Electronic Journal of Probability

Electron. J. Probab. **19** (2014), no. 73, 1–26.
ISSN: 1083-6489 DOI: 10.1214/EJP.v19-3120

Malliavin matrix of degenerate SDE and gradient estimate*

Zhao Dong‡ Xuhui Peng†

Abstract

In this article, we prove that the inverse of the Malliavin matrix belongs to $L^p(\Omega,\mathbb{P})$ for a class of degenerate stochastic differential equation(SDE). The conditions required are similar to Hörmander's bracket condition, but we don't need all coefficients of the SDE are smooth. Furthermore, we obtain a locally uniform estimate for the Malliavin matrix and a gradient estimate. We also prove that the semigroup generated by the SDE is strong Feller. These results are illustrated through examples.

Keywords: Degenerate stochastic differential equation; Gradient estimate; Strong Feller; Malliavin calculus; Hörmander's bracket condition.
AMS MSC 2010: 60H10; 60H07.
Submitted to EJP on November 7, 2013, final version accepted on August 10, 2014.

1 Introduction and Notations

In this article, we consider the following degenerate stochastic differential equations(SDE)

$$\begin{cases} x_t = x + \int_0^t a_1(x_s, y_s)ds, \\ y_t = y + \int_0^t a_2(x_s, y_s)ds + \int_0^t b(x_s, y_s)dW_s. \end{cases} \tag{1.1}$$

In the above $x \in \mathbb{R}^m,\ y \in \mathbb{R}^n, b \in \mathbb{R}^{n\times d},\ W_s$ is a d-dimensional standard Brownian motion. Eq.(1.1) is a model for many physical phenomenons. For example, x_t represents the position of an object and y_t represents the momentum of the object. When a random force affects the object, first the momentum of the object changes, then that will lead to the change of position. To understand the long time behavior of the movement of the object, we need to study the ergodicity of Eq.(1.1). For this reason, the gradient estimate of the semigroup and the strong Feller property associated to the solution

*Supported by 973 Program, No. 2011CB808000 and Key Laboratory of Random Complex Structures and Data Science, No.2008DP173182, NSFC, No.:10721101, 11271356, 11371041.

†Corresponding author. College of Mathematics and Computer Science, Hunan Normal University, Changsha, 410081, P.R.China. E-mail: pengxuhui@amss.ac.cn

‡Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, P.R.China. E-mail: dzhao@amt.ac.cn

should be considered. Furthermore, the solution is ergodic if one also knows that the solution is topologically irreducible and has an invariant probability measure.

Let $\mathbb{P}_{x,y}$ be the law of the solution to Eq.(1.1) with initial value (x,y), and P_t be the transition semigroup of Eq.(1.1)

$$P_t f(x,y) := \mathbb{E}_{x,y} f(x_t, y_t),\ f \in \mathscr{B}_b(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}),$$

where $\mathscr{B}_b(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R})$ denotes the collection of bounded Borel measurable functions.

Consider general SDE

$$X_t = x + \int_0^t V_0(X_s)ds + \sum_{j=1}^d \int_0^t V_j(X_s) \circ dW_j(s),\quad x \in \mathbb{R}^{m+n}. \tag{1.2}$$

Let $V = (V_1, \cdots, V_d)$. The Hörmander's bracket condition $(\mathbf{H})$ for (V_0, V) is that the vector space spanned by the vector fields

$$(\mathbf{H})\quad V_1, \cdots, V_d,\ [V_i, V_j], 0 \leqslant i,j \leqslant d,\ [[V_i, V_j], V_k],\ 0 \leqslant i,j,k \leqslant d,\ \cdots,$$

at point x is $\mathbb{R}^{m+n}$. The coefficients are infinitely differentiable functions with bounded partial derivatives of all order. If the Hörmander's bracket condition $(\mathbf{H})$ holds for any $x \in \mathbb{R}^{m+n}$, then X_t has smooth density and the transition semigroup of Eq.(1.2) is strong Feller (see [10], [12], [15], [18] etc). Bell and Mohammed [1, Theorem 1.0 and Theorem 1.1] proved the hypoellipticity of a large class of highly degenerate second order differential operators, where the Hörmander's bracket condition may fail on a collection of hypersurfaces.

Let $P_t(x, \cdot)$ be the transition probabilities of the X_t in Eq.(1.2). When VV^* is uniformly elliptic, two-sided bounds of the density for $P_t(x, \cdot)$ were given in Sheu [19] by using stochastic control tools. There are many other excellent works in such direction.

Also, there are many works in the hypoelliptic setting. For the case $V_0 \equiv 0$, Kusuoka and Stroock [12] gave the two-sided bounds of the density for $P_t(x, \cdot)$ under some conditions which require certain uniformity on $V_1, \cdots, V_d$. Recently, Delarue and Menozzi [5] considered the following SDE,

$$\begin{cases} X_t^1 = x_1 + \int_0^t F_1(s, X_s^1, \cdots, X_s^n)ds + \int_0^t b(s, X_s^1, \cdots, X_s^n)dW_s, \\ X_t^2 = x_2 + \int_0^t F_2(s, X_s^1, \cdots, X_s^n)ds, \\ X_t^3 = x_3 + \int_0^t F_3(s, X_s^2, \cdots, X_s^n)ds, \\ \quad \vdots \\ X_t^n = x_n + \int_0^t F_n(s, X_s^{n-1}, X_s^n)ds. \end{cases} \tag{1.3}$$

If the spectrum of the $A(t,x) = [bb^*](t,x)$ is included in $[\Lambda^{-1}, \Lambda]$ for some $\Lambda \geqslant 1$ and $D_{x_{i-1}} F_i(t, x_{i-1}, x_i, \cdots, x_n)$ is non-degenerate, uniformly in space and time, they gave two-sided bounds of the density for to the solution to Eq.(1.3). Cattiaux et al. [3] considered the SDE as

$$X_t^i = x_i + W_t^i,\ \ \forall i \in [1,n],\quad X_t^{n+1} = x_{n+1} + \int_0^t |X_s^{1,n}|^k ds, \tag{1.4}$$

here $X_s^{1,n} = (X_s^1, \cdots, X_s^n)$ and they gave two-sided bounds estimation for the transition function $p(t, x, \cdot)$. There are also many other results on the special case of Eq.(1.1), such as [14], [11], [21] etc.

In most of the above works, the coefficients are smooth or some uniform conditions are needed. Since our aim in this article is to prove the strong Feller property and give a gradient estimate of the semigroup, we don't need the smooth conditions for all the coefficients or some uniform conditions. Instead of the Hörmander's bracket conditions, we give some new conditions, which are equivalent to the Hörmander's bracket condition if the coefficients are smooth. We prove that the inverse of the Malliavin matrix is L^p integrable for any $p \geq 0$. Furthermore, our conditions also ensure that we can obtain a gradient estimate and the strong Feller property. We haven't investigated the smoothness of density or the two-sided bounds of density when smoothness or some uniform conditions on the coefficients are absent.

We introduce some notations. For $j \in \mathbb{N}$, let $C^j(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^l)$ be the collection of functions which have continuous derivatives up to order j, $C_b^j(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^l)$ the collection of functions in $C^j(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^l)$ with bounded derivatives. Sometimes, we will use C_b^j and C^j for convenience. For $l \in \mathbb{N},\ k = (k_1(x,y), \cdots, k_l(x,y))^* \in C^1(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^l)$, $x = (x_1, \cdots, x_m)^*,\ y = (y_1, \cdots, y_n)^*$,

$$\begin{aligned} \nabla_{x_i} k &= \left(\frac{\partial k_1}{\partial x_i}, \cdots, \frac{\partial k_l}{\partial x_i}\right)^*,\ i = 1, \cdots, m,\ \nabla_x k = (\nabla_{x_1} k, \cdots, \nabla_{x_m} k), \\ \nabla_{y_j} k &= \left(\frac{\partial k_1}{\partial y_j}, \cdots, \frac{\partial k_l}{\partial y_j}\right)^*,\ j = 1, \cdots, n,\ \nabla_y k = (\nabla_{y_1} k, \cdots, \nabla_{y_n} k), \end{aligned}$$

and $\nabla k = (\nabla_x k, \nabla_y k)$, where "*" denotes the transpose of vector or matrix. If $a_1 \in C^{j_0}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m)$ for some $j_0 \in \mathbb{N}$, we define vector fields:

$$\begin{aligned} &\mathcal{A}_1 = \left\{\nabla_{y_j} a_1,\ j = 1, \cdots, n\right\}, \\ &\mathcal{A}_l = \left\{\nabla_{y_j} k,\ j = 1, \cdots, n,\ -\nabla_x a_1 \cdot k + \nabla_x k \cdot a_1 : k \in \mathcal{A}_{l-1}\right\},\ l = 2, \cdots, j_0. \end{aligned}$$

Assume $a_1 = (a_1^1, \cdots, a_1^m)^*,\ a_2 = (a_2^1, \cdots, a_2^n)^*,\ a = (a_1^*, a_2^*)^*$. $\mathbb{N} = \{1, \cdots\}$. Let $\det(A)$ be the determinant of the matrix $A = (a_{ij})$, $\|A\|^2 = \sum_{ij} a_{ij}^2$. Let $\langle \cdot, \cdot \rangle$ be the Euclidean inner product and $|\cdot|$ be the Euclidean norm. For any $x_0 \in \mathbb{R}^{m+n}$ and $R > 0$, $\overline{B}(x_0, R) = \{x \in \mathbb{R}^{m+n}, |x - x_0| \leqslant R\}$, $B^\circ(x_0, R) = \{x \in \mathbb{R}^{m+n}, |x - x_0| < R\}$ and $\overline{B}_R := \overline{B}(0, R), B_R^\circ := B^\circ(0, R)$. $\|\cdot\|_\infty$ denotes the essential supreme norm. We use $C(d)$ or $\epsilon_0(d)$ to denote a positive and finite constant depending on $d,\ \|\nabla a\|_\infty$ and $\|\nabla b\|_\infty$. This constant may change from line to line. Sometimes, we will use C instead of $C(d)$ for the convenience of writing. Without specified, (x_t, y_t) is the solution for Eq.(1.1) and (x, y) is its initial value. Let M_t be the Malliavin matrix of (x_t, y_t). It is known that (c.f. [15])

$$M_t = J_t \int_0^t J_s^{-1} \begin{pmatrix} 0 \\ b(x_s, y_s) \end{pmatrix} \begin{pmatrix} 0 \\ b(x_s, y_s) \end{pmatrix}^* (J_s^{-1})^* ds J_t^*, \tag{1.5}$$

here J_t^{-1} satisfies

$$\begin{aligned} J_t^{-1} = I_{m+n} &- \int_0^t J_s^{-1} \begin{pmatrix} 0 & 0 \\ \nabla_x b_j & \nabla_y b_j \end{pmatrix} (x_s, y_s) dW_j(s) \\ &- \int_0^t J_s^{-1} \Bigg[\begin{pmatrix} \nabla_x a_1 & \nabla_y a_1 \\ \nabla_x a_2 & \nabla_y a_2 \end{pmatrix} (x_s, y_s) \\ &\qquad - \sum_{j=1}^d \begin{pmatrix} 0 & 0 \\ \nabla_y b_j \nabla_x b_j & \nabla_y b_j \nabla_y b_j \end{pmatrix} (x_s, y_s) \Bigg] ds, \end{aligned} \tag{1.6}$$

and J_t satisfies

$$\begin{aligned} J_t = I_{m+n} &+ \int_0^t \begin{pmatrix} \nabla_x a_1 & \nabla_y a_1 \\ \nabla_x a_2 & \nabla_y a_2 \end{pmatrix} (x_s, y_s) J_s ds \\ &+ \sum_{j=1}^d \int_0^t \begin{pmatrix} 0 & 0 \\ \nabla_x b_j & \nabla_y b_j \end{pmatrix} (x_s, y_s) J_s dW_j(s). \end{aligned} \tag{1.7}$$

Our article is organized as follows. In section 2, we prove $\det M_t^{-1} \in L^p(\Omega, \mathbb{P}), \forall\, p > 1$ in Theorem 2.2 under Hypothesis 2.1. In the Hypothesis 2.1, we need $a_2 \in C_b^1,\ b \in C^2 \cap C_b^1$ and $a_1 \in C^{j_0+2} \cap C_b^1$ for some $j_0 \in \mathbb{N}$. Compared with Hörmander's bracket condition, the ąğ functions a_2 and b are only required to be C_b^1 and $C_b^1 \cap C^2$ respectively. Our approach is mainly along the lines of [15], but has some differences and needs more complicated computation. These differences depend heavily on the special form of the Eq.(1.1). In [15], J_t^{-1} is regarded as a whole. In our proof, we divide J_t^{-1} into $\begin{pmatrix} A_t & B_t \\ C_t & D_t \end{pmatrix}$ and do more elaborate estimates.

