X_s,\tilde\xi_s)|g^{n_0}(s,\tilde X_s,\tilde\xi_s)|\,ds. \tag{25}$$

We want to show that the right hand side (the last term) in (25) goes to zero, firstly, as $n_0 \to \infty$, and secondly, as $R\to\infty$. (Strangely, the indicators $1\left(\gamma_{n,\tilde R}\wedge\gamma_{\tilde R} > T\right)$ are not of a real help here, although, new similar ones will appear shortly, see below.) The values of the function $g^{n_0}(s,x,\xi)$ outside the set $\{(x,\xi) : (|x|\vee|\xi|) \le \tilde R\}$ are not relevant for the evaluation of the expression in the right hand side of (25). So, without losing of generality we may assume that $g^{n_0}(s,x,\xi)$ vanishes outside this ball. Thus, in particular, we can also accept that $g^{n_0}(s,x,\xi)$ is uniformly bounded. Take any $\epsilon > 0$ and choose and fix for a while R so large that

$$TC(1+\mathbb{E}|x_0|^2)\tilde R^{-1} < \epsilon. \tag{26}$$

The reason for this choice will be clarified shortly. For any function $g(s,x,\xi) \in \mathcal{L} := \left(g \text{ (Borel measurable)} : \sup\limits_{s,x,\xi} \dfrac{|g(s,x,\xi)|}{1+|x|} \le C\right)$ for a fixed $C>0$ let us show the bound,

$$\mathbb{E}\int_0^T w(\tilde X_s, \tilde\xi_s)|g(s,\tilde X_s,\tilde\xi_s)|\,ds \le C\sqrt{\epsilon} + N\|g\|_{L_{2d+1}([0,T]\times B_R\times B_R)}, \tag{27}$$

with some $C>1$. First of all note that it suffices to establish this inequality with $(C-1)\sqrt{\epsilon}$ instead of $C\sqrt{\epsilon}$ only for continuous functions g vanishing outside the set $[0,T]\times B_R\times B_R)$, of course, assuming that the constant N does not depend on the regularity of g. Indeed, such (continuous) functions are, clearly, dense in the class $\mathcal{L}_R$ in the L_{2d+1} norm. So, choosing $\tilde g \in C$ and $\sup\limits_{s,x,\xi} \dfrac{|g(s,x,\xi)|}{1+|x|} \le C$ so that $\|\tilde g - g\|_{L_{2d+1}([0,T]\times B_R\times B_R)} < \sqrt{\epsilon}/N$, and provided that the desired estimate with $C\sqrt{\epsilon}$ replaced by $(C-1)\sqrt{\epsilon}$ is valid for $\tilde g$, we immediately get (27) for g with $C\sqrt{\epsilon}$.

In order to establish the bound with some C and N

$$\mathbb{E}\int_0^T w(\tilde X_s, \tilde\xi_s)|g(s,\tilde X_s,\tilde\xi_s)|\,ds \le C\sqrt{\epsilon} + N\|g\|_{L_{2d+1}([0,T]\times B_R\times B_R)} \tag{28}$$

for $g \in C([0,T]\times B_R\times B_R)\bigcap \mathcal{L})$ such that $\sup_{s,x,\xi}\frac{|g(s,x,\xi)|}{1+|x|} \le C$, return to the pre-limiting "smoothed" diffusions $(\tilde X^k, \tilde\xi^k)$. We estimate, using for one of the terms the replacement $\hat X$ and $\hat \xi$ (see (23)),

$$
\begin{aligned}
& \mathbb{E} \int_0^T w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)|\, ds \\
& \le \mathbb{E} 1(\gamma_{k,\tilde R} > T) \int_0^T w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)|\, ds \\
& + \mathbb{E} 1(\gamma_{k,\tilde R} \le T) \int_0^T w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)|\, ds \\
& = \mathbb{E} 1(\gamma_{k,\tilde R} > T) \int_0^T w(\hat X_s^k, \hat \xi_s^k) |g(s, \hat X_s^k, \hat \xi_s^k)|\, ds \\
& + \mathbb{E} 1(\gamma_{k,\tilde R} \le T) \int_0^T w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)|\, ds \\
& \le N \|g\|_{L_{2d+1}([0,T]\times B_R \times B_R)} \\
& + \left(\mathbb{P}(\gamma_{k,\tilde R} \le T)\right)^{1/2} \left(\mathbb{E} \left(\int_0^T w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)|\, ds \right)^2 \right)^{1/2},
\end{aligned}
$$

where the first integral with $1(\gamma_{k,\tilde R} > T)$ and $\hat X, \hat \xi$ was estimated by Krylov's bound while to the second one with $1(\gamma_{k,\tilde R} \le T)$ and $\tilde X, \tilde \xi$ we applied the Cauchy–Bunyakovsky–Schwarz inequality. Finally, for R large enough

$$
\mathbb{P}(\gamma_{k,\tilde R} \le T) \le \epsilon
$$

(see (18)), while by Jensen's inequality,

$$
\begin{aligned}
\mathbb{E} \left(\int_0^T w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)|\, ds \right)^2 & \le T \int_0^T \mathbb{E} \left(w(\tilde X_s^k, \tilde \xi_s^k) |g(s, \tilde X_s^k, \tilde \xi_s^k)| \right)^2 ds \\
& \le TC \int_0^T \mathbb{E} \left(1 + |\tilde X_s^k|^2 + |\tilde \xi_s^k)|^2 \right) ds \le \tilde C < \infty,
\end{aligned}
$$

with some $\tilde C$ due to (15) and (26). The bound (28) for $g \in \mathcal{LC}$ follows now by virtue of the Fatou Lemma as $k \to \infty$. Hence, (27) for any $g \in \mathcal{L}$ holds true. This implies that

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

The convergence (16) is, thus, proved.

5. Now let us consider convergence of stochastic integrals in (17). Our goal is an estimate similar to that for the drift and Lebesgue integrals above:

$$
\mathbb{P}(|\int_0^t \mathbb{E}^3\sigma^n(s,\tilde X_s^n,\tilde\xi_s^n)dW_s^n - \int_0^t \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s)d\tilde W_s| > c) < C\epsilon, \tag{29}
$$

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 2 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 the estimate (18) which reduces the problem to bounded processes.

By virtue of [19, 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,

$$
\begin{aligned}
&\mathbb{P}(|\int_0^t \mathbb{E}^3\sigma^n(s,\tilde X_s^n,\tilde\xi_s^n)d\tilde W_s^n - \int_0^t \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s)d\tilde W_s| > c) \\
&\le c^{-2}\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^n(s,\tilde X_s^n,\tilde\xi_s^n)d\tilde W_s^n - \int_0^t \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s)d\tilde W_s\right|^2 \\
&\le C\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^n(s,\hat X_s^n,\hat\xi_s^n)d\tilde W_s^n - \int_0^t \mathbb{E}^3\sigma(s,\hat X_s,\hat\xi_s)d\tilde W_s\right|^2 \\
&+C\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^n(s,\hat X_s^n,\hat\xi_s^n)d\tilde W_s^n - \int_0^t \mathbb{E}^3\sigma^n(s,\tilde X_s^n,\tilde\xi_s^n)d\tilde W_s^n\right|^2 \\
&+C\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma(s,\hat X_s,\hat\xi_s)d\tilde W_s - \int_0^t \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s)d\tilde W_s\right|^2 \\
&=: S^1+S^2+S^3.
\end{aligned}
$$

