Electron. J. Probab. **19** (2014), no. 91, 1–34.
ISSN: 1083-6489 DOI: 10.1214/EJP.v19-3293

Strong completeness for a class of stochastic differential equations with irregular coefficients*

Xin Chen† Xue-Mei Li‡

Abstract

We prove the strong completeness for a class of non-degenerate SDEs, whose coefficients are not necessarily uniformly elliptic nor locally Lipschitz continuous nor bounded. Moreover, for each $p > 0$ there is a positive number $T(p)$ such that for all $t < T(p)$, the solution flow $F_t(\cdot)$ belongs to the Sobolev space $W^{1,p}_{\text{loc}}$. The main tool for this is the approximation of the associated derivative flow equations. As an application a differential formula is also obtained.

Keywords: strong completeness; stochastic differential equation; derivative flow equation; approximation; differential formula.
AMS MSC 2010: 60H10.
Submitted to EJP on February 4, 2014, final version accepted on September 15, 2014.

1 Introduction

Throughout the paper $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space with complete and right continuous filtration $(\mathcal{F}_t)$, and $W_t = \{W_t^1, ..., W_t^m\}$ is an m-dimensional Brownian motion. Let $X : \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^d$ be a Borel measurable map such that for each $x \in \mathbb{R}^d$ the map $X(x, \cdot) : \mathbb{R}^m \to \mathbb{R}^d$ is linear and let $X_0 : \mathbb{R}^d \to \mathbb{R}^d$ be a Borel measurable vector field on $\mathbb{R}^d$. We study the following SDE,

$$dx_t = X(x_t)\, dW_t + X_0(x_t)\, dt. \tag{1.1}$$

Let $X^*(x)$ denote the transpose of $X(x) : \mathbb{R}^m \to \mathbb{R}^d$. We say that the diffusion coefficient X or the SDE (1.1) is uniformly elliptic if there exists a $\delta > 0$ such that $|(X^*X)(x)(\xi)| \geqslant \delta|\xi|$ for every $x, \xi \in \mathbb{R}^d$. It is elliptic if $X(x)$ is a surjection for each x.

*This research is supported by an EPSRC grant (EP/E058124/1) and the Portuguese Science Foundation (FCT) project PTDC/MAT/104173/2008.

†Grupo de Fisica Matematica, Universidade de Lisboa, Av Prof Gama Pinto 2, Lisbon 1649-003, Portugal. Current Address: Department of Mathematics, Shanghai JiaoTong University, 200240, Shanghai, China.
E-mail: chenxin_217@hotmail.com
‡Mathematics Institute, The University of Warwick, Coventry, CV4 7AL, U.K..
E-mail: xue-mei.li@warwick.ac.uk

Fixing an orthonormal basis $\{e_1, ..., e_m\}$ of $\mathbb{R}^m$, for $1 \leqslant k \leqslant m$ and $x \in \mathbb{R}^d$ we define $X_k(x) = X(x)(e_k)$. Then $\{X_0, X_1, \dots, X_m\}$ is a family of Borel measurable vector fields on $\mathbb{R}^d$ and the SDE (1.1) has the following expression,

$$dx_t = \sum_{k=1}^{m} X_k(x_t)\, dW_t^k + X_0(x_t)\, dt. \tag{1.2}$$

Throughout the paper we assume that there is a unique strong solution to (1.2) and we denote by $(F_t(x,\omega), 0 \leqslant t < \zeta(x,\omega))$ the strong solution with a (non-random) initial value $x \in \mathbb{R}^d$ and explosion time $\zeta(x,\omega) > 0$. The differential of X_k at x is denoted by $(DX_k)_x$ or $DX_k(x)$.

The SDE (1.2), or its solution, is said to be complete if the unique strong solution does not explode, i.e. $\zeta(x) = \infty$, $\mathbb{P}$-a.s. for every $x \in \mathbb{R}^d$. The SDE (1.2), or its solution, is said to be strongly complete if it is complete and there is a $\mathbb{P}$-null set Ω_0 such that for every $\omega \notin \Omega_0$, the function $(t,x) \mapsto F_t(x,\omega)$ is jointly continuous on $[0,\infty) \times \mathbb{R}^d$. For further discussion on this, see the books: K. D. Elworthy [7] and H. Kunita [22].

If the SDE is strongly complete, the corresponding stochastic dynamics has the perfect cocycle property, which is often the basic assumption in the study of stochastic dynamical systems. Continuous dependence on the initial data is also an essential assumption for successful numerical simulation of the solutions. It turns out that smoothness and boundedness of the coefficients are not sufficient for the strong completeness. In X.-M. Li and M. Scheutzow [26], a SDE on $\mathbb{R}^2$ of the form $dx_t = \sigma(x_t, y_t)dB_t$, $dy_t = 0$ (here both x_t and y_t are scalar valued process) is constructed with the property that although $\sigma : \mathbb{R}^2 \to \mathbb{R}$ is bounded and C^∞ smooth, the SDE is not strongly complete. See also M. Hairer, M. Hutzenthaler and A. Jentzen [16] on the loss of regularity for Kolmogorov equations.

It is well known, proved by J. N. Blagovescenskii and M. I. Freidlin [1], that the SDE (1.2) is strongly complete if its coefficients are (globally) Lipschitz continuous. Suppose that $\{X_k\}_{k=0}^m$ are C^2 and $\{DX_k\}_{k=0}^m$ are not necessarily bounded, a sufficient condition for the strong completeness of (1.2) is given in X.-M. Li [23]. In particular, the core condition in [23] is on the mild growth rate of $\{|DX_k|\}_{k=0}^m$, and the crucial estimate is on the integrability of the norm of the solution to the derivative flow equation which is controlled by the growth rate at infinity of the vector fields $\{X_k,\ DX_k\}_{k=0}^m$. We would remark that the SDEs studied in [23] are on Riemannian manifolds; specific computations for SDEs on $\mathbb{R}^d$ are given in [23, Section 6]. See also S. Z. Fang, P. Imkeller and T. S. Zhang [10] and X. C. Zhang [35] for different methods to obtain such sufficient conditions.

As mentioned above, a control on the derivatives of the coefficients is useful in estimating the moments of the solution to the derivative flow equation. The latter also appears to be useful for the study of the convergence rates in numerical schemes, see M. Hairer, M. Hutzenthaler and A. Jentzen [16], where they construct some SDEs with smooth bounded coefficients whose solutions fall into one of the following cases: (1) the map $x \mapsto \mathbb{E}(F_t(x))$ is continuous but not locally Hölder continuous; (2) for any $t \geqslant 2$, $C > 0$, $\alpha > 0$, and $h_0 > 0$, there is a step size $h \in (0, h_0)$ with the property that the rate of convergence for the Euler-Maruyama method is slower than Ch^α.

Let us consider the case that the coefficients of SDE (1.2) are not Lipschitz continuous. If X is uniformly elliptic, $\{X_k\}_{k=0}^m$ are bounded, and $X_k \in W^{1,2d}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ for each $k \geqslant 1$, it is established in A. Veretennikov [33] that there is a unique strong solution to (1.2). Letting $m = d$ and $X(x)$ be the identity matrix, in [21], N. V. Krylov and M. Röckner prove that there is a unique global strong solution provided that $X_0 \in L^q([0,T]; L^p(\mathbb{R}^d; \mathbb{R}^d))$ for some $p > 1$, $q > 2$ satisfying the condition $\frac{d}{p} + \frac{2}{q} < 1$. The strong completeness for such SDE is obtained by E. Fedrizzi and F. Flandoli [13]. See

also related works by I. Gyöngy and T. Martinez [15] and A. M. Davie [6]. Similar results hold for the multiplicative noise case: suppose that X is uniformly elliptic and uniformly continuous with $|DX_k| \in L^q([0,T];L^p(\mathbb{R}^d)),\ 1\leqslant k\leqslant m,\ |X_0| \in L^q([0,T];L^p(\mathbb{R}^d))$ for p,q as above, then (1.2) is shown to be strongly complete by X. C. Zhang [34, 36]. If X is uniformly elliptic, $X_0 \in C^{0,\delta}(\mathbb{R}^d;\mathbb{R}^d)$ and $\{X_k\}_{k=1}^m \subseteq C_b^{3,\delta}(\mathbb{R}^d;\mathbb{R}^d)$ for some $0<\delta<1$, it is proved by F. Flandoli, M. Gubinelli and E. Priola [14] that (1.2) is strongly complete and the solution flow $F_t(\cdot,\omega)$ is differentiable with respect to the space variable. For bounded measurable drifts, see also the Ph.D. thesis of X. Chen [4] and a recent paper of S. E. A. Mohammed, T. Nilsen and F. Proske [27] where the the noise is essentially additive and the solution flow of (1.2) is shown to belong to a (weighted) Sobolev space, which generalises the result in N. Bouleau and F. Hirsch [2] where the coefficients are Lipschitz continuous. We also refer to readers to S. Z. Fang and T. S. Zhang [12], S. Z. Fang and D. J. Luo [11], and S. Cox, M. Hutzenthaler and A. Jentzen [5] on the study of strong completeness for a SDE whose coefficients are not (locally) Lipschitz continuous nor elliptic.

In all the results mentioned earlier, concerning with the strong completeness of a SDE whose coefficients are not restricted to the class of (locally) Lipschitz continuous vector fields, some uniform conditions are assumed, such as the uniform continuity condition, or the L^p integrability, or the uniform ellipticity, which are quite different from the mild growth conditions in [10], [23], [35] for the SDEs with locally Lispchitz continuous coefficients. In this paper we are specially interested in SDEs whose coefficients are not locally Lipschitz continuous nor necessarily satisfying some uniform conditions.

Some preliminary results in this paper appeared in our earlier work, [3], we have strengthened the results there by removing the boundedness condition and the uniform ellipticity condition on the diffusion coefficients.

Throughout this paper the components of the vector fields X_k are denoted by $X_k = (X_{k1},\dots,X_{kd})$, $0\leqslant k\leqslant m$. Let $X^*X = (a_{i,j})_{i,j=1}^d$ be the $d\times d$ diffusion matrix with entries $a_{i,j}(x) = \sum_{k=1}^m X_{ki}(x)X_{kj}(x)$.

For $x,\xi\in\mathbb{R}^d$, let

$$H_p(x)\big(\xi,\xi\big) := 2p\big\langle DX_0(x)(\xi),\xi\big\rangle + (2p-1)p\sum_{k=1}^m \big|DX_k(x)(\xi)\big|^2$$

and we define the real valued function

$$K_p(x) := \sup_{|\xi|=1}\big\{H_p(x)\big(\xi,\xi\big)\big\}. \tag{1.3}$$

Assumption 1.1. *(1) There exist positive constants p_1, C_1, such that,*

$$\sum_{i,j=1}^d a_{i,j}(x)\xi_i\xi_j \geqslant \frac{C_1}{1+|x|^{p_1}}|\xi|^2,\quad \forall x\in\mathbb{R}^d,\ \xi=(\xi_1,...,\xi_d)\in\mathbb{R}^d. \tag{1.4}$$

(2) There exist positive constants C_2, p_2, such that for all $0\leqslant k\leqslant m$,

$$|X_k(x)|\leqslant C_2(1+|x|^{p_2}). \tag{1.5}$$

There is a constant $0<\delta\leqslant 1$, such that for every $p>0$,

$$\sup_{|y|\leqslant\delta}\Big(\sum_{k=1}^m p|X_k(x+y)|^2 + \langle x, X_0(x+y)\rangle\Big)\leqslant C(p)(1+|x|^2) \tag{1.6}$$

for some positive constant $C(p)>0$.

(3) *There are constants* $p_3 > 2(d+1)$, $p_4 > d+1$ *such that* $X_k \in W^{1,p_3}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$, $1 \leqslant k \leqslant m$ *and* $X_0 \in W^{1,p_4}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$. *For every* $p > 1$, *there exists a constant* $\kappa(p) > 0$, *such that for every* $R > 0$,

$$\int_{\{|x| \leqslant R\}} e^{\kappa(p) K_p(x)} dx < \infty. \tag{1.7}$$

Here $K_p(x)$ *is defined by (1.3).*

(4) *There exist positive constants* R_1, C_3, p_5, *such that for all* $0 \leqslant k \leqslant m$

$$|DX_k(x)| \leqslant C_3(1 + |x|^{p_5}), \quad \forall\, |x| > R_1. \tag{1.8}$$

For every $p > 0$, *there exists a constant* $C(p) > 0$, *such that,*

$$K_p(x) \leqslant C(p) \log(1 + |x|^2), \quad \forall\, |x| > R_1. \tag{1.9}$$

The main theorem of the paper is as following:

Theorem 1.2. *Under Assumption 1.1 the SDE (1.2) is strongly complete. Furthermore, for every* $p > 0$ *there is a positive constant* $T_1(p)$ *such that for each* $t \in [0, T_1(p)]$, $F_t(\cdot,\omega) \in W^{1,p}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$, $\mathbb{P}$*-a.s..*

Remark 1.3. *By carefully tracking the proof for Theorem 1.2, for every* $p > 0$ *large enough, we have the estimate* $T_1(p) \leqslant \frac{\kappa(p)}{d+1}$, *where* $\kappa(p)$ *is the constant in (1.7) of Assumption 1.1.*

Remark 1.4. *Theorem 1.2 is concerned only with the local regularity property of the solution flow* $F_t(\cdot,\omega)$. *In particular, given any* $\alpha < 1$, *there is a* $t_\alpha > 0$ *such that for* $t < t_\alpha$, $F_t(\cdot,\omega)$ *is locally Hölder continuous of order* α, $\mathbb{P}$*-a.s.. It is not easy to derive global property for the solution flow with only Assumption 1.1. For example for the SDE* $dx_t = dW_t$, *the solution flow has the expression* $F_t(x,\omega) = x + W_t(\omega)$, *obviously* $F_t(\cdot,\omega) \notin \bigcup_{p>1} W^{1,p}(\mathbb{R}^d;\mathbb{R}^d)$, $\mathbb{P}$*-a.s. (we only have* $F_t(\cdot,\omega) \in \bigcap_{p>1} W^{1,p}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$*). In order to obtain the uniform Hölder continuity property, we could usually check the classical uniform difference condition,*

$$\sup_{t\in[0,T_0(p)]} \mathbb{E}\left(\left|F_t(x) - F_t(y)\right|^p\right) \leqslant C(p)|x-y|^p, \ \ \forall\, x, y \in \mathbb{R}^d, \tag{1.10}$$

where $p > d$ *and* $C(p) > 0$ *is a positive constant which may depend on* p.

We will see later, in section 7, that if the following uniform integrability condition (with respect to the initial point)

$$\sup_{x\in\mathbb{R}^d, |v|\leqslant 1} \sup_{t\in[0,T_0(p)]} \mathbb{E}\left(|V_t(x,v)|^p\right) < \infty \tag{1.11}$$

holds, then (1.10) follows. Here $V_t(x,v)$ *denotes the "generalized derivative flow" which will be constructed in section 6. However, under Assumption 1.1 we can only prove the following local integrability condition for the generalized derivative flow: for any compact set* $K \subseteq \mathbb{R}^d$ *and* $p > 0$

$$\sup_{x\in K, |v|\leqslant 1} \sup_{t\in[0,T_0(p)]} \mathbb{E}\left(|V_t(x,v)|^p\right) < \infty,$$

see e.g. Lemma 6.2 below. So it is natural to believe that the local conditions on the coefficients of (1.2) as stated in Assumption 1.1 are not sufficient to obtain the difference condition (1.10). In [23, 24] a formula is given for $\mathbb{E}(|V_t(x,v)|^p)$ *in case of* C^2 *coefficients. See also the integrability condition in [13, 36] and the (global) monotone condition (2.2) below which will ensure (1.10).*

We comment on Assumption 1.1. Condition (1.6) is a technical condition that is used for approximating (1.13) by a family of SDEs with smooth coefficients satisfying

$$p\sum_{k=1}^{m}|X_k(x)|^2+\langle x,X_0(x)\rangle\leqslant C(p)(1+|x|^2). \tag{1.12}$$

A SDE with coefficients satisfying condition (1.12) is complete, see e.g. [23]. The constant $\kappa(p)$ in (1.7) is allowed to decrease with p. In conditions (1.6-1.7), the restrictions on X_0 are only one-sided. In particular condition (1.7) does not imply that $\exp\left(p|DX_k|^2\right)$ is locally integrable. In fact, if $\sup_{|\xi|=1}\langle DX_0(\xi),\xi\rangle$ is negative enough, it compensates the contribution of the norms of the derivatives of the diffusion coefficients to K_p, c.f. Example 2.1 below. The conditions (1) and (3) of Assumption 1.1 imply that there is a unique strong solution to (1.2). Indeed since X is elliptic, $X_k\in W^{1,p_3}_{\text{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ for $1\leqslant k\leqslant d$, and $X_0\in W^{1,p_4}_{\text{loc}}(\mathbb{R}^d;\mathbb{R}^d)$, we may apply [36, Theorem 1.3]. Moreover, under condition (1.6), the SDE (1.2) is complete. Roughly speaking, Assumption 1.1 means that the coefficients are contained in some Sobolev space and satisfy some local integrability condition in a compact set, in particular, the coefficients may not be Lipschitz continuous in this compact set, while outside such compact set, the mild growth rate for the derivatives of the coefficients are needed.

We also comment on the proof of the theorem. In N. V. Krylov and M. Röckner [21] and X. C. Zhang [34, 36], a transformation, first introduced in A. K. Zvonkin [38], are applied to transform (1.2) to a SDE without drift. In order to apply the Zvonkin transformation, global estimates for the solution to the associated parabolic PDE are required. Such estimates are usually obtained under the assumption that the diffusion coefficients are uniformly elliptic and uniformly continuous, see e.g. N. V. Krylov [20]. In Assumption 1.1, we do not assume the diffusion coefficients to be uniformly elliptic or to be uniformly continuous, nor the derivatives satisfy some L^p integrability conditions, no suitable estimates for the corresponding PDE is available. We therefore have to assume the drift coefficients to be more regular than that in the reference mentioned above.

We adapt the philosophy in [23] and study the strongly completeness of (1.2) by investigating the corresponding derivative flow equation. But the methods here are however quite different due to the irregularity of the coefficients. In fact, the derivative flow equation is

$$\begin{cases} dx_t=\sum_{k=1}^m X_k(x_t)dW_t^k+X_0(x_t)dt,\\ dv_t=\sum_{k=1}^m DX_k(x_t)(v_t)dW_t^k+DX_0(x_t)(v_t)dt. \end{cases} \tag{1.13}$$

Here v_t is a $\mathbb{R}^d$-valued process. Since the coefficients $\{X_k\}_{k=0}^m$ are not necessarily locally Lipschitz continuous, at this stage, the derivative flow equation, whose coefficients are not necessarily locally bounded, is only a formal expression. We must establish firstly the pathwise uniqueness and the existence of a strong solution to the derivative flow equation (1.13).

Let $(F_t(x),V_t(x,v))$ be the strong solution to (1.13) with initial point $x_0=x\in\mathbb{R}^d$, $v_0=v\in\mathbb{R}^d$. In case of $\{X_k\}_{k=0}^m$ belonging to $C_b^2(\mathbb{R}^d;\mathbb{R}^d)$, it is well known that $D_xF_t(x)(v)=V_t(x,v)\ \mathbb{P}-a.s.$, see. e.g. H. Kunita [22]. In this paper, we use the approximating Theorem (Theorem 6.5) to establish such a result, c.f. Theorem 1.2. Furthermore letting DX_k and $\tilde{D}X_k$ be two different version of the weak derivative of X_k, we show that

$$\int_0^T|DX_k(x_t)-\tilde{D}X_k(x_t)|^2dt=0,\ \ \mathbb{P}-a.s..$$

It follows that the Itô integral $\int_0^T DX_k(x_t)(v_t)dW_t^k$ is independent of the choice of versions of DX_k.

