The Annals of Probability
2023, Vol. 51, No. 2, 442–477
https://doi.org/10.1214/22-AOP1599

LOGARITHMIC HEAT KERNEL ESTIMATES WITHOUT CURVATURE RESTRICTIONS

BY XIN CHEN[1,a], XUE-MEI LI[2,b] AND BO WU[3,c]

[1]*School of Mathematical Sciences, Shanghai Jiao Tong University,* [a]*chenxin217@sjtu.edu.cn*
[2]*Imperial College London and EPFL,* [b]*xue-mei.li@epfl.ch*
[3]*School of Mathematical Sciences, Fudan University,* [c]*wubo@fudan.edu.cn*

The main results of the article are short time estimates and asymptotic estimates for the first two order derivatives of the logarithmic heat kernel of a complete Riemannian manifold. We remove all curvature restrictions and also develop several techniques.

A basic tool developed here is intrinsic stochastic variations with prescribed second order covariant differentials, allowing to obtain a path integration representation for the second order derivatives of the heat semigroup P_t on a complete Riemannian manifold, again without any assumptions on the curvature. The novelty is the introduction of an ϵ^2 term in the variation allowing greater control. We also construct a family of cut-off stochastic processes adapted to an exhaustion by compact subsets with smooth boundaries, each process is constructed path by path and differentiable in time. Furthermore, the differentials have locally uniformly bounded moments with respect to the Brownian motion measures, allowing to bypass the lack of continuity of the exit time of the Brownian motions on its initial position.

1. Introduction. Let (M, g) be an n-dimensional connected and complete Riemannian manifold endowed with the Levi–Civita connection ∇. Let Δ denote the Laplace–Beltrami operator, and let $p(t, x, y)$ denote its heat kernel, by which we mean the minimal positive fundamental solution to the equation $\frac{\partial}{\partial t} = \frac{1}{2}\Delta$. The objective of this article is to provide estimates on the first and the second order gradients of $\log p(t, x, \cdot)$ without imposing any curvature conditions on M. For a fixed $x \in M$, we use the abbreviation $\log p$ for the logarithmic heat kernel $\log p(t, x, \cdot)$ and use $\nabla \log p$ and $\nabla^2 \log p$ for its first and second order derivatives, respectively.

We begin with explaining some of the motivations and potential applications. Let $o \in M$ be fixed; we denote

$$P_o(M) := \{\gamma \in C([0, 1]; M) : \gamma(0) = o\}$$

the based path space over M. Likewise, let $L_o(M)$ denote the based loop space over M,

$$L_o(M) := \{\gamma \in P_o(M) : \gamma(0) = \gamma(1) = o\}.$$

A classical problem is to seek a suitable probability measure on $P_o(M)$ or $L_o(M)$, with which analysis on these infinite dimensional nonlinear spaces can be made and understanding of the path spaces can be furthered. If M is compact or more generally with bounded geometry, a natural candidate for the probability measure on $L_o(M)$ is the probability distribution of the diffusion process with the infinitesimal operator

$$L := \frac{1}{2}\Delta + \nabla \log p(1 - t, \cdot, o)$$

Received June 2021; revised August 2022.
MSC2020 subject classifications. Primary 60Gxx, 60Hxx; secondary 58J65, 58J70.
Key words and phrases. Riemannian manifold, heat kernel estimate, curvature, gradient formula.

and the initial value o. This is the Brownian bridge measure. Since there is no analogue of a Lebesque measure, translation invariant, on $L_o(M)$, the Brownian bridge measure is essentially the canonical measure to use. Indeed, for $M = \mathbb{R}^n$, the Brownian bridge measure is a Gaussian measure and it is quasi-invariant under translations of Cameron–Martin vectors. To construct such a diffusion process, which is usually called the Ornstein–Uhlenbeck process, we define a pre-Dirichlet form. This form will be called the Ornstein–Uhlenbeck (O-U) Dirichlet form. To verify that the pre-Dirichlet form yields a Markov process, it is necessary to show it is closed—a property following readily once we have an integration by parts (IBP) formula. The key ingredient for such an IBP formula is suitable short time estimates on $\nabla \log p$ and $\nabla^2 \log p$. We refer the reader to Aida [1, 2], Airault and Malliavin [4], Driver [22], Hsu [43] and Li [52] for more detail.

Another interesting problem is to establish functional inequalities for the O-U Dirichlet form. This includes the Poincaré inequality and logarithmic Sobolev inequality. They describe the long-time behaviours of the associated diffusion process. The logarithmic Sobolev inequality for Gaussian measures was obtained by Gross in the celebrated paper [39]. However, this is not known to hold for loop space over a general manifold M. When M was the hyperbolic space, Poincaré inequality was shown to hold on $L_0(M)$ by the authors of the article [16] and Aida [3]. If M was compact simply connected with strictly positive Ricci curvature, a weak Poincaré inequality with explicit rate function was also established by the authors of the article [17]. It was shown in Gross [40] that the Poincaré inequality for O-U Dirichlet form did not hold on $L_o(M)$ when M was not simply connected. Soon after, Eberle [24] constructed a simply connected compact manifold for which the Poincaré inequality for O-U Dirichlet form did not hold on $L_o(M)$. When the based manifold M was compact, Aida [1], Eberle [23], Gong and Ma [36], Gong, Röckner and Wu [35] and Gross [40] have obtained weighted log-Sobolev inequalities or other different versions of modified log-Sobolev inequalities on $L_o(M)$. In all the results mentioned above, the crucial ingredient was again the asymptotic estimates for $\nabla \log p$ and $\nabla^2 \log p$.

We want to stress that all the results mentioned above have been established for the base manifold M compact or with some bounded geometry conditions, since the short time or asymptotic estimates for $\nabla \log p$ and $\nabla^2 \log p$ were only known for manifolds with such restrictions. Our immediate concern is to study the construction of diffusion processes and functional inequalities on $L_o(M)$ without any bounded geometry conditions on M. We will obtain short time or asymptotic estimates for $\nabla \log p$ and $\nabla^2 \log p$ in this paper. These estimates will be applied to study several problems on $L_o(M)$ in a forthcoming paper [15].

It is intriguing that estimates for $\nabla \log p$ and $\nabla^2 \log p$ are also main tools for proving the continuous counterpart of Talagrand's conjecture for the hypercube $\Omega_n = \{-1, 1\}^n$, which we explain below. Let σ^i denote the configuration with the ith coordinate of σ flipped, and let σ_i denote the ith component of $\sigma \in \Omega_n$. Let $\mu_n \equiv 2^{-n}$ be the uniform measure on Ω_n which is reversible associated with the generator $Lf(\sigma) := \frac{1}{2}\sum_{i=1}^n (f(\sigma^i) - f(\sigma))$, where $\sigma \in \Omega_n$. Setting $T_s f(\sigma) := \int_{\Omega_n} f(\eta)\Pi_{i=1}^n (1 + e^{-s}\sigma_i\eta_i)\, d\mu_n(\eta)$, then Talagrad's conjecture states that, for any $s > 0$, there exists a constant c_s independent of the dimension n such that $\mu_n(\{\sigma : T_s f(\sigma) \geq t\}) \leq c_s \frac{1}{t\sqrt{\log t}}$ for $t > 1$. The value c_s is uniformly in the function f with $\|f\|_{L^1(\mu_n)} = 1$ and in the dimension. The continuous counterpart of the conjecture is for the Ornstein–Uhlenbeck semigroup T_t with generator $\Delta - x \cdot \nabla$

$$\sup_{f \geq 0, \|f\|_{L^1(\gamma_n)} = 1} \gamma_n\big(\{\sigma : T_s f(\sigma) \geq t\}\big) \leq c_s \frac{1}{t\sqrt{\log t}}, \quad t \geq 2,$$

where $\gamma_n \sim N(0, I_{n\times n})$ is the standard n-dimensional normal distribution. This was proven to be affirmative in Ball, Barthe, Bednorz, Oleszkiewicz and Wolff [9]. The dimension free best constants were given in Eldan and Lee [26] and Lehec [47] where the key ingredients are:

(1) For any $g \in L^1(\gamma_n)$ and any $s > 0$, $\nabla^2(\log T_s g) \geq -c_s^2 \,\mathrm{Id}$.

