Derivative Formula and Applications for Degenerate Diffusion Semigroups *

Feng-Yu Wang$^{a),b)}$ **and Xi-Cheng Zhang**$^{c)}$

a) School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

b) Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK

c) School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

July 7, 2011

Abstract

By using the Malliavin calculus and solving a control problem, Bismut type derivative formulae are established for a class of degenerate diffusion semigroups with non-linear drifts. As applications, explicit gradient estimates and Harnack inequalities are derived.

AMS subject Classification: 60J75, 60J45.
Keywords: Derivative formula, Gradient estimate, Harnack inequality, Stochastic differential equation.

1 Introduction

The Bismut derivative formula introduced in [4], also known as Bismut-Elworthy-Li formula due to [5], is a powerful tool to derive regularity estimates on diffusion semigroups. In the elliptic case this formula can be expressed by using the intrinsic curvature induced by the generator. But in the degenerate case the required curvature lower bound is no longer available. Of course, the Malliavin calculus works also for the hypoelliptic case as shown in e.g. [1] on Riemannian manifolds. In this case the pull-back operator involved

*Supported in part by WIMCS, NNSFC, SRFDP and the Fundamental Research Funds for the Central Universities.

in the formula is normally less explicit, so that it is hard for one to derive explicit gradient estimates. Nevertheless, as shown in [1, §6], in some concrete degenerate cases the derivative formula can be explicitly established by solving certain control problems.

Recently, explicit derivative formulae for damping stochastic Hamiltonian systems have been established in [10] and [3] by using Malliavin calculus and coupling respectively, where the degenerate part is linear. In this case successful couplings with control can be constructed in a very explicit way, so that some known arguments developed in the elliptic setting can be applied. However, when the degenerate part is non-linear, the study becomes much more complicated. The main purpose of this paper is to extend results derived in [10, 3] to the non-linear degenerate case.

Consider the following degenerate stochastic differential equation on $\mathbb{R}^m \times \mathbb{R}^d$:

$$\begin{cases} \mathrm{d}X_t^{(1)} = Z^{(1)}(X_t^{(1)}, X_t^{(2)})\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = Z^{(2)}(X_t^{(1)}, X_t^{(2)})\mathrm{d}t + \sigma \mathrm{d}B_t, \end{cases} \tag{1.1}$$

where $X_t^{(1)}$ and $X_t^{(2)}$ take values in $\mathbb{R}^m$ and $\mathbb{R}^d$ respectively, σ is an invertible $d \times d$-matrix, B_t is a d-dimensional Brownian motion, $Z^{(1)} \in C^2(\mathbb{R}^{m+d}; \mathbb{R}^m)$ and $Z^{(2)} \in C^1(\mathbb{R}^{m+d}; \mathbb{R}^d)$. Let $X_t = (X_t^{(1)}, X_t^{(2)}), Z = (Z^{(1)}, Z^{(2)})$. Then the equation can be formulated as

$$\mathrm{d}X_t = Z(X_t)\mathrm{d}t + (0, \sigma \mathrm{d}B_t). \tag{1.2}$$

We assume that the solution is non-explosive, which is ensured by (H1) below. Our purpose is to establish an explicit derivative formula for the associated Markov semigroup P_t:

$$P_t f(x) = \mathbb{E} f(X_t(x)), \quad t > 0, x \in \mathbb{R}^{m+d}, f \in \mathscr{B}_b(\mathbb{R}^{m+d}),$$

where $X_t(x)$ is the solution of (1.2) with $X_0 = x$, and $\mathscr{B}_b(\mathbb{R}^{m+d})$ is the set of all bounded measurable functions on $\mathbb{R}^{m+d}$.

To compare the present equation with those investigated in [10, 3] where $Z^{(1)}$ is linear, let us recall some simple notations. Firstly, we write the gradient operator on $\mathbb{R}^{m+d}$ as $\nabla = (\nabla^{(1)}, \nabla^{(2)})$, where $\nabla^{(1)}$ and $\nabla^{(2)}$ stand for the gradient operators for the first and the second components respectively, so that $\nabla f : \mathbb{R}^{m+d} \to \mathbb{R}^{m+d}$ for a differentiable function f on $\mathbb{R}^{m+d}$. Next, for a smooth function $\xi = (\xi_1, \cdots, \xi_k) : \mathbb{R}^{m+d} \to \mathbb{R}^k$, let

$$\nabla \xi = \begin{pmatrix} \nabla \xi_1 \\ \vdots \\ \nabla \xi_k \end{pmatrix}, \quad \nabla^{(i)} \xi = \begin{pmatrix} \nabla^{(i)} \xi_1 \\ \vdots \\ \nabla^{(i)} \xi_k \end{pmatrix}, \quad i = 1, 2.$$

Then $\nabla \xi, \nabla^{(1)} \xi, \nabla^{(2)} \xi$ are matrix-valued functions of orders $k \times (m+d), k \times m, k \times d$ respectively. Moreover, for an $l \times k$-matrix $M = (M_{ij})_{1 \le i \le l, 1 \le j \le k}$ and $v = (v_i)_{1 \le i \le k} \in \mathbb{R}^k$, let $Mv \in \mathbb{R}^l$ with $(Mv)_i = \sum_{j=1}^k M_{ij} v_j,\ 1 \le i \le l$. Finally, we will use $\| \cdot \|$ to denote the operator norm for linear operators, for instance, $\|M\| = \sup_{|v|=1} |Mv|$.

When $Z^{(1)}(x^{(1)}, x^{(2)})$ depends only on $x^{(2)}$ and $\nabla^{(2)} Z^{(1)}$ is a constant matrix with rank m, then the equation (1.1) reduces back to the one studied in [3] (and also in [10] for $m = d$). In this case we are able to construct very explicit successful couplings with control, which imply the desired derivative formula and Harnack inequalities as in the elliptic case. But when $Z^{(1)}$ is non-linear, it seems very hard to construct such couplings. The idea of this paper is to split $Z^{(1)}$ into a linear term and a non-linear term, and to derive an explicit derivative formula by controlling the non-linear part using the linear part in a reasonable way. More precisely, let

$$\nabla^{(2)} Z^{(1)} = B_0 + B,$$

where B_0 is a constant $m \times d$-matrix. We will be able to establish derivative formulae for P_t provided B is dominated by B_0 in the sense that

$$\langle BB_0^* a, a\rangle \geq -\varepsilon |B_0^* a|^2, \quad \forall a \in \mathbb{R}^m \tag{1.3}$$

holds for some constant $\varepsilon \in [0, 1)$.

To state our main result, we first briefly recall the integration by parts formula for the Brownian motion. Let $T > 0$ be fixed. For an Hilbert space H, let

$$\mathbb{H}(H) = \left\{ h \in C([0,T]; H) : \; h_0 = 0, \|h\|^2_{\mathbb{H}(H)} := \int_0^T |\dot{h}_t|_H^2 \mathrm{d}t < \infty \right\}$$

be the Cameron-Martin space over H. Let $\mathbb{H} = \mathbb{H}(\mathbb{R}^d)$ and, without confusion in the context, simply denote $\|\cdot\|_{\mathbb{H}} = \|\cdot\|_{\mathbb{H}(H)}$ for any Hilbert space H.