Here the term S^3 can be evaluated as follows (notations taken from the previous step):

$$S^3 \le 2C\mathbb{E}\left|\int_0^t \mathbb{E}^3 1(\gamma_{\tilde R} < s)\sigma(s,\tilde X_s,\tilde \xi_s)d\tilde W_s\right|^2 + 2C\mathbb{E}\left|\int_0^t \mathbb{E}^3 1(\gamma_{\tilde R} < s)\sigma(s,\hat X_s,\hat \xi_s)d\tilde W_s\right|^2$$

$$= 2C\mathbb{E}\int_{t\wedge\gamma^X_{\tilde R}}^t \mathrm{Tr}\left(\mathbb{E}^3 1(\gamma^\xi_{\tilde R} < s)\sigma(s,\tilde X_s,\tilde \xi_s)\right)\left(\mathbb{E}^3 1(\gamma^\xi_{\tilde R} < s)\sigma(s,\tilde X_s,\tilde \xi_s)\right)^* ds$$

$$+2C\mathbb{E}\int_{t\wedge\gamma^X_{\tilde R}}^t \mathrm{Tr}\left(\mathbb{E}^3 1(\gamma^\xi_{\tilde R} < s)\sigma(s,\hat X_s,\hat \xi_s)\right)\left(\mathbb{E}^3 1(\gamma^\xi_{\tilde R} < s)\sigma(s,\hat X_s,\hat \xi_s)\right)^* ds$$

$$\le C\mathbb{E}\int_{t\wedge\gamma^X_{\tilde R}}^t (1+|\tilde X_s|^2)\,ds + C\mathbb{E}\int_{t\wedge\gamma^X_{\tilde R}}^t (1+|\hat X_s|^2)\,ds \to 0, \quad \tilde R\to\infty,$$

as in the previous step for the drift. Quite similarly, for S^2 we have,

$$S^2 \le C\mathbb{E}\left|\int_0^t \mathbb{E}^3 1(\gamma_{n,\tilde R} < s)\sigma(s,\tilde X^n_s,\tilde \xi^n_s)d\tilde W^n_s\right|^2 + C\mathbb{E}\left|\int_t^t \mathbb{E}^3 1(\gamma_{n,\tilde R} < s)\sigma(s,\hat X^n_s,\hat \xi^n_s)d\tilde W^n_s\right|^2$$

$$\le C\mathbb{E}\int_{t\wedge\gamma^X_{n,\tilde R}}^t (1+|\tilde X^n_s|^2)\,ds + C\mathbb{E}\int_{t\wedge\gamma^X_{n,\tilde R}}^t (1+|\hat X^n_s|^2)\,ds \to 0, \quad \tilde R\to\infty.$$

For the term S^1 which only contains diffusions $\hat X$ and $\hat \xi$ with bounded coefficients, we finally estimate,

$$\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^n(s,\hat X^n_s,\hat \xi^n_s)d\tilde W^n_s - \int_0^t \mathbb{E}^3\sigma(s,\hat X_s,\hat \xi_s)d\tilde W_s\right|^2$$

$$\le 3\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^n(s,\hat X^n_s,\hat \xi^n_s)d\tilde W^n_s - \int_0^t \mathbb{E}^3\sigma^{n_0}(s,\hat X^n_s,\hat \xi^n_s)d\tilde W^n_s\right|^2$$

$$+3\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^{n_0}(s,\hat X^n_s,\hat \xi^n_s)d\tilde W^n_s - \int_0^t \mathbb{E}^3\sigma^{n_0}(s,\hat X_s,\hat \xi_s)d\tilde W_s\right|^2$$

$$+3\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^{n_0}(s,\hat X_s,\hat \xi_s)d\tilde W_s - \int_0^t \mathbb{E}^3\sigma(s,\hat X_s,\hat \xi_s)d\tilde W_s\right|^2$$

$$= 3\mathbb{E}\int_0^t \mathrm{Tr}\left(\mathbb{E}^3(\sigma^n(s,\hat X^n_s,\hat \xi^n_s) - \sigma^{n_0}(s,\hat X^n_s,\hat \xi^n_s))\right)\left(\mathbb{E}^3(\sigma^n(s,\hat X^n_s,\hat \xi^n_s) - \sigma^{n_0}(s,\hat X^n_s,\hat \xi^n_s))\right)^* ds$$

$$+3\mathbb{E}\left|\int_0^t \mathbb{E}^3\sigma^{n_0}(s,\hat X^n_s,\hat \xi^n_s)d\tilde W^n_s - \int_0^t \mathbb{E}^3\sigma^{n_0}(s,\hat X_s,\hat \xi_s)d\tilde W_s\right|^2$$

$$+3\mathbb{E}\int_0^t \mathrm{Tr}\left(\mathbb{E}^3(\sigma^{n_0}(s,\hat X_s,\hat \xi_s) - \sigma(s,\hat X_s,\hat \xi_s))\right)\left(\mathbb{E}^3(\sigma^{n_0}(s,\hat X_s,\hat \xi_s) - \sigma(s,\hat X_s,\hat \xi_s))\right)^* ds$$

$$=: S^{11} + S^{12} + S^{13}.$$

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 $\gamma_{n,\tilde R} \to \infty$ uniformly in probability as $\tilde R \to \infty$. We have skipped the functions similar to g^{n,n_0} from the previous step which work here in a totally similar way. Finally, the term S^{12} goes to zero as $n \to \infty$ by the Skorokhod Lemma 2 (see the Appendix) along with the Lebesgue's integrable convergence Theorem. This proves the desired bound (29).

So, we have established both (16) and (17), and thus, weak solution of the equation (1)–(2) exists in the case of $d_1 = d$ and under the assumption (5) instead of (4). Recall that once μ_t is the distribution of ξ_t, and distributions of ξ_t and X_t coincide, then μ_t is also the distribution of X_t.

6. Now we will show under the general assumption of continuity of the coefficients with respect to μ how to drop the assumption (5) and, in particular, how to drop the condition $d_1 = d$. We will use a hint from [31, section 4]; however, due to a more involved structure of the equation and its coefficients in this paper, it is desirable to repeat the details here.

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. However, this view is not precise. Firstly, for strong solutions this is certainly not true because regularity such as Lipschitz condition or even a simple continuity may fail for 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. Recall that existing results about weak solutions and weak uniqueness – see, e.g., [4, 9] – impose conditions on σ and not on $\sigma\sigma^*$. Hence, we find it not sufficient to refer to the "common knowledge" and have to show the calculus.

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,

$$\tilde X_t = x + \int_0^t B[s,\tilde X_s,\mu_s]ds + \int_0^t \tilde\Sigma[s,\tilde X_s,\mu_s]d\tilde W_s, \qquad (30)$$

with some d-dimensional Wiener process $(\tilde W_t,\, t \ge 0)$ on some probability space and where μ_s stands for the distribution of $\tilde X_s$.

Existence of this weak solution will be justified in the next step of the proof with the help of the Riesz – Dunford (Cauchy) formula for a function of a positive self-adjoint matrix (see, e.g., [7, VII.3.9]),

$$\sqrt{A[t,x,\mu]} = \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}(\lambda - A[t,x,\mu])^{-1}\, d\lambda. \tag{31}$$

where the contour $\Gamma\subset\mathbb{C}$ is to be chosen in a way so that its interior contains all the eigenvalues of the (elliptic) matrix $A[s,x,\cdot]$, say, for x from some compact set; due to the locally uniform ellipticity it is possible to choose Γ in a unique way for all (s,x,μ), at least, for all x from any compact and then the desired weak continuity follows directly from the right hand side of (31) with the help of the standard stopping times.

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

$$p[s,x,\mu] := \tilde\Sigma[s,x,\mu]^{-1}\,\Sigma[s,x,\mu]. \tag{32}$$

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

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

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 [12]),

$$\int_0^t \Sigma[s, \tilde X_s, \mu_s]\, dW_s^0 = \int \Sigma p^*[s, \tilde X_s, \mu_s]\, d\tilde W + \int \Sigma (I - p^* p)[s, \tilde X_s, \mu_s]\, d\bar W$$
$$= \int (A)^{1/2}[s, \tilde X_s, \mu_s]\, d\tilde W = \int \tilde\Sigma[s, \tilde X_s, \mu_s]\, d\tilde W = \tilde X_t - x - \int_0^t B[s, \tilde X_s, \mu_s]\, ds. \tag{33}$$

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.