(2) For any $g \in L^1(\gamma_n)$ nonnegative and with $\nabla^2(\log g) \geq -\beta \,\mathrm{Id}$ with a $\beta > 0$, one has $\gamma_n(g \geq t) \leq \frac{C_\beta}{t\sqrt{\log t}}$ for any $t > 1$.

Here, Id is the identity operator. Such estimates for non-Gaussian measures and also for the $M/M/\infty$ queue on $\mathbb{N}$ were obtained by Gozlan, Li, Madiman, Roberto and Samson [37].

2. Main results. The short time and asymptotic estimates are presented in (2.1)–(2.5) below. To the best of our knowledge, such estimates were obtained only for a Riemannian manifold with bounded geometry, including a compact Riemannian manifold. Gradient and Hessian estimates of the form (2.1)–(2.2) were proved by Sheu [64] for $\mathbb{R}^n$ with a nontrivial Riemannian metric where the objective was a nondegenerate parabolic PDEs with bounded derivatives up to order three, and (2.1) for a compact Riemannian manifold can be found in Driver [22], obtained using a result of Hamilton [41], Corollary 1.3 and the Gaussian bounds on heat kernels, see, for example, Li and Yau [48], Cheeger and Yau [14], Davies [20], Setti [63] and Varopoulos [72, 73]. The estimate (2.2) was shown in Hsu [42] again for the compact case. For a noncompact Riemannian manifold with nonnegative Ricci curvature, (2.1) was obtained by Kotschwar [46]. Under a bounded geometry condition together with a volume noncollapsing condition, similar estimates were obtained by Souplet and Zhang [65] and Engoulatov [33]. For the heat kernel associated with the Witten Laplacian operator, these estimates were proved by X.D. Li [49] under a bounded geometry condition on the Bakry–Emery Ricci curvature. In addition, in all the references mentioned above, suitable bounded geometry conditions were required. Likewise, the bounded geometry restrictions are used to derive differential Harnack inequalities and global heat kernel estimates, by Cheeger, Gromov and Taylor [13], Cheng, P. Li and Yau [19], Hamilton [41], P. Li and Yau [48], they provide an important step toward (2.1)–(2.2). Meanwhile, the asymptotic gradient estimate (2.4) was first shown in Bismut [12] for a compact Riemannian manifold. It was extended to the hypoelliptic heat kernel and the heat kernel on a vector bundle, for M with bounded geometry, respectively, by Ben Arous [10], Ben Arous and Léandre [11] and Norris [62], cf. also Azencott [8].

The asymptotic second order gradient estimate (2.5) was established by Malliavin and Stroock [59] for a compact Riemannian manifold. For "asymptotically flat" Riemannian manifolds with poles and bounded geometry this can be found in Aida [2]. On cut-locus estimates was studied by Neel [61].

A natural question is then whether the estimates (2.1)–(2.5) still hold for a general noncompact Riemannian manifold? Note that in Azencott [7], it was illustrated that Gaussian type heat kernel estimates could not be automatically extended to an arbitrary manifold and may fail if the completeness of the Riemannian metric was removed.

We state the main estimate. For any $y \in M$, let $\mathrm{Cut}(y)$ be the cut locus of y and $i(y)$ the injectivity radius of y.

THEOREM 2.1 (Theorems 6.7 and 6.10). *Suppose that M is a complete Riemannian manifold with Riemannian distance d*:

(1) *For every compact subset K of M, the following statements hold*:

(a) *There exists a positive constant $C(K)$, which may depend on K, such that*

$$\left|\nabla_x \log p(t,x,y)\right|_{T_xM} \leq C(K)\left(\frac{1}{\sqrt{t}} + \frac{d(x,y)}{t}\right), \tag{2.1}$$

$$\left|\nabla_x^2 \log p(t,x,y)\right|_{T_xM\otimes T_xM} \leq C(K)\left(\frac{d^2(x,y)}{t^2} + \frac{1}{t}\right) \tag{2.2}$$

for any $x, y \in K$ and for any $t \in (0,1]$.

(b) *For any* $y \in M$ *and* $\delta < i(y)$ *there exist positive constants* t_0 *and* C_1 *such that*

$$\begin{aligned} &\left|t\nabla_x^2 \log p(t,x,y) + \mathbf{I}_{T_xM}\right|_{T_xM\otimes T_xM} \\ &\quad \le C_1\big(d(x,y)+\sqrt{t}\big), \quad x \in B_y(\delta), t \in (0,t_0], \end{aligned} \tag{2.3}$$

where $\mathbf{I}_{T_xM}$ *is the identical map on* T_xM.

(2) *Let* $y \in M$, *and assume that* $\tilde{K} \subset M \setminus \mathrm{Cut}(y)$ *is a compact set. Then*

$$\lim_{t\downarrow 0}\sup_{x\in\tilde{K}}\left|t\nabla_x \log p(t,x,y) + \nabla_x\left(\frac{d^2(x,y)}{2}\right)\right|_{T_xM} = 0, \tag{2.4}$$

$$\lim_{t\downarrow 0}\sup_{x\in\tilde{K}}\left|t\nabla_x^2 \log p(t,x,y) + \nabla_x^2\left(\frac{d^2(x,y)}{2}\right)\right|_{T_xM\otimes T_xM} = 0. \tag{2.5}$$

Remarks on the main theorem. As explained in Section 1, these estimates are crucial for the stochastic analysis of the loop space $L_o(M)$. Despite of the collective efforts, so far, these type of results have been largely proved only for based manifolds with bounded geometry. While in this paper, we only need to assume that the based manifold M is complete and stochastically complete. For analysis on the path space $P_o(M)$ over a general complete Riemanian manifold without curvature conditions, some work have already been done by Chen and Wu [18] and Hsu and Ouyang [44]. For $P_o(M)$, the content of Theorem 2.1 is not essential. In a forthcoming paper [15], we shall apply these to obtain integration by parts formula and construct of O-U Dirichlet form on $L_o(M)$ and to prove several functional inequalities on $L_o(M)$.

Our main idea is to obtain localised asymptotic comparison theorems for the first and the second order gradients of logarithmic heat kernel (see Proposition 6.6 and 6.9 below). One novelty is a new second order derivative formula via a new type of (second order) stochastic variation for Brownian paths on the orthonormal frame bundles which is, in particular, different from that used by Bismut [12] or Stroock [66]. The idea of stochastic variation was initiated in [12] for obtaining an integration by part formula. While the choice of the variation in [66] will produce a term with (the time reverse of) a nonrandom vector field on $L^{2,1}(\Omega;\mathbb{R}^n)$, see also Malliavin and Stroock [59, (1.5)], it seems not possible to replace the nonrandom vector field in their paper by a random one (otherwise the time reversed field is not adapted, hence, Itô's integral is not well defined) which prevents the extension of the formula in [59] to a general noncompact M by a suitable localisation argument. We shall choose a variation (see Section 4 below) with desired properties which, in particular, ensures that the formula for the second order gradient of heat semigroup can take a random vector fields. This is the key step for us to extend the new formula to a general complete M (see, e.g., Theorem 3.1 below). The expression we obtain for the second order gradient of heat semigroup is different from that by Elworthy and Li [32], Li [53, 54], or from that in Arnaudon, Plank and Thalmaier [6] or that in Thompson [69]. We prove the formula by combining the second order stochastic variation (shown to hold for a compact manifold) and approximation arguments (for a noncompact manifold) which is totally different from that in [6, 69]. This new method is adapted for both the proof of Proposition 6.9 here and the integration by parts formula in our forthcoming paper [15].

3. Expression for the second order gradient of heat semigroup. Throughout the paper, $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$ denotes a filtered probability space satisfying the standard assumptions, and $B_t = (B_t^1, B_t^2, \dots, B_t^n)$ is a standard $\mathbb{R}^n$-valued Brownian motion. Let $L(\Omega;\mathbb{R}^n)$ denote the collection of all stochastic processes $h : \mathbb{R}_+ \times \Omega \to \mathbb{R}^n$ which are $\mathscr{F}_t$-adapted. Let

$h'(\cdot,\omega)$ denote the time derivative of $h(\cdot,\omega)$. We define the Cameron–Martin space on the Wiener space as follows:

$$L^{2,1}(\Omega;\mathbb{R}^n) := \Big\{h \in L(\Omega;\mathbb{R}^n) : h(\cdot,\omega) \text{ is absolutely continuous for a.s. } \omega\in\Omega,\\ \text{and } \mathbb{E}\Big[\int_0^1 |h'(s,\omega)|^2\,\mathrm{d}s\Big] < \infty\Big\}.$$

Elements of $L^{2,1}(\Omega;\mathbb{R}^n)$ are usually called (random) Cameron–Martin vectors. Let $C_b(M)$ and $C_c(M)$ denote the collection of all real valued bounded and continuous functions on M and continuous functions with compact supports in M, respectively. Let $\mathfrak{so}(n)$ denote the set of of antisymmetric $n\times n$ matrices, and let $SO(n)$ denote the collection of orthonormal $n\times n$ matrices.