Let μ be the distribution of $\{B_t\}_{t\in[0,T]}$, which is a probability measure (i.e. Wiener measure) on the path space $\Omega = C([0,T]; \mathbb{R}^d)$. The probability space (Ω, μ) is endowed with the natural filtration of the coordinate process $B_t(w) := w_t, t \in [0,T]$. A function $F \in L^2(\Omega; \mu)$ is called differentiable if for any $h \in \mathbb{H}$, the directional derivative

$$D_h F := \lim_{\varepsilon \to 0} \frac{F(\cdot + \varepsilon h) - F(\cdot)}{\varepsilon}$$

exists in $L^2(\Omega; \mu)$. If the map $\mathbb{H} \ni h \mapsto D_h F \in L^2(\Omega; \mu)$ is bounded, then there exists a unique $DF \in L^2(\Omega \to \mathbb{H}; \mu)$ such that $\langle DF, h\rangle_{\mathbb{H}} = D_h F$ holds in $L^2(\Omega; \mu)$ for all $h \in \mathbb{H}$. In this case we write $F \in \mathscr{D}(D)$ and call DF the Malliavin gradient of F. It is well known that $(D, \mathscr{D}(D))$ is a closed operator in $L^2(\Omega; \mu)$, whose adjoint operator $(\delta, \mathscr{D}(\delta))$ is called the divergence operator. That is,

$$\mathbb{E}(D_h F) = \int_\Omega D_h F \mathrm{d}\mu = \int_\Omega F\delta(h)\mathrm{d}\mu = \mathbb{E}(F\delta(h)), \quad F \in \mathscr{D}(D), h \in \mathscr{D}(\delta). \tag{1.4}$$

For any $s \geq 0$, let $\{K(t,s)\}_{t\geq s}$ solve the following random ODE on $\mathbb{R}^m \otimes \mathbb{R}^m$:

$$\frac{\mathrm{d}}{\mathrm{d}t} K(t,s) = (\nabla^{(1)} Z^{(1)})(X_t) K(t,s), \quad K(s,s) = I_{m\times m}. \tag{1.5}$$

We assume

(H) The matrix $\sigma \in \mathbb{R}^d \otimes \mathbb{R}^d$ is invertible, and there exists $W \in C^2(\mathbb{R}^{m+d})$ with $W \geq 1$ and $\lim_{|x|\to\infty} W(x) = \infty$ such that for some constants $C, l_1, l_3 \geq 0$ and $l_2 \in [0,1]$,

(H1) $LW \leq CW,\ |\nabla^{(2)} W|^2 \leq CW$, where $L = \frac{1}{2}\mathrm{Tr}(\sigma\sigma^* \nabla^{(2)}\nabla^{(2)}) + Z \cdot \nabla$;

(H2) $\|\nabla Z^{(1)}\| \leq C, \quad \|\nabla^2 Z^{(1)}\| \leq CW^{l_1}$;

(H3) $\|\nabla Z^{(2)}\| \leq CW^{l_2}, \quad \|\nabla^2 Z^{(2)}\| \leq CW^{l_3}$.

For any $v = (v^{(1)}, v^{(2)}) \in \mathbb{R}^{m+d}$ with $|v| = 1$, we aim to search for $h = h(v) \in \mathscr{D}(\delta)$ such that

$$\nabla_v P_T f(x) = \mathbb{E}\big[f(X_T(x))\delta(h)\big], \quad f \in C_b^1(\mathbb{R}^{m+d}) \tag{1.6}$$

holds. To construct h, for an $\mathbb{H}$-valued random variable $\alpha = (\alpha_s)_{s\in[0,T]}$, let

$$\begin{aligned} g_t &= K(t,0)v^{(1)} + \int_0^t K(t,s)\nabla^{(2)} Z^{(1)}(X_s(x))\alpha_s \mathrm{d}s, \\ h_t &= \int_0^t \sigma^{-1}\big(\nabla Z^{(2)}(X_s(x))(g_s, \alpha_s) - \dot{\alpha}_s\big)\mathrm{d}s, \quad t \in [0,T]. \end{aligned} \tag{1.7}$$

We will show that h satisfies (1.6) provided it is in $\mathscr{D}(\delta)$ and $\alpha_0 = v^{(2)}, \alpha_T = 0, g_T = 0$, see Theorem 2.1 below for details. In particular, it is the case for α_s given in the following result.

Theorem 1.1. *Assume* **(H)** *and let* $\nabla^{(2)} Z^{(1)} = B_0 + B$ *for some constant matrix* B_0 *such that* (*1.3*) *holds for some constant* $\varepsilon \in [0,1)$. *If there exist an increasing fcuntion* $\xi \in C([0,T])$ *and* $\phi \in C^1([0,T])$ *with* $\xi(t) > 0$ *for* $t \in (0,T]$, $\phi(0) = \phi(T) = 0$ *and* $\phi(t) > 0$ *for* $t \in (0,T)$ *such that*

$$\int_0^t \phi(s) K(T,s) B_0 B_0^* K(T,s)^* \mathrm{d}s \geq \xi(t) I_{m\times m}, \quad t \in (0,T]. \tag{1.8}$$

Then

(1) $Q_t := \int_0^t \phi(s)K(T,s)\nabla^{(2)}Z^{(1)}(X_s)B_0^*K(T,s)^*\mathrm{d}s$ *is invertible for* $t \in (0,T]$ *with*

$$\|Q_t^{-1}\| \leq \frac{1}{(1-\varepsilon)\xi(t)}, \quad t \in [0,T]. \tag{1.9}$$

(2) *Let* h *be determined by* (*1.7*) *for*

$$\begin{aligned} \alpha_t := &\frac{T-t}{T} v^{(2)} - \phi(t) B_0^* K(T,t)^* Q_T^{-1} \int_0^T \frac{T-s}{T} K(T,s)\nabla^{(2)} Z^{(1)}(X_s) v^{(2)} \mathrm{d}s \\ &- \frac{\phi(t)B_0^*K(T,t)^*}{\int_0^T \xi(s)^2 \mathrm{d}s} \int_t^T \xi(s)^2 Q_s^{-1} K(T,0) v^{(1)} \mathrm{d}s. \end{aligned} \tag{1.10}$$

Then for any $p \geq 2$, *there exists a constant* $T_p \in (0,\infty)$ *if* $l_2 = 1$ *and* $T_p = \infty$ *if* $l_2 < 1$, *such that for any* $T \in (0,T_p)$, (*1.6*) *holds with* $\mathbb{E}|\delta(h)|^p < \infty$.

(3) *For any $p>1$ there exist constants $c_1(p), c_2(p)\ge 0$, where $c_2(p)=0$ if $l_1=l_2=l_3=0$, such that*

$$|\nabla P_T f| \le c_1(p)(P_T|f|^p)^{1/p}\frac{\sqrt{T\wedge 1}\{(T\wedge 1)^2+\xi(T\wedge 1)\}\mathrm{e}^{c_2(p)W}}{\int_0^{T\wedge 1}\xi(s)^2\mathrm{d}s} \tag{1.11}$$

holds for all $T>0$ and $f\in\mathscr{B}_b(\mathbb{R}^{m+d})$.

The remainder of the paper is organized as follows. In Section 2 we present a general result on the derivative formula by using Malliavin calculus, from which we are able to prove Theorem 1.1 in Sections 3. In Section 4 we will verify (1.8) for the following two cases respectively:

(I) $\nabla^{(1)}Z^{(1)}$ is non-constant but $\mathrm{Rank}[B_0]=m$.

(II) $A:=\nabla^{(1)}Z^{(1)}$ is constant such that $\mathrm{Rank}[B_0, AB_0,\cdots,A^kB_0]=m$ holds for some $0\le k\le m-1$.

In both cases the L^p-gradient estimate (1.11) is derived with specific ξ, while in Case (II) the Harnack inequality introduced in [8] is established provided $\nabla Z^{(1)}$ is constant, which extends the corresponding Harnack inequality obtained in [3] for $\nabla^{(1)}Z^{(1)}=0$ and $\nabla^{(2)}Z^{(1)}$ is constant with rank m. This type Harnack inequality has been applied in the study of heat kernel estimates and contractivity properties of Markov semigroups, see e.g. [3] and references therein.

2 A General Result

In this section we will make use of the following assumption.

(H′) The function

$$U(x):=\mathbb{E}\exp\left[2\int_0^T\|\nabla Z(X_t(x))\|\mathrm{d}t\right],\quad x\in\mathbb{R}^{m+d}$$

is locally bounded.