In section 3, we give a local uniform estimate for Malliavin matrix under Hypothesis 3.1, and then give a gradient estimate in Theorem 3.2. The local uniform estimate for Malliavin matrix is a key point to prove Theorem 3.2.

In Theorem 3.2, we prove that P_t is strong Feller under some conditions which requires all coefficients of Eq.(1.1) to be C_b^2. Since there are bounded conditions on the coefficients and their derivatives, it seems too strong to apply, for example, the Hamiltonian systems, so we weaken this bounded conditions in Theorem 4.2 in section 4. In Theorem 4.2, we use the localization method to prove P_t is strong Feller under Hypothesis 4.1, which doesn't need bounded conditions on the coefficients and their derivatives. Actually, we prove that the law of (x_t, y_t) is continuous in initial value (x, y) with respect to the total variation distance in Theorem 4.2.

In section 5, we apply the above results to examples, such as the Lagevin SDEs, the stochastic Hamiltonian systems and high order stochastic differential equations.

2 The L^p Integrability of the Inverse of Malliavin Matrix

In this section, (x_t, y_t) is the solution for Eq.(1.1) with initial value (x, y), M_t is the Malliavin matrix of (x_t, y_t).

2.1 The Main Theorem and Its Relations with Hörmander Theorem

Hypothesis 2.1. $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$ *and there exists a* $j_0 := j_0(x, y) \in \mathbb{N}$ *such that:*

(i) $a_1 \in C_b^1(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m) \cap C^{j_0+2}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m), a_2 \in C_b^1(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n)$;

(ii) $b \in C_b^1(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n \times \mathbb{R}^d) \cap C^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n \times \mathbb{R}^d)$, $\det(b(x,y) \cdot b^*(x,y)) \neq 0$*;*

(iii) *The vector space spanned by* $\cup_{k=1}^{j_0} \mathcal{A}_k$ *at point* (x, y) *has dimension* m*.*

Theorem 2.2. *Assume Hypothesis* 2.1*,* $T > 0$*, then* $\det(M_T^{-1}) \in L^p(\Omega, \mathbb{P}_{x,y})$ *for any* $p > 0$*.*

Remark 2.3. **(i)** *The condition* $\det(b(x,y) \cdot b^*(x,y)) \neq 0$ *in Hypothesis* 2.1 *is necessary. For example, consider SDE*

$$\begin{cases} dx_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} y_t dt, \\ dy_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} y_t dt + \begin{pmatrix} 1 \\ 1 \end{pmatrix} dW_t. \end{cases}$$

That is

$$a_1(x,y) = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},\ a_2(x,y) = \begin{pmatrix} y_1 + y_2 \\ y_2 \end{pmatrix},\ b(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and furthermore

$$\mathcal{A}_1 = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Therefore, the (i) (iii) in the Hypothesis 2.1 *hold. But Malliavin matrix of* (x_t, y_t) *is singular almost surely(For details, please to see Appendix A).*

(ii) *A natural but difficult question is, can we replace the condition* $\det(b(x,y)\cdot b^*(x,y)) \neq 0$ *by some type of Hörmander's bracket condition? Higher regularity on* b *may be needed.*

Remark 2.4. *If the coefficients* a_1, a_2, b *in Eq.*(1.1) *also depend on* t *and for any* $T > 0$*,* $t \to (a_1(t,0), a_2(t,0))$ *and* $t \to b(t,0)$ *are bounded on* $[0,T]$*, then Theorems* 2.2*,* 3.2 *and* 4.2 *also hold.*

There is a natural relation between Hörmander's bracket condition $(\mathbf{H})$ and Hypothesis 2.1 from the well-known geometric interpretation of Hörmander's bracket condition. Also, it can be proved directly by calculations.

Remark 2.5. *Assume* $a_1 \in C^\infty(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m)$*,* $a_2 \in C^\infty(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n)$*,* $b \in C^\infty(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n \times \mathbb{R}^d)$*,* $\det(b(x,y) \cdot b^*(x,y)) \neq 0$*. Then the Hörmander's bracket condition* $(\mathbf{H})$ *is equivalent to Hypothesis* 2.1*.*

Hypothesis 2.1 is weaker than Hörmander's bracket condition, the followings are two examples.

Example 2.6. *Consider the following stochastic differential equation*

$$\begin{cases} dx_1(t) = x_2(t)dt + y_t dt \\ dx_2(t) = x_1(t)dt \\ dx_3(t) = x_2(t)dt + x_3(t)dt \\ dy_t = a_2(x_t, y_t)dt + bdW_t \end{cases},$$

where $x_t = (x_1(t), x_2(t), x_3(t))^* \in \mathbb{R}^3$*,* $y_t \in \mathbb{R}^1$*,* $a_2(x_1, x_2, x_3, y)$ *only has one order derivatives and* $b \in \mathbb{R}^1 \setminus \{0\}$ *is a constant, then Hypothesis* 2.1 *holds, but the Hörmander's bracket conditions* $(\mathbf{H})$ *can't be applied directly.*

Proof. Set $a_1(x_1, x_2, x_3, y) = (x_2 + y, x_1, x_2 + x_3)^*$, then

$$\nabla_x a_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix},\ \nabla_y a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

In this example, by calculation,

$$\mathcal{A}_1 = \nabla_y a_1,\ \mathcal{A}_2 = -\nabla_x a_1 \nabla_y a_1,\ \mathcal{A}_3 = +(\nabla_x a_1)^2 \nabla_y a_1,$$

and

$$\nabla_y a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ -\nabla_x a_1 \nabla_y a_1 = -\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ +(\nabla_x a_1)^2 \nabla_y a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Therefore the vector space spanned by $\{\mathcal{A}_j,\ j = 1,2,3\}$ at any point (x,y) is $\mathbb{R}^3$. □

The following example is a special case for the SDE considered in [5] with $n = 3$.

Example 2.7. *Consider the SDE (1.3) with* $n = 3$*. If* $\det(\sigma(0, x_1, x_2, x_3)\sigma^*(0, x_1, x_2, x_3)) \neq 0$*, by calculating,*

$$\mathcal{A}_1 = \left\{ \begin{pmatrix} \nabla_{x_1} F_2 \\ 0 \end{pmatrix} \right\}, \;\; \mathcal{A}_2 = \left\{ \begin{pmatrix} \nabla_{x_1 x_1} F_2 \\ 0 \end{pmatrix}, \begin{pmatrix} G(x_1, x_2, x_3) \\ \nabla_{x_2} F_3 \cdot \nabla_{x_1} F_2 \end{pmatrix} \right\},$$

for some function G*. The condition in [5] is* $\nabla_{x_1} F_2 \cdot \nabla_{x_2} F_3 \neq 0$*. So the* $(\mathbf{iii})$ *in Hypothesis* 2.1 *is the same as that in [5]. Because of lack of smooth in [5], the Hörmander's bracket condition* $(\mathbf{H})$ *can't be applied directly.*

2.2 Proof of Theorem 2.2

In [15], the inverse of Jacobian matrix J_t^{-1} is regarded as a whole. In this subsection, we divide J_t^{-1} into four parts $\begin{pmatrix} A_t & B_t \\ C_t & D_t \end{pmatrix}$ and do more elaborate estimates, we obtain $\det(M_T^{-1}) \in L^p(\Omega, \mathbb{P}_{x,y})$, $\forall p, T > 0$ under Hypothesis 2.1. Our approach is mainly along the lines of [15], but has some differences and needs more complicated computation. The differences depend heavily on the special form of the Eq.(1.1). Before we give the proof of Theorem 2.2, we introduce some notations and list the Lemmas which will be used in the proof of Theorem 2.2.

Assume $J_t^{-1} = \begin{pmatrix} A_t & B_t \\ C_t & D_t \end{pmatrix}$, A_t is a matrix with dimension $m \times m$, then

$$\begin{cases} dA_t = -\sum_{j=1}^{d} B_t \nabla_x b_j dW_j(t) - (A_t \nabla_x a_1 + B_t \nabla_x a_2) dt + \sum_{j=1}^{d} B_t \nabla_y b_j \nabla_x b_j dt, \\ dB_t = -\sum_{j=1}^{d} B_t \nabla_y b_j dW_j(t) - (A_t \nabla_y a_1 + B_t \nabla_y a_2) dt + \sum_{j=1}^{d} B_t \nabla_y b_j \nabla_y b_j dt, \\ dC_t = -\sum_{j=1}^{d} D_t \nabla_x b_j dW_j(t) - (C_t \nabla_x a_1 + D_t \nabla_x a_2) dt + \sum_{j=1}^{d} D_t \nabla_y b_j \nabla_x b_j dt, \\ dD_t = -\sum_{j=1}^{d} D_t \nabla_y b_j dW_j(t) - (C_t \nabla_y a_1 + D_t \nabla_y a_2) dt + \sum_{j=1}^{d} D_t \nabla_y b_j \nabla_y b_j dt. \end{cases} \tag{2.1}$$

For the vector space spanned by $\cup_{k=1}^{j_0} \mathcal{A}_k$ at point (x, y) has dimension m, then there exist two positive constants R_1 and c such that

$$\sum_{j=1}^{j_0} \sum_{V \in \mathcal{A}_j} (v^* V(x', y'))^2 \geqslant c \tag{2.2}$$

holds for all $v \in \mathbb{R}^m$, $|v| = 1$ and $|(x', y') - (x, y)| \leqslant R_1$.

Fix $R_2 = \frac{1}{100}$, define the stopping time as

$$S = S(x, y) := \inf \big\{ s \geqslant 0 : \sup_{0 \leqslant u \leqslant s} |(x_u, y_u) - (x, y)| \geqslant R_1 \text{ or } \sup_{0 \leqslant u \leqslant s} |J_u^{-1} - I_{m+n}| \geqslant R_2 \big\}, \tag{2.3}$$

here I_{m+n} denotes the identity matrix with dimension $m+n$. Define the adapted process

$$\lambda(s) = \inf_{|v|=1} \{ v^* b(x_s, y_s) b^*(x_s, y_s) v \}. \tag{2.4}$$

For $|\inf_v a_v - \inf_v b_v| \leqslant \sup_v |a_v - b_v|$, we have

$$|\lambda(s) - \lambda(t)| \leqslant \|b(x_s, y_s)b^*(x_s, y_s) - b(x_t, y_t)b^*(x_t, y_t)\|. \tag{2.5}$$

Then $\lambda(s)$ is continuous with respect to s. Since $\det(b(x,y)b^*(x,y)) \neq 0$, $\lambda(0) > 0$. For $R_3 = \lambda(0)/2$, we define the stopping times

$$\tau' = \inf\{s > 0 : |\lambda(s) - \lambda(0)| \geqslant R_3\}, \tag{2.6}$$
$$\tau = \tau' \wedge S \wedge T. \tag{2.7}$$

Let j_0 be as in Hypothesis 2.1. $v = (v_1^*, v_2^*)^* \in \mathbb{R}^m \times \mathbb{R}^n$ with $|v| = 1$. Fix $q > 8$ and set

$$F = \left\{ \sum_{j=1}^{j=d} \int_0^T |(v_1^* B_s + v_2^* D_s) b_j|^2 ds \leqslant \epsilon^{q^{3j_0+6}} \right\},$$
$$E_j = \left\{ \sum_{K \in \mathcal{A}_j} \int_0^\tau |(v_1^* A_s + v_2^* C_s) K(x_s, y_s)|^2 ds \leqslant \epsilon^{q^{3j_0+3-3j}} \right\},\ j = 1, \cdots, j_0,$$
$$E = F \cap E_1 \cap E_2 \cdots \cap E_{j_0}.$$

Remark 2.8. *In the definition of S in (2.3) , $R_2 = \frac{1}{100}$ is chosen only for technical convenience, there are other possible choices. In the Lemma 2.22, we essentially need R_2 small enough, and be finite in other places. Here, R_1, R_3 and c depend on (x, y).*

Due to (2.2) and the definition of S, it holds that for any $s \leqslant S$ and $v \in \mathbb{R}^m$ with $|v| = 1$,

$$\sum_{j=1}^{j_0} \sum_{V \in \mathcal{A}_j} (v^* V(x_s, y_s))^2 \geqslant c. \tag{2.8}$$

Lemma 2.9. *([9, Lemma 6.14]). Let $f : [0, T_0] \to \mathbb{R}$ be continuous differentiable and $\alpha \in (0, 1]$. Then*

$$\|\partial_t f\|_\infty = \|f\|_1 \leqslant 4 \parallel f \parallel_\infty \max\left\{\frac{1}{T_0},\ \parallel f \parallel_\infty^{-\frac{1}{1+\alpha}} \parallel \partial_t f \parallel_\alpha^{\frac{1}{1+\alpha}} \right\},$$

where $\|f\|_\alpha = \sup\limits_{s,t \in [0,T_0], s \neq t} \frac{|f(t) - f(s)|}{|t-s|^\alpha}$.