The curvature. Let $\mathbf{R}_x$ denote the sectional curvature tensor, and let Ric_x denote the Ricci curvature tensor at $x\in M$, respectively. Thus, both $\mathbf{R}_x : T_xM\times T_xM \to T_xM\times T_xM$ and $\mathrm{Ric}_x^\sharp : T_xM\to T_xM$ are linear map, the latter is given by the duality

$$\langle \mathrm{Ric}_x^\sharp(v_1), v_2\rangle_{T_xM} = \mathrm{Ric}_x(v_1,v_2), \quad \forall v_1,v_2\in T_xM.$$

The horizontal Brownian motion. Given a point $x\in M$, let O_xM denote the space of linear isometries from $\mathbb{R}^n$ to T_xM. Let $OM := \bigcup_{x\in M} O_xM$, which is the orthonormal frame bundle over M, and let $\pi : OM\to M$ denote the canonical projection which takes a frame $u\in O_xM$ to its base point x. For every $u\in OM$, we define $\mathrm{R}_u : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}^n\times\mathbb{R}^n$ and $\mathrm{ric}_u : \mathbb{R}^n\to\mathbb{R}^n$ by

$$\mathrm{R}_u(e_1,e_2) := u^{-1}\big(\mathbf{R}_{\pi(u)}(ue_1,ue_2)\big),$$
$$\mathrm{ric}_u(e_1) := u^{-1}\big({\mathrm{Ric}^\sharp}_{\pi(u)}(ue_1)\big)$$

for every $e_1,e_2\in\mathbb{R}^n$.

Given a vector $e\in\mathbb{R}^n$, we denote by H_e the associated canonical horizontal vector field on OM with the property that $(T\pi)_u(H_e) = ue\in T_{\pi(u)}M$. Thus, the solution of the ODE

$$u'(t) = H_e\big(u(t)\big)$$

projects to the geodesic on M with the initial position x and the initial speed $u(0)(e)$.

We choose an orthonormal basis $\{e_i\}_{i=1}^n$ of $\mathbb{R}^n$. Suppose $\{U_t\}_{t\ge 0}$ is the solution of following OM-valued Stratonovich stochastic differential equation

$$\mathrm{d}U_t = \sum_{i=1}^n H_{e_i}(U_t)\circ \mathrm{d}B_t^i, \tag{3.1}$$

where the initial value U_0 is a fixed orthonormal basis of T_xM. We usually call $\{U_t\}_{0\le t<\zeta}$ the canonical horizontal Brownian motion, where $\zeta:\Omega\to\mathbb{R}_+$ is the lifetime for U_t. Let $X_t^x := \pi(U_t)$, $0\le t<\zeta(x)$, then X_t^x is a Brownian motion on M with initial value x and life time $\zeta(x)$. This is the celebrated intrinsic construction of M-valued Brownian motion by Eells and Elworthy [25] and Elworthy [28], see also Malliavin [58]. It is well known that the Brownian motion on M does not explode if and only if the horizontal Brownian motion U_t on OM does not explode. In particular, it does not rely on the choice of an isometrically embedding from M to an ambient Euclidean space. Let

$$P_tf(x) := \mathbb{E}\big[f\big(X_t^x\big)\mathbf{1}_{\{t<\zeta(x)\}}\big]$$

be the heat semigroup associated to Brownian motion $X_\cdot$.

The superscript x may be omitted if there is no risk of confusion.

3.1. *Second order gradient of the heat semigroup.* Let $\{U_t\}_{0\le t<\zeta(x)}$ denote the horizontal Brownian motion on M, and $\{X_t^x=\pi(U_t)\}_{0\le t<\zeta(x)}$ is the Brownian motion on M with initial value x and life time $\zeta(x)$. For any $h\in L^{2,1}(\Omega;\mathbb{R}^n)$, we set

$$\Gamma_t^h:=\int_0^t \mathrm{R}_{U_s}(\circ\,\mathrm{d}B_s,h(s)),\qquad \Theta_t^h:=h'(t)+\frac{1}{2}\operatorname{ric}_{U_t}(h(t)). \tag{3.2}$$

It is easy to see that Γ_t^h is an $\mathfrak{so}(n)$-valued process. For $t\ge 0$, we define

$$\Lambda_t^h:=\Gamma_t^h h'(t)+\frac{1}{2}U_t^{-1}\nabla\operatorname{Ric}^\sharp_{X_t}(U_th(t),U_th(t))-\frac{1}{2}\Gamma_t^h\operatorname{ric}_{U_t}(h(t))+\frac{1}{2}\operatorname{ric}_{U_t}(\Gamma_t^h h(t)). \tag{3.3}$$

We are now ready to state one of our main tools, the second order gradient formula on a general complete M.

THEOREM 3.1. *Suppose that M is a complete Riemannian manifold. Let $\{D_m\}_{m=1}^\infty$ denote the increasing family of exhaustive relatively compact open sets of M, and let $\{l_m\}_{m=1}^\infty$ denote the cut-off vector fields, as constructed in Lemma 5.1. Assume that $x\in M$, and there exists $m_0\in\mathbb{N}$ such that $x\in D_{m_0+1}$.*

For every $m>m_0$, $v\in T_xM$ and $t\in(0,1]$, we define

$$h(s):=\left(\frac{t-2s}{t}\right)^+ l_m(s,X_\cdot^x)U_0^{-1}v,\quad s\ge 0.$$

Then $h\in L^{2,1}(\Omega;\mathbb{R}^n)$. Furthermore, for any $f\in C_b(M)$ we have

$$\begin{aligned}&\langle\nabla^2P_tf(x),v\otimes v\rangle_{T_xM\otimes T_xM}\\&\quad=\mathbb{E}_x\left[\left(\left(\int_0^t\langle\Theta_s^h,\mathrm{d}B_s\rangle\right)^2-\int_0^t\langle\Lambda_s^h,\mathrm{d}B_s\rangle-\int_0^t|\Theta_s^h|^2\,\mathrm{d}s\right)f(X_t^x)\mathbf{1}_{\{t<\zeta(x)\}}\right].\end{aligned} \tag{3.4}$$

In particular, the processes $l_m(t,\gamma)$ equals to 1 at any time before γ exits D_{m-1} and equals to zero after it exits D_m for the first time. So it is obvious to see that $h(t,\gamma)=U_0^{-1}v$ at $t=0$ and vanishes after the first exit time of γ from D_m.

3.2. *Comments.* The main idea for proving the second order gradient of the heat semigroup P_t is to approximate the formula on M by those for a family of specific compact manifolds. We first use a result of Greene and Wu [38] to construct a family of relatively compact exhausting open subsets $\{D_m\}_{m=1}^\infty$ which is valid for a complete Riemannian manifold M. This allows to construct a series of random cut-off vector fields $l_m\in L^{2,1}(\Omega;\mathbb{R}^n)$ vanishing, as soon as the sample path exits D_m for the first time, with the necessary quantitative estimates needed for the localisation, see Lemma 5.1 below for details. The lemma is partly inspired by the work of Thalmaier [67] and Thalmaier and Wang [68], where geodesic balls are used. For the purpose of embedding into compact manifolds, we make sure that each D_m having a smooth boundary which, because of the cut locus, can not be taken as granted of geodesic balls on arbitrary Riemannian manifolds.

We want to remark that this offers a more powerful (and also a more reliable) alternative to localisation with stopping times, the latter has been commonly used in stochastic calculus and occasionally incorrectly used. The stopping time argument relies on a continuity assumption on the Brownian motion with respect to the initial value. Such continuity condition seems not easy to verify (for stopping times) and ought not be applied casually, see, for example, Elworthy [27], Li and Sheutzow [55], and Li [51] for more details. Note, however, that exit times from regular domains do have good regularity properties in the sense of Malliavin calculus, we refer the reader to the work of Airault, Maillian, and Ren [5] for more details.