Theorem 2.1. *Assume* **(H′)** *for some $T>0$. For $v=(v^{(1)},v^{(2)})\in\mathbb{R}^{m+d}$, let $(\alpha_s)_{0\le s\le T}$ be an $\mathbb{H}$-valued random variable such that $\alpha_0=v^{(2)}$ and $\alpha_T=0$, and let g_t and h_t be given in (1.7). If $g_T=0$ and $h\in\mathscr{D}(\delta)$, then (1.6) holds.*

Proof. For simplicity, we will drop the initial data of the solution by writing $X_t(x)=X_t$. By **(H′)** and (1.2) we have $X_t\in\mathscr{D}(D)$, and due to the chain rule and the definition of h_t,

$$\begin{aligned} D_hX_t &= \int_0^t\nabla Z(X_s)D_hX_s\mathrm{d}s+\int_0^t(0,\sigma\dot h_s)\mathrm{d}s\\ &=(0,v^{(2)}-\alpha_t)+\int_0^t\nabla Z(X_s)D_hX_s\mathrm{d}s+\int_0^t\big(0,\nabla Z^{(2)}(X_s)(g_s,\alpha_s)\big)\,\mathrm{d}s \end{aligned}\tag{2.1}$$

holds for $t \in [0, T]$. Next, it is easy to see that

$$g_t = v^{(1)} + \int_0^t \nabla Z^{(1)}(X_s)(g_s, \alpha_s)\mathrm{d}s, \quad t \in [0, T].$$

Combining this with (2.1) we obtain

$$D_h X_t + (g_t, \alpha_t) = v + \int_0^t \nabla Z(X_s)\{D_h X_s + (g_s, \alpha_s)\}\mathrm{d}s, \quad t \in [0, T].$$

On the other hand, the directional derivative process

$$\nabla_v X_t := \lim_{\varepsilon \to 0} \frac{X_t(x + \varepsilon v) - X_t(x)}{\varepsilon}$$

satisfies the same equation, i.e.

$$\nabla_v X_t = v + \int_0^t \nabla Z(X_s)\nabla_v X_s \mathrm{d}s, \quad t \in [0, T]. \tag{2.2}$$

Thus, by the uniqueness of the ODE we conclude that

$$D_h X_t + (g_t, \alpha_t) = \nabla_v X_t, \quad t \in [0, T].$$

In particular, since $(g_T, \alpha_T) = 0$, we have

$$D_h X_T = \nabla_v X_T \tag{2.3}$$

and due to **(H′)** and (2.2),

$$\mathbb{E}|D_h X_T|^2 = \mathbb{E}|\nabla_v X_T|^2 \le |v|^2 \mathbb{E} \exp\left[2\int_0^T \|\nabla Z\|(X_s)\mathrm{d}s\right]. \tag{2.4}$$

Combining this with (1.4) and letting $f \in C_b^1(\mathbb{R}^{m+d})$, we are able to adopt the dominated convergence theorem to obtain

$$\nabla_v P_T f = \mathbb{E}\langle \nabla f(X_T), \nabla_v X_T \rangle = \mathbb{E}\langle \nabla f(X_T), D_h X_T \rangle = \mathbb{E} D_h f(X_T) = \mathbb{E}[f(X_T)\delta(h)].$$

□

Remark 2.1. Using the same argument as above, we also have the following derivative formula:

$$\mathbb{E}\nabla_v f(X_T) = \mathbb{E}\left(f(X_T)\sum_{i,k}\left[\delta(h(e_k))(\nabla X_T)^{-1}_{ki} - D_{h(e_k)}(\nabla X_T)^{-1}_{ki}\right]v^i\right), \tag{2.5}$$

where (e_j) is the canonical basis of $\mathbb{R}^{m+d}$, and $h(e_j)$ is defined by (1.7) with $v = e_j$. In fact, since

$$\sum_k (\partial_k X_T^j)(\nabla X_T)^{-1}_{ki} = 1_{i=j}$$

and by (2.3)

$$D_{h(e_k)} X_T^j = \nabla_{e_k} X_T^j = \partial_k X_T^j,$$

we have

$$\begin{aligned}
\nabla_v f(X_T) = \sum_i (\partial_i f)(X_T) v^i &= \sum_{i,j,k} (\partial_j f)(X_T)(\partial_k X_T^j)(\nabla X_T)^{-1}_{ki} v^i \\
&= \sum_{i,j,k} (\partial_j f)(X_T)(D_{h(e_k)} X_T^j)(\nabla X_T)^{-1}_{ki} v^i \\
&= \sum_{i,k} \{D_{h(e_k)} f(X_T)\}(\nabla X_T)^{-1}_{ki} v^i,
\end{aligned}$$

which implies (2.5) by the integration by parts formula.

Remark 2.2. For the higher order derivative formula, under further regularity assumptions, for any $v_1, \cdots, v_j \in \mathbb{R}^{m+d}$ and $f \in C_b^1(\mathbb{R}^{m+d})$, we have

$$\langle \nabla^j \mathbb{E} f(X_T(x)), v_1 \otimes \cdots \otimes v_j \rangle = \mathbb{E}\left[f(X_T(x)) J_j(T, v_1, \cdots, v_j)\right], \tag{2.6}$$

where $J_1(v) := \delta(h(v))$ and

$$\begin{aligned}
J_j(v_1, \cdots, v_j) := {} & J_{j-1}(v_1, \cdots, v_{j-1})\delta(h(v_j)) + \nabla_{v_j} J_{j-1}(v_1, \cdots, v_{j-1}) \\
& - D_{h(v_j)} J_{j-1}(v_1, \cdots, v_{j-1}),
\end{aligned}$$

where $h(v)$ is defined by (1.7). In fact, as in the proof of Theorem 2.1, we have

$$\begin{aligned}
&\langle \nabla^2 \mathbb{E} f(X_T), v_1 \otimes v_2 \rangle = \nabla_{v_2} \nabla_{v_1} \mathbb{E} f(X_T) = \nabla_{v_2} \mathbb{E}[f(X_T)\delta(h(v_1))] \\
&\quad = \mathbb{E}\left[(\nabla f)(X_T) \cdot \nabla_{v_2} X_T \cdot \delta(h(v_1))\right] + \mathbb{E}\left[f(X_T)\nabla_{v_2}\delta(h(v_1))\right] \\
&\quad = \mathbb{E}\left[(\nabla f)(X_T) \cdot D_{h(v_2)} X_T \cdot \delta(h^{v_1})\right] + \mathbb{E}\left[f(X_T(x))\nabla_{v_2}\delta(h(v_1))\right] \\
&\quad = \mathbb{E}\left[D_{h(v_2)}[f(X_T)]\delta(h(v_1))\right] + \mathbb{E}\left[f(X_T(x))\nabla_{v_2}\delta(h(v_1))\right] \\
&\quad = \mathbb{E}\left[f(X_T(x))\left[\delta(h(v_1))\delta(h_T^{v_2}) - D_{h(v_2)}\delta(h(v_1)) + \nabla_{v_2}\delta(h(v_1))\right]\right].
\end{aligned}$$

The higher derivatives can be obtained by induction.

3 Proof of Theorem 1.1

The idea of the proof is to apply Theorem 2.1 for the given process α_s. Obviously, (H1) implies that for any $l \geq 1$, there exists a constant C_l such that $LW^l \leq C_l W^l$, so that

$\mathbb{E}W(X_t(x))^l \le e^{C_l t}W(x)^l$ and thus, the process is non-explosive; while (H2) and (H3) imply that $\|\nabla Z\| + \|\nabla\nabla^{(1)}Z^{(1)}\| \le CW^{l_1\vee l_2}$ holds for some $C>0$, so that

$$\mathbb{E}\big(\|\nabla Z\|^p + \|\nabla\nabla^{(1)}Z^{(1)}\|^p + \|\nabla^2 Z^{(2)}\|^p\big)(X_t) \le e^{c(p)t}W^{p(l_1\vee l_2\vee l_3)},\quad t\ge 0 \tag{3.1}$$

holds for any $p\ge 1$ with some constant $c(p)>0$. The following lemma ensures that **(H)** implies **(H′)** for all $T>0$ if $l_2<1$ and for small $T>0$ if $l_2=1$.