Lemma 2.10. *([15, Corollary 2.2.1]). Assume Hypothesis 2.1, then for any $p, T > 0$, there exists a finite constant $C(T, p, x, y)$ such that*

$$\mathbb{E}\Big\{ \sup_{0 \leqslant t \leqslant T} |(x_t, y_t)|^p \Big\} \leqslant C(T, p, x, y).$$

Lemma 2.11. *Assume Hypothesis 2.1, then for any $p, T > 0$, there exists a finite constant $C(T, p, x, y)$ such that*

$$\mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant T} \|J_s^{-1}\|^p \Big\} \leqslant C(T, p, x, y),$$
$$\mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant T} \|J_s\|^p \Big\} \leqslant C(T, p, x, y).$$

Proof. It directly follows from (1.6) (1.7) and [15, Lemma 2.2.1]. □

Lemma 2.12. *Assume Hypothesis 2.1, then for any $p > 0$, there exists a finite constant $C(p, x, y)$ such that*

$$\mathbb{P}\{S < \epsilon\} \leqslant C(p, x, y)\epsilon^p,\ \forall \epsilon > 0.$$

Proof. With the same argument as the estimation of $\mathbb{P}\{S < \varepsilon^\beta\}$ in [15, Theorem 2.3.3], one can get its proof. For more details, please to see [15, Page139]. □

Lemma 2.13. *Assume Hypothesis* 2.1*, then for any* $p > 0$*, there exists a finite constant* $C(p, T, x, y)$ *such that*

$$\mathbb{P}\{\tau < \epsilon\} \leqslant C(p, T, x, y)\epsilon^p, \ \forall \epsilon > 0.$$

Proof. According to Lemma 2.12 and the fact

$$\mathbb{P}\{\tau < \epsilon\} \quad \leqslant \quad \mathbb{P}\{S < \epsilon\} + \mathbb{P}\{\tau' < \epsilon\} + \mathbb{P}\{T < \epsilon\},$$

we only need to estimate $\mathbb{P}\{\tau' < \epsilon\}$. For any $p > 0$

$$\begin{aligned} \mathbb{P}\{\tau' < \epsilon\} &\leqslant \mathbb{P}\Big\{ \sup_{0 \leqslant r \leqslant \epsilon} |\lambda(s) - \lambda(0)| \geqslant R_3 \Big\} \\ &\leqslant C(p, x, y)\mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant \epsilon} |\lambda(s) - \lambda(0)|^{2p} \Big\}. \end{aligned} \tag{2.9}$$

Due to inequality $|\inf_v a_v - \inf_v b_v| \leqslant \sup_v |a_v - b_v|$,

$$\begin{aligned} &\mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant \epsilon} |\lambda(s) - \lambda(0)|^{2p} \Big\} \\ &\leqslant C(p) \sum_{\substack{i,k=1,\cdots,n \\ j=1,\cdots,d}} \mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant \epsilon} \Big| b_{kj}(x_s, y_s) b_{ij}(x_s, y_s) - b_{kj}(x, y) b_{ij}(x, y) \Big|^{2p} \Big\}. \end{aligned} \tag{2.10}$$

Noting that

$$\begin{aligned} & b_{kj}(x_s, y_s) b_{ij}(x_s, y_s) - b_{kj}(x, y) b_{ij}(x, y) \\ &= (b_{kj}(x_s, y_s) - b_{kj}(x, y))(b_{ij}(x_s, y_s) - b_{ij}(x, y)) \\ &\quad + b_{kj}(x, y)(b_{ij}(x_s, y_s) - b_{ij}(x, y) \\ &\quad + b_{ij}(x, y)(b_{kj}(x_s, y_s) - b_{kj}(x, y)), \end{aligned}$$

and by (2.9)(2.10),

$$\mathbb{P}\{\tau' < \epsilon\} \leqslant C(p, x, y) \left[\mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant \epsilon} |b_{ij}(x_s, y_s) - b_{ij}(x, y)|^{2p} \Big\} + \mathbb{E}\Big\{ \sup_{0 \leqslant s \leqslant \epsilon} |b_{ij}(x_s, y_s) - b_{ij}(x, y)|^{4p} \Big\} \right].$$

Hence this Lemma follows from Burkholder's and Hölder's inequalities and the fact

$$b_{ij}(x_s, y_s) - b_{ij}(x, y) = \langle \nabla b_{ij}(\xi, \eta), (x_s, y_s) - (x, y) \rangle,$$

here (ξ, η) is some point depending on (x_s, y_s) and (x, y). □

Lemma 2.14. *Let* σ *be a finite stopping time with bound* $c_\sigma < \infty$*, and there exists* $\tilde{p} > 0$ *such that*

$$\mathbb{P}\{\sigma < \epsilon\} \leqslant C(c_\sigma, \tilde{p})\epsilon^{\tilde{p}}, \ \forall \epsilon > 0,$$

holds for some constant $C(c_\sigma, \tilde{p})$*. Assume* $\gamma(t) = (\gamma_1(t), ..., \gamma_d(t)),\ u(t) = (u_1(t), ...u_d(t))$ *are continuous adapted processes,* $W(t) = (W_1(t), \cdots, W_d(t))^*$ *is a d-dimensional standard Wiener process,* $a(t), \tilde{y}(t) \in \mathbb{R}$ *and for* $t \in [0, c_\sigma]$*,*

$$\begin{aligned} a(t) &= \alpha + \int_0^t \beta(s) ds + \int_0^t \gamma(s) dW(s), \\ \tilde{y}(t) &= \tilde{y} + \int_0^t a(s) ds + \int_0^t u(s) dW(s). \end{aligned}$$

Suppose for some $p, \tilde{c} > 0$,

$$\mathbb{E}\Big\{ \sup_{0 \leqslant t \leqslant \sigma} (|\beta(t)| + |\gamma(t)| + |a(t)| + |u(t)|)^p \Big\} \leqslant \tilde{c} < \infty. \tag{2.11}$$

Then for any three positive numbers (q, r, v) *satisfying* $2q - 36r - 9v > 16$, *there exists* $\epsilon_0 = \epsilon_0(c_\sigma, q, r, v)$ *such that for any* $\epsilon < \epsilon_0$,

$$\mathbb{P}\left(\int_0^\sigma \tilde{y}(t)^2 dt < \epsilon^q, \int_0^\sigma (|a(t)|^2 + |u(t)|^2) dt \geqslant \epsilon \right) \leqslant \tilde{c}\epsilon^{rp} + \exp(-\epsilon^{-\frac{v}{4}}) + C(c_\sigma, \tilde{p})\epsilon^{\tilde{p}}.$$

The proof of Lemma 2.14 is postponed to Appendix B.

Lemma 2.15. *Let* σ *be a finite stopping time with bound* $c_\sigma < \infty$, *and there exists* $\tilde{p} > 2$, *such that*

$$\mathbb{P}\{\sigma < \epsilon\} \leqslant C(c_\sigma, \tilde{p})\epsilon^{\tilde{p}}, \ \forall \epsilon > 0$$

holds for some constant $C(c_\sigma, \tilde{p})$. *Consider the following one dimensional stochastic differential equation*

$$\tilde{y}(t) = \tilde{y} + \int_0^t a(s) ds + \int_0^t u(s) dW(s), \ t \in [0, c_\sigma],$$

where $u(s) = (u_1(s), \cdots, u_d(s))$ *is a continuous adapted process,* $W(t) = (W_1(t), \cdots, W_d(t))^*$ *is a d-dimensional standard Wiener process.* $a(t), u(t)$ *satisfy*

$$\mathbb{E}\Big\{ \sup_{0 \leqslant t \leqslant \sigma} (|a(t)| + |u(t)|)^p \Big\} \leqslant \tilde{c} < \infty,$$

for some $p, \tilde{c} > 0$.

Then for any three positive numbers (q, r, v) *satisfying* $2q > 8 + 20r + v$, *there exists* $\epsilon_0 = \epsilon_0(c_\sigma, q, r, v)$ *such that for any* $\epsilon \leqslant \epsilon_0$,

$$\mathbb{P}\left(\int_0^\sigma \tilde{y}(t)^2 dt < \epsilon^q, \ \int_0^\sigma |u(t)|^2 dt \geqslant \epsilon \right) \leqslant \tilde{c}\epsilon^{rp} + \exp\{-\epsilon^{-\frac{v}{4}}\} + C(c_\sigma, \tilde{p})\epsilon^{\tilde{p}}.$$

The proof of Lemma 2.15 is postponed to Appendix B. Denote

$$\|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2 := \sup_{s, r \in [0, \tau]} \frac{\Big| \big(|v_1^* B_s + v_2^* D_s|^2 - |v_1^* B_r + v_2^* D_r|^2 \big) \Big|}{|s - r|^{\frac{1}{4}}}. \tag{2.12}$$

Lemma 2.16. *Assume Hypothesis* 2.1 *and denote* $C_0 = 2/\lambda(0)$, *then for any* $p > 0$, *there exists a constant* $C = C(p, T, x, y, q)$ *such that*

$$\mathbb{P}\Big\{ \|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2 > \frac{1}{4^{\frac{5}{4}} C_0^{\frac{1}{4}}} \epsilon^{-\frac{q^{3j_0+6}}{8}} \Big\} \leqslant C(p, T, x, y, q)\epsilon^p, \ \forall \epsilon > 0.$$

Proof. By (2.1) and Itô's formula,

$$\begin{aligned} d|v_1^* B_s + v_2^* D_s|^2 = & -2\langle (v_1^* B_s + v_2^* D_s), (v_1^* B_s + v_2^* D_s)\nabla_y a_2 \rangle ds \\ & - 2\langle (v_1^* B_s + v_2^* D_s), (v_1^* A_s + v_2^* C_s)\nabla_y a_1 \rangle ds \\ & - 2\sum_{j=1}^d \langle (v_1^* B_s + v_2^* D_s), (v_1^* B_s + v_2^* D_s)\nabla_y b_j \rangle dW_j(s) \\ & + 2\sum_{j=1}^d \langle (v_1^* B_s + v_2^* D_s), (v_1^* B_s + v_2^* D_s)\nabla_y b_j \nabla_y b_j \rangle ds \\ & + \sum_{j=1}^d \langle (v_1^* B_s + v_2^* D_s)\nabla_y b_j, (v_1^* B_s + v_2^* D_s)\nabla_y b_j \rangle ds. \end{aligned}$$

By Burkholder-Davis-Gundy inequality and Lemma 2.11, the above equation implies that for any $p > 0$, there exists a constant $C = C(p, T, x, y)$ such that for any $s, r \in [0, T]$,

$$\mathbb{E}\left[|v_1^* B_s + v_2^* D_s|^2 - |v_1^* B_r + v_2^* D_r|^2\right]^{2p} \leqslant C|s - r|^p.$$

Set $\gamma = 2p,\ \epsilon = p - 1$ and $T_0 = T$ in [17, Theorem 2.1], then for any $p > 2$,

$$C_{p,T,x,y} := \mathbb{E}\left[\|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2\right]^{2p} < \infty.$$

Thus, $\forall \epsilon > 0,\ \forall p' > 0$

$$\begin{aligned} &\mathbb{P}\Big\{\|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2 > \frac{1}{4^{\frac{5}{4}} C_0^{\frac{1}{4}}} \epsilon^{-\frac{q^{3j_0+6}}{8}}\Big\} \\ &\leqslant C(p')\epsilon^{\frac{q^{3j_0+6}}{8}p'} \mathbb{E}\left[\|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2\right]^{p'}. \end{aligned} \tag{2.13}$$

Then this Lemma is obtained by setting $p' = \frac{8p}{q^{3j_0+6}}$ in (2.13), □

Lemma 2.17. *Assume Hypothesis* 2.1*, then for any* $p > 0$ *there exists a constant* $C(p, T, x, y, q)$ *such that*

$$\mathbb{P}\Big\{F \cap \big\{\sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s|^2 > \epsilon^{\frac{q^{3j_0+6}}{10}}\big\}\Big\} \leqslant C(p, T, x, y, q)\epsilon^p,\ \forall \epsilon > 0.$$

Proof. Due to $\tau \leqslant \tau'$ and $\omega \in F$, for the constant $C_0 = 2/\lambda(0)$,

$$\int_0^\tau |v_1^* B_s + v_2^* D_s|^2(\omega) ds \leqslant C_0 \epsilon^{q^{3j_0+6}}.$$

Set $f(s) = \int_0^s |v_1^* B_u + v_2^* D_u|^2 du,\ T_0 = \tau(\omega)$ and $\alpha = \frac{1}{4}$ in Lemma 2.9, then

$$\begin{aligned} &\sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s|^2 \leqslant \\ &\max\left\{\frac{4}{\tau}\int_0^\tau |v_1^* B_u + v_2^* D_u|^2 du,\ 4\Big\{\int_0^\tau |v_1^* B_u + v_2^* D_u|^2 du\Big\}^{\frac{1}{5}} \left(\|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2\right)^{\frac{4}{5}}\right\}. \end{aligned}$$