Cut-off vector fields have been previously applied by Arnaudon, Plank and Thalmaier [6], Thompson [69], Thalmaier [67] and Thalmaier and Wang [68] to provide a *localised* differential formula for heat semigroups. As explained earlier, we use a new type of (second order) stochastic variation argument to construct the global second order gradient formula given below. In particular, the expression here is different from that of Elworthy and Li [32], Arnaudon, Plank, and Thalmaier [6], Li [53, 54] and Thompson [69]. And we do not use the doubly parallel translation operators used in [53, 54].

3.3. *Comparison theorems*. The outline of the proof is as follows. We first show that the formula holds for a compact Riemannian manifold, this proof is given in Section 4 using a new stochastic variation. To pass from a compact manifold to a noncompact manifold, we use a suitable isometric embedding from D_m into a compact Riemannian manifold $\tilde{M}_m$ as well as the quantitative cut-off process l_m constructed by Lemma A.1 and Lemma 5.1, respectively.

Denote by $p_{\tilde{M}_m}(t,x,y)$ the heat kernel on $\tilde{M}_m$. Although the heat kernel of a Riemannian manifold is determined in a global manner by the Riemannian metric, we obtain below short time comparison theorems between $\nabla \log p_{\tilde{M}_m}$, $\nabla^2 \log p_{\tilde{M}_m}$ and $\nabla \log p$, $\nabla^2 \log p$. These are used for proving (2.1)–(2.5).

The comparison theorem below allows us to obtain estimates for $\nabla \log p$ and $\nabla^2 \log p$, with the successive applications of first order and second order gradient formula as well as comparison estimates for functionals of the Brownian motions on M and those on $\tilde{M}_m$.

PROPOSITION 3.2 (Propositions 6.6 and 6.9). *Suppose K is a compact subset of M. For any constant $L > 1$, there exists a $m_0 = m_0(K, L) \in \mathbb{N}$, which may depend on K and L, such that, for all $m \geq m_0$, we could find a positive time $t_0 = t_0(K, L, m)$ such that*

$$\sup_{x,y\in K} \left|\nabla_x \log p(t,x,y) - \nabla_x \log p_{\tilde{M}_m}(t,x,y)\right|_{T_xM} \leq C(m)e^{-\frac{L}{t}}, \quad \forall t \in (0, t_0],$$

$$\sup_{x,y\in K} \left|\nabla_x^2 \log p(t,x,y) - \nabla_x^2 \log p_{\tilde{M}_m}(t,x,y)\right|_{T_xM\otimes T_xM} \leq C(m)e^{-\frac{L}{t}}, \quad \forall t \in (0, t_0],$$

where $C(m)$ is a positive constant depending on m.

4. Second order variation on a compact manifold. *Throughout this section, M is an n-dimensional compact Riemannian manifold*. In Proposition 4.4 below, we shall establish (3.4) for a compact manifold which is a fundamental step toward Theorem 3.1.

The first second order differential formula for the heat semigroup P_t was obtained by Elworthy and Li [30] for a noncompact manifold, however, with restrictions on their curvature. Another disadvantage of the formula was its involvement of a nonintrinsic curvature which was due to the application of the derivative flow of gradient stochastic differential equations as well as a martingale approach developed in Li [50]. An intrinsic formula for $\nabla^2 P_t f$ was given by Stroock [66] for a compact Riemannian manifold, while a localised intrinsic formula was obtained by Arnaudon, Plank and Thalmaier [6] with the martingale approach. The study of the second order gradient of the Feynman–Kac semigroup of an operator $\Delta + V$ with a potential function was pioneered by Li [53, 54], where a path integration formula was obtained with the help of doubly damped stochastic parallel transport equation. (The first order gradient formula was previously obtained in Li and Thompson [56]; cf. [30, 31].) A localised version of the Hessian formula (still with doubly stochastic damped parallel translations) for the Feynman–Kac semigroup was derived by Thompson [69].

However, all the expressions mentioned earlier do not seem to lead to our application, such as the proof of Proposition 3.2. To overcome this problem, we introduce a quantitative localisation procedure and obtain a second order gradient formula to which this localisation method can be applied.

One of our main tools is to extend Bismut's idea to perturb the M-valued Brownian motion with initial value $\xi(\varepsilon)$ (where $\xi(\varepsilon)$ is a smooth curve in M), they will be constructed as solutions of a family of SDEs with the driving Brownian motion $\{B_t\}_{t\geq 0}$ rotated and translated appropriately. The rotation and translation exerted on $\{B_t\}_{t\geq 0}$ transmits the variation in the initial value of the Brownian motion on the manifold to variations, in the same parameter, of the Radon Nikodym derivatives of a family of probability measures with respect to which the solutions are Brownian motions on M. This simple and elegant idea was applied in Bismut [12] for deducing an integration by parts formula. Incidentally, such integration by parts formula and the first order gradient formula of the heat semigroup were proved to be equivalent on a compact manifold by Elworthy and Li [32]. In Stroock [66], by calculating the concrete form of the second variation introduced by Bismut, this idea was adapted for obtaining the second order derivative formula for the heat semigroup on a compact manifold. As explained earlier, the choice of stochastic variation in [66] (see also Malliavin and Stroock [59, (1.5)] will produce a term coming from the time reverse of a nonrandom vector field on $L^{2,1}(\Omega;\mathbb{R}^n)$, and it seems not possible to replace the nonrandom vector field by a random one (otherwise, the time reversed field is not adapted, hence, Itô's integral is not well defined). Therefore, the formula obtained in Stroock [66] may not be extended to the one with a random vector field and so is not suitable to for extension to noncompact manifolds with the localisation technique we introduce shortly.

One crucial ingredient for our choice of the stochastic variation is that it ensures (4.10) which implies that the second variation vanishes at time t when we choose a vector field h in the translated part satisfying $h(t)\equiv 0$. This allows us to derive a second order gradient formula with localised vector fields and to extend it to a general (noncompact) complete Riemannian manifold.

4.1. *A novel stochastic variation with a second order term.* As before, $\{U_t\}_{0\leq t<\zeta(x)}$ is the solution of equation (3.1) with initial point U_0 and $\pi(U_0)=x$. In Bismut [12] the following classical perturbation for the driving force B_t was used:

$$\hat{B}_t^\varepsilon=\int_0^t e^{-\varepsilon\Gamma_s^h}\,\mathrm{d}B_s+\varepsilon\int_0^t\left(h'(s)+\frac{1}{2}\mathrm{ric}_{U_s}h(s)\right)\mathrm{d}s,$$

where $h\in L^{2,1}(\Omega;\mathbb{R}^n)$ is a chosen Cameron–Martin vector and $\Gamma_t^h:=\int_0^t \mathrm{R}_{U_s}(\circ\,\mathrm{d}B_s,h(s))$. This perturbation of the noise works well with the first variation for which one needs to ensure that $\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\pi(U_t^\varepsilon)=U_t h(t)$ and has been the popular and standard perturbation, as used also in Driver [21], Fang and Malliavin [34]. Other variation of the noise are also of first order perturbations.

However, with the above mentioned variation, $\frac{\partial^2}{\partial\varepsilon^2}|_{\varepsilon=0}\pi(U_t^\varepsilon)\neq 0$ as long as $h(t)\not\equiv 0$. To solve this problem, we will introduce a second order variation (such perturbation is not unique, and we may find a slightly different choice). Unlike the case with the classical perturbation, this time we cannot avoid differentiating the structure equation so have to choose a connection on the frame bundle. Our approach is inspired by the theory of linear connections induced by a SDE developed by Elworthy, LeJan and Li [29]. We believe that the same method can also be used for higher order variations.