Lemma 3.1. *If* (H1) *holds, then for any* $T>0$,

$$\mathbb{E}\exp\left[\frac{2}{T^2C\|\sigma\|^2e^{4+2CT}}\int_0^T W(X_t)\mathrm{d}t\right] \le \exp\left[\frac{2W}{TC\|\sigma\|^2e^{2+CT}}\right],\quad T>0.$$

Consequently, (H2) *and* (H3) *imply that* $U := \mathbb{E}\exp[2\int_0^T \|\nabla Z\|(X_t)\mathrm{d}t]$ *is locally bounded on* $\mathbb{R}^{m+d}$ *if either* $l_2<1$ *or* $l_2=1$ *but* $T^2C^2\|\sigma\|^2e^{4+2CT}\le 1$.

Proof. It suffices to prove the first assertion. By the Itô formula and (H1), we have

$$\mathrm{d}W(X_t) = \langle\nabla^{(2)}W(X_t),\sigma\mathrm{d}B_t\rangle + LW(X_t)\mathrm{d}t \le \langle\nabla^{(2)}W(X_t),\sigma\mathrm{d}B_t\rangle + CW(X_t)\mathrm{d}t.$$

So, for $t\in[0,T]$,

$$\mathrm{d}\big\{e^{-(C+2/T)t}W(X_t)\big\} \le e^{-(C+2/T)t}\langle\nabla^{(2)}W(X_t),\sigma\mathrm{d}B_t\rangle - \frac{2}{T}e^{-CT-2}W(X_t)\mathrm{d}t.$$

Thus, letting $\tau_n = \inf\{t\ge 0: W(X_t)\ge n\}$, for any $n\ge 1$ and $\lambda>0$ we have

$$\begin{aligned}
&\mathbb{E}\exp\left[\frac{2\lambda}{Te^{CT+2}}\int_0^{T\wedge\tau_n} W(X_t)\mathrm{d}t\right]\\
&\le e^{\lambda W}\mathbb{E}\exp\left[\lambda\int_0^{T\wedge\tau_n} e^{-(C+2/T)t}\langle\nabla^{(2)}W(X_t),\sigma\mathrm{d}B_t\rangle\right]\\
&\le e^{\lambda W}\left(\mathbb{E}\exp\left[2\lambda^2C\|\sigma\|^2\int_0^{T\wedge\tau_n} W(X_t)\mathrm{d}t\right]\right)^{1/2},
\end{aligned}$$

where the second inequality is due to the exponential martingale and (H1). By taking

$$\lambda = \frac{1}{TC\|\sigma\|^2e^{CT+2}},$$

we arrive at

$$\mathbb{E}\exp\left[\frac{2}{T^2C\|\sigma\|^2e^{4+2CT}}\int_0^{T\wedge\tau_n} W(X_t)\mathrm{d}t\right] \le \exp\left[\frac{2W}{TC\|\sigma\|^2e^{2+CT}}\right].$$

This completes the proof by letting $n\to\infty$. □

To ensure that $\mathbb{E}|\delta(h)|^p < \infty$, we need the following two lemmas.

Lemma 3.2. *Assume* **(H)**. *Then there exists a constant* $c > 0$ *such that*

$$\|DX_t\|_{\mathbb{H}} \le \sqrt{t}\|\sigma\|\mathrm{e}^{c\int_0^t W^{l_2}(X_s)\mathrm{d}s}, t \ge 0. \tag{3.2}$$

Consequently, if $l_2 < 1$, *then for any* $p \ge 1$,

$$\mathbb{E} \sup_{t\in[0,T]} \|DX_t\|_{\mathbb{H}}^p < \infty, \quad T \ge 0,$$

and if $l_2 = 1$, *then for any* $p \ge 1$ *there exists a constant* $T_p > 0$ *such that*

$$\mathbb{E} \sup_{t\in[0,T]} \|DX_t\|_{\mathbb{H}}^p < \infty, \quad T \in (0, T_p).$$

Proof. Due to Lemma 3.1, it suffices to prove (3.2). Obviously, DX_t solves the following $\mathbb{H}$-valued random ODE:

$$DX_t = \int_0^t (\nabla Z)(X_s)DX_s\mathrm{d}s + (0, \sigma)(t \wedge \cdot).$$

Combining this with (H2) and (H3) we obtain

$$\|DX_t\|_{\mathbb{H}} \le C\int_0^t W^{l_2}(X_s)\|DX_s\|_{\mathbb{H}}\mathrm{d}s + \sqrt{t}\|\sigma\|.$$

This implies (3.2) by Gronwall's inequality. □

Lemma 3.3. *Assume* **(H)**. *Then*

$$\|K(T,s)\| \le C\mathrm{e}^{CT}, \quad \|\partial_s K(T,s)\| \le C\mathrm{e}^{CT}, \quad s \in [0,T], \tag{3.3}$$

and

$$\|DK(T,s)\|_{\mathbb{H}} \le (C^2T\mathrm{e}^{CT} + C)\mathrm{e}^{CT}\int_s^T \|DX_r\|_{\mathbb{H}}W^{l_1}(X_r)\mathrm{d}r, \quad s \in [0,T]. \tag{3.4}$$

Consequently, for any $p > 1$ *there exists* $T_p \in (0,\infty)$ *if* $l_2 = 1$ *and* $T_p = \infty$ *if* $l_2 < 1$ *such that*

$$\mathbb{E} \sup_{t\in[0,T]} \|DK(T,t)\|_{\mathbb{H}}^p < \infty, \quad T \in (0, T_p).$$

Proof. By Lemma 3.2 and $\sup_{t\in[0,T]} \mathbb{E}W^l(X_t) < \infty$ for any $l > 0$ as observed in the beginning of this section, it suffices to prove (3.3) and (3.4). First of all, by (1.5) and (H2), we have

$$\|K(t,s)\| \le 1 + \int_s^t \|\nabla^{(1)}Z^{(1)}(X_r)\| \, \|K(r,s)\|\mathrm{d}r \le 1 + C\int_s^t \|K(r,s)\|\mathrm{d}r,$$

which yields the first estimate in (3.3) by Gronwall's inequality. Moreover, noticing that

$$\partial_s K(t,s) = \int_s^t (\nabla^{(1)} Z^{(1)})(X_r)\partial_s K(r,s)\mathrm{d}r - (\nabla^{(1)} Z^{(1)})(X_s),$$

by (H2) we have

$$\|\partial_s K(t,s)\| \le C \int_s^t \|\partial_s K(r,s)\|\mathrm{d}r + C.$$

The second estimate in (3.3) follows. As for (3.4), since

$$\frac{\mathrm{d}}{\mathrm{d}t} DK(t,s) = (\nabla_{DX_t}\nabla^{(1)} Z^{(1)})(X_t)K(t,s) + (\nabla^{(1)} Z^{(1)})(X_t)DK(t,s),$$

with $DK(s,s) = 0$, it follows from (H2) and (3.3) that

$$\begin{aligned}\|DK(t,s)\|_{\mathbb{H}} &\le \int_s^t \|\nabla\nabla^{(1)} Z^{(1)}(X_r)\| \; \|DX_r\|_{\mathbb{H}} \|K(r,s)\|\mathrm{d}r \\ &\quad + \int_s^t \|\nabla^{(1)} Z^{(1)}(X_r)\| \; \|DK(r,s)\|_{\mathbb{H}}\mathrm{d}r \\ &\le C(CT\mathrm{e}^{CT}+1)\int_s^t \|DX_r\|_{\mathbb{H}} W^{l_1}(X_r)\mathrm{d}r \\ &\quad + C\int_s^t \|DK(r,s)\|_{\mathbb{H}}\mathrm{d}r.\end{aligned}$$

This implies (3.4). □

Proof of Theorem 1.1. (1) Let $a \in \mathbb{R}^m$. By (1.3), (1.8) and $\nabla^{(2)} Z^{(1)} = B_0 + B$ we have

$$\begin{aligned}\langle Q_t a, a\rangle &= \int_0^t \phi(s)\Big(\langle K(T,s)B_0B_0^*K(T,s)^*a,a\rangle + \langle K(T,s)B(X_s)B_0^*K(T,s)^*a,a\rangle\Big)\mathrm{d}s \\ &\ge (1-\varepsilon)\int_0^t \phi(s)|B_0^*K(T,s)^*a|^2\mathrm{d}s \ge (1-\varepsilon)\xi(t)|a|^2.\end{aligned}$$

This implies that Q_t is invertible and (1.9) holds.