Thus

$$\begin{aligned} &\mathbb{P}\Big\{F \cap \big\{\sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s|^2 > \epsilon^{\frac{q^{3j_0+6}}{10}}\big\}\Big\} \\ &\leqslant \mathbb{P}\Big\{\int_0^\tau |v_1^* B_s + v_2^* D_s|^2 ds \leqslant C_0 \epsilon^{q^{3j_0+6}},\ \sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s|^2 > \epsilon^{\frac{q^{3j_0+6}}{10}}\Big\} \\ &\leqslant \mathbb{P}\Big\{\|v_1^* B_\cdot + v_2^* D_\cdot\|_{\frac{1}{4}}^2 > \frac{1}{4^{\frac{5}{4}} C_0^{\frac{1}{4}}} \epsilon^{-\frac{q^{3j_0+6}}{8}}\Big\} + \mathbb{P}\left(\tau < 4C_0 \epsilon^{\frac{9}{10} q^{3j_0+6}}\right). \end{aligned} \tag{2.14}$$

Due to (2.14), Lemma 2.13 and Lemma 2.16, for any $p > 0$, there exists a constant $C(p, T, x, y, q)$ such that

$$\mathbb{P}\Big\{F \cap \big\{\sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s|^2 > \epsilon^{\frac{q^{3j_0+6}}{10}}\big\}\Big\} \leqslant C(p, T, x, y, q)\epsilon^p,\ \forall \epsilon > 0.$$

□

Lemma 2.18. *Assume Hypothesis* 2.1*, then for any* $p > 0$*, there exists a positive constant* $C(p,T,x,y,q)$ *such that*

$$\begin{aligned}\mathbb{P}\Big\{\sum_{j=1}^{d}\int_0^T |(v_1^*B_s+v_2^*D_s)b_j|^2ds &\leqslant \epsilon^{q^{3j_0+6}}, \int_0^\tau |(v_1^*A_s+v_2^*C_s)\nabla_y a_1|^2 ds > \epsilon^{q^{3j_0}}\Big\}\\ &\leqslant C(p,T,x,y,q)\epsilon^p,\ \forall \epsilon>0.\end{aligned}$$

Proof. By (2.1),

$$\begin{aligned}&d(v_1^*B_s+v_2^*D_s)\\ &= -(v_1^*B_s+v_2^*D_s)\nabla_y a_2 ds - (v_1^*A_s+v_2^*C_s)\nabla_y a_1 ds\\ &\quad -\sum_{j=1}^{d}(v_1^*B_s+v_2^*D_s)\nabla_y b_j dW_j(t) + \sum_{j=1}^{d}(v_1^*B_s+v_2^*D_s)\nabla_y b_j\nabla_y b_j ds.\end{aligned}$$

By noting that $\det(b(x,y)b^*(x,y))\neq 0$ and the definition of τ, if

$$\sum_{j=1}^{d}\int_0^T |(v_1^*B_s+v_2^*D_s)b_j|^2(\omega)ds \leqslant \epsilon^{q^{3j_0+6}},$$

then for constant $C=\frac{2}{\lambda(0)}$,

$$\int_0^\tau |v_1^*B_s+v_2^*D_s|^2(\omega)ds \leqslant C\epsilon^{q^{3j_0+6}}. \tag{2.15}$$

Define

$$\tilde{y}(s) := (v_1^*B_s+v_2^*D_s)+\int_0^s (v_1^*B_u+v_2^*D_u)\nabla_y a_2 du - \sum_{j=1}^{d}\int_0^s (v_1^*B_u+v_2^*D_u)\nabla_y b_j\nabla_y b_j du, \tag{2.16}$$

then

$$d\tilde{y}(s) = -(v_1^*A_s+v_2^*C_s)\nabla_y a_1 ds - \sum_{j=1}^{d}(v_1^*B_s+v_2^*D_s)\nabla_y b_j dW_j(s).$$

Due to Hölder inequality, (2.15) and (2.16), there exists a constant $C(T,x,y)$ such that

$$\int_0^\tau |\tilde{y}(s)|^2 ds \quad\leqslant\quad C(T,x,y)\int_0^\tau |v_1^*B_s+v_2^*D_s|^2 ds$$

This implies that

$$\begin{aligned}&\left\{\sum_{j=1}^{d}\int_0^T |(v_1^*B_s+v_2^*D_s)b_j|^2 ds \leqslant \epsilon^{q^{3j_0+6}},\ \int_0^\tau |v_1^*A_s+v_2^*C_s)\nabla_y a_1||^2 ds > \epsilon^{q^{3j_0}}\right\}\\ &\subseteq \left\{\int_0^\tau |\tilde{y}(s)|^2 ds \leqslant C(T,x,y)\epsilon^{q^{3j_0+6}},\ \int_0^\tau |v_1^*A_s+v_2^*C_s)\nabla_y a_1|^2 ds \geqslant \epsilon^{q^{3j_0}}\right\}.\end{aligned}$$

The probability of the above event can be estimated by Lemma 2.14 and Lemma 2.13. □

Lemma 2.19. *Assume Hypothesis* 2.1*, then for any* $p>0$*, there exists constants* $C = C(p,T,x,y,q),\ \epsilon_0=\epsilon_0(q,x,y)$ *such that for* $j=1,\cdots,j_0-1$*,*

$$\mathbb{P}\{F\cap E_j\cap E_{j+1}^c\} \leqslant C(p,T,x,y,q)\epsilon^p,\ \forall\epsilon\leqslant\epsilon_0.$$

From the definitions of F, E_j, the sets F, E_j depend on ε actually. In order to simplify the proof of Lemma 2.19, first, we recall some definitions given in [9, Page40] and then give a proposition.

Definition 2.20. *Given a collection $H = \{H^\varepsilon\}_{\varepsilon \leqslant 1}$ of subsets of the probability space Ω, we will say that "H is a family of negligible events" if, for every $p \geqslant 1$ there exists a constant C_p such that $\mathbb{P}(H^\varepsilon) \leqslant C_p \varepsilon^p$ for every $\varepsilon \leqslant 1$.*

Given events $\{\Phi_j^\varepsilon\}_{1 \leqslant j \leqslant \ell, 0 \leqslant \varepsilon}$, and for each $j, \varepsilon, \Phi_j^\varepsilon \subseteq \Omega$. We will say: the implication

$$\Phi_j^\varepsilon \Rightarrow \Phi_{j+1}^\varepsilon$$

holds modulo a family of negligible events, if

$$\mathbb{P}\Big(\Phi_j^\varepsilon \cap \ (\Phi_{j+1}^\varepsilon)^c\Big) \leqslant C_p \varepsilon^p, \quad \forall \varepsilon \leqslant 1, \forall p > 1.$$

Proposition 2.21. *Given events $\{\Phi_j^\varepsilon\}_{1 \leqslant j \leqslant \ell, 0 \leqslant \varepsilon}$. If for $j = 1, \cdots, \ell - 1$, the implication*

$$\Phi_j^\varepsilon \Rightarrow \Phi_{j+1}^\varepsilon$$

holds modulo a family of negligible events, then the implication

$$\Phi_1^\varepsilon \Rightarrow \Phi_\ell^\varepsilon$$

holds modulo a family of negligible events.

We are now in a position to give

Proof. **The Proof of Lemma 2.19:** For any $K \in \mathcal{A}_j$, by calculating,

$$\begin{aligned}
& d(v_1^* A_s + v_2^* C_s) K(x_s, y_s) \\
& = \Big[- \sum_{i=1}^d \langle (v_1^* B_s + v_2^* D_s) \nabla_x b_i, \nabla_y K b_i \rangle + \sum_{i=1}^d (v_1^* B_s + v_2^* D_s) \nabla_y b_i \nabla_x b_i K(x_s, y_s) \Big] ds \\
& \quad + \sum_{i=1}^d \Big((v_1^* A_s + v_2^* C_s) \nabla_y K(x_s, y_s) b_i - (v_1^* B_s + v_2^* D_s) \nabla_x b_i K(x_s, y_s) \Big) \cdot dW_i(s) \\
& \quad + \Big[(v_1^* A_s + v_2^* C_s) \nabla_y K(x_s, y_s) a_2(x_s, y_s) - (v_1^* B_s + v_2^* D_s) \nabla_x a_2 K(x_s, y_s) \Big] ds \\
& \quad + (v_1^* A_s + v_2^* C_s) \big(- \nabla_x a_1(x_s, y_s) K(x_s, y_s) + \nabla_x K(x_s, y_s) a_1(x_s, y_s) \big) ds \\
& \quad + \frac{1}{2} (v_1^* A_s + v_2^* C_s) \sum_{i=1}^d \big(\nabla_y (\nabla_y K \cdot b_i) b_i \big) ds. \\
& := I_1(s) ds + \sum_{i=1}^d H_i(s) dW_i(s) + I_2(s) ds + I_3(s) ds + I_4(s) ds,
\end{aligned}$$

and denote $I(s) = \sum_{\ell=1}^4 I_\ell(s), H(s) = (H_1(s), \cdots, H_d(s))$. By Lemma 2.14 and definitions of F and E_j, the implication

$$\begin{aligned}
& F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau |H(s)|^2 ds \leqslant \epsilon^{q^{3j_0+2-3j}} \Big], \\
& F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau I(s)^2 ds \leqslant \epsilon^{q^{3j_0+2-3j}} \Big]
\end{aligned} \tag{2.17}$$

holds modulo a family of negligible events. In the following, we will prove: for some constant C, the following implications hold modulo a family of negligible events,

$$F \cap E_j \Rightarrow \Big[\int_0^\tau I_1(s)^2 ds \leqslant C\epsilon^{q^{3j_0+6}} \Big], \tag{2.18}$$

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau |(v_1^* A_s + v_2^* C_s)\nabla_y K|^2 ds \leqslant C\epsilon^{q^{3j_0+2-3j}} \Big], \tag{2.19}$$

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau I_2(s)^2 ds \leqslant C\epsilon^{q^{3j_0+1-3j}} \Big], \tag{2.20}$$

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau I_4(s)^2 ds \leqslant C\epsilon^{q^{3j_0+1-3j}} \Big]. \tag{2.21}$$

If these have been proved, then due to $I_3(s)^2 \leqslant 2[I(s)^2 + I_1(s)^2 + I_2(s)^2 + I_4(s)^2]$ and (2.17) (2.18) (2.20) (2.21), the implication

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau I_3(s)^2 ds \leqslant \epsilon^{q^{3j_0-3j}} \Big] \tag{2.22}$$

holds modulo a family of negligible events. Hence, combining (2.19) and (2.22), we get the desired result.

(i) The proof of (2.18). For the constant $C = \frac{2}{\lambda(0)}$,

$$\omega \in F \Rightarrow \int_0^\tau |(v_1^* B_s + v_2^* D_s)|^2(\omega) ds \leqslant C\epsilon^{q^{3j_0+6}}. \tag{2.23}$$

Hence, for some constant C, the following implication holds

$$\omega \in F \Rightarrow \int_0^\tau I_1(s)^2 ds \leqslant C\epsilon^{q^{3j_0+6}}.$$

(ii) The proof of (2.19). Noting that, for some constant C

$$|(v_1^* A_s + v_2^* C_s)\nabla_y K b|^2 \leqslant 2|H(s)|^2 + C|v_1^* B_s + v_2^* D_s|^2,$$

and combining it with (2.17)(2.23), the implication

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau \sum_{i=1}^d |(v_1^* A_s + v_2^* C_s)\nabla_y K \cdot b_i|^2 ds \leqslant 3\epsilon^{q^{3j_0+2-3j}} \Big] \tag{2.24}$$

holds modulo a family of negligible events. Due to the definition of τ and (2.24), for the constant $C = 3 \cdot \frac{2}{\lambda(0)} = \frac{6}{\lambda(0)}$, the implication

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau |(v_1^* A_s + v_2^* C_s)\nabla_y K|^2 ds \leqslant C\epsilon^{q^{3j_0+2-3j}} \Big]$$

holds modulo a family of negligible events.

(iii) The proof of (2.20). Combining (2.19) with (2.23), for some constant C, the implication

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau I_2(s)^2 ds \leqslant C\epsilon^{q^{3j_0+1-3j}} \Big]$$

holds modulo a family of negligible events.