For any $h\in L^{2,1}(\Omega;\mathbb{R}^n)$, we have defined an $\mathfrak{so}(n)$-valued process Γ_t^h and $\mathbb{R}^n$-valued process Θ_t^h, Λ_t^h by (3.2) and (3.3), respectively. We first introduce the translation and define the $\mathbb{R}^n$-valued process $B_t^{\varepsilon,h}$ as follows:

$$B_t^{\varepsilon,h}:=B_t+\varepsilon\int_0^t h'(s)\,\mathrm{d}s+\frac{\varepsilon^2}{2}\int_0^t\Phi_s^h\,\mathrm{d}s,\tag{4.1}$$

where $\Phi_t^h := \Gamma_t^h h'(t)$. We then introduce a rotation for $\mathbb{R}^n$-valued Brownian motion. Let us first set

$$\begin{aligned}\Gamma_t^{(2),h} &:= \int_0^t U_s^{-1} \nabla \mathbf{R}_{\pi(U_s)}\big(U_s h(s), U_s \circ \mathrm{d}B_s, U_s h(s)\big) - \int_0^t \Gamma_s^h \mathrm{R}_{U_s}\big(\circ \mathrm{d}B_s, h(s)\big) \\ &\quad + \int_0^t \mathrm{R}_{U_s}\big(h'(s), h(s)\big)\,\mathrm{d}s + \int_0^t \mathrm{R}_{U_s}\big(\circ \mathrm{d}B_s, \Gamma_s^h h(s)\big).\end{aligned} \tag{4.2}$$

It is easy to see that $\Gamma_t^{(2),h}$ is an $\mathfrak{so}(n)$-valued process. Then, for every $\varepsilon > 0$, we define $SO(n)$-valued process $G_t^{\varepsilon,h}$ as follows:

$$G_t^{\varepsilon,h} := \exp\left(-\varepsilon \Gamma_t^h - \frac{\varepsilon^2}{2} \Gamma_t^{(2),h}\right),$$

where $\exp : \mathfrak{so}(n) \to SO(n)$ is the exponential map in the Lie algebra $\mathfrak{so}(n)$ of $SO(n)$.

We can now introduce $\tilde{B}_t^{\varepsilon,h}$, the variation of B_t as well as the corresponding equation on OM.

DEFINITION 4.1. Let $\xi(\varepsilon)$, $\varepsilon \in (-1, 1)$, be a geodesic with $\xi(0) = x$. Let $\{U_0^{\varepsilon,h} : \varepsilon \in (-1, 1)\}$ be a parallel orthonormal frame along $\xi(\varepsilon)$ with $\pi(U_0^{\varepsilon,h}) = \xi(\varepsilon)$. Let U_t^ϵ denote the solution of the following equation with initial condition $U_0^{\varepsilon,h}$,

$$\begin{aligned}\mathrm{d}U_t^{\varepsilon,h} &= \sum_{i=1}^n H_{e_i}\big(U_t^{\varepsilon,h}\big) \circ \mathrm{d}\tilde{B}_t^{\varepsilon,h,i}, \\ \mathrm{d}\tilde{B}_t^{\varepsilon,h} &= G_t^{\varepsilon,h} \circ \mathrm{d}B_t^{\varepsilon,h}, \qquad \tilde{B}_0^{\varepsilon,h} = 0.\end{aligned} \tag{4.3}$$

We define $X_t^{\varepsilon,\xi(\varepsilon),h} = \pi(U_t^{\varepsilon,h})$. If $\varepsilon = 0$, then $X_t^{0,x,h} = X_t^x$ with $X_t^x = \pi(U_t)$.

We remark that the perturbation in $U_t^{\varepsilon,h}$ has a translation part $B_t^{\varepsilon,h}$ and a rotation part $G_t^{\varepsilon,h}$. The rotation $G_t^{\varepsilon,h}$ is chosen to offset precisely the twisting effects induced by the second order stochastic variation.

For simplicity we omit the subscript h, in Θ_t^h, Λ_t^h, $X_t^{\varepsilon,h}$, Γ_t^h, $\Gamma_t^{(2),h}$, $G_t^{\varepsilon,h}$, $B_t^{\varepsilon,h}$ and $U_t^{\varepsilon,h}$, from time to time.

Let ϖ and θ denote, respectively, the $\mathfrak{so}(n)$-valued connection 1-form and the $\mathbb{R}^n$-valued solder 1-form, respectively. Set

$$\varpi_t^\varepsilon := \varpi\left(\frac{\partial}{\partial \varepsilon} U_t^\varepsilon\right), \qquad \theta_t^\varepsilon := \theta\left(\frac{\partial}{\partial \varepsilon} U_t^\varepsilon\right).$$

Through this paper we use D_t, d_t to denote the stochastic covariant differential for vector fields and stochastic differential on M along a semimartingale, respectively, and $\frac{D}{\partial \varepsilon}$ denotes the covariant derivative for vector fields on M with respect to the variable ε.

LEMMA 4.1. *If we choose $h \in L^{2,1}(\Omega; \mathbb{R}^n)$ such that $h(0) = U_0^{-1}(\frac{\partial}{\partial \varepsilon}|_{\varepsilon=0}\xi(\varepsilon))$, then*

$$\varpi_t^\varepsilon = \int_0^t \mathrm{R}_{U_s^\varepsilon}\big(G_s^\varepsilon \circ \mathrm{d}B_s^\varepsilon, \theta_s^\varepsilon\big). \tag{4.4}$$

And θ_t^ε satisfies the following equation:

$$\begin{cases}\mathrm{d}\theta_t^\varepsilon = -\big(\Gamma_t + \varepsilon \Gamma_t^{(2)}\big) G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon + \varpi_t^\varepsilon G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon + G_t^\varepsilon\big(h'(t) + \varepsilon \Phi_t\big)\,\mathrm{d}t, \\ \theta_0^\varepsilon = \big(U_0^\varepsilon\big)^{-1} \dfrac{\mathrm{d}\xi(\varepsilon)}{\mathrm{d}\varepsilon}.\end{cases} \tag{4.5}$$

In particular, we have

$$\begin{cases} \theta_t^0 := \theta\left(\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} U_t^\varepsilon\right) = h(t), \\ \varpi_t^0 = \Gamma_t, \\ \frac{D}{\partial\varepsilon}\Big|_{\varepsilon=0} (U_t^\varepsilon G_t^\varepsilon e) = 0, \quad \forall e \in \mathbb{R}^d, \\ \frac{\partial X_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} = U_t h(t). \end{cases} \tag{4.6}$$

PROOF. We first use the structure equation

$$\begin{aligned} \mathrm{d}\varpi\left(\mathrm{d}_t U_t^\varepsilon, \frac{\partial}{\partial\varepsilon} U_t^\varepsilon\right) &= -\varpi \wedge \varpi\left(\mathrm{d}_t U_t^\varepsilon, \frac{\partial}{\partial\varepsilon} U_t^\varepsilon\right) + \mathrm{R}_{U_t^\varepsilon}\left(\theta(\mathrm{d}_t U_t^\varepsilon), \theta\left(\frac{\partial}{\partial\varepsilon} U_t^\varepsilon\right)\right) \\ &= -\sum_{i=1}^n \varpi \wedge \varpi\left(H_{e_i}(U_t^\varepsilon) \circ \mathrm{d}\tilde{B}_t^{\varepsilon,i}, \frac{\partial}{\partial\varepsilon} U_t^\varepsilon\right) + \mathrm{R}_{U_t^\varepsilon}\big(\theta(\mathrm{d}_t U_t^\varepsilon), \theta_t^\varepsilon\big) \\ &= \mathrm{R}_{U_t^\varepsilon}\big(\theta(\mathrm{d}_t U_t^\varepsilon), \theta_t^\varepsilon\big) \end{aligned}$$

to obtain

$$\mathrm{d}\varpi_t^\varepsilon = \mathrm{d}\varpi\left(\mathrm{d}_t U_t^\varepsilon, \frac{\partial}{\partial\varepsilon} U_t^\varepsilon\right) = \mathrm{R}_{U_t^\varepsilon}\big(\theta(\mathrm{d}_t U_t^\varepsilon), \theta_t^\varepsilon\big) = \mathrm{R}_{U_t^\varepsilon}\big(G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon, \theta_t^\varepsilon\big).$$

Since at time 0, the variation $\{U_0^\varepsilon; \varepsilon \in (-1,1)\}$ is parallel along the geodesic ξ, $\varpi_0^\varepsilon = 0$. Then (4.4) follows immediately.

Here we have used the Transfer Principle: on the compact manifold M we could *treat the Stratonovich integral as the ordinary derivative* (with respect to time variable) in the computation. Crucially, we could *exchange the order of differentiations and integrations*. The transfer principle is well known for *compact* manifolds, see, for example, [34] or [57], but not automatically apply to noncompact manifolds nor automatically to the less smooth case nor to the derivative processes. This is used in similar computations later in the article without further comment.