The remaining part of the paper is organized as following. In section 2, we give an example of a SDE which satisfies Assumption 1.1. In Section 3 we establish a lemma for the approximation of a strong solution to a SDE with pathwise uniqueness property. Section 4 is devoted to an estimation for the distribution of the solution to (1.2). A key step in the proof of the main theorem is presented in Section 5, where we construct an approximating sequence of smooth vector fields $\{X_k^\varepsilon\}_{k=0}^m$, which satisfy the conditions of Assumption 1.1 with the corresponding constants independent of ε. In Section 6 we give uniform estimates on the approximating derivative flow equations. The key convergence result is presented in Theorem 6.5. In section 7 we complete the proof of the main theorem. Finally a differentiation formula is established in Section 8.

Notation. The symbol C denotes a constant that may vary in different places and depend only on dimension d and the constants in Assumption 1.1. If it depends on another parameter, it will be emphasized by an index.

2 An Example

The example below satisfies Assumption 1.1. The vector fields $\{X_k\}_{k=1}^d$ constructed below are not uniformly elliptic if $q_2 < 0$; while $\{X_k\}_{k=1}^d$ are not bounded nor uniformly continuous if $q_2 > 0$.

Example 2.1. *We suppose that q_1, q_3, q_4 are positive numbers and $q_2 \in \mathbb{R}$. For a fixed orthonormal basis $\{e_1, \dots, e_d\}$ of $\mathbb{R}^d$ and $1 \leqslant k \leqslant d$ we define*

$$X_k(x) = \Big((1+|x|^{q_1})g_1(x) + |x|^{q_2}g_2(x)\Big)e_k,$$
$$X_0(x) = \Big(-(1+|x|^{-q_3})g_1(x) - |x|^{q_4}g_2(x)\Big)x,$$

where g_1, g_2 are C^∞ functions on $\mathbb{R}^d$ with the following specifications

$$g_1(x) = \begin{cases} 1, & \text{if } |x| \leqslant 2, \\ \in [0,1], & \text{if } 2 < |x| < 3, \\ 0, & \text{if } |x| \geqslant 3, \end{cases}$$

$$g_2(x) = \begin{cases} 0, & \text{if } |x| \leqslant 1, \\ \in [0,1], & \text{if } 1 < |x| < 2, \\ 1, & \text{if } |x| \geqslant 2. \end{cases}$$

Suppose that the constants q_1, q_2, q_3 and q_4 satisfy the following relations:

$$q_4 + 2 > 2q_2, \quad 1 - \frac{d}{2(d+1)} < q_1 < 1, \quad 2(1-q_1) < q_3 < \frac{d}{d+1}.$$

Then $\{X_k\}_{k=0}^d$ satisfy Assumption 1.1 and the corresponding SDE (1.2) is strongly complete.

We first check the ellipticity condition. If $q_2 \geqslant 0$,

$$\sum_{i,j=1}^d a_{i,j}(x)\xi_i\xi_j \geqslant |\xi|^2, \ \ \forall\, \xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d.$$

If $q_2 < 0$,

$$\sum_{i,j=1}^{d} a_{i,j}(x)\xi_i\xi_j \geqslant \frac{C|\xi|^2}{1+|x|^{-q_2}}, \ \ \forall\, \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d.$$

In both cases (1.4) is true.

It is obvious that (1.5) holds, and for $|x|$ sufficiently large,

$$\begin{aligned}
&\sup_{|y|\leqslant 1}\Big(\sum_{k=1}^{d} p|X_k(x+y)|^2 + \langle x, X_0(x+y)\rangle\Big)\\
&\leqslant C(p)|x|^{2q_2} + \sup_{|y|\leqslant 1}\Big(-|x+y|^{q_4}\langle x,(x+y)\rangle\Big)\\
&\leqslant C(p)|x|^{2q_2} - C(|x|^{q_4}-1)|x|^2 + C\sup_{|y|\leqslant 1}\big(|x|^{q_4+1}|y|\big)\\
&\leqslant -C(p)(1+|x|^{q_4+2}),
\end{aligned}$$

where the last step is due to the assumption $q_4 + 2 > 2q_2$. We have proved (1.6).

We prove below that $X_k \in W^{1,p}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$. Firstly for every $1 \leqslant k \leqslant d$, X_k is smooth on $\mathbb{R}^d\backslash\{0\}$, we only need to consider the domain $\{x \in \mathbb{R}^d; 0 < |x| \leqslant 1\}$. Let $\otimes$ denote the tensor product operator and let $\mathbf{I} : \mathbb{R}^d \to \mathbb{R}^d$ denote identity map. For all $x \in \mathbb{R}^d$ with $0 < |x| \leqslant 1$,

$$\begin{aligned}
DX_k(x) &= q_1|x|^{q_1-2}e_k \otimes x,\\
DX_0(x) &= q_3|x|^{-q_3-2}x\otimes x - (1+|x|^{-q_3})\mathbf{I}.
\end{aligned} \tag{2.1}$$

So for every $x \in \mathbb{R}^d$ with $0 < |x| \leqslant 1$,

$$|DX_k(x)| \leqslant C|x|^{q_1-1}, \ \ |DX_0(x)| \leqslant C|x|^{-q_3}.$$

The condition $q_3 < \frac{d}{d+1}$ and $0 < 1 - q_1 < \frac{d}{2(d+1)}$ ensure that, for $1 \leqslant k \leqslant m$, X_k belongs to $W^{1,p_3}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ and X_0 belongs to $W^{1,p_4}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ for some constants p_3 and p_4 satisfying the following relations

$$2(d+1) < p_3 < \frac{d}{1-q_1}, \qquad d+1 < p_4 < \frac{d}{q_3}.$$

For the local exponential integrability, (1.7), we again only need to consider the domain $\{x \in \mathbb{R}^d; 0 < |x| \leqslant 1\}$. From (2.1) we know that,

$$\sup_{|\xi|=1}\langle DX_0(x)\xi,\xi\rangle \leqslant -(1-q_3)|x|^{-q_3} \quad \forall\, 0 < |x| \leqslant 1.$$

Therefore for $|x|$ small enough,

$$K_p(x) \leqslant C(p)|x|^{-2(1-q_1)} - C|x|^{-q_3} \leqslant -C(p)|x|^{-q_3} \leqslant 0,$$

where we use condition $q_3 > 2(1-q_1)$. Hence (1.7) holds.

If $|x| > 3$,

$$\begin{aligned}
DX_k(x) &= q_2|x|^{q_2-2}e_k\otimes x, \quad 1\leqslant k\leqslant d,\\
DX_0(x) &= -(1+|x|^{q_4})\mathbf{I} - q_4|x|^{q_4-2}x\otimes x.
\end{aligned}$$

(1.8-1.9) of Assumption 1.1 follows from $q_4 + 2 > 2q_2$.

Remark 2.2. *The novelty of Example 2.1 is in the combination of the irregularity and the non-uniform ellipticity of the coefficients. The vector fields in Example 2.1 have otherwise nice properties. For example they satisfy the globally monotone conditions,*

$$\begin{aligned} &2p\langle x-y, X_0(x)-X_0(y)\rangle + (2p-1)p\sum_{k=1}^{m}|X_k(x)-X_k(y)|^2 \\ &\leqslant |x-y|^2\int_0^1 K_p\left(x+s\left(y-x\right)\right)ds \leqslant C(p)|x-y|^2,\ p>0, \end{aligned} \tag{2.2}$$

where $C(p)>0$ is a positive constant, from which it is easy to verify by Itô's formula that

$$\sup_{t\in[0,T]}\mathbb{E}\left(\left|F_t\left(x\right)-F_t\left(y\right)\right|^p\right)\leqslant C(p,T)|x-y|^p,\ \ \forall\, x,y\in\mathbb{R}^d,$$

for every $p>0$ and $T>0$. By Kolmogorov's continuity criterion, the SDE (1.2) satisfying the globally monotone condition is strongly continuous and with its solution flow uniformly Hölder continuous. If the coefficients of (1.2) are C^2 and satisfy (2.2), the solution flow $F_t(\cdot,\omega)$ is C^1 and uniformly continuous, see e.g. [22, 23]. But for Example 2.1, we do not know of a previously existing criterion that leads to the Sobolev differentiability of the solution flow.

3 A convergence Lemma

Let $Y_k^\varepsilon\in C^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $0\leqslant k\leqslant m$, $\varepsilon\in(0,\varepsilon_0)$ be a family of smooth vector fields, where ε_0 is a positive constant. We consider the following SDE

$$dy_t^\varepsilon=\sum_{k=1}^{m}Y_k^\varepsilon(y_t^\varepsilon)\,dW_t^k+Y_0^\varepsilon(y_t^\varepsilon)\,dt. \tag{3.1}$$

Since each Y_k^ε is smooth it is well known that (3.1) has a unique maximal strong solution. Throughout this section we also assume that (3.1) is complete for each $\varepsilon\in(0,\varepsilon_0)$ and we denote by $(\phi_t^\varepsilon(x))$ its strong solution with initial point $x\in\mathbb{R}^d$.

Let $\{Y_k\}_{k=0}^m$ be Borel measurable vector fields on $\mathbb{R}^d$. Now we do not assume any regularity assumption on the vector fields $\{Y_k\}_{k=0}^m$ and then have no information on the existence or the uniqueness of a strong solution to the following SDE

$$dy_t=\sum_{k=1}^{m}Y_k(y_t)\,dW_t^k+Y_0(y_t)\,dt. \tag{3.2}$$

One well known method for the existence of a strong solution is the Watanabe-Yamada method: if there is a weak solution and the pathwise uniqueness holds for SDE (3.2), then there exists a unique strong solution to (3.2), see e.g. [17].

In Lemma 3.2 we prove that under suitable conditions, the solutions of (3.1) converges to the unique strong solution to (3.2). As pointed in N. V. Krylov and A. K. Zvonkin [39], and H. Kaneko and S. Nakao [18], the pathwise uniqueness of (3.2) is crucial for the convergence of the strong solution of (3.1) to that of (3.2) as $\varepsilon\to0$. Lemma 3.2 is applied later for the convergence of the derivative flow equation (1.13). We first need the following lemma on the convergence of stochastic integrals, which is essentially due to A. V. Skorohod [31], see also I. Gyöngy and T. Martinez [15, Lemma 5.2].

Lemma 3.1. *([31]) Let W_t and $\{W_t^{(n)}\}_{n=1}^\infty$ be $\mathbb{R}^m$-valued Brownian motions, let $\xi(t)$ and $\{\xi_n(t)\}_{n=1}^\infty$ be $\mathbb{R}^{m\times d}$-valued stochastic processes such that for all $t\geqslant0$ the following Itô*

integrals are well defined:

$$I_n(t) := \int_0^t \xi_n(s)dW_s^{(n)}, \quad I(t) := \int_0^t \xi(s)dW_s.$$

Suppose that for some $T > 0$, $\lim_{n\to\infty} \sup_{t\in[0,T]} |\xi_n(t)-\xi(t)| = 0$ *and* $\lim_{n\to\infty} \sup_{t\in[0,T]} |W_t^{(n)} - W_t| = 0$ *with convergence in probability. Assume that for some* $\delta > 0$,

$$\sup_n \int_0^T \mathbb{E}\left(|\xi_n(t)|^{2+\delta}\right) dt < \infty. \tag{3.3}$$

Then for every $\kappa > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} |I_n(t) - I(t)| \geqslant \kappa \Big) = 0.$$

Proof. Let $R > 0$. Define $\xi_n^R(t) := \big(\xi_n(t) \wedge R\big) \vee (-R)$, $\xi^R(t) := \big(\xi(t) \wedge R\big) \vee (-R)$ and

$$I_n^R(t) := \int_0^t \xi_n^R(s)dW_s^{(n)}, \quad I^R(t) := \int_0^t \xi^R(s)dW_s,$$

where $a \wedge b := \min(a, b)$, $a \vee b := \max(a, b)$ for every $a, b \in \mathbb{R}$. Since the stochastic proceses $\{(\xi_n^R(t), t \in [0, T]), n \in \mathbb{N}_+\}$ and $\{\xi^R(t), t \in [0, T]\}$ are uniformly bounded and $\xi_n^R(t) \to \xi^R(t)$ in probability as $n \to \infty$, we may apply Lemma 5.2 in I. Gyöngy-T. Martinez [15] to obtain

$$\lim_{n\to\infty} \mathbb{P}\Big(\sup_{t\in[0,T]} |I_n^R(t) - I^R(t)| \geqslant \kappa \Big) = 0.$$

By Burkholder-Davis-Gundy inequality, Chebyshev inequality and Hölder inequality,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in[0,T]} |I_n^R(t) - I_n(t)| \geqslant \kappa \right) \leqslant \frac{1}{\kappa^2} \sup_n \mathbb{E}\left(\sup_{t\in[0,T]} |I_n^R(t) - I_n(t)|^2 \right) \\
&\leqslant \frac{C}{\kappa^2} \sup_n \mathbb{E}\left(\int_0^T |\xi_n(s)|^2 1_{\{|\xi_n(s)|>R\}} ds \right) \\
&\leqslant \frac{1}{\kappa^2 R^\delta} \sup_n \int_0^T \mathbb{E}\left(|\xi_n(s)|^{2+\delta}\right) ds.
\end{aligned} \tag{3.4}$$

By (3.3) the above term converges to zero uniformly for n as $R \to \infty$.

By taking a subsequence if necessary we know $\lim_{n\to\infty} \sup_{t\in[0,T]} |\xi_n(t) - \xi(t)| = 0$, $\mathbb{P}-$ a.s.. Therefore by Fatou lemma and (3.3) we obtain

$$\begin{aligned}
&\int_0^T \mathbb{E}\left(|\xi(s)|^{2+\delta}\right) ds \leqslant \liminf_{n\to\infty} \int_0^T \mathbb{E}\left(|\xi_n(s)|^{2+\delta}\right) ds \\
&\leqslant \sup_n \int_0^T \mathbb{E}\left(|\xi_n(s)|^{2+\delta}\right) ds < \infty.
\end{aligned} \tag{3.5}$$

So based on (3.5) and following the same procedure in (3.4) we have

$$\lim_{R\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]} |I^R(t) - I(t)| \geqslant \kappa \right) = 0.$$

Note that for every $R > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,T]} |I_n(t) - I(t)| \geqslant \kappa \Big) \leqslant \mathbb{P}\Big(\sup_{t\in[0,T]} |I^R(t) - I(t)| \geqslant \kappa \Big) \\
&+ \mathbb{P}\Big(\sup_{t\in[0,T]} |I_n^R(t) - I_n(t)| \geqslant \kappa \Big) + \mathbb{P}\Big(\sup_{t\in[0,T]} |I_n^R(t) - I^R(t)| \geqslant \kappa \Big),
\end{aligned}$$

we first take $n \to \infty$ then take $R \to \infty$ to complete the proof. □

Following the proof in [18, Theorem A] and [15, Theorem 2.2], we can show the following result about the convergence of general SDE (3.1), which is suitable for our application (to the derivative flow equation).

Lemma 3.2. *Fix a $T > 0$, let $\mu^{\varepsilon,x}$ denote the distribution of the process $(\phi^\varepsilon_\cdot(x),\, t \leqslant T)$ on the path space $\mathbf{W} := C([0,T];\mathbb{R}^d)$. Assume that pathwise uniqueness holds for (3.2). We suppose that there exist some $p > 2$ and $q > 1$ such that the following conditions hold for every compact set $K \subseteq \mathbb{R}^d$.*

(1) For all $1 \leqslant k \leqslant m$,

$$\begin{aligned} \sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)} \sup_{x\in K} \int_0^T \mathbb{E}\left(|Y^\varepsilon_k(\phi^{\tilde{\varepsilon}}_t(x))|^p\right) dt < \infty, \\ \sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)} \sup_{x\in K} \int_0^T \mathbb{E}\left(|Y^\varepsilon_0(\phi^{\tilde{\varepsilon}}_t(x))|^q\right) dt < \infty; \end{aligned} \tag{3.6}$$

(2) For all $1 \leqslant k \leqslant m$,

$$\begin{aligned} \limsup_{\varepsilon,\tilde{\varepsilon}\to 0} \sup_{x\in K} \int_0^T \mathbb{E}\left(|Y^\varepsilon_k(\phi^\varepsilon_t(x)) - Y^{\tilde{\varepsilon}}_k(\phi^\varepsilon_t(x))|^p\right) dt = 0, \\ \limsup_{\varepsilon,\tilde{\varepsilon}\to 0} \sup_{x\in K} \int_0^T \mathbb{E}\left(|Y^\varepsilon_0(\phi^\varepsilon_t(x)) - Y^{\tilde{\varepsilon}}_0(\phi^\varepsilon_t(x))|^q\right) dt = 0; \end{aligned} \tag{3.7}$$

(3) Let $\{x_n\}_{n=1}^\infty \subseteq K$ and $\{\varepsilon_n\}_{n=1}^\infty \subseteq (0,\varepsilon_0)$ such that $\lim_{n\to\infty} \varepsilon_n = 0$. If μ^{ε_n,x_n} converges weakly to a limit measure μ^0, then for every $1 \leqslant k \leqslant m$,

$$\begin{aligned} \lim_{\varepsilon\to 0} \int_0^T \int_{\mathbf{W}} |Y^\varepsilon_k(\sigma_t) - Y_k(\sigma_t)|^p \; \mu^0(d\sigma)\, dt = 0, \\ \lim_{\varepsilon\to 0} \int_0^T \int_{\mathbf{W}} |Y^\varepsilon_0(\sigma_t) - Y_0(\sigma_t)|^q \; \mu^0(d\sigma)\, dt = 0. \end{aligned} \tag{3.8}$$

Then for every $x \in \mathbb{R}^d$ there exists a unique complete strong solution $\phi_t(x)$ with initial point $x \in \mathbb{R}^d$, to (3.2). Moreover for every compact set $K \subseteq \mathbb{R}^d$,

$$\lim_{\varepsilon\to 0} \sup_{x\in K} \mathbb{E}\left(\sup_{t\in[0,T]} |\phi^\varepsilon_t(x) - \phi_t(x)|\right) = 0. \tag{3.9}$$

Proof. We suppose that there is a compact set $K_0 \subseteq \mathbb{R}^d$, such that

$$\limsup_{\varepsilon,\tilde{\varepsilon}\to 0} \sup_{x\in K_0} \mathbb{E}\left(\sup_{t\in[0,T]} |\phi^\varepsilon_t(x) - \phi^{\tilde{\varepsilon}}_t(x)|\right) > 0, \tag{3.10}$$

then there exist $\kappa > 0$, $\{x_n\}_{n=1}^\infty \subseteq K_0$, and two sequences $\{\varepsilon_{n,1}\}_{n=1}^\infty$, $\{\varepsilon_{n,2}\}_{n=1}^\infty$ contained in $(0,\varepsilon_0)$ such that

$$\begin{aligned} &\lim_{n\to\infty} \varepsilon_{n,i} = 0, \quad i = 1,2, \\ &\lim_{n\to\infty} \mathbb{E}\left(\sup_{t\in[0,T]} |\phi^{\varepsilon_{n,1}}_t(x_n) - \phi^{\varepsilon_{n,2}}_t(x_n)|\right) > \kappa. \end{aligned} \tag{3.11}$$

Let $z^n_\cdot = \left(\phi^{\varepsilon_{n,1}}_\cdot(x_n), \phi^{\varepsilon_{n,2}}_\cdot(x_n), W_\cdot\right)$ and let ν^n be the distribution of $z^n_\cdot$ on the path space $C([0,T];\mathbb{R}^{2d+m})$.