(iv) The proof of (2.21). In the process of obtaining (2.24), substitute $\nabla_y K \cdot b_i$ for K, then for some constant C and every i, the implication

$$\begin{aligned} &F \cap \Big[\int_0^\tau |(v_1^* A_s + v_2^* C_s)\nabla_y K \cdot b_i|^2 ds \leqslant 3\epsilon^{q^{3j_0+2-3j}} \Big] \\ &\Rightarrow F \cap \Big[\sum_{\ell=1}^d \int_0^\tau |(v_1^* A_s + v_2^* C_s)\nabla_y(\nabla_y K \cdot b_i) b_\ell|^2 ds \leqslant C\epsilon^{q^{3j_0+1-3j}} \Big] \end{aligned} \tag{2.25}$$

holds modulo a family of negligible events. Due to (2.24)(2.25) and Proposition 2.21, for some constant C, the implication

$$F \cap E_j \Rightarrow F \cap \Big[\int_0^\tau I_4(s)^2 ds \leqslant C\epsilon^{q^{3j_0+1-3j}} \Big]$$

holds modulo a family of negligible events.

□

Lemma 2.22. *Assume Hypothesis* 2.1*, then there exists a constant* $\epsilon_0 = \epsilon_0(q, x, y)$ *such that*

$$E \cap \{\tau \geqslant \epsilon^q\} \cap \Big\{ \sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s| \leqslant \epsilon^{\frac{q^{3j_0+6}}{10}} \Big\} = \emptyset,\ \forall \epsilon < \epsilon_0.$$

Proof. If $\omega \in E \cap \{\tau \geqslant \epsilon^q\} \cap \Big\{ \sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s| \leqslant \epsilon^{\frac{q^{3j_0+6}}{10}} \Big\}$, by (2.8), for some $c > 0$,

$$\begin{aligned} &\sum_{j=1}^{j_0} \sum_{V\in\mathcal{A}_j} \int_0^\tau \big|(v_1^* A_s + v_2^* C_s) V(x_s, y_s)(\omega)\big|^2 ds \\ &= \int_0^\tau \sum_{j=1}^{j_0} \sum_{V\in\mathcal{A}_j} \Big(\frac{(v_1^* A_s + v_2^* C_s)}{|v_1^* A_s + v_2^* C_s|} V(x_s, y_s)(\omega)\Big)^2 \cdot |v_1^* A_s + v_2^* C_s|^2\, ds \\ &\geqslant c \int_0^\tau \big|v_1^* A_s + v_2^* C_s\big|^2 ds. \end{aligned} \tag{2.26}$$

For $\omega \in \big\{ \sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s| \leqslant \epsilon^{\frac{q^{3j_0+6}}{10}} \big\}$, let $s = 0$,

$$|v_2| \leqslant \epsilon^{\frac{q^{3j_0+6}}{10}} \leqslant \frac{1}{100},\ |v_1| = \sqrt{1 - |v_2|^2} > \frac{9}{10}.$$

This, together with the fact when $s \leqslant \tau$, $\|A_s - I_m\| \leqslant \frac{1}{100}$ and $\|C_s\| \leqslant \frac{1}{100}$, implies

$$\int_0^\tau \big|v_1^* A_s + v_2^* C_s\big|^2 ds \geqslant \frac{1}{8}\tau \geqslant \frac{1}{8}\epsilon^q. \tag{2.27}$$

By (2.26) and (2.27),

$$\sum_{j=1}^{j_0} \sum_{V\in\mathcal{A}_j} \int_0^\tau \big|(v_1^* A_s + v_2^* C_s) V(x_s, y_s)(\omega)\big|^2 ds \geqslant \frac{c}{8}\epsilon^q. \tag{2.28}$$

In the following part, we will prove (2.28) is impossible when ϵ is small enough. Set $\epsilon_0 = \epsilon_0(q, x, y)$ such that when $\epsilon < \epsilon_0$,

$$\epsilon^{\frac{q^{3j_0+6}}{10}} \leqslant \frac{1}{100},\ \sum_{j=1}^{j_0} \epsilon^{q^{3j_0+3-3j}} < \frac{c}{8}\epsilon^q.$$

For $\omega \in E \subseteq E_j$, then

$$\sum_{K\in\mathcal{A}_j} \int_0^\tau \big|(v_1^* A_s + v_2^* C_s) K(x_s, y_s)(\omega)\big|^2 ds \leqslant \epsilon^{q^{3j_0+3-3j}},$$

hence when $\epsilon < \epsilon_0$,

$$\sum_{j=1}^{j_0} \sum_{V\in\mathcal{A}_j} \int_0^\tau \big|(v_1^* A_s + v_2^* C_s) V(x_s, y_s)(\omega)\big|^2 ds \leqslant \sum_{j=1}^{j_0} \epsilon^{q^{3j_0+3-3j}} < \frac{c}{8}\epsilon^q,$$

this contradicts with (2.28).

Thus $E \cap \{\tau \geqslant \epsilon\} \cap \big\{ \sup_{s\in[0,\tau]} |v_1^* B_s + v_2^* D_s|^2 \leqslant \epsilon^{\frac{q^{3j_0+6}}{10}} \big\} = \emptyset$ as $\epsilon < \epsilon_0$. □

We are now in a position to give

Proof. **The proof of Theorem** 2.2**:** Since

$$M_T = J_T \tilde{M}_T J_T^*, \tag{2.29}$$

where

$$\tilde{M}_T = \int_0^T J_s^{-1} \begin{pmatrix} 0 \\ b(x_s, y_s) \end{pmatrix} \begin{pmatrix} 0 \\ b(x_s, y_s) \end{pmatrix}^* (J_s^{-1})^* ds, \tag{2.30}$$

we only need to prove the L^p integrability of $\det(\tilde{M}_T^{-1})$. For this purpose, we need to prove for any $p > 0$, there exists constant $C(p)$, such that

$$\sup_{|v|=1} \mathbb{P}\{v^* \tilde{M}_T v \leqslant \epsilon\} \leqslant C(p)\epsilon^p, \quad \forall \epsilon > 0. \tag{2.31}$$

It is easy to check that (2.31) is equivalent to for any $p > 0, v \in \mathbb{R}^{m+n}, |v| = 1$, there exists positive constants $\epsilon_0(p),\ C(p)$ such that

$$\mathbb{P}\{v^* \tilde{M}_T v \leqslant \epsilon\} \leqslant C(p)\epsilon^p, \quad \forall \epsilon \leqslant \epsilon_0(p). \tag{2.32}$$

Assume $v = (v_1^*, v_2^*)^* \in \mathbb{R}^m \times \mathbb{R}^n$, $J^{-1}(s) = \begin{pmatrix} A_s & B_s \\ C_s & D_s \end{pmatrix}$. Due to (2.30),

$$v^* \tilde{M}_T v = \sum_{j=1}^d \int_0^T |(v_1^* B_s + v_2^* D_s) b_j|^2 ds.$$

Here, we recall the definitions of E, F, E_j, τ which are given in the beginning of subsection 2.2. Then (2.32) is equivalent to for any $p > 0$ and $v \in \mathbb{R}^{m+n}, |v| = 1$, there exists constants $C(p)$ and $\epsilon_0(p)$ such that

$$\mathbb{P}(F) \leqslant C(p)\epsilon^p,\ \forall \epsilon \leqslant \epsilon_0(p).$$

Since

$$F \subseteq (F \cap E_1^c) \cup (F \cap E_1 \cap E_2^c) \cup (F \cap E_2 \cap E_3^c) \cup \cdots \cup (F \cap E_{j_0-1} \cap E_{j_0}^c) \cup E,$$

it holds that

$$\mathbb{P}(F) \leqslant \mathbb{P}(E) + \sum_{j=1}^{j=j_0-1} \mathbb{P}(F \cap E_j \cap E_{j+1}^c) + \mathbb{P}(F \cap E_1^c). \tag{2.33}$$

By Lemma 2.18 and Lemma 2.19, for any $p > 0$ and $v \in \mathbb{R}^{m+n}, |v| = 1$, there exists positive constants $C(p, T, x, y, q), \epsilon_0(q, x, y)$ such that for any $\epsilon \leqslant \epsilon_0(q, x, y)$,

$$\sum_{j=1}^{j=j_0-1} \mathbb{P}(F \cap E_j \cap E_{j+1}^c) + \mathbb{P}(F \cap E_1^c) \leqslant C(p, T, x, y, q)\epsilon^p. \tag{2.34}$$

For estimating $\mathbb{P}(E)$, we note that

$$\begin{aligned}
\mathbb{P}(E) &\leqslant \mathbb{P}\big(E \cap \{\tau \geqslant \epsilon^q\}\big) + \mathbb{P}(\tau < \epsilon^q) \\
&\leqslant \mathbb{P}\Big(F \cap \{\tau \geqslant \epsilon^q\} \cap \big\{\sup_{s \in [0,\tau]} |v_1^* B_s + v_2^* D_s| > \epsilon^{\frac{q^{3j_0+6}}{10}}\big\}\Big) \\
&\quad + \mathbb{P}\Big(E \cap \{\tau \geqslant \epsilon^q\} \cap \big\{\sup_{s \in [0,\tau]} |v_1^* B_s + v_2^* D_s| \leqslant \epsilon^{\frac{q^{3j_0+6}}{10}}\big\}\Big) \\
&\quad + \mathbb{P}\left(\tau < \epsilon^q\right).
\end{aligned}$$

Therefore, due to Lemma 2.17, Lemma 2.22 and Lemma 2.13, there exists constants $C(p,T,x,y,q)$ and $\epsilon_0 = \epsilon_0(q,x,y)$ such that

$$\mathbb{P}(E) \leqslant C(p,T,x,y,q)\epsilon^p, \ \forall \epsilon \leqslant \epsilon_0. \tag{2.35}$$

Due to $(2.33), (2.34)$ and (2.35), for any $p > 0$ and $v \in \mathbb{R}^{m+n}$ with $|v| = 1$, there exists constants $C(p,T,x,y,q)$ and $\epsilon_0(q,x,y)$ such that

$$\mathbb{P}(F) \leqslant C(p,T,x,y,q)\epsilon^p, \ \forall \epsilon \leqslant \epsilon_0(q,x,y).$$

Since T, x, y, q are fixed, this theorem has been proved. □

3 Gradient Estimate

In this section, we give a gradient estimate. The Hypothesis and Theorem in this section is

Hypothesis 3.1. *There exists $j_0 \in \mathbb{N}$ and $R > 0$ such that*

(i) $a_1 \in C_b^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m) \cap C^{j_0+2}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m), \ a_2 \in C_b^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n)$;

(ii) $b \in C_b^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n \times \mathbb{R}^d)$, $\det(b(x,y) \cdot b^*(x,y)) \neq 0$, $\forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^n$ *with* $|(x,y)| \leqslant R$;

(iii) $\forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^n, |(x,y)| \leqslant R$, *the vector space spanned by* $\cup_{k=1}^{j_0} \mathcal{A}_k$ *at point* (x,y) *has dimension* m.

Theorem 3.2. *Assume Hypothesis* 3.1, *then for any* $t > 0$, *then there exists a constant* $C = C(R,t)$ *such that for any* $f \in C_b^1(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R})$, $(x,y) \in \mathbb{R}^m \times \mathbb{R}^n$ *with* $|(x,y)| \leqslant R$,

$$|\nabla P_t f(x,y)| \leqslant C(R,t)\|f\|_\infty.$$

In order to prove this Theorem, we need the following Lemmas. These Lemmas give some estimates of J_t, J_t^{-1}, (x_t, y_t) and M_t. Especially, we give a uniform estimate of M_t in Lemma 3.5. In the end of this section, we give the proof of Theorem 3.2. The method to prove Theorem 3.2 is standard.

We introduce some notations first. $D(x_t, y_t)$ denotes the Malliavin derivative of (x_t, y_t) and $H = L^2([0,\infty), ds)$. δ denotes the divergence operator.

Lemma 3.3. *Assume Hypothesis* 3.1, *then for any* $T, p > 0$,

$$\sup_{|(x,y)|\leqslant R} \mathbb{E}_{x,y}\Big\{ \sup_{t\in[0,T]} |(x_t,y_t)|^p \Big\} < \infty, \tag{3.1}$$

$$\sup_{|(x,y)|\leqslant R} \mathbb{E}_{x,y}\Big\{ \sup_{t\in[0,T]} \|J_t\|^p \Big\} < \infty, \tag{3.2}$$

$$\sup_{|(x,y)|\leqslant R} \mathbb{E}_{x,y}\Big\{ \sup_{t\in[0,T]} \|J_t^{-1}\|^p \Big\} < \infty, \tag{3.3}$$

$$\sup_{|(x,y)|\leqslant R} \mathbb{E}_{x,y}\|M_T\|^p < \infty, \tag{3.4}$$

Proof. For any (x,y) fixed, $\mathbb{E}_{x,y}\big\{\sup_{s\in[0,T]} |(x_s,y_s)|^p\big\} < \infty$. Since function

$$h(x,y) := \mathbb{E}_{x,y}\big\{ \sup_{t\in[0,T]} |(x_t,y_t)|^p \big\}$$

is continuous with respect to (x,y), (3.1) holds.