(2) According to Lemma 3.1, **(H)** implies **(H′)** for all $T>0$ if $l_2<1$ and for small $T>0$ if $l_2=1$. Next, we intend to prove that $h \in \mathscr{D}(\delta)$ and $\mathbb{E}|\delta(h)|^p < \infty$ for small $T>0$ if $l_2 = 1$ and for all $T>0$ if $l_2<1$. Indeed, by Lemmas 3.2, 3.3, (3.1), and the fact that

$$DQ_t^{-1} = -Q_t^{-1}(DQ_t)Q_t^{-1},$$

there exists $T_p > 0$ if $l_2 = 1$ and $T_p = \infty$ if $l_2 < 1$ such that

$$\sup_{t\in[0,T]} \mathbb{E}|DQ_t|^p < +\infty, \quad T \in (0,T_p),$$

and by (1.9),

$$\Big(\mathbb{E}\|DQ_t^{-1}\|_{\mathbb{H}}^p\Big)^{1/p} \le \frac{\big(\mathbb{E}|DQ_t|^p\big)^{1/p}}{[(1-\epsilon)\xi(t)]^2}, \quad t\in(0,T], \tag{3.5}$$

$$\sup_{t\in[0,T]}\Big(\mathbb{E}\|D\alpha_t\|_{\mathbb{H}}^p+\mathbb{E}\|Dg_t\|_{\mathbb{H}}^p\Big)^{1/p}<\infty, \quad T\in(0,T_p). \tag{3.6}$$

Since

$$\begin{aligned}&\dot h_t=\sigma^{-1}\big\{(\nabla Z^{(2)})(X_t)(g_t,\alpha_t)-\dot\alpha_t\big\},\\ &\|D\dot h_t\|_{\mathbb{H}}\le\|\sigma^{-1}\|\big\{\|\nabla^2 Z^{(2)}(X_t)\|\cdot\|DX_t\|_{\mathbb{H}}|(g_t,\alpha_t)|\\ &\qquad\qquad+\|\nabla Z^{(2)}(X_t)\|\|(Dg_t,D\alpha_t)\|_{\mathbb{H}}+\|D\dot\alpha_t\|_{\mathbb{H}}\big\}.\end{aligned} \tag{3.7}$$

we conclude from (H2), (H3), (3.1) and (3.6) that

$$\mathbb{E}\bigg(\int_0^T\|D\dot h_t\|_{\mathbb{H}}^2\mathrm{d}t\bigg)^{p/2}+\mathbb{E}\|h\|_{\mathbb{H}}^p<\infty, \quad T\in(0,T_p).$$

Therefore, according to e.g. [6, Proposition 1.5.8], we have $h\in\mathscr{D}(\delta)$ and $\mathbb{E}|\delta(h)|^p<\infty$ provided $T\in(0,T_p)$.

Now, to prove (1.6), it remains to verify the required conditions of Theorem 2.1 for α_t given by (1.10). Since $\phi(0)=\phi(T)=0$, we have $\alpha_0=v^{(2)}$ and $\alpha_T=0$. Moreover, noting that

$$\begin{aligned}I_1&:=\frac{1}{\int_0^T\xi(t)^2\mathrm{d}t}\int_0^T\phi(t)K(T,t)\nabla^{(2)}Z^{(1)}(X_t)B_0^*K(T,t)^*\mathrm{d}t\int_t^T\xi(s)^2Q_s^{-1}K(T,0)v^{(1)}\mathrm{d}s\\ &=\frac{1}{\int_0^T\xi(t)^2\mathrm{d}t}\int_0^T\dot Q_t\mathrm{d}t\int_t^T\xi(s)^2Q_s^{-1}K(T,0)v^{(1)}\mathrm{d}s\\ &=\frac{1}{\int_0^T\xi(t)^2\mathrm{d}t}\int_0^T\xi(t)^2Q_tQ_t^{-1}K(T,0)v^{(1)}\mathrm{d}t=K(T,0)v^{(1)}\end{aligned}$$

and

$$\begin{aligned}I_2&:=\bigg(\int_0^T\phi(t)K(T,t)\nabla^{(2)}Z^{(1)}(X_t)B_0^*K(T,t)^*\mathrm{d}t\bigg)Q_T^{-1}\int_0^T\frac{T-s}{T}K(T,s)\nabla^{(2)}Z^{(1)}(X_s)v^{(2)}\mathrm{d}s\\ &=Q_TQ_T^{-1}\int_0^T\frac{T-s}{T}K(T,s)\nabla^{(2)}Z^{(1)}(X_s)v^{(2)}\mathrm{d}s=\int_0^T\frac{T-s}{T}K(T,s)\nabla^{(2)}Z^{(1)}(X_s)v^{(2)}\mathrm{d}s,\end{aligned}$$

we obtain by (1.10)

$$\begin{aligned}g_T&=K(T,0)v^{(1)}+\int_0^T K(T,t)\nabla^{(2)}Z^{(1)}(X_t)\alpha_t\mathrm{d}t\\ &=K(T,0)v^{(1)}-I_1+\int_0^T\frac{T-t}{T}K(T,t)\nabla^{(2)}Z^{(1)}(X_t)v^{(2)}\mathrm{d}t-I_2=0.\end{aligned}$$

(3) By an approximation argument, it suffices to prove the desired gradient estimate for $f \in C_b^1(\mathbb{R}^{m+d})$. Moreover, by the semigroup property and the Jensen inequality, we only have to prove for $p \in (1,2]$ and $T \in (0, T_p \wedge 1)$. In this case we obtain from (1.6) that

$$|\nabla P_T f| \le (P_T|f|^p)^{1/p} (\mathbb{E}|\delta(h)|^q)^{1/q},$$

where $q := \frac{p}{p-1} \ge 2$. Therefore, it remains to find constants $c_1, c_2 \ge 0$, where $c_2 = 0$ if $l_1 = l_2 = l_3 = 0$, such that

$$(\mathbb{E}|\delta(h)|^q)^{1/q} \le \frac{c_1\sqrt{T}(T^2 + \xi(T))\mathrm{e}^{c_2 W}}{\int_0^T \xi(s)^2 \mathrm{d}s}. \tag{3.8}$$