Note that z_t^n is a semi-martingale, we will apply [37, Theorem 3] or [30] to show that the family of probability measures $\{\nu^n\}_{n=1}^{\infty}$ on $C([0,T];\mathbb{R}^{2d+m})$ is tight. In particular, as in [37, Theorem 3] or [30], it suffices to verify the uniformly bounded property for the variational processes and the drift processes of the semi-martingales $\{z_\cdot^n\}_{n=1}^{\infty}$.

Note that $\phi_t^{\varepsilon_{n,i}}(x_n) = x_n + M_t^{n,i} + A_t^{n,i}$, $i = 1, 2$, where

$$M_t^{n,i} := \sum_{k=1}^{m} \int_0^t Y_k^{\varepsilon_{n,i}}\big(\phi_s^{\varepsilon_{n,i}}(x_n)\big) dW_s^k, \quad A_t^{n,i} := \int_0^t Y_0^{\varepsilon_{n,i}}\big(\phi_s^{\varepsilon_{n,i}}(x_n)\big) ds.$$

Let $\langle M^{n,i}\rangle_t$ be the variational process for $M^{n,i}$. We define

$$u_t^{n,i} := \sum_{k=1}^{m} |Y_k^{\varepsilon_{n,i}}\big(\phi_t^{\varepsilon_{n,i}}(x_n)\big)|^2, \quad a_t^{n,i} := Y_0^{\varepsilon_{n,i}}\big(\phi_t^{\varepsilon_{n,i}}(x_n)\big).$$

Hence

$$\langle M^{n,i}\rangle_t = \int_0^t u_s^{n,i} ds, \quad A_t^{n,i} = \int_0^t a_s^{n,i} ds.$$

From (3.6) we know for $p' := \min\{\frac{p}{2}, q\} > 1$,

$$\sup_n \mathbb{E}\Big(\int_0^T |u_t^{n,i}|^{p'} dt\Big) < \infty, \ \sup_n \mathbb{E}\Big(\int_0^T |a_t^{n,i}|^{p'} dt\Big) < \infty, \ i = 1, 2,$$

which implies that the following random variables

$$\Big\{x_n, \int_0^T |u_t^{n,i}|^{p'} dt, \int_0^T |a_t^{n,i}|^{p'} dt, \ n \in \mathbb{N}_+, \ i = 1, 2\Big\}$$

are uniformly bounded in probability. Therefore according to [37, Theorem 3], $\{\nu^n\}_{n=1}^{\infty}$ is tight.

By the Skorohod theorem, see e.g. Theorem 2.7 of Chapter 1 in [17], we can find a subsequence of $\{z_\cdot^n\}_{n=1}^{\infty}$ which will also be denoted by $\{z_\cdot^n\}_{n=1}^{\infty}$ for simplicity, and there exists a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ on which there is a sequence of $\mathbb{R}^{2d+m}$-valued stochastic processes $\tilde{z}_\cdot^n := \big(\tilde{y}_\cdot^{n,1}, \tilde{y}_\cdot^{n,2}, \tilde{W}_\cdot^n\big)$ with the property that $\tilde{z}_\cdot^n$ has the same distribution with $z_\cdot^n$, and

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \big|\tilde{z}_t^n - \tilde{z}_t\big| = 0, \quad \tilde{\mathbb{P}} - a.s. \tag{3.12}$$

for some $\mathbb{R}^{2d+m}$-valued process $\tilde{z}_\cdot = \big(\tilde{y}_\cdot^1, \tilde{y}_\cdot^2, \tilde{W}_\cdot\big)$.

Condition (3.6) implies that $\{\sup_{t\in[0,T]} |\tilde{z}_t^n|\}_{n=1}^{\infty}$ is uniformly integrable which follows from a round of BDG inequality and Hölder inequality, therefore we have

$$\lim_{n\to\infty} \tilde{\mathbb{E}}\left(\sup_{t\in[0,T]} \big|\tilde{z}_t^n - \tilde{z}_t\big|\right) = 0.$$

By (3.11) we also obtain that

$$\tilde{\mathbb{E}}\left(\sup_{t\in[0,T]} \big|\tilde{y}_t^1 - \tilde{y}_t^2\big|\right) > \kappa. \tag{3.13}$$

Since $\tilde{z}_\cdot^n \overset{law}{=} z_\cdot^n$, for every $0 \leqslant s < t \leqslant T$, $\tilde{W}_t^n - \tilde{W}_s^n$ is independent of the σ-algebra $\mathcal{G}_s^n := \sigma\{\tilde{z}_r^n; \ 0 \leqslant r \leqslant s\}$. Hence for every $j \in \mathbb{N}_+$ and $f \in C_b(\mathbb{R}^{(2d+m)j})$, $g \in C_b(\mathbb{R}^m)$, $0 < s_1 < s_2 < \cdots < s_j < s < t \leqslant T$,

$$\tilde{\mathbb{E}}\left(g(\tilde{W}_t^n - \tilde{W}_s^n) f(\tilde{z}_{s_1}^n, \ldots, \tilde{z}_{s_j}^n)\right) = \tilde{\mathbb{E}}\left(g(\tilde{W}_t^n - \tilde{W}_s^n)\right) \tilde{\mathbb{E}}\left(f(\tilde{z}_{s_1}^n, \ldots, \tilde{z}_{s_j}^n)\right).$$

Set $\mathcal{G}_s = \sigma\{\tilde{z}_r;\ 0 \leqslant r \leqslant s\}$. Taking $n \to \infty$ in the above identity and using (3.12) we obtain that

$$\tilde{\mathbb{E}}\left(g(\tilde{W}_t - \tilde{W}_s) f(\tilde{z}_{s_1}, \dots, \tilde{z}_{s_j})\right) = \tilde{\mathbb{E}}\left(g(\tilde{W}_t - \tilde{W}_s)\right) \tilde{\mathbb{E}}\left(f(\tilde{z}_{s_1}, \dots, \tilde{z}_{s_j})\right),$$

which implies that $\tilde{W}_t - \tilde{W}_s$ is independent of the σ-algebra $\mathcal{G}_s$. Since $\tilde{W}_\cdot$ is the limit of the family of Brownian motions $\tilde{W}^n_\cdot$, it has the same finite dimensional distribution as $W_\cdot$, therefore $\tilde{W}_\cdot$ is a Brownian motion with respect to the filtration $(\mathcal{G}_s, 0 \leqslant s \leqslant T)$.

In the computation below we will drop the index 1, so $\tilde{y}^{n,1}_t$, $\tilde{y}^1_t$, $\varepsilon_{n,1}$ will be denoted by $\tilde{y}^n_t$, $\tilde{y}_t$ and ε_n respectively. We use again the fact that $\tilde{z}^n_\cdot \overset{law}{=} z^n_\cdot$ to observe that $(\tilde{y}^n_t, \tilde{W}^n_t)$ is a strong solution to SDE (3.1) with $\varepsilon = \varepsilon_n$, i.e.

$$\tilde{y}^n_t = x_n + \sum_{k=1}^m \int_0^t Y_k^{\varepsilon_n}(\tilde{y}^n_s)\, d\tilde{W}^{n,k}_s + \int_0^t Y_0^{\varepsilon_n}(\tilde{y}^n_s)\, ds, \tag{3.14}$$

where $\tilde{W}^n_t = (\tilde{W}^{n,1}_t, \dots, \tilde{W}^{n,m}_t)$ denotes the components of $\tilde{W}^n_t$. Next we will take the limit $n \to \infty$ in (3.14) to prove that $(\tilde{y}_t, \tilde{W}_t)$ is a strong solution to (3.2).

For a fixed $n_0 \in \mathbb{N}_+$, we define

$$\begin{aligned}
I_1^{n,n_0}(t) &= \sum_{k=1}^m \int_0^t \left(Y_k^{\varepsilon_n}(\tilde{y}^n_s) - Y_k^{\varepsilon_{n_0}}(\tilde{y}^n_s)\right) d\tilde{W}^{n,k}_s \\
I_2^{n,n_0}(t) &= \sum_{k=1}^m \left(\int_0^t Y_k^{\varepsilon_{n_0}}(\tilde{y}^n_s)\, d\tilde{W}^{n,k}_s - \int_0^t Y_k^{\varepsilon_{n_0}}(\tilde{y}_s)\, d\tilde{W}^k_s\right), \\
I_3^{n_0}(t) &= \sum_{k=1}^m \int_0^t \left(Y_k^{\varepsilon_{n_0}}(\tilde{y}_s) - Y_k(\tilde{y}_s)\right)\, d\tilde{W}^k_s.
\end{aligned}$$

We use condition (3.7), BDG inequality and Hölder inequality to obtain the following estimate for I_1^{n,n_0},

$$\begin{aligned}
&\limsup_{n_0\to\infty} \limsup_{n\to\infty} \tilde{\mathbb{E}}\left(\sup_{t\in[0,T]} |I_1^{n,n_0}(t)|^p\right) \\
&\leqslant C(p) \sum_{k=1}^m \limsup_{n_0\to\infty} \limsup_{n\to\infty} \tilde{\mathbb{E}}\left(\int_0^T |Y_k^{\varepsilon_n}(\tilde{y}^n_t) - Y_k^{\varepsilon_{n_0}}(\tilde{y}^n_t)|^2 dt\right)^{\frac{p}{2}} \\
&\leqslant C(p,T) \sum_{k=1}^m \limsup_{\varepsilon,\tilde{\varepsilon}\to 0} \sup_{x\in K_0} \int_0^T \mathbb{E}\big(|Y_k^{\varepsilon}(\phi^{\varepsilon}_t(x)) - Y_k^{\tilde{\varepsilon}}(\phi^{\varepsilon}_t(x))|^p\big)\, dt = 0.
\end{aligned} \tag{3.15}$$

Now we work on the second integral. Since $Y_k^{\varepsilon_{n_0}} \in C^\infty(\mathbb{R}^d; \mathbb{R}^d)$, by (3.12), we know for every fixed n_0,

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |Y_k^{\varepsilon_{n_0}}(\tilde{y}^n_t) - Y_k^{\varepsilon_{n_0}}(\tilde{y}_t)| = 0, \ \ \tilde{\mathbb{P}} - a.s..$$

Due to condition (3.6), we may apply the convergence Lemma 3.1 for stochastic integrals and conclude that for every fixed n_0, $\sup_{t\in[0,T]} |I_2^{n,n_0}(t)|$ converges to 0 in probability as $n \to \infty$. In an analogous way to (3.15), by condition (3.6), we can show that $\{\sup_{t\in[0,T]} |I_2^{n,n_0}(t)|^2\}_{n=1}^\infty$ is uniformly integrable, therefore for every fixed n_0,

$$\limsup_{n\to\infty} \tilde{\mathbb{E}}\left(\sup_{t\in[0,T]} |I_2^{n,n_0}(t)|^2\right) = 0.$$

From (3.12) the distribution μ^0 of $\tilde{y}_\cdot$ is a weak limit of μ^{ε_n,x_n}, therefore the condition (3.8) can be applied to the third integral and we have

$$\limsup_{n_0\to\infty}\tilde{\mathbb{E}}\left(\sup_{t\in[0,T]}|I_3^{n_0}(t)|^2\right)=0.$$

Combing all the estimates above for I_1^{n,n_0}, I_2^{n,n_0}, $I_3^{n_0}$, we first take $n\to\infty$ then take $n_0\to\infty$ to obtain

$$\lim_{n\to\infty}\sum_{k=1}^m\tilde{\mathbb{E}}\left(\sup_{t\in[0,T]}\left|\int_0^t Y_k^{\varepsilon_n}(\tilde{y}_s^n)\,d\tilde{W}_s^{n,k}-\int_0^t Y_k(\tilde{y}_s)\,d\tilde{W}_s^k\right|^2\right)=0.$$

By the same method we also prove that

$$\lim_{n\to\infty}\tilde{\mathbb{E}}\left(\sup_{t\in[0,T]}\left|\int_0^t Y_0^{\varepsilon_n}(\tilde{y}_s^n)ds-\int_0^t Y_0(\tilde{y}_s)ds\right|\right)=0.$$

Finally we take $n\to\infty$ in (3.14) to see that

$$\tilde{y}_t=x_0+\sum_{k=1}^m\int_0^t Y_k(\tilde{y}_s)d\tilde{W}_s^k+\int_0^t Y_0(\tilde{y}_s)ds.$$

The above argument applies equally to $\tilde{y}_t^{n,2}$ and we prove that both $(\tilde{y}_t^1,\tilde{W}_t)$ and $(\tilde{y}_t^2,\tilde{W}_t)$ are $\mathcal{G}_t$ adapted strong solution to (3.2) with initial value x_0. Consequently by the pathwise uniqueness for (3.2), for every $t\in[0,T]$, $\tilde{y}_t^1=\tilde{y}_t^2$, $\tilde{\mathbb{P}}-a.s.$, and

$$\tilde{\mathbb{E}}\left(\sup_{t\in[0,T]}|\tilde{y}_t^1-\tilde{y}_t^2|\right)=0,$$

which contradicts with (3.13). So the assumption (3.10) is not true, the sequence $\sup_{t\in[0,T]}|\phi_t^\varepsilon(x)\ -\phi_t^{\tilde{\varepsilon}}(x)|$ must be a Cauchy sequence as $\varepsilon,\tilde{\varepsilon}\to 0$, and there exists a stochastic process $\phi_\cdot(x)$, such that the convergence in (3.9) holds. By the same approximation argument above, $(\phi_\cdot(x),W_\cdot)$ is the unique complete strong solution to (3.2) with initial point x. □

4 An estimate for the probability distribution

Let $\mathcal{L}=\frac{1}{2}\sum_{i,j=1}^d a_{i,j}\ \frac{\partial^2}{\partial x_i\partial x_j}+\sum_{i=1}^d X_{0i}\frac{\partial}{\partial x_i}$. If $A(x)$ is strictly elliptic, $\{X_k\}_{k=0}^m$ are bounded and uniformly Hölder continuous, there is a Gaussian type upper and lower bound for the fundamental solution to the parabolic PDE $\frac{\partial u_t}{\partial t}=\mathcal{L}u_t$. Such estimates are used in our earlier work [3], an unpublished notes. But under Assumption 1.1, we are not sure whether such estimate is true, so we will apply Lemma 4.4 instead.

We first cite a lemma on the distributions of continuous semi-martingales, which is a special case of that in N. V. Krylov [19, Lemma 5.1], see also I. Gyöngy and T. Martinez [15, Lemma 3.1]. Let $\det(A)$ and $\mathrm{tr}(A)$ denote respectively the determinant and the trace of a $d\times d$ matrix A.

Lemma 4.1. *([19]) Suppose that $F_t(x)$ is a strong solution to (1.2) with initial point $x\in\mathbb{R}^d$, set $\tilde{F}_t(x):=F_t(x)-x$. For every $q\geqslant d+1$, $T>0$, $R>0$ and Borel measurable function $f:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}_+$, letting $\tau_R(x):=\inf\{t\geqslant 0,\ |\tilde{F}_t(x)|>R\}$, we have*

$$\begin{aligned}&\mathbb{E}\left(\int_0^{T\wedge\tau_R(x)}f(t,\tilde{F}_t(x))\big(\det A(F_t(x))\big)^{\frac{1}{q}}\,dt\right)\\&\leqslant C(d)\,e^T\left(\mathbf{A}(R)+\mathbf{B}(R)^2\right)^{\frac{d}{2q}}\left(\int_0^T\int_{|x|\leqslant R}f^q(t,x)dx\,dt\right)^{\frac{1}{q}},\end{aligned}\tag{4.1}$$

where $C(d)$ is a constant depending only on d and

$$\begin{aligned}\mathbf{A}(R) &= \mathbb{E}\left(\int_0^{T\wedge\tau_R(x)} \mathrm{tr}A(F_t(x))dt\right),\\ \mathbf{B}(R) &= \mathbb{E}\left(\int_0^{T\wedge\tau_R(x)} |X_0(F_t(x))|dt\right).\end{aligned} \tag{4.2}$$

Proof. In [15, Lemma 3.1], we take $X(t) = \tilde{F}_t(x)$, $A(t) = t$, $d\,m(t) = X(F_t(x))\,dW_t$, $dB(t) = X_0(F_t(x))dt$, $\gamma = T$, $r(t) = 1$, $c(t) = 1_{[0,T]}(t)$, $p = q - 1$, and the conclusion follows. □

We also cite the following lemma, [23, Lemma 6.1], which is concerned with the moment estimates for (1.2), the regularity condition imposed on $\{X_k\}$ in [23] will not be needed.

Lemma 4.2. *([23]) Suppose that $\{X_k\}_{k=0}^{\infty}$ are locally bounded vector fields. Let $g : \mathbb{R}^d \to \mathbb{R}_+$ be a positive C^2 function. For any $\lambda > 0$ let*

$$\Theta_g(\lambda) = \sup_{x\in\mathbb{R}^d}\Big\{\big(Dg(X_0)\big)(x) + \frac{1}{2}\sum_{k=1}^{m}\big(\lambda|Dg(X_k)|^2 + D^2g(X_k,X_k)\big)(x)\Big\}. \tag{4.3}$$

If furthermore $\Theta_g(\lambda) < \infty$, then for every $t > 0$ and stopping time $\tau < \zeta(x)$, we have

$$\mathbb{E}(\mathrm{e}^{\lambda g(F_{t\wedge\tau}(x))}) \leqslant \mathrm{e}^{\lambda(g(x)+\Theta_g(\lambda)t)},$$

where $F_t(x)$ is a strong solution to (1.2) with initial point $x \in \mathbb{R}^d$, and $\zeta(x)$ is the explosion time of $F_t(x)$.

Proof. The conclusion is just that of [23, Lemma 6.1]. In [23, Lemma 6.1] the coefficients are assumed to be C^1, by carefully tracking the proof, we observe that the regularity condition in [23, Lemma 6.1] is not needed.