For any $p > 2$, set $f(t) = \mathbb{E}_{x,y}\big\{\sup_{s\in[0,t]}\|J_s\|^p\big\}$. Due to (1.7), there exists constants $C(p), C(p,T)$ such that

$$f(t) \leqslant C(p) + C(p,T)\int_0^t f(s)ds,\ \forall t \in [0,T].$$

Then, (3.2) follows from the Gronwall inequality and the proof of (3.3) is similar.

(3.4) follows from (3.1) (3.2) (3.3) and (1.5) . □

Lemma 3.4. *Assume Hypothesis* 3.1*, then for any* $T, p > 0$*,*

$$\sup_{|(x,y)|\leqslant R}\sup_{r\in[0,T]}\mathbb{E}_{x,y}\Big\{\sup_{t\in[r,T]}\|D_r(x_t,y_t)\|^p\Big\} < \infty, \tag{3.5}$$

$$\sup_{|(x,y)|\leqslant R}\sup_{r\in[0,T]}\mathbb{E}_{x,y}\Big\{\sup_{t\in[r,T]}\big\|D_rJ_t^{-1}\big\|^p\Big\} < \infty, \tag{3.6}$$

$$\sup_{|(x,y)|\leqslant R}\sup_{r\in[0,T]}\mathbb{E}_{x,y}\Big\{\sup_{t\in[r,T]}\big\|D_rJ_t\big\|^p\Big\} < \infty, \tag{3.7}$$

$$\sup_{|(x,y)|\leqslant R}\sup_{r_1,r_2\in[0,T]}\mathbb{E}_{x,y}\Big\{\sup_{r_1\vee r_2\leqslant t\leqslant T}\|D_{r_1,r_2}X(t)\|^p\Big\} < \infty. \tag{3.8}$$

Proof. (3.5), (3.8) are given in [15, Theorem 2.2.1, Theorem 2.2.2]. The other two estimations are similar. □

Lemma 3.5. *Assume Hypothesis* 3.1*, then for any* $p, T > 0$*, there exists a constant* $C(T,p,R)$ *such that*

$$\sup_{|(x,y)|\leqslant R}\mathbb{E}_{x,y}\big|\det(M_T^{-1})\big|^p \leqslant C(p,R,T) < \infty.$$

Proof. Due to (2.30), it only need to show

$$\sup_{|(x,y)|\leqslant R}\sup_{|v|=1}\mathbb{P}_{x,y}\{v^*\tilde{M}_Tv \leqslant \epsilon\} \leqslant C(p,R,T)\epsilon^p, \quad \forall\epsilon > 0, \forall p > 0. \tag{3.9}$$

The proof of (3.9) is similar to the proof of Theorem 2.2 in subsection 2.2, but it also has some changes. In the following paragraphs, we will list these changes. These changes come from that we need to show the constants appeared in subsection 2.2 depending on R but independent of the $(x,y) \in \overline{B}_R$ under the Hypothesis 3.1.

(1) R_1 in (2.2)(2.3), c in (2.2) and Lemma 2.22. Define

$$\Lambda(x,y) := \inf_{|v|=1}\Big(\sum_{j=1}^{j_0}\sum_{V\in\mathcal{A}_j}(v^*V(x,y)V^*(x,y)v)\Big).$$

For $\Lambda(x,y)$ is continuous with respect to (x,y) (c.f.(2.5)) and $\Lambda(x,y) > 0, \forall(x,y) \in \overline{B}_R$, then there exists $R_1 > 0$ such that

$$\inf_{|(x,y)|\leqslant R+R_1}\Lambda(x,y) > c := \frac{1}{2}\inf_{|(x,y)|\leqslant R}\Lambda(x,y) > 0. \tag{3.10}$$

Then the following inequality holds,

$$\sum_{j=1}^{j_0}\sum_{V\in\mathcal{A}_j}\big(v^*V(x',y')\big)^2 \geqslant c,\ \forall(x',y') \in \overline{B}((x,y),R_1),\ \forall(x,y) \in \overline{B}_R,\ \forall|v| = 1.$$

(2) R_3 in (2.6), C_0 in Lemma 2.17 and Lemma 2.16. Set

$$R_3 = \frac{1}{C_0} := \frac{1}{2}\inf_{(x,y)\in\overline{B}_R}\inf_{|v|=1}(v^*b(x,y)b^*(x,y)v) > 0.$$

By the choosing of R_3, definition of τ' and Hypothesis 3.1, it holds that

$$|v|^2 \leqslant C_0 \sum_{j=1}^{d} |v^* b_j(x_s, y_s)|^2, \ \forall v \in \mathbb{R}^n, \forall (x,y) \in \overline{B}_R, \forall s \leqslant \tau',$$

here $(x_s, y_s)|_{s=0} = (x,y)$. Hence, the equalities (2.15)(2.19) hold with $C = \frac{3}{C_0}$ actually.

(3) Lemma 2.13. Due to Lemma 3.3, for the constants appeared in Lemma 2.12 and Lemma 2.13, we can choose them depending on R but independent of the $(x,y) \in \overline{B}_R$.

(4) The using of Lemma 2.14 and Lemma 2.15. For example, in Lemma 2.18, we need to estimate the following probability for some constant $C(R)$,

$$\mathbb{P}\left\{\int_0^\tau \|\tilde{y}(s)\|^2 ds \leqslant (1 + T^2 C(R))\epsilon^{q^{3j_0+6}}, \int_0^\tau |(v_1^* A_s + v_2^* C_s)\nabla_y a_1|^2 ds \geqslant \epsilon^{q^{3j_0}}\right\},$$

here

$$d\tilde{y}(s) = -(v_1^* A_s + v_2^* C_s)\nabla_y a_1 ds - \sum_{j=1}^{d} (v_1^* B_s + v_2^* D_s)\nabla_y b_j \cdot dW_j(s).$$

When we use the Lemma 2.14, it needs to show that we can choose the constant $\tilde{c}$ in (2.11) depending on R but independent of the $(x,y) \in \overline{B}_R$. This is due to Lemma 3.3 and the fact when $s \leqslant \tau$, $|(x,y)| \leqslant R$ and $|(x_s, y_s)| \leqslant R + R_1$.

(5) Other constants appeared in subsection 2.2. We can also choose them depending on R but independent of the $(x,y) \in \overline{B}_R$. □

We are now in a position to give

Proof. **The proof of Theorem 3.2:** For any $\xi \in \mathbb{R}^{m+n}$,

$$\langle \nabla P_t f(x,y), \xi\rangle \quad = \quad \mathbb{E}_{x,y} \nabla f(x_t, y_t) J_t \xi.$$

Assume $x_t = (x_t^1, \cdots, x_t^m)$ and $y_t = (y_t^1, \cdots, y_t^n)$, then by [15, (2.29),(2.30)],

$$\begin{aligned}&\mathbb{E}_{x,y}\Big\{\nabla_i f(x_t, y_t) J_t \xi\Big\} \\ &= \sum_{k=1}^{m} \mathbb{E}\Big\{f(x_t,y_t)\delta\big(J_t\xi(M_t^{-1})^{i,k} x_t^k\big)\Big\} + \sum_{k=m+1}^{m+n} \mathbb{E}\Big\{f(x_t,y_t)\delta\big(J_t\xi(M_t^{-1})^{i,k} y_t^{k-m}\big)\Big\}.\end{aligned}$$

Then, this Theorem follows from Lemma 3.3, Lemma 3.4, Lemma 3.5 and [15, Proposition 1.5.8]. □

In the end of this section, we give a Proposition which is supplementary to this article.

Proposition 3.6. *Assume Hypothesis* 2.1 *and* $a_1, a_2, b \in C_b^2$*, then the law of* (x_t, y_t) *with initial value* (x,y) *is absolutely continuous with respect to Lebesgue measure and its density function* $p(t,(u,v))$ *is continuous with respect to* $(u,v) \in \mathbb{R}^m \times \mathbb{R}^n$ *for fixed t. Furthermore, the following estimation holds*

$$\sup_{(u,v)\in\mathbb{R}^m\times\mathbb{R}^n} |p(t,(u,v))| < \infty.$$

Proof. It directly follows from Theorem 2.2 and [18, Theorem 5.9]. □

4 Strong Feller Property

In this section, we prove that the semigroup P_t associated with Eq.(1.1) is strong Feller without the bounded conditions on the coefficients and their derivatives. By Theorem 3.2, P_t is strong Feller under some conditions which need all coefficients for Eq.(1.1) are in C_b^2. But in the Hamiltonian systems, the diffusion and drift part are polynomial growth, therefore the Theorem 3.2 can't apply directly. But if the SDE has global solution, we can also prove P_t is strong Feller without the bounded conditions.

The followings are the Hypothesis and Theorem in this section.

Hypothesis 4.1. *There exists $j_0 \in \mathbb{N}$ such that:*

(i) $a_1 \in C^{j_0+2}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^m),\ a_2 \in C^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n)$;

(ii) $b \in C^2(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^n \times \mathbb{R}^d),\ \det(b(x,y) \cdot b^*(x,y)) \neq 0,\ \forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^n$;

(iii) $\forall (x,y) \in \mathbb{R}^m \times \mathbb{R}^n$, *the vector space spanned by $\cup_{k=1}^{j_0} \mathcal{A}_k$ at point (x,y) has dimension m;*

(iv) *The solution to Eq.(1.1) exists globally for any initial value $(x,y) \in \mathbb{R}^m \times \mathbb{R}^n$.*

Theorem 4.2. *Assume Hypothesis 4.1 and let (x_t, y_t) be the solution to Eq.(1.1) with initial value (x,y), then the law of (x_t, y_t) is continuous in variable (x,y) with respect to the total variation distance. In particular, the semigroup $(P_t)_{t>0}$ has the strong Feller property, i.e., for any $t>0$ and $f \in \mathcal{B}_b(\mathbb{R}^{m+n})$,*

$$(x,y) \in \mathbb{R}^{m+n} \mapsto \mathbb{E}_{x,y} f(x_t, y_t) \text{ is continuous.}$$

Remark 4.3. *If there exists a Liapunov function W such that $LW \leqslant cW$ for some $c>0$, then the $(\mathbf{iv})$ in Hypothesis 4.1 holds by [13, Theorem 5.9]. Here*

$$L = \sum_{i=1}^m a_1^i \frac{\partial}{\partial x_i} + \sum_{i=1}^n a_2^i \frac{\partial}{\partial y_i} + \frac{1}{2} \sum_{i,j=1}^n (b \cdot b^*)_{i,j} \frac{\partial^2}{\partial y_i \partial y_j}.$$

Remark 4.4. *Under Hörmander's bracket condition, it is well known that P_t is strong Feller. In Theorem 4.2, we give a stronger result: the law of (x_t, y_t) is continuous in variable (x,y) with respect to the total variation distance.*

For the convenience of writing, we will use x instead of (x,y) in the rest of this section. Let $X_t^x = (x_t, y_t)$ be the solution of Eq.(1.1) with initial value $x \in \mathbb{R}^m \times \mathbb{R}^n$. In the following part, we will use the localization to prove Theorem 4.2.

For any fixed $\ell \in \mathbb{N}$, set $a(x) = (a_1^*(x), a_2^*(x))^*$, $a_\ell(x) = h_\ell(x) a(x)$, $b_\ell(x) = h_\ell(x) b(x)$, $h_\ell(x) \in \mathbb{R}$ is a smooth function with compact support and $h_\ell(x) = 1$ on B_ℓ°. Let $X_s^\ell(x)$ be the solution to the following equation,

$$X_s^\ell(x) = x + \int_0^s a_\ell(X_r^\ell(x)) dr + \int_0^s \begin{pmatrix} 0 \\ b_\ell(X_r^\ell(x)) \end{pmatrix} dW_r. \tag{4.1}$$

Define a sequence of stopping time

$$S_\ell(x) \quad = \quad \inf\{s > 0,\ X_s^\ell(x) \notin B_\ell^\circ\},\ \ell \geqslant 1.$$

If Hypothesis 4.1 holds, then for any $x \in \mathbb{R}^{m+n}$, the following properties holds a.s.

$$S_\ell(x) < S_{\ell+1}(x), \tag{4.2}$$

$$X_s^\ell(x) = X_s^{\ell+1}(x),\ \forall s \in [0, S_\ell(x)), \tag{4.3}$$

$$X_s^x = X_s^\ell(x), \qquad \forall s \in [0, S_\ell(x)), \tag{4.4}$$

$$\sup_\ell S_\ell(x) = \infty. \tag{4.5}$$

In order to prove Theorem 4.2, we also need the following Lemmas.