7. Now it remains to show independently of the previous step existence of weak solution to the equation (30), with $\tilde\Sigma[t,x,\mu] := \sqrt{A[t,x,\mu]}$, where

$$A[t,x,\mu] := \Sigma[t,x,\mu]\Sigma^*[t,x,\mu] = \left(\int \sigma(t,x,y)\mu(dy)\right)\left(\int \sigma^*(t,x,y)\mu(dy)\right).$$

Assume for a minute that all coefficients σ and b are bounded. Then a unique contour Γ in (31) may be chosen such that the equation (30) can be rewritten as

$$\tilde X_t = x + \int_0^t B[s, \tilde X_s, \mu_s] ds + \int_0^t \left(\frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2} (\lambda - A[s, \tilde X_s, \mu_s])^{-1}\, d\lambda\right) d\tilde W_s. \tag{34}$$

Correspondingly, when we mollify σ in the last variable, the smoothed equations look

$$\tilde X_t^n = x + \int_0^t B[s, \tilde X_s^n, \mu_s^n] ds + \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2} (\lambda - A^n[s, \tilde X_s^n, \mu_s^n])^{-1}\, d\lambda\, d\tilde W_s^n,$$

or, equivalently,

$$\tilde X^n_t = x + \int_0^t \mathbb{E}^3 b(s, \tilde X^n_s, \tilde\xi^n_s) ds$$

$$+\int_0^t \left(\frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\left(\lambda - \int \mathbb{E}^3\sigma(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_n(z)dz \times \right.\right.$$

$$\left.\left. \times \int \mathbb{E}^3\sigma^*(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_n(z)dz\right)^{-1} d\lambda\right) d\tilde W^n_s, \tag{35}$$

where, as usual, $(\tilde\xi^n)$ are independent and equivalent to $(\tilde X^n)$ processes and $\mathbb{E}^3$ means expectation with respect to the "third variable". The equation (35) has a (strong) solution due to smoothness of the coefficients. Note that here the matrix σ is mollified, not $\sigma\sigma^*$, and only with respect to the third variable. This is important because the smoothed diffusion remains non-degenerate which is used in Krylov's bounds. Because of this smoothing (and due to the non-degeneracy), the equation (35) has a weak solution according to the part II of the Theorem which is already proved; in fact, smoothing of the drift could have been also performed here but it is not necessary in this variant. Convergence (over a subsequence) of the term $\int_0^t B[s,\tilde X^n_s,\mu^n_s]\,ds = \int_0^t \mathbb{E}^3 b[s,\tilde X^n_s,\tilde\xi^n_s]\,ds$ to the limiting one $\int_0^t B[s,\tilde X_s,\mu_s]\,ds$ follows from the same calculus as earlier in the step 1.4, based on the non-degeneracy, Krylov's estimates, and stopping times. A bit more involved is the stochastic term,

$$\int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\left(\lambda - \left(\int \mathbb{E}^3\sigma(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_n(z)dz \times \right.\right.$$

$$\left.\left. \times \int \mathbb{E}^3\sigma^*(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_n(z)dz\right)\right)^{-1} d\lambda\, d\tilde W^n_s.$$

We will evaluate the difference between this term and its desirable limit by splitting into the following three parts analogous to the step 1.5:

$$I^1 = \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\left(\lambda - \left(\int \mathbb{E}^3\sigma(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_n(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_n(z)dz\right)\right)^{-1} d\lambda\, d\tilde W^n_s$$

$$-\int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\left(\lambda - \left(\int \mathbb{E}^3\sigma(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde X^n_s,\tilde\xi^n_s - z)\phi_{n_0}(z)dz\right)\right)^{-1} d\lambda\, d\tilde W^n_s,$$

$$I^2 = \int_0^t \frac{1}{2\pi i} \oint_\Gamma \lambda^{1/2} \left(\lambda - \left(\int \mathbb{E}^3 \sigma(s, \tilde{X}_s^n, \tilde{\xi}_s^n - z) \phi_{n_0}(z) dz \int \mathbb{E}^3 \sigma^*(s, \tilde{X}_s^n, \tilde{\xi}_s^n - z) \phi_{n_0}(z) dz\right)\right)^{-1} d\lambda \, d\tilde{W}_s^n$$

$$- \int_0^t \frac{1}{2\pi i} \oint_\Gamma \lambda^{1/2} \left(\lambda - \left(\int \mathbb{E}^3 \sigma(s, \tilde{X}_s, \tilde{\xi}_s - z) \phi_{n_0}(z) dz \int \mathbb{E}^3 \sigma^*(s, \tilde{X}_s, \tilde{\xi}_s - z) \phi_{n_0}(z) dz\right)\right)^{-1} d\lambda \, d\tilde{W}_s,$$

and

$$I^3 = \int_0^t \frac{1}{2\pi i} \oint_\Gamma \lambda^{1/2} \left(\lambda - \left(\int \mathbb{E}^3 \sigma(s, \tilde{X}_s, \tilde{\xi}_s - z) \phi_{n_0}(z) dz \int \mathbb{E}^3 \sigma^*(s, \tilde{X}_s, \tilde{\xi}_s - z) \phi_{n_0}(z) dz\right)\right)^{-1} d\lambda \, d\tilde{W}_s$$

$$- \int_0^t \frac{1}{2\pi i} \oint_\Gamma \lambda^{1/2} \left(\lambda - \left(\mathbb{E}^3 \sigma(s, \tilde{X}_s, \tilde{\xi}_s) \mathbb{E}^3 \sigma^*(s, \tilde{X}_s, \tilde{\xi}_s)\right)\right)^{-1} d\lambda \, d\tilde{W}_s.$$

Let us start with the term I^1, probably the most instructive and intuitive, although, not the most complicated. Using the formula for the resolvent difference

$$(\lambda - A_1)^{-1} - (\lambda - A_2)^{-1} = (\lambda - A_1)^{-1}(A_2 - A_1)(\lambda - A_2)^{-1}, \quad \lambda \notin \mathrm{sp}(A_1) \cup \mathrm{sp}(A_2),$$

and choosing the contour Γ so that $\|(\lambda - A_i)^{-1}\|$ (as well as $|\lambda|$ itself) is uniformly bounded on it for $i = 1, 2$, where A_1 stands for the matrix

$$\mathbb{E}^3 \left(\int \sigma\sigma^*(s, \tilde{X}_s^n, \tilde{\xi}_s^n - z) \phi_n(z) dz\right),$$

and A_2 for

$$\mathbb{E}^3 \left(\int \sigma\sigma^*(s, \tilde{X}_s^n, \tilde{\xi}_s^n - z) \phi_{n_0}(z) \, dz\right),$$

we estimate (constants C may change from line to line),

$$\mathbb{E}\|\int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde{W}^n_s$$

$$-\int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde{W}^n_s\|^2$$

$$\le \mathbb{E}\int_0^t \|\frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz\Big)\Big)^{-1} d\lambda$$

$$-\int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\|^2 ds$$

$$\le C\mathbb{E}\int_0^t |\oint_\Gamma \| \Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz\Big)\Big)^{-1}$$

$$-\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} \|d\lambda|^2 ds$$

$$\le C\mathbb{E}\int_0^t \| \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz\Big)$$

$$-\Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz\Big)\|^2_{L_{2d+1}}\, ds$$

$$\le C\mathbb{E}\int_0^t \| \Big(\int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_n(z)dz - \int \mathbb{E}^3\sigma(s,\tilde{X}^n_s,\tilde{\xi}^n_s - z)\phi_{n_0}(z)dz\Big)\|_{L_{2d+1}}\, ds$$

$$\le N\left(\|\int \sigma(s,x,\xi - z)\phi_n(z)dz - \sigma(s,x,\xi)\|_{L_{2d+1}} + \|\int \sigma(s,x,\xi - z)\phi_{n_0}(z)dz - \sigma(s,x,\xi)\|_{L_{2d+1}}\right),$$

where L_{2d+1} stands in both lines for the integral norms of order $2d+1$ in the variables (s,x,ξ) of the expressions $\|\ldots\|_{L_{2d+1}}$. Hence, again with the help of Krylov's estimates and usual stopping times – which should be added in a usual way to the calculus above – the issue is reduced to the convergence in L_p norms of the differences between the mollified functions and their originals. It is true that the corresponding norm of these differences tends to zero on any compact domain, while probability that our solutions exit this domain on a finite interval of time is small if the domain is chosen large enough. Overall, this justifies the convergence $\mathbb{E}\|I^1\|^2 \to 0$, $n, n_0 \to \infty$.