In fact, it suffices to show the case where $\lambda = 1$. Since $F_t(x)$ is a strong solution to (1.2), by definition it is also a semi-martingale. By Itô formula, we have for each $t > 0$ and stopping time $\tau < \zeta(x)$,

$$g(F_{t\wedge\tau}(x)) = g(x) + N_{t\wedge\tau} - \frac{\langle N\rangle_{t\wedge\tau}}{2} + b_{t\wedge\tau},$$

where

$$\begin{aligned}N_t &= \int_0^t Dg(F_s(x))(X(F_s(x)))dW_s,\\ b_t &= \int_0^t \Big(\frac{1}{2}\sum_{k=1}^{m}|Dg(F_s(x))(X_k(F_s(x)))|^2 + Dg(F_s(x))(X_0(F_s(x)))\Big)ds\\ &\quad + \frac{1}{2}\sum_{k=1}^{m}\int_0^t D^2g(F_s(x))(X_k(F_s(x)),X_k(F_s(x)))ds,\end{aligned}$$

and $\langle N\rangle_t$ denotes the variational process of N_t. By the definition of $\Theta_g(1)$, $b_t \leqslant t\Theta_g(1)$ and we have,

$$\exp\left(g(F_{t\wedge\tau_R\wedge\tau}(x))\right) \leqslant \exp\left(g(x) + \Theta_g(1)t\right)\exp\left(N_{t\wedge\tau_R\wedge\tau} - \frac{1}{2}\langle N\rangle_{t\wedge\tau_R\wedge\tau}\right),$$

where $\tau_R := \inf\{t\geqslant 0;\ |F_t(x)-x| > R\}$. Since $\{X_k\}_{k=0}^{\infty}$ are locally bounded, $\exp\left(N_{t\wedge\tau_R\wedge\tau} - \frac{1}{2}\langle N\rangle_{t\wedge\tau_R\wedge\tau}\right)$ is a martingale for each $R > 0$, we take expectations of both sides of the inequality above and let $R \to \infty$, then the required conclusion follows from Fatou's lemma. □

Example 4.3. *Suppose that $\{X_k\}_{k=0}^{\infty}$ are locally bounded vector fields. Assume that for every $p>0$ there is $C(p)>0$ such that,*

$$\sum_{k=1}^{m} p|X_k(x)|^2 + \langle x, X_0(x)\rangle \leqslant C(p)(1+|x|^2). \tag{4.4}$$

We apply Lemma 4.2 to $g(x)=\log(1+|x|^2)$. Since

$$\Theta_g(\lambda) \leqslant C(\lambda) \sup_{x\in\mathbb{R}^d} \frac{1}{1+|x|^2}\left(p(\lambda)\sum_{k=1}^{m}|X_k(x)|^2 + \langle x, X_0(x)\rangle\right) < \infty,$$

we have for every $p>0$, $R>0$,

$$\mathbb{E}\big(|F_{t\wedge\tau_R}(x)|^p\big) \leqslant C(p)\mathrm{e}^{C(p)t}(|x|^p+1)$$

for some constant $C(p)>0$ independent of R. Therefore let $R\to\infty$, we obtain,

$$\mathbb{E}\big(|F_t(x)|^p\big) \leqslant C(p)\mathrm{e}^{C(p)t}(|x|^p+1). \tag{4.5}$$

In particular, SDE (1.2) is complete if (4.4) holds.

Lemma 4.4. *Let $F_t(x)$ be a strong solution to (1.2) with initial value $x\in\mathbb{R}^d$. Suppose that the conditions (1.4), (1.5) and (4.4) hold. Then for every $p>d+1$, $T>0$ and non-negative measurable function $f:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}_+$, we have*

$$\mathbb{E}\left(\int_0^T f(t,F_t(x))\,dt\right) \leqslant Q_1(T)Q_2(x)\left(\int_0^T\int_{\mathbb{R}^d} f^p(t,y)dydt\right)^{\frac{1}{p}}, \tag{4.6}$$

where $Q_1:\mathbb{R}_+\to\mathbb{R}_+$, $Q_2:\mathbb{R}^d\to\mathbb{R}_+$ are positive Borel measurable functions which only depend on d, p and the constants in (1.4), (1.5) and (4.4), such that $\sup_{T\in[0,\tilde{T}_0]} Q_1(T)<\infty$ and $\sup_{x\in K} Q_2(x)<\infty$ for every $\tilde{T}_0>0$ and compact set $K\subseteq\mathbb{R}^d$.

Proof. Let $\tilde{f}(t,y):=f(t,y+x),\ y\in\mathbb{R}^d$, $\tilde{F}_t(x):=F_t(x)-x$ and $\alpha:=\frac{p}{d+1}$. Note that from Example 4.3, we know the solution $F_t(x,\omega)$ exists for all time $\mathbb{P}$-a.s.. We apply Lemma 4.1 with $q=d+1$ and Hölder's inequality with exponent $\alpha>1$ and let $R\to\infty$ in (4.1). By Fatou's lemma,

$$\begin{aligned}
&\mathbb{E}\left(\int_0^T f(t,F_t(x))dt\right) = \mathbb{E}\left(\int_0^T \tilde{f}(t,\tilde{F}_t(x))dt\right)\\
&\leqslant \left(\mathbb{E}\left(\int_0^T (\det A(F_t(x)))^{\frac{1}{d+1}}\,\tilde{f}^{\alpha}(t,\tilde{F}_t(x))dt\right)\right)^{\frac{1}{\alpha}}\\
&\cdot\left(\mathbb{E}\left(\int_0^T \big(\det A(F_t(x))\big)^{-\frac{1}{(d+1)(\alpha-1)}}dt\right)\right)^{\frac{\alpha-1}{\alpha}}\\
&\leqslant (C(d)e^T)^{\frac{1}{\alpha}}\sup_{R>0}\big(\mathbf{A}(R)+\mathbf{B}(R)^2\big)^{\frac{d}{2(d+1)\alpha}}\left(\int_0^T\int_{\mathbb{R}^d}|f|^p(t,y)dydt\right)^{\frac{1}{p}}\\
&\cdot\left(\mathbb{E}\left(\int_0^T \big(\det A(F_t(x))\big)^{-\frac{1}{(d+1)(\alpha-1)}}dt\right)\right)^{\frac{\alpha-1}{\alpha}},
\end{aligned} \tag{4.7}$$

where we use the translation invariant property for the Lebesgue integral, i.e. $\int_{\mathbb{R}^d}|\tilde{f}(t,y)|^p dy=\int_{\mathbb{R}^d}|f(t,y)|^p dy$, and the constant $\mathbf{A}(R)$, $\mathbf{B}(R)$ are defined by (4.2).

Since (4.4) holds, by Example 4.3 we know that the moment estimate (4.5) is true. From (1.5) we have the following estimate,

$$\begin{aligned}&\sup_{R>0}\mathbf{B}(R)\leqslant\mathbb{E}\left(\int_0^T|X_0(F_t(x))|dt\right)\\&\leqslant C\mathbb{E}\left(\int_0^T(1+|F_t(x)|^{p_2})dt\right)\leqslant C\mathrm{e}^{CT}T(1+|x|^{p_2}).\end{aligned}\tag{4.8}$$

For $\mathrm{tr}(A)=\mathrm{tr}(X^*X)$, we apply again (1.5) and (4.5) to obtain

$$\sup_{R>0}\mathbf{A}(R)\leqslant\mathbb{E}\left(\int_0^T\mathrm{tr}A(F_t(x))dt\right)\leqslant C\mathrm{e}^{CT}T(1+|x|^{2p_2}).\tag{4.9}$$

Similarly, by the ellipticity condition (1.4), $\det(A(x))^{-\frac{1}{(d+1)(\alpha-1)}}\leqslant C(1+|x|^{\frac{dp_1}{(d+1)(\alpha-1)}})$, and we have,

$$\mathbb{E}\left(\int_0^T\big(\det A(F_t(x))\big)^{-\frac{1}{(d+1)(\alpha-1)}}dt\right)\leqslant C\mathrm{e}^{CT}T(1+|x|^{\frac{dp_1}{(d+1)(\alpha-1)}}).\tag{4.10}$$

In particular, it is easy to check that all the constants C above only depend on the constants in (1.4), (1.5) and (4.4).

Putting the estimates (4.8)-(4.10) into (4.7), we can show (4.6) with

$$Q_1(T)=\mathrm{e}^{C(1+T)}T^{\frac{\alpha-1}{\alpha}}(T+T^2)^{\frac{d}{2(d+1)\alpha}},\ Q_2(x)=1+|x|^{\frac{d(p_1+p_2)}{(d+1)\alpha}}.$$

□

5 Construction of the approximation vector fields

We will construct a class of approximation SDEs with smooth and elliptic coefficients for (1.13). Let $\eta:\mathbb{R}^d\to\mathbb{R}_+$ be the smooth mollifier defined by $\eta(x)=C\mathrm{e}^{\frac{1}{|x|^2-1}}\mathbf{1}_{\{|x|<1\}}$, where C is a normalizing constant such that $\int_{\mathbb{R}^d}\eta(x)dx=1$. For every $\varepsilon>0$, set $\eta_\varepsilon(x):=\varepsilon^{-d}\eta(\frac{x}{\varepsilon})$. For $f\in L^1_{\mathrm{loc}}(\mathbb{R}^d)$, we let $f*\eta_\varepsilon$ denote the convolution of f with η_ε,

$$f*\eta_\varepsilon(x):=\int_{\mathbb{R}^d}\eta_\varepsilon(x-y)f(y)dy=\int_{|y-x|\leqslant\varepsilon}\eta_\varepsilon(x-y)f(y)dy,\quad x\in\mathbb{R}^d.$$

It is natural to approximate each X_k by C^∞ smooth vector field $X_k*\eta_\varepsilon$. However, since we do not make the assumption that X_k are bounded, the approximating systems $\{X_k*\eta_\varepsilon\}_{k=1}^m$ may loose ellipticity if ε is small enough.

Suppose that Assumption 1.1 holds, in particular, the condition that $X_k\in W^{1,p_3}_{loc}(\mathbb{R}^d;\mathbb{R}^d)$, $1\leqslant k\leqslant m$ and $X_0\in W^{1,p_4}_{loc}(\mathbb{R}^d;\mathbb{R}^d)$ for some constants $p_3>2(d+1)$, $p_4>d+1$ ensures that X_k, $0\leqslant k\leqslant m$ are continuous. Let

$$B_R=\{x\in\mathbb{R}^d;\ |x|<R\},\qquad S_R=\{x\in\mathbb{R}^d;\ |x|=R\}.$$

Let $\pi_R:\mathbb{R}^d\to S_R$ be the shortest distance projection so $\pi_R(z):=(R,\theta(z))$ for every $z=(|z|,\theta(z))\in\mathbb{R}^d$, where $(|z|,\theta(z))\in\mathbb{R}_+\times\mathbb{S}^{d-1}$ denotes the spherical coordinate of z. Then for every $R\geqslant R_1+1$ we may define the truncated vector field $\tilde{X}_{k,R}$ as following,

$$\tilde{X}_{k,R}(x):=X_k(x)\mathbf{1}_{\{x\in B_R\}}(x)+X_k(\pi_R(x))\mathbf{1}_{\{x\notin B_R\}},\tag{5.1}$$

where R_1 is the constant in Assumption 1.1 (4). We first state a technical lemma for $\{\tilde{X}_{k,R}\}_{k=0}^\infty$.

Lemma 5.1. *If Assumption 1.1 holds for $\{X_k\}_{k=0}^m$, then for every $R \geqslant R_1+1$, Assumption 1.1 holds for $\{\tilde{X}_{k,R}\}_{k=0}^m$ with the corresponding constants independent of R.*

Proof. Since the right hand side of (1.4-1.5) depends only on $|x|$, it is clear from the definition (5.1) that (1.4-1.5) hold true for $\{\tilde{X}_{k,R}\}_{k=0}^m$ with the same constants C_1, C_2.

Suppose that (1.6) holds with a constant $0 < \delta \leqslant 1$. For every $x, y \in \mathbb{R}^d$ such that $|y| \leqslant \frac{\delta}{2}$ and $|x+y| \leqslant R$, $\tilde{X}_{k,R}(x+y) = X_k(x+y)$ and so by (1.6) we have

$$\begin{aligned} &\sup_{\{y\in\mathbb{R}^d;\ |y|\leqslant\frac{\delta}{2}, |y+x|\leqslant R\}} \left(p\sum_{k=1}^m |\tilde{X}_{k,R}(x+y)|^2 + \langle x, \tilde{X}_{0,R}(x+y)\rangle \right) \\ &\leqslant \sup_{|z|\leqslant\delta} \left(p\sum_{k=1}^m |X_k(x+z)|^2 + \langle x, X_0(x+z)\rangle \right) \leqslant C(p)(1+|x|^2). \end{aligned} \tag{5.2}$$

If $|x+y| > R$, then by definition $\tilde{X}_{k,R}(x+y) = X_k(\pi_R(x+y)) = X_k\big((R, \theta(x+y))\big)$, and we obtain

$$\begin{aligned} &p\sum_{k=1}^m |\tilde{X}_{k,R}(x+y)|^2 + \langle x, \tilde{X}_{0,R}(x+y)\rangle \\ &= p\sum_{k=1}^m \Big| X_k\Big(\pi_R(x+y)\Big)\Big|^2 + \frac{|x|}{R}\Big\langle \pi_R(x), X_0(\pi_R(x+y))\Big\rangle. \end{aligned} \tag{5.3}$$

It is clear that $\pi_R : \mathbb{R}^d \to \mathbb{R}^d$ is (globally) Lipschitz continuous under the spherical coordinates with the product metric. By its definition and the change of coordinates, we will see that $\pi_R : \mathbb{R}^d \to \mathbb{R}^d$ is still Lipschitz continuous under the Euclidean coordinates in $\mathbb{R}^d$ with the Lipschitz constant independent of R. Therefore, there exists a constant $0 < \delta_1 < 1$(independent of R), such that $\Big|\pi_R(x) - \pi_R(x+y)\Big| \leqslant \delta$ for every $|y| \leqslant \frac{\delta_1}{2}$ and $x \in \mathbb{R}^d$ with $|x+y| \geqslant R > 2$. To the right hand side of (5.3) we apply (1.6) for the system $\{X_k\}_{k=0}^m$ at the point $\pi_R(x)$ to obtain that

$$\begin{aligned} &\sup_{\{y\in\mathbb{R}^d;\ |y|\leqslant\frac{\delta_1}{2}, |y+x|>R\}} \left(p\sum_{k=1}^m |\tilde{X}_{k,R}(x+y)|^2 + \langle x, \tilde{X}_{0,R}(x+y)\rangle \right) \\ =&\frac{|x|}{R} \sup_{\{y\in\mathbb{R}^d;\ |y|\leqslant\frac{\delta_1}{2}, |y+x|>R\}} \\ &\left(p\frac{R}{|x|}\sum_{k=1}^m \Big| X_k\Big(\pi_R(x+y)\Big)\Big|^2 + \Big\langle \pi_R(x), X_0(\pi_R(x+y))\Big\rangle \right) \\ &\leqslant \frac{|x|}{R} \sup_{|z|\leqslant\delta} \left(2p\sum_{k=1}^m \Big| X_k\Big(\pi_R(x)+z\Big)\Big|^2 + \Big\langle \pi_R(x), X_0(\pi_R(x)+z)\Big\rangle \right) \\ &\leqslant \frac{2pC|x|}{R}(1+R^2) \leqslant C(p)(1+|x|^2). \end{aligned}$$

Note that $|x| > R-1$. Here the first inequality is due to the property $\frac{|x|}{R} \geqslant \frac{1}{2}$ and the last step is due to the property that $\frac{1+R^2}{R} \geqslant C(1+|x|)$. Together with (5.2), this shows that (1.6) holds with the corresponding constant δ replaced by $\tilde{\delta} := \min\{\frac{\delta}{2}, \frac{\delta_1}{2}\}$.

Now we move on to item (3) of Assumption 1.1 and prove first that there is a constant $p_3 > 2(d+1)$, such that $X_{k,R} \in W^{1,p_3}_{\text{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ for $1 \leqslant k \leqslant m$.

Since $\tilde{X}_{k,R}(x) = X_k(x)$ for every $|x| \leqslant R$, $\tilde{X}_{k,R} \in W^{1,p_3}(B_R; \mathbb{R}^d)$ by Assumption 1.1 (3), also note that the boundary ∂B_R is C^1 and X_k is continuous, we may apply the

integration by parts formula to $\tilde{X}_{k,R}$, therefore for every $\psi \in C_0^\infty(\mathbb{R}^d)$ and $1 \leqslant i \leqslant d$,

$$\int_{B_R} D_i X_k(x)\psi(x)dx = -\int_{B_R} \tilde{X}_{k,R}(x) D_i\psi(x)dx + \int_{S_R} \tilde{X}_{k,R}\psi\nu_i dS, \tag{5.4}$$

where $D_i\psi = \partial_{x_i}\psi$, $\nu = (\nu_1, \dots \nu_d)$ denotes the outward normal vector field on S_R, dS denotes integration with respect to the area measure on S_R.

By (1.8), DX_k is locally bounded on the complement $B^c_{R_1}$ of B_{R_1}, hence X_k is locally Lipschitz continuous on $B^c_{R_1}$ and belongs to $W^{1,\infty}_{\mathrm{loc}}(B^c_{R_1};\mathbb{R}^d)$, see e.g. [9, Theorem 4 in Section 5.8.2]). For every $x = (|x|, \theta(x))$ and $y = (|y|, \theta(y))$ with $R \leqslant |x| \leqslant |y|$ and $\theta(x), \theta(y) \in \mathbb{S}^{d-1}$, we have,

$$\begin{aligned}&|\tilde{X}_{k,R}(x) - \tilde{X}_{k,R}(y)| = \big|X_k\big(\pi_R(x)\big) - X_k\big(\pi_R(y)\big)\big| \\ &\leqslant C_3(1+|R|^{p_5})\big|\pi_R(x) - \pi_R(y)\big| \leqslant CC_3(1+|R|^{p_5})|x-y|\end{aligned}$$

where first inequality is due to the Lipschitz continuity of X_k on S_R and (1.8), and the second inequality is by the Lipschitz continuity of π_R . We conclude that the truncated vector field $\tilde{X}_{k,R}$ is globally Lipschitz continuous on B^c_R, and $\tilde{X}_{k,R} \in W^{1,\infty}(B^c_R;\, \mathbb{R}^d)$. Applying again integration by parts formula to $\tilde{X}_{k,R}$, for every $\psi \in C_0^\infty(\mathbb{R}^d)$,

$$\int_{B^c_R} D_i\tilde{X}_{k,R}(x)\psi(x)dx = -\int_{B^c_R} \tilde{X}_{k,R}(x) D_i\psi(x)dx - \int_{S_R} \tilde{X}_{k,R}\psi\nu_i dS, \tag{5.5}$$

where we use the property that the outward normal vector on ∂B^c_R is $-\nu$. From (5.4) and (5.5) we see that for every $\psi \in C_0^\infty(\mathbb{R}^d)$ and $1 \leqslant i \leqslant d$,

$$\begin{aligned}&\int_{\mathbb{R}^d} \big(D_i X_k(x)\mathbf{1}_{\{x\in B_R\}} + D_i\tilde{X}_{k,R}(x)\mathbf{1}_{\{x\in B^c_R\}}\big)\psi(x)dx \\ &= -\int_{\mathbb{R}^d} \tilde{X}_{k,R}(x) D_i\psi(x)dx,\end{aligned}$$

which means that $\tilde{X}_{k,R}$ is weakly differentiable with the differential $D\tilde{X}_{k,R}$, and $D\tilde{X}_{k,R}(x) = DX_k(x)\mathbf{1}_{\{x\in B_R\}} + D\tilde{X}_{k,R}(x)\mathbf{1}_{\{x\in B^c_R\}}$, then we conclude $\tilde{X}_{k,R} \in W^{1,p_3}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ from the fact that $X_k \in W^{1,p_3}(B_R;\mathbb{R}^d)$ and $\tilde{X}_{k,R} \in W^{1,\infty}(B^c_R;\mathbb{R}^d)$. As the same way we can show $\tilde{X}_{0,R} \in W^{1,p_4}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$.

Let $\nu(\theta)$ be the unit outward normal vector of S_R at the point (R,θ), by the definition of $\tilde{X}_{k,R}$, for almost every $x = (|x|, \theta(x)) \in \mathbb{R}^d$ with $|x| > R$ we obtain,

$$D\tilde{X}_{k,R}(x)\big(\nu(\theta(x))\big) = 0. \tag{5.6}$$

In particular, we want to remark that throughout this paper, we always use $\nu(\theta(x))$ to denote the vector in tangent space $T_x\mathbb{R}^d$ which has the same direction with $\nu(\theta(x))$ at $T_{\theta(x)}S_R$.