Lemma 4.5. *Assume Hypothesis* 4.1, *then for any* $x_0 \in \mathbb{R}^{m+n}$, $\ell \geqslant 2,\ t > 0$

$$\limsup_{x \to x_0} I_{\{t > S_\ell(x)\}} \leqslant I_{\{t \geqslant S_{\ell-1}(x_0)\}},\ a.s.$$

Proof. There exists a measurable set Γ such that $\mathbb{P}(\Gamma^c) = 0$ and for any ℓ and $\omega \in \Gamma$, $X_s^\ell(x, \omega)$ is continuous with respect to x and s. We will prove that for $\omega \in \Gamma$,

$$\limsup_{x \to x_0} I_{\{t > S_\ell(x)\}}(\omega) = 0 \text{ or } I_{\{t \geqslant S_{\ell-1}(x_0)\}}(\omega) = 1,$$

which implies the desired result. Assume that $\omega \in \Gamma$ and

$$\limsup_{x \to x_0} I_{\{t > S_\ell(x)\}}(\omega) = 1 \text{ and } I_{\{t \geqslant S_{\ell-1}(x_0)\}}(\omega) = 0. \tag{4.6}$$

We will prove that this is impossible. If $\omega \in \Gamma$ and satisfies (4.6), then

$$t < S_{\ell-1}(x_0), \tag{4.7}$$

and there exist $\{x_n\} \subset \mathbb{R}^{m+n}$ with $x_n \to x_0$ as $n \to \infty$, such that for n large enough

$$\sup_{s \in [0,t]} |X_s^\ell(x_n, \omega)| \geqslant \ell. \tag{4.8}$$

By (4.7),

$$\sup_{s \in [0,t]} |X_s^{\ell-1}(x_0, \omega)| \leqslant \ell - 1.$$

By $t < S_{\ell-1}(x_0)$, (4.3) and $\sup_{s \in [0,t]} |X_s^{\ell-1}(x_0, \omega)| \leqslant \ell - 1$,

$$\sup_{s \in [0,t]} |X_s^\ell(x_0, \omega)| \leqslant \ell - 1. \tag{4.9}$$

For $X_s^\ell(x, \omega)$ is continuous with respect to s and x and $[0,t] \times \overline{B}(0,1) \subseteq [0, \infty) \times \mathbb{R}^{m+n}$ is a compact set, therefor for $\epsilon_0 = \frac{1}{2}$ there exists $\delta_0 > 0$ such that for any $|x - x_0| \leqslant \delta_0$ and $s \in [0,t]$

$$|X_s^\ell(x_0, \omega) - X_s^\ell(x, \omega)| \leqslant \frac{1}{2}.$$

This means that when $|x - x_0| \leqslant \delta_0$,

$$\sup_{s \in [0,t]} |X_s^\ell(x, \omega)| \leqslant \sup_{s \in [0,t]} |X_s^\ell(x_0, \omega)| + \frac{1}{2},$$

Therefore, by (4.9), for any x with $|x - x_0| \leqslant \delta_0$,

$$\sup_{s \in [0,t]} |X_s^\ell(x, \omega)| \leqslant \ell - 1 + \frac{1}{2} = \ell - \frac{1}{2},$$

which contradicts with (4.8). □

Let $\{P_t^\ell\}_{t \geqslant 0}$ be the transition semigroup of (4.1).

Lemma 4.6. *Assume Hypothesis* 4.1, *then for any* $x \in \overline{B}_\ell$,

$$\lim_{y \to x} \sup_{\|f\|_\infty \leqslant 1} \mathbb{E}\Big[f(X_t^\ell(y)) - f(X_t^\ell(x))\Big] = 0.$$

Proof. Since $a_\ell = a := (a_1, a_2)$ and $b_\ell = b$ on $\overline{B}_\ell$, hence (a_ℓ, b_ℓ) satisfies Hypothesis 3.1 with $R = \ell$. In the proof of Theorem 3.2, let $R = \ell$, substitute a_ℓ for a and b_ℓ for b, then

$$|\nabla P_t^\ell f(y)| \leqslant C(\ell, t)\|f\|_\infty, \quad \forall f \in C_b, \forall y \in \overline{B}_\ell. \tag{4.10}$$

From (4.10), we can obtain this Lemma. □

We are now in a position to give

Proof. **The proof of Theorem** 4.2**:** For any $x, y \in B_\ell^o$

$$\begin{aligned}\mathbb{E}[f(X_t(x)) - f(X_t(y))] &\leqslant \mathbb{E}[f(X_t(x))1_{t<\tau_\ell(x)} - f(X_t(y))1_{t<\tau_\ell(y)}] \\ &\quad + \|f\|_\infty \mathbb{P}(t \geqslant \tau_\ell(x)) + \|f\|_\infty \mathbb{P}(t \geqslant \tau_\ell(y)) \\ &\leqslant \mathbb{E}[f(X_t^\ell(x))1_{t<\tau_\ell(x)} - f(X_t^\ell(y))1_{t<\tau_\ell(y)}] \\ &\quad + \|f\|_\infty \mathbb{P}(t \geqslant \tau_\ell(x)) + \|f\|_\infty \mathbb{P}(t \geqslant \tau_\ell(y)) \\ &\leqslant \mathbb{E}[f(X_t^\ell(x)) - f(X_t^\ell(y))] \\ &\quad + 2\|f\|_\infty \mathbb{P}(t \geqslant \tau_\ell(x)) + 2\|f\|_\infty \mathbb{P}(t \geqslant \tau_\ell(y)).\end{aligned}$$

By Lemma 4.6

$$\lim_{y \to x} \sup_{\|f\|_\infty \leqslant 1} \mathbb{E}[f(X_t^\ell(x)) - f(X_t^\ell(y))] = 0. \tag{4.11}$$

By (4.11) and Lemma 4.5,

$$\begin{aligned}\lim_{y \to x} \sup_{\|f\|_\infty \leqslant 1} \mathbb{E}[f(X_t(x)) - f(X_t(y))] &\leqslant 2\mathbb{P}(t \geqslant \tau_\ell(x)) + 2 \limsup_{y \to x} \mathbb{P}(t \geqslant \tau_\ell(y)) \\ &\leqslant 2\mathbb{P}(t \geqslant \tau_\ell(x)) + 2\mathbb{P}(t \geqslant \tau_{\ell-1}(x)),\end{aligned}$$

let $\ell \to \infty$ in the above inequality, we obtain that

$$\lim_{y \to x} \sup_{\|f\|_\infty \leqslant 1} \mathbb{E}[f(X_t(x)) - f(X_t(y))] = 0.$$

□

Remark 4.7. *For the Eq.(1.3), [5, Thoerem 1.1] proved that X_t has a density $p(t, x, y)$ and gave the upper and lower bounds of $p(t, x, y)$ if the spectrum of the matrix-valued function $A = \sigma \cdot \sigma^*$ is included in $[\Lambda^{-1}, \Lambda]$ for some $\Lambda \geqslant 1$. In this article, we can't obtain such strong results since our condition is $\det\big(\sigma(x) \cdot \sigma^*(x)\big) \neq 0$, which is weaker than that in [5].*

5 Some Applications

The strong Feller property is very useful when we prove the uniqueness of invariant measure. If $X_t \in \mathbb{R}^n,\ t \in [0, +\infty),\ n \in \mathbb{N}$ is a continuous Markov process. The following theorem is classical.

Theorem 5.1. *(c.f. [20] [6] [7] etc.) Let P_t be the semigroup associated with X_t, and*

(i) *the Markov process X_t is irreducible, i.e,*

$$P_t(x, A) > 0, \ \textit{for all } t > 0, \ x \in \mathbb{R}^n, \ \textit{open set } A,$$

(ii) *P_t is strong Feller,*

then P_t exists at most one invariant measure.

Remark 5.2. *The conditions in Theorem 5.1 can be weaken, such as [9, Corollary 1.4].*

5.1 The Langevin Equation

This example is extended from the one in [14]. Let $W_t, t \geq 0$ be a standard d-dimensional Brownian Motion and $F : \mathbb{R}^d \to R$, $\sigma \in \mathbb{R}^{d\times d}$ invertible. Consider the Langevin SDE for $q, p \in \mathbb{R}^d$ the position and momenta of particle of unit mass, namely

$$\begin{cases} dq = pdt, \\ dp = -\gamma pdt - \nabla F(q)dt + \sigma dW_t. \end{cases} \tag{5.1}$$

Proposition 5.3. *Assume that $F \in C^3(\mathbb{R}^d, \mathbb{R})$ and satisfies*

(i) $F(q) \geqslant 0$ *for all* $q \in \mathbb{R}^d$;

(ii) *There exists an $\alpha > 0$ and $\beta \in (0, 1)$ such that*

$$\frac{1}{2}\langle \nabla F(q), q\rangle \geqslant \beta F(q) + \gamma^2 \frac{\beta(2-\beta)}{8(1-\beta)}||q||^2 - \alpha.$$

Then the semigroup P_t associated with the Langevin SDE is strong Feller and has a unique invariant measure.

Proof. First, Hypothesis 4.1 holds for $j_0 = 1$, hence P_t is strong Feller by Theorem 4.2. Second, P_t is irreducible by [14, Lemma 3.4] . Therefore, P_t has at most one invariant measure. Third, the invariant measure for P_t exists by [14, Corollary A.5]. □

5.2 Stochastic Hamiltonian Systems

This example is extended from the one in [21]. Consider a stochastic differential system of the type

$$\begin{cases} X_t = X_0 + \displaystyle\int_0^t \partial_y H(X_s, Y_s)ds, \\ Y_t = Y_0 - \displaystyle\int_0^t \big[\partial_x H(X_s, Y_s) + F(X_s, Y_s)\partial_y H(X_s, Y_s)\big]ds + W_t, \end{cases} \tag{5.2}$$

where X_t, Y_t, W_t belong to $\mathbb{R}^d$.

In the following Proposition, we don't need F and $H \in C^\infty$ as in [21].

Proposition 5.4. *Assume there exists strictly positive numbers ν, M, δ, and there exits a function $R(x, y)$ on $\mathbb{R}^{2d}$ with second derivatives having polynomial growth at infinity, such that*

(i) $F \in C^2, H \in C^4$;

(ii) $0 < \nu|\xi|^2 \leqslant \sum_{i,j=1}^d \partial_{y_i y_j} H(x, y)\xi_i\xi_j$, $\forall x, y, \xi$;

(iii) $H(x, y) + R(x, y) + M \geqslant \delta(|x|^\nu + |y|^\nu)$;

(iv) $LH(x, y) + LR(x, y) \leqslant -\delta(H(x, y) + R(x, y)) + M$;

(v) $|\partial_y H(x, y) + \partial_y R(x, y)|^2 \leqslant M(H(x, y) + R(x, y) + 1)$.

then the semigroup P_t associated with the equation (5.2) is strong Feller and has a unique invariant measure.

Proof. First, Hypothesis 4.1 holds for $j_0 = 1$ by (ii). Thus P_t is strong Feller by Theorem 4.2. Second, P_t is irreducible by [21, Lemma 2.2]. Hence the invariant measure for P_t is at most one. Third, the invariant measure for P_t exists by [21, Lemma 2.1 and Corollary 2.1]. □

5.3 High Order Stochastic Differential Equations

Consider the following Stochastic Differential Equations with order n,

$$x_t^{(n)} = f(x_t^{(n-1)}, \cdots, x_t) + b(x_t^{(n-1)}, \cdots, x_t)\dot{B}_t, \tag{5.3}$$

where $x_t^{(k)} = \frac{d^k x_t}{dt^k},\ k = 1, \cdots, n,\ x_t \in \mathbb{R}^m,\ b \in \mathbb{R}^{m \times d},\ B_t \in \mathbb{R}^d$.

Set $y_i(t) = x_t^{(i-1)},\ 1 \leqslant i \leqslant n$, then $y_t = (y_1(t), \cdots, y_n(t))$ satisfy the following stochastic differential equation:

$$\begin{cases} dy_1(t) = y_2(t)dt, \\ \vdots \\ dy_{n-1}(t) = y_n(t)dt, \\ dy_n(t) = f(y_n, y_{n-1}, \cdots, y_1)dt + b(y_n, y_{n-1} \cdots, y_1)dB_t. \end{cases} \tag{5.4}$$

Proposition 5.5. *Let x_t^x be the solution of equation (5.3) with initial value $x = (x_0, \cdots, x_0^{(n-1)}) \in \mathbb{R}^{m \times n}$, P_t be the semigroup associated with (5.3),*

(1) *If $f \in C_b^2(\mathbb{R}^{m\times n}; \mathbb{R}^m), b \in C_b^2(\mathbb{R}^{m\times n}; \mathbb{R}^m)$ and $\det(b(x)b^*(x)) \neq 0$, then the law of x_t^x is absolutely continuous with respect to Lebesgue measure, and its density $p(t, x, y)$ is continuous with respect to y and $\sup_y |p(t, x, y)| < \infty$.*

(2) *If $f \in C^2(\mathbb{R}^{m\times n}; \mathbb{R}^m), b \in C^2(\mathbb{R}^{m\times n}; \mathbb{R}^m)$ and for any $x \in \mathbb{R}^{m\times n}$, $\det(b(x)b^*(x)) \neq 0$ and the solution to equation (5.3) with initial value x is globally exists, then the semigroup P_t is strong Feller.*

Proof. Hypothesis 4.1 holds for $j_0 = 1$, hence (1) follows from Proposition 3.6. And (2) follows from Theorem 4.2,. □

Specially, if we consider the following stochastic differential equation

$$x_t^{(n)} + a_{n-1}(x_t)x_t^{(n-1)} + \cdots + a_0(x_t)x_t + c(x_t) + \frac{b(x_t)dB_t}{dt} = 0, \tag{5.5}$$

where $x_t^{(k)} = \frac{d^k x_t}{dt^k},\ x_t \in \mathbb{R}^m,\ B_t \in \mathbb{R}^d,\ b(x_t) \in \mathbb{R}^{m\times d},\ c \in \mathbb{R}^m,\ a_0, \cdots, a_{n-1} \in \mathbb{R}^{m\times m}$.