To this end, we take $\phi(t) = \frac{t(T-t)}{T^2}$ such that $0 \le \phi \le 1$ and $|\dot\phi(t)| \le \frac{1}{T}$ for $t \in [0,T]$. Since ξ is increasing, by (3.3) and (1.8), we have for some constant $C > 0$,

$$\int_0^t \xi(s)^2 \mathrm{d}s \le \xi(t)^2 \le Ct^2, \quad t \in [0,1].$$

Thus, by Lemmas 3.1, 3.2, 3.3 and (3.1), it is easy to see that for any $\theta \ge 2$ there exist constants $c_1, c_2 \ge 0$, where $c_2 = 0$ if $l_1 = l_2 = l_3 = 0$, such that for all $0 < t \le T \le T_p \wedge 1$,

$$\begin{aligned}
&\Big(\mathbb{E}\|DX_t\|_{\mathbb{H}}^\theta\Big)^{1/\theta} \le c_1\sqrt{T}\mathrm{e}^{c_2 W}, \quad \big(\mathbb{E}\|DK(T,t)\|_{\mathbb{H}}^\theta\big)^{1/\theta} \le c_1 T^{3/2}\mathrm{e}^{c_2 W} \\
&(\mathbb{E}\|DQ_t^{-1}\|_{\mathbb{H}}^\theta)^{1/\theta} \le \big\{\mathbb{E}(\|Q_t^{-1}\|\|DQ_t\|_{\mathbb{H}}\|Q_t^{-1}\|)^\theta\big\}^{1/\theta} \le \frac{c_1 t\sqrt{T}}{\xi(t)^2}\mathrm{e}^{c_2 W}, \\
&(\mathbb{E}\|D\alpha_t\|_{\mathbb{H}}^\theta)^{1/\theta} \le \frac{c_1 T^{5/2}\mathrm{e}^{c_2 W}}{\int_0^T \xi(s)^2\mathrm{d}s}, \quad (\mathbb{E}\|Dg_t\|_{\mathbb{H}}^\theta)^{1/\theta} \le \frac{c_1 T^{7/2}\mathrm{e}^{c_2 W}}{\int_0^T \xi(s)^2\mathrm{d}s}, \\
&(\mathbb{E}\|D\dot\alpha_t\|_{\mathbb{H}}^\theta)^{1/\theta} \le \frac{c_1 T^{3/2}\mathrm{e}^{c_2 W}}{\int_0^T \xi(s)^2\mathrm{d}s}, \quad (\mathbb{E}|\dot h_t|^\theta)^{1/\theta} \le \frac{c_1\xi(T)\mathrm{e}^{c_2 W}}{\int_0^T \xi(s)^2\mathrm{d}s}.
\end{aligned}$$

Combining these with (3.7), (H2), (H3) and (3.1), we obtain

$$\begin{aligned}
\|h\|_{\mathbb{D}^{1,q}} &:= \big(\mathbb{E}\|Dh\|_{\mathbb{H}\otimes\mathbb{H}}^q\big)^{1/q} + \|\mathbb{E}h\|_{\mathbb{H}} \\
&\le \sqrt{T}\bigg\{\mathbb{E}\bigg(\frac{1}{T}\int_0^T \|D\dot h_t\|_{\mathbb{H}}^2\mathrm{d}t\bigg)^{q/2}\bigg\}^{1/q} + \mathbb{E}\|h\|_{\mathbb{H}} \\
&\le \sqrt{T}\bigg(\frac{1}{T}\int_0^T \mathbb{E}\|D\dot h_t\|_{\mathbb{H}}^q\mathrm{d}t\bigg)^{1/q} + \bigg(\mathbb{E}\int_0^T |\dot h_t|^2\mathrm{d}t\bigg)^{1/2} \\
&\le \frac{c_1\sqrt{T}(T^{3/2} + \xi(T))\mathrm{e}^{c_2 W}}{\int_0^T \xi(s)^2\mathrm{d}s}.
\end{aligned}$$

This implies (3.8) since $\delta : \mathbb{D}^{1,q} \to L^q$ is bounded, see e.g. Proposition 1.5.8 in [6]. □

4 Two Specific Cases

As indicated in the end of Section 1, we intend to apply Theorem 1.1 to Case (I) and Case (II) respectively with concrete functions ξ satisfying (1.8).

4.1 Case (I): Rank$[B_0] = m$

Theorem 4.1. *Assume* **(H)** *and (1.3) for some* $\varepsilon \in [0,1)$. *If* $\mathrm{Rank}[B_0] = m$, *then there exist constants* $c_1, c_2 > 0$ *such that (1.8) holds for*

$$\xi(t) = c_1 \int_0^t \phi(s) \mathrm{e}^{-c_2(T-s)} \mathrm{d}s, \quad t \in [0,T].$$

Consequently, for any $p > 1$ *there exist two constants* $c_1(p), c_2(p) \geq 0$, *where* $c_2(p) = 0$ *if* $l_1 = l_2 = l_3 = 0$, *such that*

$$|\nabla P_T f| \leq \frac{c_1(p)(P_T|f|^p)^{1/p}}{(T \wedge 1)^{3/2}} \mathrm{e}^{c_2(p)W}, \quad T > 0.$$

Proof. It is easy to see that the desired gradient estimate follow from (1.11) for the claimed ξ with $\phi(t) = \frac{t(T-t)}{T^2}$, we only prove the first assertion. Since $\nabla^{(1)} Z^{(1)}$ is bounded, there exists a constant $C > 0$ such that

$$|K(T,s)^* a| \geq \mathrm{e}^{-C(T-s)}|a|, \quad a \in \mathbb{R}^m.$$

If $\mathrm{Rank}[B_0] = m$, then $|B_0^* a| \geq c'|a|$ holds for some constant $c' > 0$ and all $a \in \mathbb{R}^m$. Therefore,

$$M_t := \int_0^t \phi(s) K(T,s) B_0 B_0^* K(T,s)^* \mathrm{d}s$$

satisfies

$$\langle M_t a, a \rangle = \int_0^t \phi(s) |B_0^* K(T,s)^* a|^2 \mathrm{d}s \geq c'^2 \int_0^t \phi(s) \mathrm{e}^{-2C(T-s)} |a|^2 \mathrm{d}s.$$

This completes the proof. □

To illustrate this result, let us consider an example where $\nabla^{(2)} Z^{(1)}$ is either uniformly positively definite or uniformly negatively definite. This is especially related to the stochastic Hamiltonian system: Letting $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a C^2-Hamilton function such that that $x^{(2)} \mapsto H(x^{(1)}, x^{(2)})$ is strictly convex/concave uniformly with respect to $x^{(1)}$, then

$$Z = (Z^{(1)}, Z^{(2)}) = (\nabla^{(2)} H, -\nabla^{(1)} H)$$

meets the requirement.

Example 4.1. Let $m = d$ and $\nabla^{(2)} Z^{(1)}$ be symmetric such that for some $C > 0$,

$$CI_{d\times d} \le \nabla^{(2)} Z^{(1)}|_{\mathbb{R}^d}, \text{ or } \nabla^{(2)} Z^{(1)}|_{\mathbb{R}^d} \le -CI_{d\times d}.$$

Then the assertion in Theorem 4.1 holds. Indeed, take $B_0 = CI_{d\times d}$ if $\nabla^{(2)} Z^{(1)}|_{\mathbb{R}^d} \ge CI_{d\times d}$, while $B_0 = -CI_{d\times d}$ if $\nabla^{(2)} Z^{(1)}|_{\mathbb{R}^d} \le -CI_{d\times d}$. Then it is trivial to see that $\text{rank}[B_0] = d = m$ and (1.3) holds for $\varepsilon = 0$.

4.2 Case (II): $A := \nabla^{(1)} Z^{(1)}$ is constant

Throughout this subsection we assume that

(A) (Kalman condition) $A := \nabla^{(1)} Z^{(1)}$ is constant and there exists an integer number $0 \le k \le m - 1$ such that

$$\text{Rank}[B_0, AB_0, \cdots, A^k B_0] = m. \tag{4.1}$$

When $k = 0$, (4.1) means $\text{Rank}[B_0] = m$ which has been considered in Theorem 4.1.