Let $T_\theta S_R$ be the tangent space to the sphere S_R at the point (R,θ), since $\tilde{X}_{k,R}$ is Lipschitz continuous on S_R, by Rademacher's theorem the derivative $D\tilde{X}_{k,R}$ in the directions of $T_\theta S_R$ is almost everywhere well defined with respect to the area measure on S_R. For every $\xi \in T_{\theta(x)}S_{|x|}$, by a standard isomorphism, we can also assume $\xi \in T_{\theta(x)}S_R$. And by definition (5.1), for almost every $x = (|x|, \theta(x)) \in \mathbb{R}^d$ with $|x| > R$ and every $\xi \in T_{\theta(x)}S_{|x|}$,

$$D\tilde{X}_{k,R}(x)(\xi) = \frac{R}{|x|} DX_k\big(\pi_R(x)\big)(\xi). \tag{5.7}$$

For every $p > 1$, let $\tilde{K}_{p,R}(x) := \sup_{|\xi|=1} \tilde{H}_{p,R}(x)\big(\xi,\xi\big)$, where

$$\tilde{H}_{p,R}(x)\big(\xi,\xi\big) = 2p\langle D\tilde{X}_{0,R}(x)(\xi),\xi\rangle + (2p-1)p\sum_{k=1}^m |D\tilde{X}_{k,R}(x)(\xi)|^2. \tag{5.8}$$

By (5.6), for almost every $x=(|x|,\theta(x))\in\mathbb{R}^d$ with $|x|>R\geqslant R_1+1$, $\tilde{H}_{p,R}(x)\big(\nu(\theta(x)),\nu(\theta(x))\big)=0$, so by (5.7) and (1.9), we have

$$\begin{aligned}&\tilde{K}_{p,R}(x)=\max\Big\{0,\sup_{\xi\in T_{\theta(x)}S_{|x|},|\xi|=1}H_p\big(\pi_R(x)\big)\big(\xi,\xi\big)\Big\}\leqslant 0\vee K_p\big(\pi_R(x)\big)\\&\leqslant C(p)\log(1+|R|^2)\leqslant C(p)\log(1+|x|^2).\end{aligned}$$

On the other hand, it is obvious that $\tilde{K}_{p,R}(x)=K_p(x)$ for almost every $x\in\mathbb{R}^d$ with $|x|<R$. So we obtain that (1.9) holds for $\tilde{X}_{k,R}$ with the same constants $C(p)$ and R_1 as that for X_k.

So for the constant $\kappa(p)$ in (1.7) and every $\tilde{R}>0$,

$$\begin{aligned}&\sup_{R>R_1}\int_{\{|x|\leqslant\tilde{R}\}}e^{\kappa(p)\tilde{K}_{p,R}(x)}dx\\&\leqslant\sup_{R>R_1}\Big(\int_{\{|x|\leqslant R_1\}}e^{\kappa(p)K_p(x)}dx+\int_{\{R_1<|x|\leqslant R\}}e^{\kappa(p)K_p(x)}dx\\&+\int_{\{R<|x|\leqslant\tilde{R}\}}e^{\kappa(p)\left(0\vee K_p\left(\pi_R(x)\right)\right)}dx\Big)\\&\leqslant\int_{\{|x|\leqslant R_1\}}e^{\kappa(p)K_p(x)}+\int_{\{R_1<|x|\leqslant\tilde{R}\}}e^{\kappa(p)C(p)\log(1+|x|^2)}dx<\infty,\end{aligned}\tag{5.9}$$

which means (1.7) is true for $\{\tilde{X}_{k,R}\}_{k=0}^m$ with the corresponding constants independent of R. Similarly, we can show (1.8) holds for $\{\tilde{X}_{k,R}\}_{k=0}^m$ with the corresponding constants independent of R. □

For every $\varepsilon>0$ we define the approximating vector fields $\{X_k^\varepsilon\}_{k=0}^m$ by $X_k^\varepsilon:=\tilde{X}_{k,\varepsilon^{-\lambda}}*\eta_\varepsilon$, where the constant $\lambda>0$ will be chosen later in Lemma 5.2. Since for every $\varepsilon>0$, $\tilde{X}_{k,\varepsilon^{-\lambda}}$ is bounded, it is obvious that $X_k^\varepsilon\in C_b^\infty(\mathbb{R}^d;\mathbb{R}^d)$. Following result concerns about the properties of $\{X_k^\varepsilon\}_{k=0}^m$ which are uniformly for ε.

Lemma 5.2. *Suppose Assumption 1.1 holds. There exist $\lambda_0>0$, $\varepsilon_0>0$, such that if we define $X_k^\varepsilon:=\tilde{X}_{k,\varepsilon^{-\lambda_0}}*\eta_\varepsilon$, then for every $\varepsilon\in(0,\varepsilon_0)$, (1.4), (1.5), (1.7)-(1.9) hold for $\{X_k^\varepsilon\}_{k=0}^m$ with the corresponding constants independent of ε. Furthermore, for every $p>0$, there exists a $C(p)>0$, such that*

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\Big(\sum_{k=1}^m p|X_k^\varepsilon(x)|^2+\langle x,X_0^\varepsilon(x)\rangle\Big)\leqslant C(p)(1+|x|^2).\tag{5.10}$$

Proof. In the proof we fix a $\lambda>0$ which will be determined later, we set $\varepsilon_1(\lambda):=\min((R_1+2)^{-\frac{1}{\lambda}},\frac{\delta}{4})$.

Since $\varepsilon_1^{-\lambda}\geqslant R_1+2$, from Lemma 5.1 we have,

$$\sup_{\varepsilon\in(0,\varepsilon_1)}\sup_{|y|\leqslant\frac{\delta}{2}}\Big(p\sum_{k=1}^m|\tilde{X}_{k,\varepsilon^{-\lambda}}(x+y)|^2+\langle x,\tilde{X}_{0,\varepsilon^{-\lambda}}(x+y)\rangle\Big)\leqslant C(p)(1+|x|^2).$$

For every $\varepsilon<\varepsilon_1<\frac{\delta}{2}$, we apply this to $X_k^\varepsilon(x)=\int_{|y|\leqslant\varepsilon}\tilde{X}_{k,\varepsilon^{-\lambda}}(x-y)\eta_\varepsilon(y)dy$ and by Jensen's inequality we obtain

$$\begin{aligned}&\sup_{\varepsilon\in(0,\varepsilon_1)}\Big(p\sum_{k=1}^m|X_k^\varepsilon(x)|^2+\langle x,X_0^\varepsilon(x)\rangle\Big)\\&\leqslant\sup_{\varepsilon\in(0,\varepsilon_1)}\Big(p\sum_{k=1}^m\int_{|y|\leqslant\frac{\delta}{2}}|\tilde{X}_{k,\varepsilon^{-\lambda}}(x-y)|^2\eta_\varepsilon(y)dy\\&+\int_{|y|\leqslant\frac{\delta}{2}}\big\langle x,\tilde{X}_{0,\varepsilon^{-\lambda}}(x-y)\big\rangle\eta_\varepsilon(y)dy\Big)\leqslant C(p)(1+|x|^2),\end{aligned}\tag{5.11}$$

which means (5.10) holds. Similarly, we can show (1.5) holds for $\{X_k^\varepsilon\}_{k=0}^m$ with the corresponding constants independent of ε.

Let $K_p^\varepsilon(x) := \sup_{|\xi|=1} H_p^\varepsilon(x)\big(\xi,\xi\big)$ where

$$H_p^\varepsilon(x)\big(\xi,\xi\big) := 2p\langle DX_0^\varepsilon(x)(\xi),\xi\rangle + (2p-1)p\sum_{k=1}^m |DX_k^\varepsilon(x)(\xi)|^2. \tag{5.12}$$

The local integrability (1.7) is trivial for the smooth functions X_k^ε. Now we try to give an uniform bounds for ε. As the same argument for (5.11), according to Jensen's inequality we have $K_p^\varepsilon \leqslant \tilde{K}_{p,\varepsilon^{-\lambda}} * \eta_\varepsilon$, where $\tilde{K}_{p,\varepsilon^{-\lambda}}$ is defined by (5.8). Letting $\kappa(p)$ be the constant in (1.7), by Jensen's inequality and (5.9), for every $p>1$, $R>0$,

$$\begin{aligned}
&\sup_{\varepsilon\in(0,\varepsilon_1)}\int_{\{|x|\leqslant R\}} \exp\big(\kappa(p)K_p^\varepsilon(x)\big)dx\\
&\leqslant \sup_{\varepsilon\in(0,\varepsilon_1)}\int_{\{|x|\leqslant R\}} \exp\Big(\kappa(p)\tilde{K}_{p,\varepsilon^{-\lambda}} * \eta_\varepsilon(x)\Big)dx\\
&\leqslant \sup_{\varepsilon\in(0,\varepsilon_1)}\int_{\{|x|\leqslant R\}} \Big(\exp\big(\kappa(p)\tilde{K}_{p,\varepsilon^{-\lambda}}\big) * \eta_\varepsilon(x)\Big)dx\\
&\leqslant \sup_{\varepsilon\in(0,\varepsilon_1)}\int_{\{|x|\leqslant R+1\}} \exp\Big(\kappa(p)\tilde{K}_{p,\varepsilon^{-\lambda}}(x)\Big)\ dx<\infty.
\end{aligned} \tag{5.13}$$

In the last step, to verify

$$\sup_{\varepsilon\in(0,\varepsilon_1)}\int_{\{|x|\leqslant R+1\}} \exp\Big(\kappa(p)\tilde{K}_{p,\varepsilon^{-\lambda}}(x)\Big)\ dx<\infty,$$

we used the property that $\varepsilon^{-\lambda}>R_1$ for every $\varepsilon\in(0,\varepsilon_1)$ and the inequality (5.9).

Hence (1.7) holds for $\{X_k^\varepsilon\}_{k=0}^m$ with the corresponding constants independent of ε. As the similar way, we can check (1.8) and (1.9) hold for $\{X_k^\varepsilon\}_{k=0}^m$ with the corresponding constants independent of ε.

Finally we study the ellipticity condition (1.4). By (5.6) and (5.7), for every $\varepsilon\in(0,\varepsilon_1)$, $1\leqslant k\leqslant m$,

$$\sup_{|y|\geqslant R_1}|D\tilde{X}_{k,\varepsilon^{-\lambda}}(y)|\leqslant \sup_{R_1\leqslant|y|\leqslant\varepsilon^{-\lambda}}|DX_k(y)|\leqslant C(1+\varepsilon^{-\lambda p_5}).$$

Therefore we have,

$$\big|\tilde{X}_{k,\varepsilon^{-\lambda}}(x)-\tilde{X}_{k,\varepsilon^{-\lambda}}(y)\big|\leqslant C(1+\varepsilon^{-\lambda p_5})|x-y|,\ \ x,y\in B_{R_1}^c. \tag{5.14}$$

On the other hand, by (5.1), for every $\varepsilon\in(0,\varepsilon_1)$ and $x\in\mathbb{R}^d$ with $|x|\leqslant R_1+2\leqslant\varepsilon^{-\lambda}$, we know that $\tilde{X}_{k,\varepsilon^{-\lambda}}(x)=X_k(x)$. Since $X_k\in W^{1,p_3}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$ for some constant $p_3>2(d+1)$, according to the Sobolev embedding lemma we have,

$$\sup_{\varepsilon\in(0,\varepsilon_1)}\big|\tilde{X}_{k,\varepsilon^{-\lambda}}(x)-\tilde{X}_{k,\varepsilon^{-\lambda}}(y)\big|\leqslant C|x-y|^\iota,\ x,y\in B_{R_1+2} \tag{5.15}$$

for some constant $\iota\in(0,1)$, which is independent of ε. Then by (5.14) and (5.15), for every $\varepsilon\in(0,\varepsilon_1)$,

$$\begin{aligned}
&\big|X_k^\varepsilon(x)-X_{k,\varepsilon^{-\lambda}}(x)\big|\ \leqslant \int_{|y|\leqslant\varepsilon}\big|\tilde{X}_{k,\varepsilon^{-\lambda}}(x+y)-\tilde{X}_{k,\varepsilon^{-\lambda}}(x)\big|\eta_\varepsilon(y)dy\\
&\leqslant C\varepsilon^\iota\mathbf{1}_{\{|x|\leqslant R_1+1\}}+C(1+\varepsilon^{-\lambda p_5})\varepsilon\mathbf{1}_{\{|x|>R_1+1\}}.
\end{aligned} \tag{5.16}$$

We write the components of X_k^ε as $X_k^\varepsilon = (X_{k1}^\varepsilon, \cdots, X_{kd}^\varepsilon)$ and for every $1 \leqslant i, j \leqslant d$ we define

$$a_{i,j}^\varepsilon(x) := \sum_{k=1}^m X_{ki}^\varepsilon(x) X_{kj}^\varepsilon(x), \quad \tilde{a}_{i,j}^\varepsilon(x) := \sum_{k=1}^m \tilde{X}_{ki,\varepsilon^{-\lambda}}(x) \tilde{X}_{kj,\varepsilon^{-\lambda}}(x).$$

By (1.5) and definition (5.1), for every $\varepsilon \in (0, \varepsilon_1)$ and $x \in \mathbb{R}^d$,

$$|\tilde{X}_{k,\varepsilon^{-\lambda}}(x)| \leqslant \sup_{|x| \leqslant \varepsilon^{-\lambda}} |X_k(x)| \leqslant C(1 + \varepsilon^{-\lambda p_2}),$$

therefore we have

$$|X_k^\varepsilon(x)| \leqslant \int_{|y| \leqslant \varepsilon} |\tilde{X}_{k,\varepsilon^{-\lambda}}(x+y)| \eta_\varepsilon(y) dy \leqslant C(1 + \varepsilon^{-\lambda p_2}).$$

Combing this with (5.16) we get

$$\begin{aligned} &|a_{i,j}^\varepsilon(x) - \tilde{a}_{i,j}^\varepsilon(x)| \\ &\leqslant C \sup_{1 \leqslant k \leqslant m} \left| X_k^\varepsilon(x) - \tilde{X}_{k,\varepsilon^{-\lambda}}(x) \right| \left(|X_k^\varepsilon(x)| + |\tilde{X}_{k,\varepsilon^{-\lambda}}(x)| \right) \\ &\leqslant C\varepsilon^{\iota - \lambda p_2} \mathbf{1}_{\{|x| \leqslant R_1 + 1\}} + C\varepsilon^{1-\lambda(p_2+p_5)} \mathbf{1}_{\{|x| > R_1+1\}}. \end{aligned} \tag{5.17}$$

By definition (5.1), and ellipticity condition (1.4), for every $\varepsilon \in (0, \varepsilon_1(\lambda))$ and $\xi = (\xi_1, \cdots, \xi_d) \in \mathbb{R}^d$ with $|\xi| = 1$,

$$\sum_{i,j=1}^d \tilde{a}_{i,j}^\varepsilon(x) \xi_i \xi_j \geqslant \frac{C}{1 + |x|^{p_1}} \mathbf{1}_{\{|x| \leqslant \varepsilon^{-\lambda}\}} + \frac{C}{2} \varepsilon^{\lambda p_1} \mathbf{1}_{\{|x| > \varepsilon^{-\lambda}\}}. \tag{5.18}$$

We will prove below that the error made by convolution does not affect the ellipticity of $\{a_{i,j}^\varepsilon\}$. In fact, according to (5.17) and (5.18),

$$\begin{aligned} &\sum_{i,j=1}^d a_{i,j}^\varepsilon(x) \xi_i \xi_j \\ &\geqslant \sum_{i,j=1}^d \tilde{a}_{i,j}^\varepsilon(x) \xi_i \xi_j - d^2 \max_i |\xi_i|^2 \sup_{\varepsilon \in (0,\varepsilon_1)} \sup_{i,j} |a_{i,j}^\varepsilon(x) - \tilde{a}_{i,j}^\varepsilon(x)| \\ &\geqslant C \left(\frac{\mathbf{1}_{\{|x| \leqslant \varepsilon^{-\lambda}\}}}{1 + |x|^{p_1}} - \varepsilon^{\iota - \lambda p_2} \mathbf{1}_{\{|x| \leqslant R_1+1\}} - \varepsilon^{1-\lambda(p_2+p_5)} \mathbf{1}_{\{R_1+1 < |x| \leqslant \varepsilon^{-\lambda}\}} \right) \\ &\quad + C \left(\varepsilon^{\lambda p_1} - \varepsilon^{1-\lambda(p_2+p_5)} \right) \mathbf{1}_{\{|x| > \varepsilon^{-\lambda}\}}. \end{aligned} \tag{5.19}$$

We choose a constant $\lambda_0 > 0$ small enough satisfying $\lambda_0 p_1 < \iota - \lambda_0 p_2$ and $\lambda_0 p_1 < 1 - \lambda_0(p_2 + p_5)$. Hence for such λ_0, there exists a positive constant $\varepsilon_0(\lambda_0) < \varepsilon_1(\lambda_0)$, such that for every $\varepsilon \in (0, \varepsilon_0)$,

$$\varepsilon^{\iota - \lambda_0 p_2} \leqslant \frac{\varepsilon^{\lambda_0 p_1}}{4(1 + \varepsilon^{\lambda_0 p_1})}, \quad \varepsilon^{1-\lambda_0(p_2+p_5)} \leqslant \frac{\varepsilon^{\lambda_0 p_1}}{4(1 + \varepsilon^{\lambda_0 p_1})} \leqslant \frac{\varepsilon^{\lambda_0 p_1}}{4}.$$

So for every $\varepsilon \in (0, \varepsilon_0)$, $x \in \mathbb{R}^d$ with $|x| \leqslant \varepsilon^{-\lambda_0}$,

$$\begin{aligned} &\frac{1}{1 + |x|^{p_1}} - \varepsilon^{\kappa - \lambda_0 p_2} \mathbf{1}_{\{|x| \leqslant R_1+1\}} - \varepsilon^{1-\lambda_0(p_2+p_5)} \mathbf{1}_{\{R_1+1 < |x| \leqslant \varepsilon^{-\lambda_0}\}} \\ &\geqslant \frac{1}{1 + |x|^{p_1}} - \frac{\varepsilon^{\lambda_0 p_1}}{2(1 + \varepsilon^{\lambda_0 p_1})} \geqslant \frac{1}{2(1 + |x|^{p_1})}. \end{aligned}$$

Now we fix the constant λ_0 and $\varepsilon_0(\lambda_0)$ obtained above, putting above estimates together into (5.19), we have for every $\varepsilon \in (0, \varepsilon_0)$,

$$\begin{aligned}\sum_{i,j=1}^{d} a^{\varepsilon}_{i,j}(x)\xi_i\xi_j &\geqslant \frac{C}{2(1+|x|^{p_1})}\mathbf{1}_{\{|x|\leqslant \varepsilon^{-\lambda_0}\}} + \frac{C}{2}|\varepsilon|^{\lambda_0 p_1}\mathbf{1}_{\{|x|>\varepsilon^{-\lambda_0}\}} \\ &\geqslant \frac{C}{2}\left(\frac{1}{1+|x|^{p_1}}\right),\end{aligned}$$

which means (1.7) holds for $\{X^{\varepsilon}_k\}_{k=1}^m$, $\varepsilon \in (0, \varepsilon_0)$ with the corresponding constants independent of ε. □

From now on we take the constants λ_0 and ε_0 to be that obtained in Lemma 5.2, and for every $\varepsilon \in (0, \varepsilon_0)$, we define $X^{\varepsilon}_k(x) := X_{k,\varepsilon^{-\lambda_0}} * \eta_\varepsilon$.