Corollary 5.6. *Let x_t^x be the solution of equation (5.5) with initial value $x = (x_0, \cdots, x_0^{(n-1)}) \in \mathbb{R}^{m\times n}$.*

(1) *If $a_0, \cdots, a_{n-1} \in C_b^2(\mathbb{R}^m; \mathbb{R}^{m\times m}),\ b \in C_b^2(\mathbb{R}^m; \mathbb{R}^{m\times d}),\ c \in C_b^2(\mathbb{R}^m; \mathbb{R}^m)$, and $\det(b(x_0)b^*(x_0)) \neq 0$, then the law of x_t^x is absolutely continuously with respect to Lebesgue measure, and its density $p(t, x, y)$ is continuous with respect to y and $\sup_y |p(t, x, y)| < \infty$.*

(2) *If $a_0, \cdots, a_{n-1} \in C^2(\mathbb{R}^m; \mathbb{R}^{m\times m}),\ b \in C^2(\mathbb{R}^m; \mathbb{R}^{m\times d}),\ c \in C^2(\mathbb{R}^m; \mathbb{R}^m)$, and for any $x = (x_0, \cdots, x_0^{(n-1)}) \in \mathbb{R}^{m\times n}$, $\det(b(x_0)b^*(x_0)) \neq 0$ and x_t^x is globally exists, then the semigroup P_t is strong Feller.*

Proof. It can be obtained by Proposition 5.5. □

Appendix A

If we set

$$V = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 2 & 2 & 1 & -1 \\ 2 & -3 & 1 & -1 \\ 1 & 0 & -1 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} \overline{X}_t^1 \\ \overline{X}_t^2 \\ \overline{Y}_t^1 \\ \overline{Y}_t^2 \end{pmatrix} = V \begin{pmatrix} X_t^1 \\ X_t^2 \\ Y_t^1 \\ Y_t^2 \end{pmatrix},$$

then it can be shown that $\overline{Y}_t^2 \equiv 0$, hence the Malliavin matrix for $(\overline{X}_t^1, \overline{X}_t^2, \overline{Y}_t^1, \overline{Y}_t^2)$ singular a.s.. For V is invertible, therefor the the Malliavin matrix for (X_t, Y_t) is singular also.

Appendix B: Proof of Lemma 2.14 and Lemma 2.15

The proof of Lemma 2.14 and Lemma 2.15 are very similar to the Norris's proof of Kusuoka-Stroock Lemma (c.f. [15, Lemma 2.3.1]), we only give the proof of Lemma 2.15 here.

Proof. **Proof of Lemma 2.15:** Define stopping time as

$$\zeta = \inf\left\{t \geqslant 0 : \sup_{0\leqslant s\leqslant t}(|a(s)| + |u(s)|) > \epsilon^{-r}\right\} \wedge \sigma,$$

then

$$B = \left\{\int_0^\sigma \tilde{y}(t)^2 dt < \epsilon^q, \int_0^\sigma |u(t)|^2 dt \geqslant \epsilon\right\} \subseteq A_1 \cup A_2 \cup A_3,$$

where

$$\begin{aligned} A_1 &= \left\{\int_0^\sigma \tilde{y}(t)^2 dt < \epsilon^q,\ \int_0^\sigma |u(t)|^2 dt \geqslant \epsilon,\ \zeta = \sigma,\ \sigma \geqslant \epsilon\right\}, \\ A_2 &= \{\zeta < \sigma\}, \\ A_3 &= \{\sigma < \epsilon\}. \end{aligned}$$

Obviously,

$$\mathbb{P}\{A_2\} \leqslant \tilde{c}\epsilon^{rp},\ \mathbb{P}\{A_3\} \leqslant C(c_\sigma, \tilde{p})\epsilon^{\tilde{p}},$$

therefore we only need to estimate $\mathbb{P}(A_1)$.

Introduce the following notation

$$\begin{aligned} N_t &= \sum_{i=1}^d \int_0^t \tilde{y}(s) u_i(s) dW_i(s), \\ M_t &= \sum_{i=1}^d \int_0^t u_i(s) dW_i(s), \\ B &= \left\{\langle N\rangle_\sigma < \rho_1,\ \sup_{0\leqslant s\leqslant \sigma} |N_s| \geqslant \delta_1\right\}, \end{aligned}$$

where $\rho_1 = \epsilon^{q-2r},\ \delta_1 = \epsilon^{\frac{q}{2}-r-\frac{v}{4}}$.

We will prove that there exists $\epsilon_0 = \epsilon_0(c_\sigma, q, r, v)$, such that

$$A_1 \subseteq B,\ \text{for all } \epsilon \leqslant \epsilon_0.$$

If this has been proved, then

$$\mathbb{P}\{A_1\} \leqslant \mathbb{P}\{B\} \leqslant 2\exp\{-\frac{\delta_1^2}{2\rho_1}\} \leqslant \exp\{-\epsilon^{-\frac{v}{4}}\},$$

and this Lemma holds.

In the below, we will to prove: there exists $\epsilon_0 = \epsilon_0(c_\sigma, q, r, v)$, such that

$$A_1 \subseteq B,\ \text{for all } \epsilon \leqslant \epsilon_0.$$

Set $\epsilon_0 = \epsilon_0(c_\sigma, q, r, v)$, such that for $\epsilon \leqslant \epsilon_0(c_\sigma, q, r, v)$, the following inequalities hold.

$$\begin{aligned} &\epsilon^q + 2c_\sigma(\sqrt{c_\sigma}\epsilon^{\frac{q}{2}-r} + \delta_1) \leqslant \epsilon^{\frac{q}{2}-r-\frac{v}{4}}(1 + 2c_\sigma), \\ &\epsilon^{\frac{q}{4}-\frac{r}{2}-\frac{v}{8}}(1 + 2c_\sigma) + \epsilon^{\frac{5q}{4}-\frac{5r}{2}-\frac{v}{8}} < \epsilon. \end{aligned}$$

We only need to prove for any $\epsilon \leqslant \epsilon_0(c_\sigma, q, r, v)$, $\omega \in B^c$ implies $\omega \in A_1^c$.

Let $\omega \in B^c$, $\int_0^\sigma \tilde{y}(t)^2 dt < \epsilon^q$, $\sigma(\omega) = \zeta(\omega) \geqslant \epsilon$, then similar to the estimations of $\sup\limits_{0 \leqslant s \leqslant T} \left| \int_0^t Y_s dY_s \right|$, $\int_0^T \langle M \rangle_t dt$ and $\langle M \rangle_T$ in the proof of Lemma 2.3.2 in [15], we can obtain

$$\begin{aligned}
&\sup_{t \leqslant \sigma} \left| \int_0^t \tilde{y}_s d\tilde{y}_s \right| \leqslant \sqrt{c_\sigma} \epsilon^{\frac{q}{2} - r} + \delta_1, \\
&\int_0^\sigma \langle M \rangle_t dt \leqslant \epsilon^{\frac{q}{2} - r - \frac{v}{4}} (1 + 2c_\sigma), \\
&\langle M \rangle_\sigma \leqslant \gamma^{-1} \epsilon^{\frac{q}{2} - r - \frac{v}{4}} (1 + 2c_\sigma) + \gamma \epsilon^{-2r}, \quad \forall \gamma \in (0, \sigma).
\end{aligned} \tag{5.6}$$

Let $\gamma = \epsilon^{\frac{1}{2}\left(\frac{q}{2} - r - \frac{v}{4}\right)} < \epsilon \leqslant \sigma$ in (5.6). Since $2q > 8 + 20r + v$, we have $\langle M \rangle_\sigma < \epsilon$, i.e. $\omega \in A_1^c$. □

References

[1] Bell, D.R. and Mohammed, S.E.A.: An Extension of Hörmander's Theorem For Infinitely Degenerate Second-Order Operators[J]. *Duke Mathematical Journal* **78(3)**, (1995), 453-475.

[2] Cattiaux, P., Leon, J.R. and Prieur, C.: Estimation for Stochastic Damping Hamiltonian Systems under Partial Observation. *I. Invariant density[J]*, (2012).

[3] Cinti, C., Menozzi, S and Polidoro, S.: Two-sided bounds for degenerate processes with densities supported in subsets of $\mathbb{R}^N$. arXiv:1203.4918, (2012).

[4] Daprato, G. and Zabczyk, J.: Ergodicity for infinite dimensional systems. Cambridge University Press, (1996). MR-1417491.

[5] Delarue, F. and Menozzi, S.: Density estimates for a random noise propagating through a chain of differential equations. *Journal of Functional Analysis* **259(6)**, (2010), 1577-1630. MR-2659772.

[6] Flandoli, F.: Dissipativity and invariant measures for stochastic Navier-Stokes equations. *Nonlinear Differential Equations and Applications* **1**, (1994), 403-423. MR-1300150.

[7] Flandoli, F. and Maslowski, B.: Ergodicity of the 2-D Navier-Stokes Equation Under Random Perturbations. *Communications in Mathematical Physics* **172**, (1995), 119-141. MR-1346374.

[8] Hairer, M. and Mattingly, J.C.: Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Annals of Mathematics.* **164**, (2006), 993-1032. MR-2259251.

[9] Hairer, M. and Mattingly, J.C.: A theory of Hypoellipticity and Unique Ergodicity for Semilinear Stochastic PDEs. *Electronic Journal of Probability* **16**, (2011), 658-738. MR-2786645.

[10] Ichihara, Kanji. and Kunita, Hiroshi.: A classification of the second order degenerate elliptic operators and its probabilistic characterization. *Probability Theory and Related Fields* **30(3)**, (1974), 235-254. MR-0381007.

[11] Kliemann, K.: Recurrence and invariant measure for degenerate diffusions. *The Annals of Probability* **15(2)**, (1987), 690-707. MR-0885138.

[12] Kusuoka, S. and Stroock, D.: Applications of the Malliavin calculus, III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math* **34**, (1987), 391ÍC442. MR-0914028.

[13] L, Rey-Bellet.: Ergodic properties of Markov processes, Open Quantum Systems II **1881**, (2006), 1-39. MR-2248986.

[14] Mattingly, J.C., Stuartb, A.M. and Higham, D.J.: (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and Their Applications* **101**, (2002), 185-232. MR-1931266.

[15] Nualart, D.: The Malliavin Calculus and Related Topics, Springer, (2006). MR-2200233.

[16] Protter, P.E.: Stochastic Integration and Differential Equations, 2nd ed, Springer, (2005). MR-2273672.

[17] Revuz, D. and Yor, M.: Continuous Martingales and Brownian Motion, 3nd ed. Springer, (2005).

[18] Shigekawa, I.: Stochastic Analysis, *Translations of Mathematical Monographs* Vol **224**, (2004). MR-2060917.

[19] Sheu, S.J.: Some estimates of the transition density of a nondegenerate diffusion Markov process[J]. *The Annals of Probability.* **19(2)**, (1991), 538-561. MR-1106275.

[20] Stettner, L.: Remarks on Ergodic Conditions for Markov Processes on Polish Spaces. *Bulletin of the Polish Academy of Sciences - Mathematics* **42**, (1994), 103-114. MR-1810695.

[21] Talay, D.: Stochastic Hamiltonian Systems: Exponential Convergence to the Invariant Measure, and Discretization by the Implicit Euler Scheme. *Markov Processes and Related Fields* **8**, (2002), 163-198. MR-1924934.

Acknowledgments. The authors thank Tusheng Zhang, Xicheng Zhang and Jian Zhou for discussions. This article is supported by 973 Program, No. 2011CB808000 and Key Laboratory of Random Complex Structures and Data Science, No.2008DP173182, NSFC, No.:10721101, 11271356, 11371041. The authors are grateful for the reviewers' suggestions which improved the readability of the article.