Theorem 4.2. *Assume* **(H), (A)** *and (1.3) for some* $\varepsilon \in (0,1)$. *Let* $\phi(t) = \frac{t(T-t)}{T^2}$. *Then:*

(1) *There exist constants* $c_1, c_2 > 0$ *such that (1.8) holds for*

$$\xi(t) = \frac{c_1 (t\wedge 1)^{2(k+1)}}{T e^{c_2 T}}, \quad t \in [0,T].$$

(2) *For any* $p > 1$, *there exist two constants* $c_1(p), c_2(p) \ge 0$, *where* $c_2(p) = 0$ *if* $l_1 = l_2 = l_3 = 0$, *such that*

$$|\nabla P_T f| \le \frac{c_1(p)(P_T|f|^p)^{1/p}}{(T\wedge 1)^{(4k-1)\vee 0 + 3/2}} e^{c_2(p)W}, \quad T > 0.$$

(3) *If* $\nabla^{(2)} Z^{(1)} = B_0$ *is constant and* $l_2 < \frac{1}{2}$, *then there exists a constant* $c > 0$ *such that*

$$\begin{aligned} |\nabla P_T f| &\le \lambda\big\{P_T f \log f - (P_T f)\log P_T f\big\} \\ &+ \frac{c}{\lambda}\bigg\{\frac{l_2 W}{(1+\lambda^{-1})^2} + \frac{(1+\lambda^{-1})^{4l_2/(1-2l_2)}}{(T\wedge 1)^{(4k+2-2l_2)/(1-2l_2)}} + \frac{1}{(1\wedge T)^{4k+3}}\bigg\} P_T f, \ \ \lambda > 0, T > 0 \end{aligned}$$

holds for all $f \in \mathscr{B}_b^+(\mathbb{R}^{m+d})$, *the set of positive functions in* $\mathscr{B}_b(\mathbb{R}^{m+d})$.

(4) *If* $\nabla^{(2)} Z^{(1)} = B_0$ *is constant and* $l_2 = \frac{1}{2}$, *then there exist constants* $c, c' > 0$ *such that for any* $T > 0, \lambda \ge \frac{c}{(T\wedge 1)^{2k}}$ *and* $f \in \mathscr{B}_b^+(\mathbb{R}^{m+d})$,

$$|\nabla P_T f| \le \lambda\big\{P_T f \log f - (P_T f)\log P_T f\big\} + \frac{c'((1\wedge T)^2 W + 1)}{\lambda (T\wedge 1)^{4k+3}} P_T f.$$

Proof. Since (2) is a direct consequence of (1.11) and (1), we only prove (1), (3) and (4).

(1) Let

$$M_t = \int_0^t \frac{s(T-s)}{T^2} \mathrm{e}^{(T-s)A} B_0 B_0^* \mathrm{e}^{(T-s)A^*} \mathrm{d}s, \quad U_t = \int_0^t \mathrm{e}^{sA} B_0 B_0^* \mathrm{e}^{sA^*} \mathrm{d}s, \quad t \in [0, T].$$

According to [7, §3], the limit

$$Q := \lim_{t \to 0} t^{-(2k+1)} \Gamma_t U_t \Gamma_t$$

exists and is an invertible matrix, where $(\Gamma_t)_{t>0}$ is a family of projection matrices. Thus, $U_t \geq c(t \wedge 1)^{2k+1} I_{m\times m}$ holds for some constant $c > 0$ and all $t > 0$. Then there exist constants $c_1, c_2 > 0$ such that for any $t \in (0, \frac{T}{2}]$,

$$M_t \geq \frac{t}{4T} \int_{t/2}^t \mathrm{e}^{(T-s)A} B_0 B_0^* \mathrm{e}^{(T-s)A^*} \mathrm{d}s \geq \frac{t\mathrm{e}^{-2\|A\|T}}{4T} \int_0^{t/2} \mathrm{e}^{sA} B_0 B_0^* \mathrm{e}^{sA^*} \mathrm{d}s \geq \frac{c_1 t^{2(k+1)}}{4T\mathrm{e}^{c_2 T}} I_{m\times m}$$

holds. This proves the first assertion.

(3) By the semigroup property and the Jensen inequality, we assume that $T \in (0, 1]$. Let $\nabla^{(2)} Z^{(1)} = B_0$ be constant. Then h given in Theorem 1.1 is adapted such that

$$\delta(h) = \int_0^T \langle \dot{h}_t, \mathrm{d}B_t \rangle.$$

Moreover, it is easy to see that for $\xi(t)$ given in (1) and $T \in (0, 1]$,

$$|\dot{h}_t| \leq \frac{c_1(TW^{l_2}(X_t) + 1)}{T^{2(k+1)}}, \quad t \in [0, T]$$

holds for some constant $c_1 > 0$ independent of T. Thus, for any $\lambda > 0$,

$$\begin{aligned} \mathbb{E}\mathrm{e}^{\delta(h)/\lambda} &= \mathbb{E} \exp\left[\frac{1}{\lambda} \int_0^T \langle \dot{h}_t, \mathrm{d}B_t \rangle\right] \leq \left(\mathbb{E} \exp\left[\frac{2}{\lambda^2} \int_0^T |\dot{h}_t|^2 \mathrm{d}t\right]\right)^{1/2} \\ &\leq \left(\mathbb{E} \exp\left[\frac{c_2}{\lambda^2}\Big(\frac{\int_0^T W^{2l_2}(X_t)\mathrm{d}t}{T^{4k+2}} + \frac{1}{T^{4k+3}}\Big)\right]\right)^{1/2}. \end{aligned} \tag{4.2}$$

On the other hand, since $l_2 \in [0, 1]$, by Lemma 3.1 and the Jensen inequality, there exist two constants $c_3, c_4 > 0$ such that

$$\mathbb{E} \exp\left[\frac{c_3 l_2}{T} \int_0^T W(X_t) \mathrm{d}t\right] \leq \mathrm{e}^{c_4 l_2 W}, \quad T \in (0, 1]. \tag{4.3}$$

Moreover, since $2l_2 < 1$, there exists a constant $c_5 > 0$ such that

$$\frac{c_2 W^{2l_2}}{\lambda^2 T^{4k+2}} \leq \frac{c_3 l_2 W}{(1+\lambda)^2 T} + \frac{c_5 (1 + \lambda^{-1})^{4l_2/(1-2l_2)}}{\lambda^2 T^{(4k+2-2l_2)/(1-2l_2)}}, \quad \lambda, T > 0.$$

Combining this with (4.2) and (4.3), we conclude that

$$\log \mathbb{E}e^{\delta(h)/\lambda} \le \frac{cl_2 W}{(1+\lambda)^2} + \frac{c(1+\lambda^{-1})^{4l_2/(1-2l_2)}}{\lambda^2 T^{(4k+2-2l_2)/(1-2l_2)}} + \frac{c}{\lambda^2 T^{4k+3}}, \quad T \in (0,1], \lambda > 0$$

holds for some constant $c > 0$. This completes the proof of (3) by (1.6) and the Young inequality (see [2, Lemma 2.4])

$$|\nabla P_T f| = |\mathbb{E}[f(X_T)\delta(h)]| \le \lambda\big\{P_T f \log f - (P_T f)\log P_T f\big\} + \lambda(P_T f)\log \mathbb{E}e^{\delta(h)/\lambda}. \tag{4.4}$$

(4) Again, we only consider $T \in (0,1]$. Let c_2 and C be in (4.2) and Lemma 3.1 respectively. Then there exists a constant $c > 0$ be a constant such that for any $T \in (0,1]$, $\lambda \ge \frac{c}{T^{2k}}$ implies

$$\frac{c_2}{\lambda^2 T^{4k+2}} \le \frac{2}{T^2 C\|\sigma\|^2 e^{4+2CT}}.$$

Thus, by (4.2) and Lemma 3.1, if $\lambda \ge \frac{c}{T^{2k}}$ then

$$\log \mathbb{E}e^{\delta(h)/\lambda} \le \frac{c_2 T^2 C\|\sigma\|^2 e^{4+2CT}}{4\lambda^2 T^{4k+2}} \log \mathbb{E}\exp\left[\frac{2\int_0^T W(X_t)dt}{T^2 C\|\sigma\|^2 e^{4+2CT}}\right] + \frac{c_2}{\lambda^2 T^{4k+3}} \le \frac{c'(T^2 W + 1)}{\lambda^2 T^{4k+3}}$$

holds for some constant $c' > 0$ independent of T. Combining this with (4.4) we finish the proof. □