Lemma 5.3. *Suppose that Assumption 1.1 holds. For every $R > 0$ and $p > 1$,*

$$\lim_{\varepsilon\to 0}\int_{\{|x|\leqslant R\}} |X^{\varepsilon}_k(x) - X_k(x)|^p dx = 0, \quad 0 \leqslant k \leqslant m, \tag{5.20}$$

$$\lim_{\varepsilon\to 0}\int_{\{|x|\leqslant R\}} |DX^{\varepsilon}_k(x) - DX_k(x)|^{p_3} dx = 0, \quad 1 \leqslant k \leqslant m, \tag{5.21}$$

$$\lim_{\varepsilon\to 0}\int_{\{|x|\leqslant R\}} |DX^{\varepsilon}_0(x) - DX_0(x)|^{p_4} dx = 0, \tag{5.22}$$

where $p_3 > 2(d+1)$, $p_4 > d+1$ are the constants in (3) of Assumption 1.1.

Proof. For every fixed $R > 0$ and every ε small enough such that $\varepsilon^{-\lambda} > R+1$, by definition (5.1) we have $\tilde{X}_{k,\varepsilon^{-\lambda}}(x) = X_k(x)$ for all $x \in \mathbb{R}^d$ with $|x| \leqslant R+1$. Therefore for every $x \in \mathbb{R}^d$ with $|x| \leqslant R$,

$$DX^{\varepsilon}_k(x) = D\tilde{X}_{k,\varepsilon^{-\lambda}} * \eta_\varepsilon(x) = DX_k * \eta_\varepsilon(x),$$

Hence (5.21) holds since $X_k \in W^{1,p_3}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$, $1 \leqslant k \leqslant m$. As the same way we can show (5.22).

Since the $\{X_k\}_{k=0}^m$ are locally bounded by part (2) of Assumption 1.1, similarly we can prove (5.20) for any $p > 1$.

□

6 The derivative flow equation

Through this section, let $X^{\varepsilon}_k \in C^{\infty}_b(\mathbb{R}^d;\mathbb{R}^d)$, $0 \leqslant k \leqslant m$, $\varepsilon \in (0, \varepsilon_0)$ be the vector fields constructed in Lemma 5.2, we consider the following approximating SDE for (1.13),

$$\begin{cases} dx^{\varepsilon}_t = \sum_{k=1}^m X^{\varepsilon}_k(x^{\varepsilon}_t)dW^k_t + X^{\varepsilon}_0(x^{\varepsilon}_t)dt, \\ dv^{\varepsilon}_t = \sum_{k=1}^m DX^{\varepsilon}_k(x^{\varepsilon}_t)(v^{\varepsilon}_t)dW^k_t + DX^{\varepsilon}_0(x^{\varepsilon}_t)(v^{\varepsilon}_t)dt. \end{cases} \tag{6.1}$$

We denote the strong solution to (6.1) with initial point $(x, v) \in \mathbb{R}^{2d}$ by $(F^{\varepsilon}_t(x), V^{\varepsilon}_t(x, v))$.

According to Lemma 5.2, $\{X^{\varepsilon}_k\}_{k=0}^m$ satisfies (1.4), (1.5) and (4.4) with corresponding constants independent of ε, by a straightforward application of Lemma 4.4 to $F^{\varepsilon}_t(x)$, we obtain the following lemma, which will be frequently used in this section.

Lemma 6.1. *Suppose that Assumption (1.1) holds, then for every $p > d+1$, $T > 0$, compact set $K \subseteq \mathbb{R}^d$, and non-negative measurable function $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$, we have,*

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x\in K}\mathbb{E}\left(\int_0^T f(t, F^{\varepsilon}_t(x))\,dt\right) \leqslant C(K)Q(T)\left(\int_0^T\int f^p(t,y)dydt\right)^{\frac{1}{p}}, \tag{6.2}$$

where $Q:\mathbb{R}_+\to\mathbb{R}_+$ is a positive Borel measurable function such that $\sup_{T\in[0,\tilde{T}_0]}Q(T)<\infty$ for every $\tilde{T}_0>0$ and $C(K)$ is a positive constant which may depend on K.

In this section, we will prove existence and uniqueness for (1.13). We first give the following lemma about the uniform moment estimate for $V_t^\varepsilon(x,v)$.

Lemma 6.2. *Suppose that Assumption (1.1) holds. Then for every $p\geqslant 2$ and compact set $\tilde{K}\subseteq\mathbb{R}^{2d}$,*

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{(x,v)\in\tilde{K}}\sup_{t\in[0,T_0(p)]}\mathbb{E}\left(|V_t^\varepsilon(x,v)|^p\right)<\infty, \tag{6.3}$$

where $T_0(p):=\frac{\kappa(p)}{d+2}$ with the constant $\kappa(p)$ in (1.7).

Proof. Given $(x,v)\in\mathbb{R}^{2d}$ fixed, we write $(F_t^\varepsilon,V_t^\varepsilon)$ for $(F_t^\varepsilon(x),V_t^\varepsilon(x,v))$ for simplicity. We first follow some steps in [23, Theorem 5.1] (see also [24]) for the estimation. Applying Itô formula to (6.1), we derive

$$|V_t^\varepsilon|^p=|v|^p+\sum_{k=1}^m\int_0^t|V_s^\varepsilon|^p dM_s^\varepsilon+\int_0^t|V_s^\varepsilon|^p da_s^\varepsilon, \tag{6.4}$$

where

$$M_t^\varepsilon:=p\sum_{k=1}^m\int_0^t\frac{\langle DX_k^\varepsilon(F_s^\varepsilon)\big(V_s^\varepsilon\big),V_s^\varepsilon\rangle}{|V_s^\varepsilon|^2}dW_s^k,\ a_t^\varepsilon:=\frac{p}{2}\int_0^t\frac{\bar{H}_p^\varepsilon(F_s^\varepsilon)\big(V_s^\varepsilon,V_s^\varepsilon\big)}{|V_s^\varepsilon|^2}ds. \tag{6.5}$$

Here for every $x\in\mathbb{R}^d,\xi\in\mathbb{R}^d$,

$$\begin{aligned}\bar{H}_p^\varepsilon(x)\big(\xi,\xi\big)=&2\langle DX_0^\varepsilon(x)(\xi),\xi\rangle\\&+\sum_{k=1}^m\Big(|DX_k^\varepsilon(x)(\xi)|^2+(p-2)\frac{\big|\langle DX_k^\varepsilon(x)\big(\xi\big),\xi\rangle\big|^2}{|\xi|^2}\Big)\end{aligned}$$

with the convention that $\frac{0}{0}=0$.

Furthermore, we know that for every $\mathbb{R}$-valued semi-martingale N_t, the unique solution to the linear equation (in $\mathbb{R}$) $dz_t=z_t dN_t$ will have the expression $z_t=z_0\exp\big(N_t-\frac{\langle N\rangle_t}{2}\big)$, where $\langle N\rangle_t$ denotes the quadratic variational process for N_t, see e.g. [29, Proposition 2.3 in Page 361] or [23, Theorem 5.1]. So by (6.4) we have

$$|V_t^\varepsilon|^p=|v|^p\exp\Big(M_t^\varepsilon-\frac{\langle M^\varepsilon\rangle_t}{2}+a_t^\varepsilon\Big). \tag{6.6}$$

Since $\tilde{M}_t^\varepsilon:=\exp(2M_t^\varepsilon-2\langle M^\varepsilon\rangle_t)$ is a super martingale, $\mathbb{E}(\tilde{M}_t^\varepsilon)\leqslant 1$, after applying Hölder inequality to (6.6) we deduce the following estimate,

$$\begin{aligned}\mathbb{E}\left(|V_t^\varepsilon|^p\right)&\leqslant|v|^p\big(\mathbb{E}\tilde{M}_t^\varepsilon\big)^{\frac{1}{2}}\Big(\mathbb{E}\left(\exp\left(\langle M^\varepsilon\rangle_t+2a_t^\varepsilon\right)\right)\Big)^{\frac{1}{2}}\\&\leqslant|v|^p\left(\mathbb{E}\left(\exp\big(\int_0^t K_p^\varepsilon(F_s^\varepsilon)ds\big)\right)\right)^{\frac{1}{2}},\end{aligned} \tag{6.7}$$

where we use the property that $\langle M^\varepsilon\rangle_t+2a_t^\varepsilon\leqslant\int_0^t K_p^\varepsilon(F_s^\varepsilon)ds$ for K_p^ε defined by (5.12). For

every fixed $T>0$ and $t\in(0,T]$, by Jensen's inequality,

$$\begin{aligned}
&\mathbb{E}\left(\exp\Big(\int_0^t K_p^\varepsilon(F_s^\varepsilon)ds\Big)\right)=\mathbb{E}\left(\exp\left(\int_0^T K_p^\varepsilon(F_s^\varepsilon)\mathbf{1}_{\{s\in(0,t)\}}ds\right)\right)\\
&\leqslant\frac{1}{T}\left(\mathbb{E}\left(\int_0^t\exp\big(T\ K_p^\varepsilon(F_s^\varepsilon)\big)\,ds\right)+(T-t)\right)\\
&\leqslant\frac{1}{T}\mathbb{E}\left(\int_0^t\exp\big(T\ K_p^\varepsilon(F_s^\varepsilon)\big)\mathbf{1}_{\{|F_s^\varepsilon|\leqslant R_1+2\}}ds\right)\\
&+\frac{1}{T}\mathbb{E}\left(\int_0^t\exp\big(TK_p^\varepsilon(F_s^\varepsilon)\big)\,\mathbf{1}_{\{|F_s^\varepsilon|>R_1+2\}}ds\right)+1.
\end{aligned}$$

Applying Lemma 6.1 with $p=d+2$, for every compact $K\subseteq\mathbb{R}^d$,

$$\begin{aligned}
&\sup_{t\in[0,T]}\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x\in K}\mathbb{E}\left(\int_0^t\exp\Big(TK_p^\varepsilon(F_s^\varepsilon)\Big)\mathbf{1}_{\{|F_s^\varepsilon|\leqslant R_1+2\}}\,ds\right)\\
&\leqslant C(K,T)\sup_{\varepsilon\in(0,\varepsilon_0)}\left(\int_{\{|x|\leqslant R_1+2\}}\exp\Big(T(d+2)\,K_p^\varepsilon(x)\Big)\,dx\right)^{\frac{1}{d+2}}.
\end{aligned}$$

If $T=T_0(p):=\frac{\kappa(p)}{d+2}$, the above quantity is finite by (5.13) in the proof of Lemma 5.2.

Also by Lemma 5.2, there is a constant $C(p)>0$ independent of ε such that for every $x\in\mathbb{R}^d$ with $|x|>R_1+2$,

$$\sup_{\varepsilon\in(0,\varepsilon_0)}K_p^\varepsilon(x)\leqslant C(p)\log(1+|x|^2).$$

By (5.10) and Example 4.3, for every $p>0, T>0$ and compact set $K\subseteq\mathbb{R}^d$,

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x\in K}\sup_{t\in[0,T]}\mathbb{E}\big(|F_t^\varepsilon|^p\big)<\infty,\tag{6.8}$$

therefore we have

$$\begin{aligned}
&\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x\in K}\mathbb{E}\left(\int_0^{T_0(p)}\exp\big(T_0(p)K_p^\varepsilon(F_s^\varepsilon)\big)\mathbf{1}_{\{|F_s^\varepsilon|>R_1+2\}}ds\right)\\
&\leqslant\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x_0\in K}\mathbb{E}\left(\int_0^{T_0(p)}\big(1+|F_s^\varepsilon|^{2C(p)T_0(p)}\big)ds\right)<\infty.
\end{aligned}$$

We put all the estimates above back into (6.7) to complete the proof. □

Lemma 6.3. *Suppose that Assumption 1.1 holds. Then for all $p>1$ and compact set $K\subseteq\mathbb{R}^d$,*

$$\limsup_{\varepsilon,\tilde\varepsilon\to0}\sup_{x\in K}\int_0^T\mathbb{E}\left(|X_k^\varepsilon(F_t^\varepsilon(x))-X_k^{\tilde\varepsilon}(F_t^\varepsilon(x))|^p\right)dt=0,\ 0\leqslant k\leqslant m.\tag{6.9}$$

Moreover, there exist constants $\beta_1>0$ and $\beta_2>0$ such that for all $1\leqslant k\leqslant m$, $T>0$, and compact subset $K\subseteq\mathbb{R}^d$, the following holds:

$$\limsup_{\varepsilon,\tilde\varepsilon\to0}\sup_{x\in K}\int_0^T\mathbb{E}\left(|DX_k^\varepsilon(F_t^\varepsilon(x))-DX_k^{\tilde\varepsilon}(F_t^\varepsilon(x))|^{2+\beta_1}\right)dt=0,\tag{6.10}$$

$$\limsup_{\varepsilon,\tilde\varepsilon\to0}\sup_{x\in K}\int_0^T\mathbb{E}\left(|DX_0^\varepsilon(F_t^\varepsilon(x))-DX_0^{\tilde\varepsilon}(F_t^\varepsilon(x))|^{1+\beta_2}\right)dt=0.\tag{6.11}$$

$$\begin{aligned}\sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{x\in K}\mathbb{E}\left(\int_0^T\left|DX_k^{\varepsilon}\big(F_t^{\tilde{\varepsilon}}(x)\big)\right|^{2+\beta_1}dt\right)<\infty,\\ \sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{x\in K}\mathbb{E}\left(\int_0^T\left|DX_0^{\varepsilon}\big(F_t^{\tilde{\varepsilon}}(x)\big)\right|^{1+\beta_2}dt\right)<\infty.\end{aligned}\tag{6.12}$$

Proof. Given $x\in\mathbb{R}^d$ fixed, we write F_t^{ε} for $F_t^{\varepsilon}(x)$ for simplicity. We only prove (6.10), the proof for (6.9), (6.11) and (6.12) are similar. Let $p_3>2(d+1)$ be the constant in Assumption 1.1(3). We take a $\delta_1\in(d+1,\frac{p_3}{2})$ and define $\beta_1:=\frac{p_3}{\delta_1}-2>0$. In particular, we have $(2+\beta_1)\delta_1=p_3$.

Fix a $R>0$, we apply Lemma 6.1 to the function $\Big(|DX_k^{\varepsilon}(F_t^{\varepsilon})-DX_k^{\tilde{\varepsilon}}(F_t^{\varepsilon})|^{2+\beta_1}\,\mathbf{1}_{\{|F_t^{\varepsilon}|\leqslant R\}}\Big)$, and take $p=\delta_1$ in (6.2) to obtain,

$$\begin{aligned}&\limsup_{\varepsilon,\tilde{\varepsilon}\to 0}\sup_{x\in K}\int_0^T\mathbb{E}\left(|DX_k^{\varepsilon}(F_t^{\varepsilon})-DX_k^{\tilde{\varepsilon}}(F_t^{\varepsilon})|^{2+\beta_1}\mathbf{1}_{\{|F_t^{\varepsilon}|\leqslant R\}}\right)dt\\ &\leqslant\limsup_{\varepsilon,\tilde{\varepsilon}\to 0}\,C(K,T)\left(\int_{\{|x|\leqslant R\}}|DX_k^{\varepsilon}(x)-DX_k^{\tilde{\varepsilon}}(x)|^{p_3}\,dx\right)^{\frac{1}{\delta_1}}=0.\end{aligned}\tag{6.13}$$

Here in the second step we also use Lemma 5.3.

By the statement of Lemma 5.2, (1.8) in Assumption 1.1 holds for every $\{X_k^{\varepsilon}\}_{k=0}^m$ with the constants independent of ε. Thus for sufficiently large R we have

$$\sup_{\varepsilon\in(0,\varepsilon_0)}|DX_k^{\varepsilon}(x)|\mathbf{1}_{\{|x|>R\}}\leqslant C(1+|x|^{p_5})\mathbf{1}_{\{|x|>R\}}.$$

Then we obtain

$$\begin{aligned}&\sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{x\in K}\int_0^T\mathbb{E}\left(|DX_k^{\varepsilon}(F_t^{\varepsilon})-DX_k^{\tilde{\varepsilon}}(F_t^{\varepsilon})|^{2+\beta_1}\mathbf{1}_{\{|F_t^{\varepsilon}|>R\}}\right)dt\\ &\leqslant 2C\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x\in K}\int_0^T\mathbb{E}\left(\left(1+|F_t^{\varepsilon}|^{p_5(2+\beta_1)}\right)\mathbf{1}_{\{|F_t^{\varepsilon}|>R\}}\right)dt\\ &\leqslant CR^{-p_5(2+\beta_1)}\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{x\in K}\int_0^T\mathbb{E}\left(1+|F_t^{\varepsilon}|^{2p_5(2+\beta_1)}\right)dt\\ &\leqslant C(K,T)R^{-p_5(2+\beta_1)}.\end{aligned}\tag{6.14}$$

Here in the second step of inequality, we use Hölder inequality and Chebyshev inequality, and the third step is due to the estimate (6.8).

In the inequalities (6.13-6.14) we first let $\varepsilon,\tilde{\varepsilon}\to 0$, then let $R\to 0$, this gives (6.10). □

We will show the pathwise uniqueness for the solution of (1.13).

Proposition 6.4. *Under Assumption 1.1 pathwise uniqueness holds for the solution to (1.13).*

Proof. Given a Brownian motion W_t, suppose (x_t,v_t,W_t,ζ) and $(\tilde{x}_t,\tilde{v}_t,W_t,\tilde{\zeta})$ are two strong solutions to (1.13) with the same initial points, up to the explosion time ζ, $\tilde{\zeta}$. We already know that Assumption 1.1 implies that any solution to SDE (1.2) does not explode and the pathwise uniqueness holds for (1.2), see e.g. [36, Theorem 1.3], i.e. $x_t=\tilde{x}_t$ $\mathbb{P}$-*a.s.*, for every $t\geqslant 0$. Let $\bar{v}_t:=v_t-\tilde{v}_t$, it is easy to see that $\bar{v}_t$ satisfies the following linear equation,

$$d\bar{v}_t=\sum_{k=1}^m DX_k(x_t)(\bar{v}_t)dW_t^k+DX_0(x_t)(\bar{v}_t)dt,\ \ \bar{v}_0=0.$$

Since $DX_k \in L^{p_3}_{\text{loc}}(\mathbb{R}^d;\mathbb{R}^d)$, $1 \leqslant k \leqslant m$, $DX_0 \in L^{p_4}_{\text{loc}}(\mathbb{R}^d;\mathbb{R}^d)$, and by Assumption 1.1, they have polynomial growth outside of B_{R_1}, following the proof of Lemma 6.3, we apply Lemma 4.4 and Example 4.3 to see that

$$\mathbb{E}\left(\int_0^T |DX_k(x_t)|^2 dt\right) < \infty,\ \mathbb{E}\left(\int_0^T |DX_0(x_t)|dt\right) < \infty. \tag{6.15}$$

In particular the integrals in the above stochastic differential equation makes sense.

Set $\bar{\zeta} := \zeta \wedge \tilde{\zeta}$. Applying Itô's formula to $\bar{v}_t$, for every $p > 2$ and any stopping time $\tau < \bar{\zeta}$ we obtain,

$$|\bar{v}_{t\wedge\tau}|^p = |v|^p + \sum_{k=1}^m \int_0^{t\wedge\tau} |\bar{v}_s|^p dM_s + \int_0^{t\wedge\tau} |\bar{v}_s|^p da_s,$$

where the definition of the processes M_s, a_s are the same as that for M^ε_s, a^ε_s by (6.5), but with $\{X^\varepsilon_k, F^\varepsilon_t(x), V^\varepsilon_t(x,v)\}$ replaced by $\{X_k, x_t, \bar{v}_t\}$. The estimates in (6.15) ensure that M_s and a_s are well defined semi-martingales. Following the argument for (6.6), we see that

$$|\bar{v}_{t\wedge\tau}|^p = |\bar{v}_0|^p \exp\left(M_{t\wedge\tau} - \frac{\langle M\rangle_{t\wedge\tau}}{2} + a_{t\wedge\tau}\right) = 0.$$

Thus $v_t = \tilde{v}_t$ $\mathbb{P}$-$a.s.$ for every $t < \tau$. Since τ is arbitrary, we have $\zeta = \tilde{\zeta}$ $\mathbb{P}$-$a.s.$ and $v_t = \tilde{v}_t$ $\mathbb{P}$-$a.s.$ for every $t < \zeta$. By now we have completed the proof.