In the case of unbounded coefficients the countour Γ, generally speaking, may not be chosen uniform for all values of x, but only for all x from any compact. This suffices for the calculus as above with stopping times as in the step 1.5.

Now let us show that $I^2 \to 0$ in square mean as $n \to \infty$, for each n_0. This is a little more involved than usually with this kind of term because we did not smooth in the second variable (so as to keep the diffusion non-degenerate) and because no continuity is assumed with respect to it. Yet, we will use the same trick as above along with Krylov's bounds. Although the matrix function σ is already mollified with respect to the third variable, let us smooth it again in *all* the variables. We will not specify the kernels now (the old ones ψ_n, ϕ_n can be used) and just denote the result as $\sigma^\delta(s, x, \xi)$ which is continuous in all variables and assumed close to σ in the L_{2d+1} norm. So we estimate, again assuming for a minute that the contour Γ

can be chosen unique for all s, x,

$$\mathbb{E}\| \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s^n$$

$$- \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s\|^2$$

$$\le 3\mathbb{E}\| \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s^n$$

$$- \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma^\delta(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz \int \mathbb{E}^3(\sigma^\delta)^*(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s^n\|^2$$

$$+3\mathbb{E}\| \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma^\delta(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz \int \mathbb{E}^3(\sigma^\delta)^*(s,\tilde X_s^n,\tilde\xi_s^n - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s^n$$

$$- \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma^\delta(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3(\sigma^\delta)^*(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s\|^2$$

$$+3\mathbb{E}\| \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma^\delta(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3(\sigma^\delta)^*(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s$$

$$- \int_0^t \frac{1}{2\pi i}\oint_\Gamma \lambda^{1/2}\Big(\lambda - \Big(\int \mathbb{E}^3\sigma(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz \int \mathbb{E}^3\sigma^*(s,\tilde X_s,\tilde\xi_s - z)\phi_{n_0}(z)dz\Big)\Big)^{-1} d\lambda\, d\tilde W_s\|^2$$

$$=: 3(I^{21} + I^{22} + I^{23}).$$

Here by Krylov's bounds,

$$I^{21} \le N\|\sigma - \sigma^\delta\|^2_{L_{4d+2}} \to 0, \quad \delta \to 0.$$

Further,

$$I^{22} \to 0, \quad n \to \infty,$$

by Skorokhod's Lemma, see the Appendix. Finally,

$$I^{23} \le N\|\sigma - \sigma^{\delta}\|^2_{L_{4d+2}},$$

by virtue of Fatou's lemma if we firstly establish this Krylov type bound for continuous functions and $(\tilde X_n, \tilde{\xi}^n)$ and then approximate the difference $\sigma - \sigma^{\delta}$ by continuous functions in L_{4d+2}. This shows that, indeed, for each n_0

$$\mathbb{E}|I^2|^2 \to 0, \quad n \to \infty.$$

Again, in the case of unbounded coefficients the countour Γ may not be chosen uniform for all values of x, but only for all x from any compact. However, this suffices for the calculus as above with stopping times as in the step 5.

The last term I^3 is estimated quite similarly to I^{23}: we smooth the matrix function σ in all variables, write Krylov's bounds for the pre-limiting processes $(\tilde X^n, \xi^n)$, and pass to the limit by Fatou's lemma. The same remark about locally uniform choice of Γ and stopping times applies. Therefore, we also have,

$$\mathbb{E}|I^3|^2 \to 0, \quad n \to \infty,$$

as desired. Hence, we may conclude that there exists a weak solution of the equation (30). The proof of the Theorem 1 is thus completed.

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 theorem, but only strong existence in the sense of the Definition 1. We also notice for interested readers that in [29] the assumption of continuity in time was dropped in comparison to [28]; 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 nondegeneracy assumption (5), and let just σ be Lipschitz in x uniformly with respect to s and locally with respect to y,*

$$\|\sigma(t,x,y) - \sigma(t,x',y)\| \le C(1+|y|^2)|x-x'|. \tag{36}$$

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. However, the authors were unable to find an exact reference which is desirable. So, for completeness as well as for the convenience of the reader a brief sketch of the proof is presented below.

1. First of all, note that that weak solutions exist and that the a priori bounds (9)–(13) are valid.

Considerations are based on the results from [28] and [29] 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,

$$dX_t = \tilde{b}(t,X_t)dt + \tilde{\sigma}(t,X_t)dW_t, \qquad X_0 = x. \tag{37}$$

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 nondegenerate, or that both $\tilde{b}$ and $\tilde{\sigma}$ are Lipschitz in x in the second case.

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

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

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 (9),

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

The uniform nondegeneracy of σ – and, hence, also of $\sigma\sigma^*$ – follows from the inequality (5) 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 [28]. However, because weak solution is well-defined for all values of time, strong uniqueness is global. According to the Yamada–Watanabe principle ([33]), any solution of the equation (2) is strong. So, any solution of the original equation (1) is also strong.

(2) 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 2 *Notice that as a solution of the "linearized" equation (37), 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 2 below.*

Remark 3 *In the case of dimension one, Lipschitz condition may be relaxed to Hölder of order $1/2$ and, actually, a little bit further by using techniques from [33] and [27]. Under the additional assumption of boundedness of b and σ, the fourth moment of the initial value x_0 is not necessary and can be further relaxed as in the Theorem 1.*

3.2 Strong and weak uniqueness: main result

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 2 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 nondegeneracy. 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 (5); let* $d_1 = d$*, the matrix* σ *be quadratic, 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* $x \in R^d$*,*

$$\sup_{s,\mu} |\tilde{B}[s,x,\mu]| \le C(1+|x|). \tag{38}$$

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

$$dX^0_t = \sigma(t, X^0_t)\, dW_t, \quad X^0_0 = x_0, \tag{39}$$

has a unique strong solution for any x *(see [28, 29]):*

$$\sup_{t\ge 0}\ \sup_{x,x'\colon x'\neq x} \frac{\|\sigma(t,x)-\sigma(t,x')\|}{|x-x'|} < \infty. \tag{40}$$

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

Remark 4 *Just for weak uniqueness – without strong one – the assumption (40) may be relaxed to the uniform continuity for any* t *of* $\sigma\sigma^*(t,\cdot)$ *if* $d > 1$*, or even dropped completely if* $d = 1$*.*

Remark 5 *Note that under the condition (40), not only the equation (39) 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 (37).*

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 (38); 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 particuar, 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 6 *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 as in the Theorem 1 or even weaker.*

Remark 7 *Instead of Lipschitz condition (40), 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 [28, Theorem 1] and [29], and any of them can be used in our Theorem 2 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 [28]. As usual, even more relaxed conditions on sigma can be stated in the case of dimension one as in [28, Theorem 2].*

3.3 Proof of Theorem 2

Denote by X^0_t the unique (strong) solution of the Itô equation (39). Note that

$$dW_t = \sigma^{-1}(t, X^0_t) dX^0_t.$$

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 "linearised" equation (37). We will show this weak uniqueness by contradiction. Suppose there are two solutions X^1 and X^2 of the equation (1) with distributions μ^1 and μ^2 respectively in the space of trajectories $C[0,\infty;\mathbb{R}^d]$. Without loss of generality, we may and will assume that both processes X^1 and X^2 are realized on the same probability space and with the same Wiener process:

$$dX^1_t = \sigma(t, X^1_t)\, dW_t + B[t, X^1_t, \mu^1_t]\, dt, \quad X^1_0 = x, \tag{41}$$

and

$$dX^2_t = \sigma(t, X^2_t)\, dW_t + B[t, X^2_t, \mu^2_t]\, dt, \quad X^2_0 = x, \tag{42}$$

respectively. This is possible because any solution of this equation is strong and, hence, exists on any probability space with a Wiener process of the required dimension. This is not necessary for the proof and could have been avoided if we only aimed to prove weak uniqueness, see the Remark 4. 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 ([17]).