Due to the torsion free property, the time derivative and the derivative for ε could commute: $D_t \frac{\partial}{\partial\varepsilon} = \frac{D}{\partial\varepsilon} \mathrm{d}_t$. Also note that $\theta_t^\varepsilon = (U_t^\varepsilon)^{-1} T\pi(\frac{\partial}{\partial\varepsilon} U_t^\varepsilon)$, so we have,

$$\begin{aligned} \mathrm{d}\theta_t^\varepsilon &= (U_t^\varepsilon)^{-1}\left(D_t\left(\frac{\partial}{\partial\varepsilon} X_t^\varepsilon\right)\right) = (U_t^\varepsilon)^{-1}\left(\frac{D}{\partial\varepsilon} \mathrm{d}_t X_t^\varepsilon\right) \\ &= (U_t^\varepsilon)^{-1}\left(\frac{D}{\partial\varepsilon}(U_t^\varepsilon G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon)\right) \\ &= \varpi_t^\varepsilon G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon + \frac{\partial G_t^\varepsilon}{\partial\varepsilon} \circ \mathrm{d}B_t^\varepsilon + G_t^\varepsilon \circ \mathrm{d}\left(\frac{\partial}{\partial\varepsilon} B_t^\varepsilon\right) \\ &= \varpi_t^\varepsilon G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon - \big(\Gamma_t + \varepsilon\Gamma_t^{(2)}\big) G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon + G_t^\varepsilon\big(h'(t) + \varepsilon\Phi_t\big)\,\mathrm{d}t, \end{aligned} \tag{4.7}$$

where the fourth equality is due to

$$\frac{D}{\partial\varepsilon}(U_t^\varepsilon G_t^\varepsilon) = U_t^\varepsilon\left(\varpi_t^\varepsilon G_t^\varepsilon + \frac{\partial}{\partial\varepsilon} G_t^\varepsilon\right). \tag{4.8}$$

So we have obtained the first equation in (4.5). The initial condition in (4.5) follows trivially from the fact $\theta_0^\varepsilon = (U_0^\varepsilon)^{-1}\pi(\frac{\partial}{\partial\varepsilon} U_0^\varepsilon))$, $\{U_0^\varepsilon; \varepsilon \in (-1,1)\}$ is a parallel orthonormal frame bundle along $\xi(\cdot)$ and $X_0^\varepsilon = \xi(\varepsilon)$.

Based on the fact that

$$\varpi_t^0 = \int_0^t \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, \theta_s^0), \qquad \Gamma_t = \int_0^t \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, h(s))$$

and taking $\varepsilon = 0$ in (4.5), we arrive at

$$\mathrm{d}\theta_t^0 = \left(\int_0^t \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, \theta_s^0) - \int_0^t \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, h(s))\right) \circ \mathrm{d}B_t + h'(t)\,\mathrm{d}t, \qquad \theta_0^0 = h(0).$$

It is easy to verify that $\theta_t^0 = h(t)$ is the unique solution to above equation, proving the first line of (4.6). Then plugging in $\theta_t^0 = h(t)$ into (4.4) to see that $\varpi_t^0 = \Gamma_t$, so we have

$$\frac{D}{\partial\varepsilon}\bigg|_{\varepsilon=0} (U_t^\varepsilon G_t^\varepsilon e) = U_t\left(\varpi_t^0 e + \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} G_t^\varepsilon e\right) = U_t(\Gamma_t e - \Gamma_t e) = 0,$$

which is the third line of (4.6). Finally, $D_t\left(\frac{\partial X_t^\varepsilon}{\partial\varepsilon}|_{\varepsilon=0}\right) = U_t\,\mathrm{d}\theta_t^0 = U_t h'(t)\,\mathrm{d}t$, giving $\frac{\partial X_t^\varepsilon}{\partial\varepsilon}|_{\varepsilon=0} = U_t h(t)$. This completes the proof. □

In particular, we obtain the following lemma.

LEMMA 4.2. *For every $h \in L^{2,1}(\Omega; \mathbb{R}^n)$ with $h(0) \equiv v = U_0^{-1}(\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\xi(\varepsilon))$, we have*

$$\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \varpi_t^\varepsilon = \int_0^t \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, \eta_s) + \Gamma_t^{(2)}, \tag{4.9}$$

where $\eta_s := \frac{\partial\theta_s^\varepsilon}{\partial\varepsilon}|_{\varepsilon=0}$ and $\Gamma_t^{(2)}$ is defined by (4.2).

PROOF. By the first line of (4.6) we have $\theta_t^0 = h(t)$. We differentiate the integral expression (4.4) for ϖ_t^ε and apply the third line of (4.6) to obtain

$$\begin{aligned}
\frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \varpi_t^\varepsilon &= \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \int_0^t (U_s^\varepsilon)^{-1} \mathrm{R}_{X_s^\varepsilon}(U_s^\varepsilon G_s^\varepsilon \circ \mathrm{d}B_s^\varepsilon, U_s^\varepsilon \theta_s^\varepsilon) \\
&= \int_0^t \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} (G_s^\varepsilon (U_s^\varepsilon G_s^\varepsilon)^{-1} \mathrm{R}_{X_s^\varepsilon}(U_s^\varepsilon G_s^\varepsilon \circ \mathrm{d}B_s^\varepsilon, U_s^\varepsilon G_s^\varepsilon (G_s^\varepsilon)^{-1}\theta_s^\varepsilon)) \\
&= \int_0^t \left(\frac{\partial G_s^\varepsilon}{\partial\varepsilon}\bigg|_{\varepsilon=0}\right) \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, \theta_s^0) + \int_0^t U_s^{-1} \nabla \mathrm{R}_{X_s}(U_s\theta_s^0, U_s \circ \mathrm{d}B_s, U_s\theta_s^0) \\
&\quad + \int_0^t \mathrm{R}_{U_s}\left(\circ\, \mathrm{d}\frac{\partial B_s^\varepsilon}{\partial\varepsilon}\bigg|_{\varepsilon=0}, \theta_s^0\right) + \int_0^t \mathrm{R}_{U_s}\left(\circ\, \mathrm{d}B_s, \left(\frac{\partial (G_s^\varepsilon)^{-1}}{\partial\varepsilon}\bigg|_{\varepsilon=0}\right)\theta_s^0\right) \\
&\quad + \int_0^t \mathrm{R}_{U_s}\left(\circ\, \mathrm{d}B_s, \frac{\partial\theta_s^\varepsilon}{\partial\varepsilon}\bigg|_{\varepsilon=0}\right).
\end{aligned}$$

Here, the last term is $\int_0^t \mathrm{R}_{U_s}(\circ\, \mathrm{d}B_s, \eta_s)$, while the sum of the rest is $\Gamma_t^{(2)}$, so we have completed the proof. □

We observe that $\eta_s = \frac{\partial\theta_s^\varepsilon}{\partial\varepsilon}|_{\varepsilon=0}$ is essentially the second variation of $\pi(U_s^\epsilon)$.

LEMMA 4.3. *For every $h \in L^{2,1}(\Omega; \mathbb{R}^n)$ with $h(0) \equiv v = U_0^{-1}(\frac{\partial}{\partial\varepsilon}|_{\varepsilon=0}\xi(\varepsilon))$, we have $\eta_t \equiv 0$ for all $t \in [0, 1]$ and*

$$\frac{D}{\partial\varepsilon}\bigg|_{\varepsilon=0}\left(\frac{\partial X_t^\varepsilon}{\partial\varepsilon}\right) = U_t \Gamma_t h(t). \tag{4.10}$$

PROOF. We recall the first equation of (4.5),

$$\mathrm{d}\theta_t^\varepsilon = -\big(\Gamma_t + \varepsilon\Gamma_t^{(2)}\big)G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon + \varpi_t^\varepsilon G_t^\varepsilon \circ \mathrm{d}B_t^\varepsilon + G_t^\varepsilon\big(h'(t) + \varepsilon\Gamma_t h'(t)\big)\,\mathrm{d}t.$$