To derive the Harnack inequality of P_T from Theorem 4.2 (3) and (4), let us recall a result of [3]. If there exist a constant $\lambda_0 > 0$ and a positive measurable function $\gamma : [\lambda_0, \infty) \times \mathbb{R}^{m+d} \to [0,\infty)$ such that

$$|\nabla_v P_T f| \le \lambda\big\{P_T f \log f - (P_T f)\log P_T f\big\} + \gamma(\lambda,\cdot)P_T f, \quad \lambda \ge \lambda_0 \tag{4.5}$$

holds for some constant $\lambda_0 \in (0,\infty]$ and all $f \in \mathscr{B}_b^+(\mathbb{R}^{m+d})$, then by [3, Proposition 4.1],

$$P_T f(x) \le (P_T f^p)^{1/p}(x+v)\exp\left[\int_0^1 \frac{\gamma(\frac{p-1}{1+(p-1)s}, x+sv)}{1+(p-1)s}ds\right] \tag{4.6}$$

holds for all $f \in \mathscr{B}_b^+(\mathbb{R}^{m+d})$ and $p \ge 1+\lambda_0$. Then we have the following consequence of Theorem 4.2 (3) and (4).

Corollary 4.3. *Let* **(H)** *and* **(A)** *hold such that $\nabla^{(2)} Z^{(1)} = B_0$ is constant.*

(1) *If $l_2 \in [0,1/2)$, then there exists a constant $c > 0$ such that*

$$P_T f(x) \le (P_T f^p)^{1/p}(x+v)$$
$$\times \exp\left[\frac{c|v|^2}{p-1}\Big(\frac{(p-1)l_2\int_0^1 W(x+sv)ds}{p-1+|v|} + \frac{(1+\frac{p|v|}{p-1})^{4l_2/(1-2l_2)}}{(T\wedge 1)^{(4k+2-2l_2)/(1-2l_2)}} + \frac{1}{T^{4k+3}}\Big)\right]$$

holds for all $x, v \in \mathbb{R}^{m+d}, T > 0, p > 1$ and $f \in \mathscr{B}_b^+(\mathbb{R}^{m+d})$.

(2) *If $l_2=1$ then there exist two constants $c,c'>0$ such that for any $T>0, f\in\mathbb{R}^{m+d}$ and $x,v\in\mathbb{R}^{m+d}$,*

$$P_Tf(x)\le (P_Tf^p)^{1/p}(x+v)\exp\left[\frac{c'|v|^2\big\{1+(T\wedge 1)^2\int_0^1 W(x+sv)\mathrm{d}s\big\}}{(p-1)(T\wedge 1)^{4k+3}}\right]$$

holds for $p\ge 1+\frac{c|v|}{(T\wedge 1)^{2k}}$.

Proof. (1) Let $v\in\mathbb{R}^{m+d}$ with $|v|>0$. By Theorem 4.2(3), we have

$$\begin{aligned}|\nabla_v P_Tf|\le&\lambda|v|\big\{P_Tf\log f-(P_Tf)\log P_Tf\big\}\\&+\frac{c|v|}{\lambda}\bigg\{\frac{l_2W}{(1+\lambda^{-1})^2}+\frac{(1+\lambda^{-1})^{4l_2/(1-2l_2)}}{(T\wedge 1)^{(4k+2-2l_2)/(1-2l_2)}}+\frac{1}{(T\wedge 1)^{4k+3}}\bigg\}P_Tf,\ \ \lambda>0.\end{aligned}$$

Replacing λ by $\frac{\lambda}{|v|}$, we see that (4.5) holds for any $\lambda_0>0$ and

$$\gamma(\lambda,\cdot)=\frac{c|v|^2}{\lambda}\bigg\{\frac{l_2W}{(1+|v|\lambda^{-1})^2}+\frac{(1+|v|\lambda^{-1})^{4l_2/(1-2l_2)}}{(T\wedge 1)^{(4k+2-2l_2)/(1-2l_2)}}+\frac{1}{(T\wedge 1)^{4k+3}}\bigg\},\quad \lambda>0.$$

Then the desired Harnack inequality follows from (4.6) since

$$\begin{aligned}&\int_0^1\frac{\gamma(\frac{p-1}{1+(p-1)s},x+sv)}{1+(p-1)s}\mathrm{d}s\\&=\frac{c|v|^2}{p-1}\int_0^1\bigg\{\frac{l_2W(x+sv)}{1+\frac{|v|(1+(p-1)s)}{p-1}}+\frac{(1+\frac{|v|(1+(p-1)s)}{p-1})^{4l_2/(1-2l_2)}}{(T\wedge 1)^{(4k+2-2l_2)/(1-2l_2)}}+\frac{1}{(T\wedge 1)^{4k+3}}\bigg\}\mathrm{d}s\\&\le\frac{c|v|^2}{p-1}\bigg(\frac{l_2(p-1)\int_0^1 W(x+sv)\mathrm{d}s}{p-1+|v|}+\frac{(1+\frac{p|v|}{p-1})^{4l_2/(1-2l_2)}}{(T\wedge 1)^{(4k+2-2l_2)/(1-2l_2)}}+\frac{1}{(T\wedge 1)^{4k+3}}\bigg).\end{aligned}$$

(2) Let $v\in\mathbb{R}^{m+d}$ with $|v|>0$. By Theorem 4.2(4),

$$|\nabla_v P_Tf|\le|v|\lambda\big\{P_Tf\log f-(P_Tf)\log P_Tf\big\}+\frac{c'|v|((1\wedge T)^2W+1)}{\lambda(T\wedge 1)^{4k+3}}P_Tf$$

holds for $\lambda\ge\frac{c}{(T\wedge 1)^{2k}}$. Using $\frac{\lambda}{|v|}$ to replace λ, we see that (4.5) holds for $\lambda_0=\frac{c|v|}{(T\wedge 1)^{2k}}$ and

$$\gamma(\lambda,\cdot)=\frac{c'|v|^2((1\wedge T)^2W+1)}{\lambda(T\wedge 1)^{4k+3}}.$$

Then the proof is completed by (4.6). □

Finally, according to e.g. [9, §4.2], the Harnack inequalities presented above imply explicit heat kernel estimates and entropy-cost inequalities for the invariant probability measure (if exists).

References

[1] M. Arnaudon, A. Thalmaier, *The differentiation of hypoelliptic diffusion semigroups,* Arxiv preprint arXiv:1004.2174.

[2] M. Arnaudon, A. Thalmaier, F.-Y. Wang, *Gradient estimates and Harnack inequalities on non-compact Riemannian manifolds,* Stoch. Proc. Appl. 119(2009), 3653–3670.

[3] A. Guillin, F.-Y. Wang, *Degenerate Fokker-Planck equations : Bismut formula, gradient estimate and Harnack inequality,* Arxiv preprint arXiv:1103.2817

[4] J. M. Bismut, *Large Deviations and the Malliavin Calculus,* Boston: Birkhäuser, MA, 1984.

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

[6] D. Nualart, *The Malliavin Calculus and Related Topics,* Springer, Berlin, 2006.

[7] T. Seidman, *How violent are fast controls?* Mathematics of Control Signals Systems, 1(1988), 89-95.

[8] F.-Y. Wang, *On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups,* Probab. Theory Relat. Fields 108(1997), 87–101.

[9] F.-Y. Wang, *Derivative formula and Harnack inequality for jump processes,* arXiv:1104.5531.

[10] X.-C. Zhang, *Stochastic flows and Bismut formulas for stochastic Hamiltonian systems,* Stoch. Proc. Appl. 120(2010), 1929–1949.