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

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

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

It is well-known that in the case of bounded $\tilde{B}$ the random variables γ^i_T, $i = 1, 2$, are probability densities due to Girsanov's theorem (see, e.g., [19, Theorem 6.8.8]). So, till the step 4 we assume $\tilde{B}$ bounded; note that in this case we have,

$$|B[s, x, \mu] - B[s, x, \nu]| \le C\|\mu - \nu\|_{TV}. \tag{43}$$

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

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

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

$$\begin{aligned} dX^0_t &= \sigma(t, X^0_t) d\tilde{W}^1_t + \sigma(t, X^0_t)\tilde{B}[t, X^0_t, \mu^1_t] dt \\ &= \sigma(t, X^0_t) d\tilde{W}^1_t + B[t, X^0_t, \mu^1_t]\, dt, \end{aligned} \tag{44}$$

with the initial condition $X^0_0 = x_0$. In other words, the process X^0 on $[0, T]$ satisfies the equation (41), just with another Wiener process and under another probability measure. However, given μ^1_t, $0 \le t \le T$, this solution considered as a solution of Itô's – or "linearized" – equation is strongly unique [28, 29]. As a consequence, they are also weakly unique; note that this is the reason for the Remark 4 along with the calculus in the next steps. 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 (42), 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 $A \subset C[0,T; \mathbb{R}^d]$,

$$\mu^1_{0,T}(A) := \mathbb{P}(X^1 \in A) = \mathbb{P}^{\gamma^1}(X^0 \in A) = \mathbb{E}^{\gamma^1} 1(X^0 \in A) = \mathbb{E}\gamma^1_T 1(X^0 \in A), \quad (45)$$

and

$$\mu^2_{0,T}(A) := \mathbb{P}(X^2 \in A) = \mathbb{P}^{\gamma^2}(X^0 \in A) = \mathbb{E}^{\gamma^2} 1(X^0 \in A) = \mathbb{E}\gamma^2_T 1(X^0 \in A). \quad (46)$$

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

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

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

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

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

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

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

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

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

Further, due to (45) and (46) the measure μ^i is an image of $\mathbb{P}^{\gamma^i}$ under the mapping X^0 for $i = 1, 2$. So,

$$v(t) := \|\mu^1_{[0,t]} - \mu^2_{[0,t]}\|_{TV} \le \|P^{\gamma^1}|_{\mathcal{F}_t^W} - P^{\gamma^2}|_{\mathcal{F}_t^W}\|_{TV}. \tag{47}$$

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

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

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

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

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 (47),

$$v(t) \le 2\sqrt{\mathbb{E}^{\gamma^1} \rho_t^2 - 1}. \tag{48}$$

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

$$
\begin{aligned}
\mathbb{E}^{\gamma^1}\rho_T^2 &= \mathbb{E}^{\gamma^1}\exp(-2\int_0^T (\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1])d\tilde{W}_s^1 \\
&\qquad - \int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds) \\
&= \mathbb{E}^{\gamma^1}\exp(-2\int_0^T (\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1])d\tilde{W}_s^1 \\
&\qquad -4\int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds) \\
&\qquad \times\exp(+3\int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds) \\
&\le \Bigg(\mathbb{E}^{\gamma^1}\exp(-4\int_0^T (\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1])d\tilde{W}_s^1 \\
&\qquad -8\int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds)\Bigg)^{1/2} \\
&\quad \times\Bigg(\mathbb{E}^{\gamma^1}\exp(6\int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds)\Bigg)^{1/2} \\
&\le (=)\sqrt{\mathbb{E}^{\gamma^1}\exp\left(6\int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds\right)}.
\end{aligned}
\tag{49}
$$

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

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

$$
\begin{aligned}
&\mathbb{E}^{\gamma^1}\exp\left(6\int_0^T |\tilde{B}[s,X_s^0,\mu_s^2]-\tilde{B}[s,X_s^0,\mu_s^1]|^2 ds\right) \\
&\le \mathbb{E}^{\gamma^1}\exp\left(6\|\tilde{B}\|_B^2\int_0^T \|\mu_s^1-\mu_s^2\|_{TV}^2\, ds\right).
\end{aligned}
\tag{50}
$$

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

$$v(T) \le 2\sqrt{\exp\left(C\int_0^T v(s)^2 ds\right) - 1}. \tag{51}$$

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

$$\exp(\alpha) - 1 \le 2\alpha, \tag{52}$$

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

$$v(T) \le 2\sqrt{\exp\left(C\int_0^T v(s)^2\, ds\right) - 1} \le 2\sqrt{2CTv(T)^2} = 2\sqrt{2CT}v(T). \tag{53}$$

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

$$v(2T) = v(3T) = \ldots = 0. \tag{54}$$

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

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

and re-denote

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

Then $\tilde{W}^1_t$ 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 (49), (50), and (51), and having in mind the induction assumption $v(kT) = 0$, we obtain with the same constant C,

$$v((k+1)T) \le \sqrt{\exp\left(C\int_{kT}^{(k+1)T} v(s)^2 ds\right) - 1}, \tag{55}$$

which straightforward implies

$$v((k+1)T) \le \sqrt{2CTv((k+1)T)^2} = \sqrt{2CT}v((k+1)T). \tag{56}$$

As earlier, the condition $T < \min(1/(2C), 1/(\alpha C))$ (see (52)) guarantees that

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

as required. This completes the induction (54).

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 2 as well as of the Remark 6 are justified.

4. Now let us return to the inequality (49) and explain how to drop the additional assumption of boundedness of $\tilde{B}$, and also how to deal with a localised version of (43). First of all, prior to (49) 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 (49) by virtue of Cauchy–Bunyakovsky–Schwarz, this problem is reduced to the question whether or not the following expression is finite:

$$\begin{aligned}
\mathbb{E}(\gamma^i)^2 \le &\left(\mathbb{E}\exp(4\int_0^T \tilde{B}[s, X_s^0, \mu_s^i]\, dW_s - 8\int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2\, ds))\right)^{1/2} \\
&\times \left(\mathbb{E}\exp(6\int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2\, ds)\right)^{1/2} \\
&\le \left(\mathbb{E}\exp(6\int_0^T |\tilde{B}[s, X_s^0, \mu_s^i]|^2\, ds)\right)^{1/2} \\
&\le \left(\mathbb{E}\exp(C\int_0^T (1 + |X_s^0|^2)\, ds)\right)^{1/2}.
\end{aligned} \tag{57}$$

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 (57) has been shown; then, by standard induction arguments with conditional expectations it follows that both γ^i_T are, indeed, probability densities for *any* $T > 0$. Hence, the calculus leading to (48) and (49) is valid and we have,