□

Theorem 6.5. *Suppose that Assumption 1.1 holds. There exists a unique strong solution $(F_t(x), V_t(x,v))$ to (1.13) with initial value $(x,v) \in \mathbb{R}^{2d}$, which is defined for $t \in [0,\infty)$. Furthermore there is a constant $\tilde{T}_0 > 0$, such that for every compact set $\tilde{K} \subseteq \mathbb{R}^{2d}$,*

$$\lim_{\varepsilon\to 0} \sup_{(x,v)\in\tilde{K}} \mathbb{E}\left(\sup_{t\in[0,\tilde{T}_0]} (|F^\varepsilon_t(x) - F_t(x)| + |V^\varepsilon_t(x,v) - V_t(x,v)|)\right) = 0. \tag{6.16}$$

Proof. Through the proof, when the initial value $(x,v) \in \mathbb{R}^{2d}$ is fixed, we denote $(F^\varepsilon_t(x), V^\varepsilon_t(x,v))$ and $(F_t(x), V_t(x,v))$ by $(F^\varepsilon_t, V^\varepsilon_t)$ and (F_t, V_t) respectively for simplicity.

Since pathwise uniqueness for (1.13) is proved in Proposition 6.4, we only need to verify that, with (6.1) as the approximating equations for (1.13), the conditions (1)-(3) in Lemma 3.2 hold. According to Lemma 3.2, this will lead to the conclusion of the existence of a complete strong solution to (1.13) and the convergence in (6.16).

By Lemma 6.3, there exists a $\beta_1 > 0$, such that for every $T > 0$, compact set $K \subseteq \mathbb{R}^d$, $1 \leqslant k \leqslant m$,

$$\sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)} \sup_{x\in K} \mathbb{E}\left(\int_0^T \left|DX^\varepsilon_k\left(F^{\tilde{\varepsilon}}_t\right)\right|^{2+\beta_1} dt\right) < \infty. \tag{6.17}$$

For a $\gamma_1 \in (0,\beta_1)$, let $\alpha = \frac{2+\beta_1}{2+\gamma_1} > 1$ and let $\alpha' = \frac{2+\beta_1}{\beta_1-\gamma_1}$ be conjugate to α . By Lemma 6.2, there is a constant $T_1(\gamma_1,\beta_1) > 0$ such that for every compact set $\tilde{K} \subseteq \mathbb{R}^{2d}$,

$$\sup_{\varepsilon\in(0,\varepsilon_0)} \sup_{(x,v)\in\tilde{K}} \sup_{t\in[0,T_1]} \mathbb{E}\left(\left|V^\varepsilon_t\right|^{(2+\gamma_1)\alpha'}\right) < \infty. \tag{6.18}$$

By Hölder inequality,

$$\begin{aligned}
&\sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{(x,v)\in\tilde{K}}\mathbb{E}\left(\int_0^{T_1}\left|DX_k^{\varepsilon}(F_s^{\tilde{\varepsilon}})\left(V_s^{\tilde{\varepsilon}}\right)\right|^{2+\gamma_1}ds\right)\\
&\leqslant \sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{(x,v)\in\tilde{K}}\left\{\left(\mathbb{E}\left(\int_0^{T_1}\left|DX_k^{\varepsilon}\left(F_s^{\tilde{\varepsilon}}\right)\right|^{2+\beta_1}ds\right)\right)^{\frac{1}{\alpha}}\right.\\
&\qquad\left.\cdot\left(\mathbb{E}\left(\int_0^{T_1}\left|V_s^{\tilde{\varepsilon}}\right|^{(2+\gamma_1)\alpha'}ds\right)\right)^{\frac{1}{\alpha'}}\right\}<\infty.
\end{aligned}\tag{6.19}$$

As the same way, there exist constants $\gamma_2>0$ and $T_2(\gamma_2,\beta_2)>0$, such that for every $p>0$,

$$\begin{aligned}
&\sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{(x,v)\in\tilde{K}}\mathbb{E}\left(\int_0^{T_2}\left|DX_0^{\varepsilon}(F_s^{\tilde{\varepsilon}})\left(V_s^{\tilde{\varepsilon}}\right)\right|^{1+\gamma_2}ds\right)<\infty,\\
&\sup_{\varepsilon,\tilde{\varepsilon}\in(0,\varepsilon_0)}\sup_{(x,v)\in\tilde{K}}\mathbb{E}\left(\int_0^{T_2}\left|X_k^{\varepsilon}(F_s^{\tilde{\varepsilon}})\right|^{p}ds\right)<\infty,\ \forall\,0\leqslant k\leqslant m.
\end{aligned}$$

Combing this with (6.19) we know the condition (3.6) of Lemma 3.2 holds for equation (6.1) in time interval $t\in[0,\tilde{T}_0]$ with $\tilde{T}_0:=\min\{T_1,T_2\}$.

As the same argument above, according to Lemma 6.2 , 6.3 and by Hölder inequality, we conclude that condition (3.7) of Lemma 3.2 for equation (6.1) in time interval $t\in[0,\tilde{T}_0]$.

We proceed to prove the last condition, condition (3.8) in Lemma 3.2. Let $\mu^{\varepsilon,x,v}$ be the distribution of the stochastic process $(F_t^{\varepsilon}(x),V_t^{\varepsilon}(x,v))$ on $\mathbf{W}:=C([0,\tilde{T}_0];\mathbb{R}^{2d})$ and let $\sigma(t)=(\sigma_1(t),\sigma_2(t))$ be the canonical path on $\mathbf{W}$, so the distribution of $\sigma(\cdot)$ under $\mu^{\varepsilon,x,v}$ is the same as that of $(F_\cdot^{\varepsilon}(x),V_\cdot^{\varepsilon}(x,v))$ under $\mathbb{P}$. Suppose that $\{x_n,v_n\}_{n=1}^{\infty}\subseteq\tilde{K}$, $\{\varepsilon_n\}_{n=1}^{\infty}\subseteq(0,\varepsilon_0)$ are sequences such that $\mu^{\varepsilon_n,x_n,v_n}$ converges weakly to some μ^0 as $n\to\infty$. By Lemma 6.1, for every $p>d+1$ and non-negative Borel measurable function $f:\mathbb{R}^d\to\mathbb{R}_+$,

$$\sup_n\int_{\mathbf{W}}\int_0^{\tilde{T}_0}f(\sigma_1(t))dt\;\mu^{\varepsilon_n,x_n,v_n}(d\sigma)\leqslant C(\tilde{K},\tilde{T}_0)\|f\|_p,$$

where $\|f\|_p$ denotes the L^p norm with respect to the Lebesgue measure. If f is furthermore bounded and continuous,

$$\begin{aligned}
&\int_{\mathbf{W}}\int_0^{\tilde{T}_0}f(\sigma_1(t))\;dt\,\mu^0(d\sigma)=\lim_{n\to\infty}\int_0^{\tilde{T}_0}\int_{\mathbf{W}}f\big(\sigma_1(t)\big)\;\mu^{\varepsilon_n,x_n,v_n}(d\sigma)\;dt\\
&\leqslant\sup_n\int_0^{\tilde{T}_0}\int_{\mathbf{W}}f(\sigma_1(t))\;\mu^{\varepsilon_n,x_n,v_n}(d\sigma)\;dt\leqslant C(\tilde{K},\tilde{T}_0)\|f\|_p.
\end{aligned}\tag{6.20}$$

Let $O\subseteq\mathbb{R}^d$ be a bounded open set, there exists a sequence $\{g_n\}_{n=1}^{\infty}$, of non-negative continuous functions with compact supports such that $\sup_{x\in\mathbb{R}^d}|g_n(x)|\leqslant 1$ and $\lim_{n\to\infty}g_n(x)=\mathbf{1}_O(x)$ point wise. Then it follows from the dominated convergence theorem that (6.20) holds with $f(x)=\mathbf{1}_O(x)$. For every bounded measurable set $U\subseteq\mathbb{R}^d$ which is with null Lebesgue measure, from the out regularity of the Lebesgue measure, there exists a sequence of bounded open set $\{O_n\}_{n=1}^{\infty}$ containing U such that $\lim_{n\to\infty}Leb(O_n)=0$. Then putting such $\mathbf{1}_{O_n}$ into (6.20), letting $n\to\infty$, by Fatou lemma we have

$$\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\mathbf{1}_U\big(\sigma_1(t)\big)\;dt\,\mu^0(d\sigma)=0.\tag{6.21}$$

Let $f:\mathbb{R}^d\to\mathbb{R}_+$ be a non-negative bounded Borel measurable function with compact support. There is a sequence, $\{f_n\}_{n=1}^\infty$, of non-negative continuous functions with compact supports and a bounded Lebesgue-null set Q such that $\sup_n\|f_n\|_p\leqslant\|f\|_p$ for all $1\leqslant p\leqslant\infty$, and

$$\lim_{n\to\infty}f_n(x)=f(x),\quad\forall\, x\notin Q. \tag{6.22}$$

It follows that

$$\begin{aligned}
&\lim_{n\to\infty}\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\Big|f_n\big(\sigma_1(t)\big)-f\big(\sigma_1(t)\big)\Big|\,dt\,\mu^0(d\sigma)\\
&\leqslant\lim_{n\to\infty}\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\Big(\big|f_n(\sigma_1(t))-f(\sigma_1(t))\big|\Big)\mathbf{1}_{Q^c}(\sigma_1(t))\,dt\mu^0(d\sigma)\\
&+2\|f\|_{L_\infty}\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\mathbf{1}_Q(\sigma_1(t))dt\,\mu^0(d\sigma)\\
&=\lim_{n\to\infty}\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\big(|f_n(\sigma_1(t))-f(\sigma_1(t))|\big)\mathbf{1}_{Q^c}(\sigma_1(t))\,dt\,\mu^0(d\sigma)=0,
\end{aligned}$$

where in the second step above we use the property (6.21) and the last step is due to (6.22) and the dominated convergence theorem. Hence putting such f_n into (6.20) and letting $n\to\infty$, we know (6.20) holds for every non-negative bounded Borel measurable function with compact support, and by the monotone convergence theorem, (6.20) holds for every non-negative measurable function $f:\mathbb{R}^d\to\mathbb{R}_+$.

Applying (6.20) and Lemma 5.3, and following the proof in Lemma 6.3, for all $1\leqslant k\leqslant m$ we have

$$\lim_{\varepsilon\to0}\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\big|DX_k^\varepsilon(\sigma_1(t))-DX_k(\sigma_1(t))\big|^{2+\beta_1}\,dt\;\mu^0(d\sigma)=0. \tag{6.23}$$

By (6.18), as the same approximation argument for (6.20) we can prove that

$$\sup_{t\in[0,\tilde{T}_0]}\int_{\mathbf{W}}|\sigma_2(t)|^{\frac{(2+\beta_1)(2+\gamma_1)}{\beta_1-\gamma_1}}\;\mu^0(d\sigma)<\infty. \tag{6.24}$$

Following the same procedure for (6.19), by (6.23), (6.24) and Hölder inequality we obtain

$$\lim_{\varepsilon\to0}\int_{\mathbf{W}}\int_0^{\tilde{T}_0}\Big|DX_k^\varepsilon\big(\sigma_1(t)\big)(\sigma_2(t))-DX_k\big(\sigma_1(t)\big)(\sigma_2(t))\Big|^{2+\gamma_1}\,dt\,\mu^0(d\sigma)=0.$$

Similarly, we can prove the corresponding convergence in condition (3.8) of Lemma 3.2 associated with the derivative flow equation (6.1).

By now we have verified all the conditions of Lemma 3.2 hold for (6.1), so there exists a unique complete strong solution (F_t,V_t) for (1.13) in time interval $t\in[0,\tilde{T}_0]$ such that (6.16) holds. Let $\Phi_t(x,v,W_\cdot):=(F_t(x),V_t(x,v))$. For $\tilde{T}_0<t\leqslant2\tilde{T}_0$, we define

$$\Phi_t(x,v,W_\cdot):=\Phi_{t-\tilde{T}_0}\big(F_{\tilde{T}_0}(x),V_{\tilde{T}_0}(x,v),\theta_{\tilde{T}_0}(W)_\cdot\big),$$

where $\theta_{\tilde{T}_0}(W):C([0,\infty);\mathbb{R}^m)\to C([0,\infty);\mathbb{R}^m)$ defined by $\theta_{\tilde{T}_0}(W)_t=W_{t+\tilde{T}_0}-W_{\tilde{T}_0}$ is the time shift operator. By the Markov property and the pathwise uniqueness one may check that this is indeed the solution to SDE (1.13) in $t\in[\tilde{T}_0,2\tilde{T}_0]$. Repeating this procedure, we will obtain a unique global strong solution to SDE (1.13). □

Remark 6.6. *In Assumption 1.1, we assume that the elliptic constant, $|X_k|$ and $|DX_k|$ to grow at most polynomially as $|x|\to\infty$. The reason is that based on (1.6), we have to*

apply the function $g(x) := \log(1+|x|^2)$ in Lemma 4.2 to obtain the uniform integrable property (4.5). If we strengthen (1.6) slightly, see Assumption 6.7 below, we may apply the polynomial function in Lemma 4.2 (see [23, Corollary 6.3]). Moreover, following the same argument in the proof of Theorem 6.5 we will obtain Corollary 6.8.

Assumption 6.7. *Suppose there is a constant $\alpha \in (0, \frac{1}{2}]$ such that the following conditions are satisfied.*

(1) There are positive constants C_1, C_2 such that

$$\sum_{i,j=1}^{d} a_{i,j}(x)\xi_i\xi_j \geqslant \frac{C_1|\xi|^2}{1+\mathrm{e}^{C_2|x|^{2\alpha}}}, \quad \forall x \in \mathbb{R}^n,\ \xi = (\xi_1, ..., \xi_d) \in \mathbb{R}^d.$$

(2) There are positive constants C_3, C_4 such that for all $0 \leqslant k \leqslant m$,

$$|X_k(x)| \leqslant C_3(1+\mathrm{e}^{C_4|x|^{2\alpha}}).$$

There is a constant $\delta \in (0,1]$, and for every $p > 0$ there is a constant $C(p) > 0$ such that

$$\begin{aligned}&\sup_{|y|\leqslant\delta}\Big(\sum_{k=1}^{m} p(1+|x|^{2\alpha})|X_k(x+y)|^2 + \langle x, X_0(x+y)\rangle\Big)\\ &\leqslant C(p)(1+|x|^{2(1-\alpha)}).\end{aligned}$$

(3) Part (3) of Assumption 1.1 holds;

(4) There exists a positive constant $R_1 > 0$, such that for every $p > 1$,

$$K_p(x) \leqslant C(p)(1+|x|^{2\alpha}), \quad |x| > R_1,$$

for some $C(p) > 0$, where the function $K_p(x)$ is defined in part (3) of the Assumption 1.1. Moreover, for all $0 \leqslant k \leqslant m$,

$$|DX_k(x)| \leqslant C_5(1+\mathrm{e}^{C_6|x|^{2\alpha}}), \quad |x| > R_1,$$

for some positive constants C_5, C_6.

Corollary 6.8. *The conclusion of Theorem 6.5 holds with Assumption 1.1 replaced by Assumption 6.7.*

7 Proof of Theorem 1.2

Let $(F_t^\varepsilon(x), V_t^\varepsilon(x,v))$ be the solution to (6.1) with initial point $(x,v) \in \mathbb{R}^{2d}$, since $X_k^\varepsilon \in C_b^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $x \mapsto F_t^\varepsilon(x)$ is differentiable and $V_t^\varepsilon(x,v) = D_xF_t^\varepsilon(x)(v)$, $\mathbb{P}$-a.s.. For any given $R > 0$, $p > 1$, and for all $x, y \in B_R := \{x \in \mathbb{R}^d;\ |x| \leqslant R\}$, $t > 0$,

$$\begin{aligned}&\mathbb{E}\left(|F_t^\varepsilon(x) - F_t^\varepsilon(y)|^p\right) = \mathbb{E}\left(\left|\left\langle x-y, \int_0^1 D_xF_t^\varepsilon(x+s(y-x))ds\right\rangle\right|^p\right)\\ &\leqslant C|x-y|^p \sup_{x\in B_{2R}, |v|\leqslant 1}\mathbb{E}\left(|V_t^\varepsilon(x,v)|^p\right).\end{aligned} \tag{7.1}$$

Here the notation $D_xF_t^\varepsilon(x+s(y-x))$ denotes the derivative (with respect to the space variable) of the solution flow $F_t^\varepsilon(\cdot)$ at the point $x+s(y-x) \in \mathbb{R}^d$. See also the analysis in the proof of Theorem 4.1 in [23].

According to (4.5) and Lemma 5.2, for every $p>1$, $T>0$, and K compact,

$$\sup_{x\in K}\sup_{\varepsilon\in(0,\varepsilon_0)}\mathbb{E}\big(|F_t^\varepsilon(x)|^{p+1}\big)<\infty,$$

which implies that $\{|F_t^\varepsilon(x)|^p\}_{\varepsilon\in(0,\varepsilon_0),x\in K}$ is uniformly integrable. So by Theorem 6.5 we derive for every $t\in[0,\tilde{T}_0]$,

$$\lim_{\varepsilon\to 0}\mathbb{E}\left(|F_t^\varepsilon(x)-F_t(x)|^p\right)=0,\ \forall x\in\mathbb{R}^d,$$

where $\tilde{T}_0$ is the constant in Theorem 6.5.

Let $\hat{T}_0(p):=\min\{\tilde{T}_0,T_0(p)\}$, where $T_0(p)$ is the constant in Lemma 6.2. Therefore according to Lemma 6.2, we take the limit $\varepsilon\to 0$ in (7.1) to obtain for every $t\in[0,\hat{T}_0]$, $x,y\in B_R$,

$$\begin{aligned}&\mathbb{E}\left(|F_t(x)-F_t(y)|^p\right)\\&\leqslant C|x-y|^p\sup_{\varepsilon\in(0,\varepsilon_0)}\sup_{t\in[0,\hat{T}_0]}\sup_{z\in B_{2R},|v|\leqslant 1}\mathbb{E}\left(|V_t^\varepsilon(z,v)|^p\right)\\&\leqslant C(\hat{T}_0,R)|x-y|^p,\end{aligned}$$

Since X_k are polynomial growth, it is easy to show for every $0\leqslant s\leqslant t\leqslant\hat{T}_0(p)$, $x,y\in B_R$,

$$\mathbb{E}\big(|F_t(x)-F_s(y)|^p\big)\leqslant C(R,\hat{T}_0)\big(|x-y|^p+|t-s|^{\frac{p}{2}}\big).$$

In the above estimate, noting that R is arbitrary large, and we may take $p>2(d+1)$ and apply Kolmogorov's continuity criterion to conclude that there is a version of the solution flow $F_t(x,\omega)$ for SDE (1.2), such that $F_\cdot(\cdot,\omega)$ is continuous in $[0,\hat{T}_0]\times\mathbb{R}^d$.