Differentiating it at $\varepsilon = 0$, using (4.9) and the following fact:

$$\varpi_t^0 = \Gamma_t, \qquad \frac{\partial B_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} = h'(t), \qquad \frac{\partial G_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} = -\Gamma_t, \qquad \Phi_t = \Gamma_t h'(t),$$

we could obtain

$$\begin{aligned}
\mathrm{d}\eta_t &= -\left(\Gamma_t^{(2)} + \Gamma_t \frac{\partial G_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right) \circ \mathrm{d}B_t - \Gamma_t \circ \mathrm{d}\left(\frac{\partial B_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right)\\
&\quad + \left(\varpi_t^0 \frac{\partial G_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} + \frac{\partial\varpi_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right) \circ \mathrm{d}B_t\\
&\quad + \varpi_t^0\,\mathrm{d}\left(\frac{\partial B_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right) + \frac{\partial G_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} h'(t)\,\mathrm{d}t + \Gamma_t h'(t)\,\mathrm{d}t\\
&= \left(\int_0^t \mathrm{R}_{U_s}(\circ\,\mathrm{d}B_s, \eta_s)\right) \circ \mathrm{d}B_t.
\end{aligned}$$

At the same time, since $X_0^\varepsilon = \xi(\varepsilon)$, $\xi(\cdot)$ is a geodesic and also $\{U_0^\varepsilon, \varepsilon \in (-1,1)\}$ is a parallel orthonormal frame bundle along $\xi(\cdot)$, we could verify that

$$\eta_0 = \frac{\partial\theta_0^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0} = U_0^{-1}\left(\frac{D}{\partial\varepsilon}\Big|_{\varepsilon=0}\left(\frac{\partial\xi(\varepsilon)}{\partial\varepsilon}\right)\right) = 0.$$

Observe that the unique solution to following equation is $v_t \equiv 0$:

$$\mathrm{d}v_t = \left(\int_0^t \mathrm{R}_{U_s}(\circ\,\mathrm{d}B_s, v_s)\right) \circ \mathrm{d}B_t, \qquad v_0 = 0.$$

Then we derive that $\eta_t \equiv 0$ for all $t \in [0,1]$.

Moreover, note that by definition we have $\frac{\partial X_t^\varepsilon}{\partial\varepsilon} = U_t^\varepsilon\theta_t^\varepsilon$, due to the fact $\eta_t = \frac{\partial\theta_t^\varepsilon}{\partial\varepsilon}|_{\varepsilon=0} \equiv 0$, we obtain

$$\frac{D}{\partial\varepsilon}\Big|_{\varepsilon=0}\left(\frac{\partial X_t^\varepsilon}{\partial\varepsilon}\right) = \frac{D}{\partial\varepsilon}\Big|_{\varepsilon=0}(U_t^\varepsilon\theta_t^\varepsilon) = U_t\left(\varpi_t^0\theta_t^0 + \frac{\partial\theta_t^\varepsilon}{\partial\varepsilon}\Big|_{\varepsilon=0}\right) = U_t\Gamma_t h(t).$$

Now we have obtained (4.10). □

4.2. *Proof for the second order gradient formula on a compact manifold.*

PROPOSITION 4.4. *Let $t > 0$, $x \in M$ and $v \in T_xM$. Then, for any $f \in C_b(M)$ and $h \in L^{2,1}(\Omega; \mathbb{R}^n)$, satisfying that $h(0) = U_0^{-1}v$ and $h(t) = 0$ a.s., we have*

$$\langle \nabla P_t f(x), v\rangle_{T_xM} = -\mathbb{E}\left[f(X_t^x)\int_0^t \langle \Theta_s^h, \mathrm{d}B_s\rangle\right], \tag{4.11}$$

where $\Theta_t^h := h'(t) + \frac{1}{2}\mathrm{ric}_{U_t}(h(t))$. Furthermore,

$$\begin{aligned}
&\langle \nabla^2 P_t f(x), v\otimes v\rangle_{T_xM\otimes T_xM}\\
&\quad = \mathbb{E}\left[f(X_t^x)\left(\left(\int_0^t \langle \Theta_s^h, \mathrm{d}B_s\rangle\right)^2 - \int_0^t \langle \Lambda_s^h, \mathrm{d}B_s\rangle - \int_0^t |\Theta_s^h|^2\,\mathrm{d}s\right)\right].
\end{aligned} \tag{4.12}$$

PROOF. We take $\xi(\cdot)$ to be a geodesic with initial value $\xi(0)=x$ and initial velocity $\frac{\partial \xi(\varepsilon)}{\partial \varepsilon}|_{\varepsilon=0}=v$. Let $\{U_0^\varepsilon \in (-1,1)\}$ denote the parallel orthonormal frame bundle along $\xi(\cdot)$ with $U_0^\varepsilon|_{\varepsilon=0}=U_0$. In particular, it holds that $\pi(U_0^\varepsilon)=\xi(\varepsilon)$. Recall that U_t^ε is the solution to (4.3) with initial value U_0^ε chosen above. It holds that

$$\begin{aligned}\int_0^t G_s^\varepsilon \circ \mathrm{d}B_s^\varepsilon &= \int_0^t G_s^\varepsilon \circ \mathrm{d}B_s + \int_0^t G_s^\varepsilon\Big(\varepsilon h'(s) + \frac{\varepsilon^2}{2}\Gamma_s h'(s)\Big)\,\mathrm{d}s \\ &= \int_0^t G_s^\varepsilon\,\mathrm{d}B_s + \int_0^t \frac{1}{2}\,\mathrm{d}\langle G_\cdot^\varepsilon, B_\cdot\rangle_s + \int_0^t G_s^\varepsilon\Big(\varepsilon h'(s) + \frac{\varepsilon^2}{2}\Gamma_s h'(s)\Big)\,\mathrm{d}s \\ &= \int_0^t G_s^\varepsilon\,\mathrm{d}B_s + \varepsilon\int_0^t G_s^\varepsilon \Theta_s\,\mathrm{d}s + \frac{\varepsilon^2}{2}\int_0^t G_s^\varepsilon \Lambda_s\,\mathrm{d}s.\end{aligned}$$

Here we have used that

$$\begin{aligned}\mathrm{d}\langle G_\cdot^\varepsilon, B_\cdot\rangle_t &= -\varepsilon G_t^\varepsilon\,\mathrm{d}\langle \Gamma_\cdot, B_\cdot\rangle_t - \frac{\varepsilon^2}{2} G_t^\varepsilon\,\mathrm{d}\langle \Gamma_\cdot^{(2)}, B_\cdot\rangle_t \\ &= \frac{\varepsilon}{2} G_t^\varepsilon\,\mathrm{ric}_{U_t}\big(h(t)\big)\,\mathrm{d}t + \frac{\varepsilon^2}{2} G_t^\varepsilon\big[U_t^{-1}\nabla \mathrm{Ric}^\sharp_{\pi(U_t)}\big(U_t h(t), U_t h(t)\big) \\ &\quad + \Gamma_t\,\mathrm{ric}_{U_t}\big(h(t)\big) - \mathrm{ric}_{U_t}\big(\Gamma_t h(t)\big)\big].\end{aligned}$$

Note that $W_t^\varepsilon := \int_0^t G_s^\varepsilon\,\mathrm{d}B_s$ is still an $\mathbb{R}^n$-valued Brownian motion, so we have

$$\mathrm{d}U_t^\varepsilon = H\big(U_t^\varepsilon\big)\circ\Big(\mathrm{d}W_t^\varepsilon + G_s^\varepsilon\Big(\varepsilon\Theta_t + \frac{\varepsilon^2}{2}\Lambda_t\Big)\,\mathrm{d}t\Big).$$

Let

$$M_t^\varepsilon := \exp\Big(-\int_0^t\Big\langle \varepsilon\Theta_s + \frac{\varepsilon^2}{2}\Lambda_s, \mathrm{d}B_s\Big\rangle - \int_0^t\Big(\frac{\varepsilon^2}{2}\Big|\Theta_s + \frac{\varepsilon}{2}\Lambda_s\Big|^2\Big)\,\mathrm{d}s\Big).$$

Then by the Girsanov theorem, the distribution of $\{U_s^\varepsilon; s\in[0,t]\}$ under $\mathrm{d}\mathbb{Q}^\varepsilon := M_t^\varepsilon\,\mathrm{d}\mathbb{P}$ is the same as that of $\{U_s^{0,\varepsilon}; s\in[0,t]\}$, where $U_\cdot^{0,\varepsilon}$ is the solution to equation (3.1) with initial value $U_0^{0,\varepsilon}=U_0^\varepsilon$. Therefore, we obtain

$$P_t f\big(\xi(\varepsilon)\big) = \mathbb{E}\big[f\big(X_t^{\xi(\varepsilon)}\big)\big] = \mathbb{E}\big[f\big(X_t^{\varepsilon,\xi(\varepsilon)}\big)M_t^\varepsilon\big], \tag{4.13}$$

where $X_t^{\varepsilon,\xi(\varepsilon)}=\pi(U_t^\varepsilon)$, $X_t^{\xi(\varepsilon)}=\pi(U_t^{0,\varepsilon})$.