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

and

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

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

$$|\tilde{B}[s, X^0_s, \mu^2_s] - \tilde{B}[s, X^0_s, \mu^1_s]| \le \sup_y |\tilde{b}(s, X^0_s, y)| \|\mu^2_s - \mu^1_s\|_{TV}. \tag{58}$$

Due to the linear growth assumption (38), the inequality (58) implies

$$|\tilde{B}[s, X^0_s, \mu^2_s] - \tilde{B}[s, X^0_s, \mu^1_s]| \le C(1 + |X^0_s|)\|\mu^1_s - \mu^2_s\|_{TV}. \tag{59}$$

Hence, by virtue of (58) we obtain

$$\begin{aligned} \mathbb{E}^{\gamma^1}\rho^2 &\le \mathbb{E}^{\gamma^1}\exp\left(6\int_0^T [C(1 + |X^0_s|)\|\mu^1_s - \mu^2_s\|_{TV}]^2 ds\right) \\ &\le \mathbb{E}^{\gamma^1}\exp\left(6C^2 v(T)^2 \int_0^T (1 + |X^0_s|^2\, ds)\right). \end{aligned} \tag{60}$$

Recall that the process X^0 satisfies the equation (44) on $[0, T]$ with respect to the measure $\mathbb{P}^{\gamma^1}$. We want to show that given C, the right hand side in (60) is finite for any T small enough. For this end, denote $6C^2 v(T)^2 := r \ge 0$. We would like to show that for any fixed constant $K > 0$ ($K = 24C^2$ suffices), the value

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

is finite for $0 \le r < K$, 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

$$\psi(r) = \mathbb{E}\exp\left(r\int_0^T (1 + |X^1_s|^2\, ds)\right), \tag{61}$$

where X^1 solves the equation (44) 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

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

(see, for example, [1]). Further, denote

$$\beta(s,x) = \tilde{B}[s,x,\mu_s^1], \tag{62}$$

and

$$\lambda_t := \exp(-\int_0^t \beta(s,X_s^1)\, dW_s - \frac{1}{2}\int_0^t \beta^2(s,X_s^1)\, ds)$$

This λ_t is a probability density for $t \le T$ with any $T > 0$ which is, at least, small enough, since this random variable has the same distribution with respect to $\mathbb{P}$ as γ_t^{-1} with respect to the measure $\mathbb{P}^{\gamma^1}$ on $[0,T]$ and since we know that γ_t^{-1} **is** a probability density of the measure $\mathbb{P}$ with respect to $\mathbb{P}^{\gamma^1}$. Note that with respect to the measure $\mathbb{P}^{\lambda_T}$ the process X^1 solves the equation without drift (39) with a corresponding Wiener process

$$\tilde{W}_t := W_t + \int_0^t \beta(s,X_s^1)\, ds, \quad 0 \le t \le T,$$

due to Girsanov's theorem. Naturally, we have also

$$W_t = \tilde{W}_t - \int_0^t \beta(s,X_s^1)\, ds, \quad 0 \le t \le T.$$

Now let us estimate the function ψ from (61),

$$\psi(r) = \mathbb{E} \exp \left(r \int_0^T (1 + |X_s^1|^2 \, ds) \right) = \mathbb{E}^{\lambda_T} \lambda_T^{-1} \exp \left(r \int_0^T (1 + |X_s^1|^2 \, ds) \right) \tag{63}$$

$$\le \left(\mathbb{E}^{\lambda_T} \exp(2 \int_0^T \beta(s,X_s^1)\, dW_s + \int_0^T \beta^2(s,X_s^1)\, ds) \right)^{1/2} \left(\mathbb{E}^{\lambda_T} \exp \left(2r \int_0^T (1 + |X_s^1|^2 \, ds) \right) \right)^{1/2}$$

$$= \left(\mathbb{E}^{\lambda_T} \exp(2 \int_0^T \beta(s,X_s^1)\, d\tilde{W}_s - \int_0^T \beta^2(s,X_s^1)\, ds) \right)^{1/2} \left(\mathbb{E}^{\lambda_T} \exp \left(2r \int_0^T (1 + |X_s^1|^2 \, ds) \right) \right)^{1/2}.$$

Our local goal is to show that both multipliers in the last line of the last formula are finite. We have for the first multiplier (dropping the square root),

$$
\begin{aligned}
&\mathbb{E}^{\lambda_T} \exp(2 \int_0^T \beta(s, X_s^1)\, d\tilde{W}_s - \int_0^T \beta^2(s, X_s^1)\, ds) \\
&= \mathbb{E} \exp(2 \int_0^T \beta(s, X_s^0)\, dW_s - \int_0^T \beta^2(s, X_s^0)\, ds) \\
&= \mathbb{E} \exp(2 \int_0^T \beta(s, X_s^0)\, dW_s - 4 \int_0^T \beta^2(s, X_s^0)\, ds + 3 \int_0^T \beta^2(s, X_s^0)\, ds) \\
&\le \left(\mathbb{E} \exp(4 \int_0^T \beta(s, X_s^0)\, dW_s - 8 \int_0^T \beta^2(s, X_s^0)\, ds)\right)^{1/2} \left(\mathbb{E} \exp(6 \int_0^T \beta^2(s, X_s^0)\, ds)\right)^{1/2} \\
&\le \left(\mathbb{E} \exp(6 \int_0^T \beta^2(s, X_s^0)\, ds)\right)^{1/2}, \qquad (64)
\end{aligned}
$$

due to the well-known fact that for any adapted integrand β the process $\left(\exp(4 \int_0^t \beta(s, X_s^0)\, dW_s - 8 \int_0^T \beta^2(s, X_s^0)\, ds),\, t \ge 0\right)$ is a supermartingale and expectation $\mathbb{E} \exp(4 \int_0^t \beta(s, X_s^0)\, dW_s - 8 \int_0^T \beta^2(s, X_s^0)\, ds)$ may not exceed one (see, e.g., [19]).

Note that the task to show that the right hand side in (64) is finite is similar to the problem about finiteness of the last expectation in (57) for $T > 0$ small enough. So, we show both simultaneously. The idea is that after some random time change and by using comparison theorems, this task can be reduced to the problem of evaluating the expression

$$
\left(\mathbb{E} \exp(CT(1 + \sup_{0 \le s \le T} |W_s^0|^2))\right)^{1/2}
$$

for a standard Wiener process, which expression can be precisely computed.

5. Random time change. In case of $\underline{d > 1}$, let us apply Ito's formula to $|X_t^0| = \sqrt{\sum_{k=1}^d (X_t^{0,k})^2}$. For simplicity and slightly abusing notations, let us drop the index 0 in the notation for the k-component of the process X_t^0, i.e., write it – only in this small subsection – just as X_t^k instead of the full $X_t^{0,k}$; to the same end of simplicity, let us denote

$$\sigma_t := \sigma(t, X_t^0), \quad a_t := \sigma_t \sigma_t^*.$$

Note that the cases $d > 1$ and $d = 1$ require separate considerations. We have,

$$dX_t^0 = \sigma(t, X_t^0)\, dW_t \equiv \sigma_t\, dW_t;$$

so,

$$dX_t^{0,k} = \sum_j \sigma_t^{kj} dW_t^j.$$

Hence, since for each t we have $\mathbb{P}(X_t^0 = 0) = 0$, we may write,

$$\begin{aligned}
d|X_t^0| = d\sqrt{\sum_{k=1}^d (X_t^{0,k})^2} = \frac{1}{2}\left(\sum_{\ell=1}^d (X_t^{0,\ell})^2\right)^{-1/2} \left(\sum_k 2X_t^{0,k} \sum_j \sigma_t^{kj} dW_t^j\right) \\
+\frac{1}{|X_t^0|}\sum_k \left[1 - \frac{(X_t^{0,k})^2}{|X_t^0|^2}\right] \sum_j (\sigma_t^{kj})^2\, dt - \frac{1}{|X_t^0|}\sum_{k,\ell:\, k\neq\ell} \frac{X_t^{0,k} X_t^{0,\ell}}{|X_t^0|^2} \sum_j \sigma_t^{kj}\sigma_t^{\ell j}\, dt \\
= \sum_j \left(\sum_k \frac{X_t^{0,k}}{|X_t^0|}\sigma_t^{kj}\right) dW_t^j + \frac{1}{|X_t^0|}\left[\mathrm{Tr}\, a_t - \sum_{k,\ell} a_t^{k\ell} \frac{X_t^{0,k} X_t^{0,\ell}}{|X_t^0|^2}\right] dt \\
= \left(\frac{X_t^0}{|X_t^0|}\right)^* \sigma_t\, dW_t + \frac{1}{|X_t^0|}\left[\mathrm{Tr}\, a_t - \left(a_t \frac{X_t^0}{|X_t^0|}, \frac{X_t^0}{|X_t^0|}\right)\right] dt.
\end{aligned}$$