As for $t>\hat{T}_0$, let $\Psi_t(x,W_\cdot):=F_t(x,\omega)$. By the Markov property and the uniqueness of the strong solution to SDE (1.2), it is satisfied that

$$F_t(x,\omega)=\Psi_t(x,W_\cdot)=\Psi_{t-\hat{T}_0}(F_{\hat{T}_0}(x,\omega),\theta_{\hat{T}_0}(W)_\cdot),\ \mathbb{P}-a.s.$$

where $\theta_{\hat{T}_0}(W)_t=W_{t+\hat{T}_0}-W_{\hat{T}_0}$ is the time shift operator. Hence the solution flow $F_\cdot(\cdot,\omega)$ is continuous in $[0,2\hat{T}_0]\times\mathbb{R}^d$, and in $[0,\infty)\times\mathbb{R}^d$ by repeating the procedure.

Let $\{e_i\}_{i=1}^d$ be an orthonormal basis of $\mathbb{R}^d$ and $(F_t(x),V_t(x,v))$ be the strong solution to (1.13) with initial point $(x,v)\in\mathbb{R}^{2d}$. By Theorem 6.5 and the diagonal principle there exist a subsequence $\{\varepsilon_n\}_{n=1}^\infty$ with $\lim_{n\to\infty}\varepsilon_n=0$ and a set $\tilde{\Lambda}_0\subseteq\Omega$ with $\mathbb{P}(\tilde{\Lambda}_0)=0$, such that if $\omega\in\tilde{\Lambda}_0^c$, for every $R>0$, $1\leqslant i\leqslant d$,

$$\lim_{n\to\infty}\int_{\{|x|\leqslant R\}}\sup_{t\in[0,\tilde{T}_0]}|V_t^{\varepsilon_n}(x,e_i,\omega)-V_t(x,e_i,\omega)|dx=0,\tag{7.2}$$

$$\lim_{n\to\infty}\int_{\{|x|\leqslant R\}}\sup_{t\in[0,\tilde{T}_0]}|F_t^{\varepsilon_n}(x,\omega)-F_t(x,\omega)|dx=0.\tag{7.3}$$

For simplicity we write $(F_t^n(x),V_t^n(x,e_i))$ for $(F_t^{\varepsilon_n}(x),V_t^{\varepsilon_n}(x,e_i))$. As referred above, $D_xF_t^n(x)(v)=V_t^n(x,v)$ a.s., therefore there exists a $\mathbb{P}$-null set Λ_n, such that for every $\omega\in\Lambda_n^c$, the following integration by parts formula holds for every $1\leqslant i\leqslant d$, $t\in[0,\tilde{T}_0]$ and $\varphi\in C_0^\infty(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d}\frac{\partial\varphi}{\partial x_i}(x)F_t^n(x,\omega)dx=-\int_{\mathbb{R}^d}\varphi(x)V_t^n(x,e_i,\omega)dx.\tag{7.4}$$

Let $\tilde{\Lambda}:=(\bigcup_{n=1}^\infty\Lambda_n)\cup\tilde{\Lambda}_0$, then $\tilde{\Lambda}$ is a $\mathbb{P}$-null set. Taking n to infinity in (7.4) and using (7.2), (7.3) we see for every $1\leqslant i\leqslant d$, $\omega\in\tilde{\Lambda}^c$, $t\in[0,\tilde{T}_0]$,

$$\int_{\mathbb{R}^d}\frac{\partial\varphi}{\partial x_i}(x)F_t(x,\omega)dx=-\int_{\mathbb{R}^d}\varphi(x)V_t(x,e_i,\omega)dx$$

which means that $F_t(\cdot,\omega)$ is weakly differentiable in the distribution sense for almost surely all ω and $D_xF_t(x,\omega)(e_i) = V_t(x,e_i,\omega)$. Next we prove that given a $p>1$, there exist a $T_1>0$, such that for every $t\in[0,T_1]$, $F_t(\cdot,\omega)\in W^{1,p}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$, a.s..

By Lemma 6.2, Theorem 6.5 and Fatou Lemma, given a $p>1$, there is a constant $0<T_1\leqslant\tilde{T}_0$, such that for every $R>0$, $t\in[0,T_1]$,

$$\mathbb{E}\left(\int_{B_R}|V_t(x,e_i)|^p dx\right)=\int_{B_R}\mathbb{E}\left(|V_t(x,e_i)|^p\right)dx\leqslant C(R,T_1).$$

Hence for every fixed $t\in[0,T_1]$, we can find a $\mathbb{P}$-null set Λ_0 (that may depend on t), such that $\int_{B_R}|V_t(x,e_i,\omega)|^p dx<\infty$ for every $R>0$, $1\leqslant i\leqslant d$ when $\omega\in\Lambda_0^c$. As the same way, we can prove the similar integrable property for $F_t(x,\omega)$. Therefore $F_t(x,\omega),V_t(x,e_i,\omega)\in L^p_{\rm loc}(\mathbb{R}^n)$ for $\omega\in\left(\Lambda_0\cup\tilde{\Lambda}\right)^c$. In particular, $\Lambda:=\Lambda_0\cup\tilde{\Lambda}$ is a $\mathbb{P}$-null set. We proved that for every $t\in[0,T_1]$, $F_t(\cdot,\omega)\in W^{1,p}_{\rm loc}(\mathbb{R}^d;\mathbb{R}^d)$, $\mathbb{P}$-a.s..

8 The differentiation formula

Suppose that Assumption 1.1 holds, let $(F_t(x),V_t(x,v))$ be the unique strong solution of (1.13) with initial point $(x,v)\in\mathbb{R}^{2d}$. For $f\in C_b(\mathbb{R}^d)$ we define $P_tf(x):=\mathbb{E}\left(f(F_t(x))\right)$ and let $Y:\mathbb{R}^d\to L(\mathbb{R}^d,\mathbb{R}^m)$ be the right inverse of map $X:\mathbb{R}^d\to L(\mathbb{R}^m,\mathbb{R}^d)$, where

$$X(x)(\xi):=\sum_{k=1}^m\xi_kX_k(x)\qquad\text{for }\ \xi=(\xi_1,\xi_2,\ldots,\xi_m)\in\mathbb{R}^m. \tag{8.1}$$

Theorem 8.1. *Suppose that Assumption 1.1 holds. There is a positive constant T_2, such that for every $v\in\mathbb{R}^d$, $f\in C_b(\mathbb{R}^d)$, $t\in(0,T_2]$,*

$$D_x(P_tf)(v)=\frac{1}{t}\mathbb{E}\left(f\big(F_t(x)\big)\int_0^t\Big\langle Y(F_s(x))(V_s(x,v)),dW_s\Big\rangle_{\mathbb{R}^m}\right). \tag{8.2}$$

Proof. We first assume that $f\in C_b^1(\mathbb{R}^d)$. Since the coefficients of SDE (6.1) are smooth , uniformly elliptic, and with bounded derivatives, by the classical differential formula in [25] and [8], we have for every $t>0$,

$$D_x\mathbb{E}\left(f(F_t^\varepsilon(x))\right)(v)=\frac{1}{t}\mathbb{E}\left(f(F_t^\varepsilon(x))\int_0^t\langle Y^\varepsilon(F_s^\varepsilon(x))(V_s^\varepsilon(x,v)),dW_s\rangle_{\mathbb{R}^m}\right), \tag{8.3}$$

where $(F_t^\varepsilon(x),V_t^\varepsilon(x,v))$ is the strong solution to (6.1) with initial point $(x,v)\in\mathbb{R}^{2d}$, $Y^\varepsilon:\mathbb{R}^d\to L(\mathbb{R}^d,\mathbb{R}^m)$ is the right inverse of map $X^\varepsilon:\mathbb{R}^d\to L(\mathbb{R}^m,\mathbb{R}^d)$.

Since $f\in C_b^1(\mathbb{R}^d)$, by Theorem 6.5, Lemma 6.2 and Hölder inequality, there is a constant $T_2>0$, such that for any bounded set K in $\mathbb{R}^d$,

$$\lim_{\varepsilon\to0}\sup_{x\in K}\sup_{t\in[0,T_2]}\mathbb{E}\left(|f(F_t^\varepsilon(x))-f(F_t(x))|^8\right)=0 \tag{8.4}$$

$$\lim_{\varepsilon\to0}\sup_{x\in K}\sup_{t\in[0,T_2]}\mathbb{E}\left(|V_s^\varepsilon(x,v)-V_s(x,v)|^8\right)=0 \tag{8.5}$$

Let $A^\varepsilon:=(X^\varepsilon)^*X^\varepsilon$, where $*$ denotes taking the transpose. Then we have

$$Y^\varepsilon=(X^\varepsilon)^*(A^\varepsilon)^{-1}.$$

In particular, if we write $X_k^\varepsilon=(X_{k1}^\varepsilon,\ldots,X_{kd}^\varepsilon)$, $A^\varepsilon=(a_{i,j}^\varepsilon)_{i,j=1}^n$, then $a_{i,j}^\varepsilon=\sum_{k=1}^mX_{ki}^\varepsilon X_{kj}^\varepsilon$, and for every $\xi=(\xi_1,\ldots,\xi_d)\in\mathbb{R}^d$, $Y^\varepsilon(x)(\xi)=(\zeta_1^\varepsilon(x),\zeta_2^\varepsilon(x)\ ...,\zeta_m^\varepsilon(x))$, where $\zeta_k^\varepsilon(x)=\sum_{i,j=1}^dX_{ki}^\varepsilon(x)b_{i,j}^\varepsilon(x)\xi_j$, and $(b_{i,j}^\varepsilon)=(A^\varepsilon)^{-1}$.

By Lemma 5.2,

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\left|\left(A^{\varepsilon}(x)\right)^{-1}\right|\leqslant C(1+|x|^q)$$

for some $q>0$. Combining this with (6.9) and Theorem 6.5, it is easy to show for every compact set $K\subseteq\mathbb{R}^d$,

$$\lim_{\varepsilon\to 0}\sup_{x\in K}\mathbb{E}\left(\int_0^{T_2}|b^{\varepsilon}_{i,j}(F^{\varepsilon}_t(x))-b_{i,j}(F_t(x))|^8dt\right)=0.$$

This together with the convergence (8.5) leads to

$$\lim_{\varepsilon\to 0}\sup_{x\in K}\mathbb{E}\left(\int_0^{T_2}|Y^{\varepsilon}(F^{\varepsilon}_t(x))(V^{\varepsilon}_t(x,v))-Y(F_t(x))(V_t(x,v))|^4dt\right)=0.$$

Then by (8.4) and BDG inequality, we see that for every $t\in[0,T_2]$ and compact set $K\subseteq\mathbb{R}^d$,

$$\begin{aligned}\lim_{\varepsilon\to 0}\sup_{x\in K}\Big|&\mathbb{E}\left(f(F^{\varepsilon}_t(x))\int_0^t\langle Y^{\varepsilon}(F^{\varepsilon}_s(x))(V^{\varepsilon}_s(x,v)),dW_s\rangle_{\mathbb{R}^m}\right)\\&-\mathbb{E}\left(f(F_t(x))\int_0^t\langle Y(F_s(x))(V_s(x,v)),dW_s\rangle_{\mathbb{R}^m}\right)\Big|=0.\end{aligned}$$

which implies the differentiation formula (8.2) holds for each $f\in C^1_b(\mathbb{R}^d)$.

For $f\in C_b(\mathbb{R}^d)$, there is a sequence of functions $\{f_n\}_{n=1}^{\infty}\subseteq C^1_b(\mathbb{R}^d)$, such that $\sup_n||f_n||_{\infty}\leqslant||f||_{\infty}$, and for every $R>0$,

$$\lim_{n\to\infty}\sup_{\{|x|\leqslant R\}}|f_n(x)-f(x)|=0.$$

Therefore for every $R>0$, $t\in(0,T_2]$,

$$\begin{aligned}&\mathbb{E}\left(|f_n(F_t(x))-f(F_t(x))|^2\right)\\&\leqslant\sup_{\{|x|\leqslant R\}}|f_n(x)-f(x)|^2+C||f||^2_{\infty}\mathbb{P}(|F_t(x)|>R)\\&\leqslant\sup_{\{|x|\leqslant R\}}|f_n(x)-f(x)|^2+\frac{C||f||^2_{\infty}\mathbb{E}(|F_t(x)|)}{R},\end{aligned}$$

and by (4.5), first let $n\to 0$ and then $R\to\infty$, we obtain that for every compact set $K\subseteq\mathbb{R}^d$,

$$\lim_{n\to\infty}\sup_{x\in K}\mathbb{E}\left(|f_n(F_t(x))-f(F_t(x))|^2\right)=0,$$

which proves that (8.2) holds by standard approximation argument. □

References

[1] J. N. Blagoveščenskiĭ, and M. I. Freĭdlin.: Some properties of diffusion processes depending on a parameter, *Dokl. Akad. Nauk SSSR* **138** (1961), 508–511. MR-0139196

[2] N. Bouleau and F. Hirsch.: Dirichlet forms and analysis on Wiener space, *de Gruyter Studies in Mathematics* **138**, Walter de Gruyter & Co, Berlin, (1991). MR-1133391

[3] X. Chen and X.-M. Li.: An approximation scheme for SDEs with non-smooth coefficients, arXiv:1008.0899.

[4] X. Chen.: University of Warwick Ph.D. thesis (2011).

[5] S. Cox, M. Hutzenthaler and A. Jentzen.: Local Lipschitz continuity in the initial value and strong completeness for nonlinear stochastic differential equations, arXiv:1309.5595.

[6] A. M. Davie.: Uniqueness of solutions of stochastic differential equations, *Int. Math. Res. Not.* **24** (2007). MR-2377011

[7] K. D. Elworthy.: Stochastic differential equation on manifolds, *Lecture notes series* **70**, Cambridge University Press, Cambridge, (1982). MR-0675100

[8] K. D. Elworthy and X.-M. Li.: Formulae for the derivatives of heat semigroups, *J. Funct. Anal.* **125**(1) (1994), 252–286. MR-1297021

[9] L. C. Evans.: Partial differential equations, *Graduate Studies in Mathematics* **19**, American Mathematical Society, Providence, RI, (1998). MR-1625845

[10] S. Z. Fang, P. Imkeller and T. S. Zhang.: Global flows for stochastic differential equations without global Lipschitz conditions, *Ann. Probab.* **35**(1) (2007), 180–205. MR-2303947

[11] S. Z. Fang and D. J. Luo.: Flow of homeomorphisms and stochastic transport equations, *Stoch. Anal. and Appl.* **25** (2007), 1179–1208. MR-2352953

[12] S. Z. Fang and T. S. Zhang.: A study of a class of stochastic differential equations with non-Lipschitzian coefficients, *Probab. Theory Related Fields.* **132**(3) (2005), 356–390. MR-2197106

[13] E. Fedrizzi and F. Flandoli.: Pathwise uniqueness and continuous dependence of SDEs with non-regular drift, *Stochastics.* **83**(3) (2003), 241–257. MR-2810591

[14] F. Flandoli, M. Gubinelli,and E. Priola.: Flow of diffeomorphisms for SDEs with unbounded Hölder continuous drift, *Bull. Sci. Math.* **134**(4) (2010), 405–422. MR-2651899

[15] I. Gyöngy and T. Martinez.: On stochastic differential equations with locally unbounded drift, *Czechoslovak Math. J.* **51**(4) (2001), 763–783. MR-1864041

[16] M. Hairer, M. Hutzenthaler, and A. Jentzen.: Loss of regularity for Kolmogorov equations, to appear in Ann. Probab, arXiv:1209.6035.

[17] N. Ikeda and S. Watanabe.: Stochastic differential equations and diffusion processes, *North-Holland Mathematical Library* **24** North-Holland Publishing Co., Amsterdam, (1989). MR-1011252

[18] H. Kaneko and S. Nakao.: A note on approximation for stochastic differential equations, Séminaire de Probabilités, XXII *Lecture Notes in Math.* **1321** Springer, Berlin, (1988), 155–162. MR-0960522

[19] N. V. Krylov.: Estimates of the maximum of the solution of a parabolic equation and estimates of the distribution of a semimartingale, *Mat. Sb. (N.S.)* **130**(2) (1986), 207–221. MR-0854972

[20] N. V. Krylov.: Lectures on elliptic and parabolic equations in sobolev spaces, *Graduate Studies in Mathematics* **96** American Mathematical Society, Providence, RI, (2008). MR-2435520

[21] N. V. Krylov and M. Röckner.: Strong solutions of stochastic equations with singular time dependent drift, *Probab. Theory Related Fields.* **131**(2) (2005), 154–196. MR-2117951

[22] H. Kunita.: Stochastic flows and stochastic differential equations, *Cambridge Studies in Advanced Mathematics* **24** Cambridge University Press, Cambridge, (1990). MR-1070361

[23] X.-M. Li.: Strong p-completeness of stochastic differential equations and the existence of smooth flows on noncompact manifolds, *Probab. Theory Related Fields.* **100**(4) (1994), 485–511. MR-1305784

[24] X.-M. Li.: Stochastic differential equations on non-compact manifolds: moment stability and its topological consequences. *Probab. Theory Related Fields.* **100**(4), 417-428. (1994). MR-1305780

[25] X.-M. Li.: Stochastic Flows on Non-compact Manifolds, The University of Warwick Ph. D. Thesis, 1993.

[26] X.-M. Li and M. Scheutzow.: Lack of strong completeness for stochastic flows, *Ann. Probab.* **39**(4) (2001), 1407–1421. MR-2857244

[27] S. E. A. Mohammed, T. Nilssen and F. Proske.: Sobolev differentiable stochastic flows of SDEs with measurable drift and applications, arXiv:1204.3867.

[28] N. I. Portenko.: Generalized diffusion processes. Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975), pp. 500–523. *Lecture Notes in Math.* **550**, Springer, Berlin, (1976). MR-0440716

[29] D. Revuz, M. Yor.: Continuous Martingales and Brownian Motion, second ed., Springer-Verlag, (1994). MR-1303781

[30] R. Rebolledo.: La méthode des martingales appliquée á l'étude de la convergence en loi de processus, *Bull. Soc. Math. France Mém* **62** (1979). MR-0568153

[31] A. V. Skorohod.: Studies in the Theory of Random Processes, Izdat. Kiev. Univ., Kiev, (1961), English transl.: Addison Wesley Publ. Co., Reading, Mass., (1965). MR-0185620

[32] S. Taniguchi.: Stochastic flows of diffeomorphisms on an open set of $\mathbb{R}^n$, *Stochastics* **28**(4) (1989), 301–315. MR-1028536

[33] A. Veretennikov.: Strong solutions and explicit formulas for solutions of stochastic integral equations, *Mat. Sb. (N.S.)* **111**(3) (1980), 434–452. MR-0568986

[34] X. C. Zhang.: Strong solutions of SDEs with singular drift and Sobolev diffusion coefficients, *Stochastic Process. Appl.* **115**(11) (2005), 1805–1818. MR-2172887

[35] X. C. Zhang.: Stochastic flows and Bismut formulas for stochastic Hamiltonian systems, *Stochastic Process. Appl.* **120**(10) (2010), 1929–1949. MR-2673982

[36] X. C. Zhang.: Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients, *Electron. J. Probab.* **16**(38) (2011), 1096–1116. MR-2820071

[37] W. A. Zheng.: Tightness results for laws of diffusion processes application to stochastic mechanics, *Ann. Inst. H. Poincaré Probab. Statist.* **21**(2) (1985), 103–124. MR-0798890

[38] A. K. Zvonkin.: A transformation of the phase space of a diffusion process that will remove the drift, *Mat. Sb. (N.S.)* **93** (1974), 129–149. MR-0336813

[39] A. K. Zvonkin and N. V. Krylov.: Strong solutions of stochastic differential equations, *Sel. Math. Sov* **1** (1981), 19–61.