Note that here the "drift"

$$B_t := |X_t^0|^{-1}\left[\mathrm{Tr}\, a_t - \left(a_t \frac{X_t^0}{|X_t^0|}, \frac{X_t^0}{|X_t^0|}\right)\right]$$

is uniformly bounded by some non-random value on the event $(\omega :\ |X_t^0| \geq 1)$ (as well as on $(\omega :\ |X_t^0| \geq c)$ for any positive constant c), say,

$$\sup_\omega \sup_t \sup_{x:\, |x|\geq 1} |x|^{-1}\left[\mathrm{Tr}\, a_t - \left(a_t \frac{x}{|x|}, \frac{x}{|x|}\right)\right] 1(\omega :\ |X_t^0| \geq 1) \leq K,$$

while the "diffusion" $\left(\frac{X_t^0}{|X_t^0|}\right)^* \sigma_t$ is a random vector which is adapted, bounded and non-degenerate, that is, there exists (non-random) $C_0 > 0$ such that

$$C_0^{-1} \leq \left|\left(\frac{X_t^0}{|X_t^0|}\right)^* \sigma_t\right|^2 \leq C_0. \tag{65}$$

Let $\tau(t) := \int_0^t \left| \left(\frac{X^0_s}{|X^0_s|} \right)^* \sigma_s \right|^{-2} ds$ and $\chi(t) := \tau^{-1}(t)$ (the inverse function). Then, the functions τ and χ are well-defined and the process

$$\hat W_t := \int_0^{\chi(t)} \left(\frac{X^0_s}{|X^0_s|} \right)^* \sigma_s \, dW_s$$

is a one-dimensional Wiener process (see [22]). Denote

$$\hat X_t := |X^0_{\chi(t)}|, \quad \hat B_t := B_{\chi(t)}. \tag{66}$$

Both processes $\hat X_t$ and $\hat B_t$ are adapted with respect to the filtration $\mathcal{F}_{\chi(t)}$, which sigma-algebra is also well-defined because each $\chi(t)$ is a stopping time. Then the process $\hat X_t$ has a stochastic differential (see [22])

$$d\hat X_t = d\hat W_t + \hat B_t \chi'(t)\, dt \equiv d\hat W_t + \hat B_t \left| \left(\frac{X^0_t}{|X^0_t|} \right)^* \sigma_t \right|^{-2} dt. \tag{67}$$

Note that there is no local time at zero here: this is because the process which starts outside the origin in dimension $d \geq 2$ does not touch the origin on any finite interval of time.

Now, simultaneously with the process $(|\hat X_t|,\, t \geq 0)$ consider a (unique) solution of the non-sticky reflecting SDE on the half-line $[1, +\infty)$,

$$dZ_t = d\hat W_t + C_1\, dt + d\phi_t, \quad Z_0 \geq |\hat X_0| \vee 1, \tag{68}$$

where ϕ_t is a local time at one, see [22], and $C_1 \geq KC_0$, with C_0 from (65).

The processes (X^0_t), $(|X^0_t|)$, (W_t), $(\hat W^0_t)$, (Z_t) are all defined on the same probability space (recall that solution of the equation (68) is, of course, strong, and, hence, exists on any probability space with a Wiener process). An easy comparison then shows

$$\mathbb{P}(Z_t \geq \hat X_t,\ t \geq 0) = 1. \tag{69}$$

Indeed, $Z_0 \geq \hat X_0$, and Itô's formula applied to $(\hat X_t - Z_t)^2 1(\hat X_t - Z_t \geq 0)$ shows that

$$d(\hat X_t - Z_t)^2 1(\hat X_t - Z_t \geq 0)$$

$$= 2(\hat X_t - Z_t) 1(\hat X_t - Z_t \geq 0)(d\hat W_t + \hat B_t \chi'(t)\, dt - d\hat W_t - C_0\, dt - d\phi_t) \leq 0,$$

which confirms (69).

Also note that due to (65) and (66),

$$\sup_{0\le s\le t}|X_t^0| \le \sup_{0\le s\le C_0 t}\hat X_s. \tag{70}$$

Hence, by virtue of the assumption (38) and of the definition (62), the right hand side in (64) admits a bound,

$$\begin{aligned}\mathbb{E}\exp(6\int_0^T \beta^2(s,X_s^0)\,ds) &\le \mathbb{E}\exp(6C^2\int_0^T(1+|X_s^0|^2\,ds) \\ &\le \mathbb{E}\exp(6C^2T(1+\sup_{0\le s\le C_0T}|Z_s|)^2).\end{aligned} \tag{71}$$

Now the evaluation of (71) can be completed, for example, as follows. Consider an SDE on $\mathbb{R}^1$,

$$dV_t = d\hat W_t + KC_0 \mathrm{sign}(V_s)\,dt, \quad V_0 = Z_0\,(=|x_0|).$$

Here $\mathrm{sign}(a) = 1(a>0) - 1(a<0)$. By Ito's formula for the modulus [22], the process $|V_t|$ satisfies an SDE

$$d|V_t| = d\bar W_t + KC_0\,dt + d\psi_t^0$$

with a new local time ψ_t^0 at zero and a new Wiener process $\bar W_t = \int_0^t \mathrm{sign}(V_s)\,d\hat W_s$ (by Lévy's theorem since $\mathbb{P}(V_t=0)=0$ for each t and so the bracket $\langle \int_0^\cdot \mathrm{sign}(V_s)^2\,ds\rangle_t = t$ a.s.), which has a weakly unique solution. So, its distribution in the space of trajectories coincides with that of the process $(Z_t-1,\, t\ge 0)$. Hence, for any monotonic increasing Borel function g,

$$\mathbb{E}g(\sup_{0\le s\le C_0T}|Z_s^0|) \le \mathbb{E}g(\sup_{0\le s\le C_0T}(|V_s|+1)) \le \mathbb{E}g(1+KC_0T+|X_0|+\sup_{0\le s\le C_0T}|\bar W_s|).$$

Thus, we obtain,

$$\begin{aligned}\mathbb{E}\exp(6C^2T(1+\sup_{0\le s\le C_0T}|Z_s^0|)^2) &\le \mathbb{E}\exp(18C^2T(1+|X_0|^2+\sup_{0\le s\le C_0T}|\bar W_s^0|)^2) \\ &\le \exp(18C^2T)\mathbb{E}\exp(18C^2T|X_0|^2)\mathbb{E}\exp(18C^2T\sup_{0\le s\le C_0T}|\bar W_s^0|)^2),\end{aligned} \tag{72}$$

or, equivalently (since all Wiener processes are equal in distributions),

$$
\begin{aligned}
&\mathbb{E}\exp(6C^2T(1+\sup_{0\le s\le C_0T}|Z_s^0|)^2) \\
&\le \exp(18C^2T)\mathbb{E}\exp(18C^2T|x_0|^2)\mathbb{E}\exp(18C^2T\sup_{0\le s\le C_0T}|W_s^0|)^2).
\end{aligned}
\tag{73}
$$

Let us now complete this analysis by considering the case $\underline{d=1}$ which is a bit easier, although, it involves local time from the very beginning. Let C_0 be a constant such that

$$
C_0^{-1}\le \inf_{t,x}\sigma^2(t,x)\le \sup_{t,x}\sigma^2(t,x)\le C_0.
$$

Denote $\sigma_t := \sigma(t,X_t)$. Let $\tau(t) := \displaystyle\int_0^t |\sigma_s|^{-2}\,ds$ and $\chi(t) := \tau^{-1}(t)$. Then, as in the case $d>1$, the functions τ and χ are well-defined and the process

$$
\hat W_t := \int_0^{\chi(t)} \mathrm{sign}(X_s^0)\sigma_s\,dW_s
$$

is a one-dimensional Wiener process (see [22]). The process $X^0_{\chi(t)}$ has a stochastic differential,

$$
dX^0_{\chi(t)} = d\hat W_t,
$$

i.e., $\hat W_t - x_0$ is a new Wiener process [22]. Denote

$$
\hat X_t := |X^0_{\chi(t)}|. \tag{74}
$$

The process $\hat X_t$ is adapted to the filtration $\mathcal{F}_{\chi(t)}$ and it has a stochastic differential (see [22])

$$
d\hat X_t = d\hat W_t + d\phi_t, \tag{75}
$$

with a local time ϕ_t at zero. Moreover, its distribution in the space of trajectories coincides with that of $|\hat W_t|$. The inequality (70) is valid. Hence, for any monotonic increasing Borel function g,

$$
\mathbb{E}g(\sup_{0\le s\le C_0T}|X_s^0|)\le \mathbb{E}g(\sup_{0\le s\le C_0T}(|\hat W_s|+1)).
$$

Thus, we obtain a bound similar to that in the case $d > 1$:

$$\begin{aligned}\mathbb{E}\exp(6C^2T(1+\sup_{0\le s\le C_0T}|X^0_s|)^2) &\le \mathbb{E}\exp(6C^2T(1+|X_0|+\sup_{0\le s\le C_0T}|\bar W_s|)^2) \\ &\le \exp(18C^2T)\mathbb{E}\exp(18C^2T|x_0|^2)\mathbb{E}\exp(12C^2T\sup_{0\le s\le C_0T}|W_s|)^2),\end{aligned} \tag{76}$$

or, equivalently (since all Wiener processes are equal in distributions),

$$\begin{aligned}&\mathbb{E}\exp(6C^2T(1+\sup_{0\le s\le C_0T}|X^0_s|)^2) \\ &\le \exp(18C^2T)\mathbb{E}\exp(18C^2T|x_0|^2)\mathbb{E}\exp(18C^2T\sup_{0\le s\le C_0T}|W^0_s|)^2).\end{aligned} \tag{77}$$

6. Now, we have,

$$\mathbb{P}(\sup_{0\le s\le T}|W^0_s| > x) \le 4\mathbb{P}(W^0_T > x) \quad (x>0),$$

that is, the density of $\sup_{0\le s\le T}|W^0_s|$ is $f(x) = 2(2\pi T)^{-1/2}\exp(-x^2/(2T)),\ x>0$. Hence, we estimate, with $CT < 2T_1$,

$$\begin{aligned}\mathbb{E}\exp(CT(1+\sup_{0\le s\le T_1}|W^0_s|^2)) &= \int_0^\infty \exp(CT(1+y^2))\frac{2}{\sqrt{2\pi T_1}}\exp(-y^2/(2T_1))\,dy \\ &= \exp(CT)\int_0^\infty \frac{2}{\sqrt{2\pi T}}\exp(-y^2((2T)^{-1}-CT))\,dy \\ &\le \exp(CT)\int_0^\infty \frac{2}{\sqrt{2\pi T_1}}\exp(-y^2((2T_1)^{-1}))\,dy = \exp(CT) < \infty.\end{aligned}$$

For the sequel, note that for any $0 < T_2 < T_1$ and $CT < 2(T_1 - T_2)$, due to the same estimate above we have,

$$\mathbb{E}\exp(CT(1+\sup_{T_2\le s\le T_1}|W^0_s - W_{T_2}|^2)) \le \exp(CT) < \infty \tag{78}$$

As a consequence, the functions in the right hand sides of (73) and (77) are finite for $T > 0$ small enough. Similarly for the second multiplier in the last line of (63),

$$\begin{aligned}&\mathbb{E}^{\lambda_T}_{X_0}\exp\left(2r\int_0^T(1+|X^1_s|^2\,ds)\right) \\ &= \mathbb{E}_{X_0}\exp\left(2r\int_0^T(1+|X^0_s|^2\,ds)\right) \le \exp(CrT + CrT|X_0|^2),\end{aligned}$$

with some $C > 0$, if rT is small enough. Thus, the function ψ (see (61)) 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 (57) 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 (38), too. Hence, we can return to the inequalities (48) earlier established for bounded $\tilde{b}$, and by virtue of (60) 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. Note that since $\mathbb{E}\exp(c_0|x_0|^2) < \infty$ then due to the estimates (73) and (77) and the bound (78) from the previous steps, for *any* t there exists $c > 0$ such that

$$\mathbb{E}\exp(c \sup_{0\le s\le t} |X_s|^2) < \infty.$$

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 (48)) where the right hand side is clearly continuous in t. Moreover, as it follows from (60),

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

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 calculus similar to the one in the previous steps 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.

4 Appendix

Lemma 1 (Skorokhod (on unique probability space and convergence)) *Let ξ^n_t ($t \ge 0$, $n = 0, 1, \ldots$) 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 $\tilde{\xi}^{n'}_t$, $t \ge 0$ and $\tilde{\xi}_t$, $t \ge 0$, 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 $t \ge 0$,

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

See [25, Ch.1, §6].

Lemma 2 (Skorokhod) *Let $f^n : \mathbb{R} \times \Omega \to \mathbb{R}\,(n \ge 0)$ be uniformly bounded random processes on some probability space; let $(W^n\,(n \ge 0))$ be a sequence of (one-dimensional) Wiener processes on the same probability space, and let all Itô's stochastic integrals $\int_0^T f^n_s\, dW^n_s, n \ge 0$ be well-defined. Assume that for any $\varepsilon > 0$,*

$$\lim_{h\to 0} \sup_n \sup_{|s-t|\le h} \mathbb{P}\{|f^n_s - f^n_t| > \varepsilon\} = 0, \tag{79}$$

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

$$(f^n_s, W^n_s) \xrightarrow{\mathbb{P}} (f^0_s, W^0_s).$$

Then

$$\int_0^T f^n_s\, dW^n_s \xrightarrow{\mathbb{P}} \int_0^T f^0_s\, dW^0_s.$$

See [25, Ch.2, §3, Theorem], where W^n are allowed to be more general martingales with brackets converging to that of a Wiener process.

Acknowledgements

For the second author this study has been funded by the Russian Academic Excellence Project '5-100'. Some 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 anonynous referee for useful comments and remarks.

References

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

[2] 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.

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

[4] 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.

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

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

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

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

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

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

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

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

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

[18] N. V. Krylov. *Controlled diffusion processes*, Springer-Verlag, Berlin et al., 2nd edition, 2009.

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

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

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

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

[23] H. Osada, Diffusion processes with generators of generalized divergence form, J. Math. Kyoto Univ., 27 (1987), 597-619.

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

[25] 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.

[26] 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.

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

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

[29] A. Yu. Veretennikov. Parabolic equations and It's stochastic equations with coefficients discontinuous in the time variable. Math. notes, 1982, 31(4), 278-283.

[30] 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.

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

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

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

[34] 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.