We first assume $f\in C_b^2(M)$, differentiating (4.13) with respect to ε yields that

$$\begin{aligned}\langle \nabla P_t f(x), v\rangle_{T_xM} &= \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} P_t f\big(\xi(\varepsilon)\big) \\ &= \mathbb{E}\Big[\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} f\big(X_t^{\varepsilon,\xi(\varepsilon)}\big)\Big] + \mathbb{E}\Big[f(X_t)\Big(\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} M_t^\varepsilon\Big)\Big].\end{aligned} \tag{4.14}$$

Another round of differentiation gives

$$\begin{aligned}&\langle \nabla^2 P_t f(x), v\otimes v\rangle_{T_xM\otimes T_xM} \\ &\quad = \frac{\partial^2}{\partial\varepsilon^2}\Big|_{\varepsilon=0} P_t f\big(\xi(\varepsilon)\big) \\ &\quad = \mathbb{E}\Big[\frac{\partial^2}{\partial\varepsilon^2}\Big|_{\varepsilon=0} f\big(X_t^{\varepsilon,\xi(\varepsilon)}\big)\Big] + 2\mathbb{E}\Big[\Big(\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} f\big(X_t^{\varepsilon,\xi(\varepsilon)}\big)\Big)\Big(\frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0} M_t^\varepsilon\Big)\Big] \\ &\qquad + \mathbb{E}\Big[f\big(X_t^x\big)\frac{\partial^2}{\partial\varepsilon^2}\Big|_{\varepsilon=0} M_t^\varepsilon\Big].\end{aligned} \tag{4.15}$$

According to the last line of (4.6), (4.10), the definition of M_t^ε and the fact that $h(t) \equiv 0$, we derive

$$\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} f\big(X_t^{\varepsilon,\xi(\varepsilon)}\big) = \langle \nabla f\big(X_t^x\big), U_t h(t)\rangle_{T_{X_t^x} M} = 0$$

and also

$$\frac{\partial}{\partial \varepsilon}\bigg|_{\varepsilon=0} M_t^\varepsilon = -\int_0^t \langle \Theta_s, \mathrm{d}B_s\rangle.$$

Furthermore,

$$\begin{aligned}
\frac{\partial^2}{\partial \varepsilon^2}\bigg|_{\varepsilon=0} f\big(X_t^{\varepsilon,\xi(\varepsilon)}\big) &= \bigg\langle \nabla^2 f\big(X_t^x\big), \frac{\partial X_t^{\varepsilon,\xi(\varepsilon)}}{\partial \varepsilon}\bigg|_{\varepsilon=0} \otimes \frac{\partial X_t^{\varepsilon,\xi(\varepsilon)}}{\partial \varepsilon}\bigg|_{\varepsilon=0}\bigg\rangle_{T_{X_t^x} M \otimes T_{X_t^x} M} \\
&\quad + \bigg\langle \nabla f\big(X_t^x\big), \frac{D}{\partial \varepsilon}\bigg|_{\varepsilon=0}\bigg(\frac{\partial X_t^{\varepsilon,\xi(\varepsilon)}}{\partial \varepsilon}\bigg)\bigg\rangle_{T_{X_t^x} M} \\
&= \langle \nabla^2 f\big(X_t^x\big), U_t h(t) \otimes U_t h(t)\rangle_{T_{X_t^x} M \otimes T_{X_t^x} M} + \langle \nabla f\big(X_t^x\big), U_t \Gamma_t h(t)\rangle_{T_{X_t^x} M} \\
&= 0,
\end{aligned}$$

$$\frac{\partial^2}{\partial \varepsilon^2}\bigg|_{\varepsilon=0} M_t^\varepsilon = \bigg(\int_0^t \langle \Theta_s, \mathrm{d}B_s\rangle\bigg)^2 - \int_0^t \langle \Lambda_s, \mathrm{d}B_s\rangle - \int_0^t |\Theta_s|^2 \,\mathrm{d}s.$$

Crucially, this special choice of variation ensures that $\frac{\partial^2}{\partial \varepsilon^2}|_{\varepsilon=0} f(X_t^{\varepsilon,\xi(\varepsilon)})$ depends only on $h(t)$, not on the history of the process h.

Putting these back to (4.14) and (4.15) yields (4.11), (4.12) for $f \in C_b^2(M)$. By standard approximation procedure and the compact property of M, we see that these equalities still hold for any $f \in C_b(M)$. □

5. Quantitative cut-off processes. *From now on, we assume that M is an n-dimensional general complete Riemannian manifold, not necessarily compact.*

In this section we introduce a class of cut-off processes satisfying estimates crucial for the localisation procedures, which we shall apply later to (4.12) and to obtain the asymptotic gradient estimates for the logarithmic heat kernel.

Since geodesic balls have typically nonregular boundary, we first construct a family of relatively compact open sets $\{D_m\}_{m=1}^\infty$ with smooth boundary which plays the roles of geodesic balls and such that $\bigcup_{m=1}^\infty D_m = M$. Our localisation procedure crucially relies on D_m has smooth boundaries, see Lemma A.1. We first use a result in Greene and Wu [38] on the existence of a smooth approximate distance function, which is valid for complete manifold, and then construct a family of cut off vector fields adapted to $\{D_m\}_{m=1}^\infty$. Fixing an $o \in M$, denote by d the Riemannian distance function on M from o. Since M is complete, according to [38], there exists a nonnegative smooth function $\hat{d} : M \to \mathbb{R}_+$ with the property that $0 < |\nabla \hat{d}| \le 1$ and

$$\bigg|\hat{d}(x) - \frac{1}{2} d(x)\bigg| < 1, \quad \forall x \in M.$$

For every nonnegative m, define $D_m := \hat{d}^{-1}((-\infty, m)) := \{z \in M; \hat{d}(z) < m\}$, then it is easy to verify $B_o(2m-2) \subset D_m \subset B_o(2m+2)$, where $B_o(r) := \{z \in M; d(z) < r\}$ is the geodesic ball centred at o with radius r. Let $\phi : \mathbb{R} \to [0, 1]$ be a smooth function such that

$$\phi(r) = \begin{cases} 1, & r \le 1, \\ \in (0, 1), & r \in (1, 2), \\ 0, & r \ge 2. \end{cases} \tag{5.1}$$

Setting

$$f_m(z) := \phi\big(\hat{d}(z) - m + 2\big), \quad z \in M, \tag{5.2}$$

then it is easy to see that

$$f_m(z) = \begin{cases} 1, & \text{if } z \in \overline{D}_{m-1}, \\ 0, & \text{if } z \in D_m^c, \\ \in (0,1), & \text{otherwise} \end{cases}$$

and $D_m = \{z \in M; f_m(z) > 0\}$. Without loss of generality, we can assume that D_m is a bounded connected open set (otherwise, we could take the connected component of D_m containing $B_o(2m-2)$). Moreover, since $\partial D_m = \{z \in M; \hat{d}(z) = m\}$ and $|\nabla \hat{d}(z)| \neq 0$ for all $z \in M$, we know ∂D_m is a smooth $n-1$ dimensional submanifold of M.

As before, we suppose that $\{U_t\}_{0 \le t < \zeta(x)}$ is the solution to the canonical horizontal equation (3.1) with $\zeta(x)$ denoting its explosion time, and $\{X_t^x := \pi(U_t)\}_{0 \le t < \zeta(x)}$ is a Brownian motion on M with initial value $x := \pi(U_0)$.

Let ∂ denote the cemetery state for M, and set $\bar{M} = M \cup \{\partial\}$. Given a $x \in M$, we let

$$P_x(\bar{M}) := \{\gamma \in C\big([0,1]; \bar{M}\big) : \gamma(0) = x\}$$

denote the collection of all $\bar{M}$-valued continuous paths with initial vale x. Let μ_x denote the Brownian motion measure on $P_x(\bar{M})$. We also refer the natural filtration of the canonical process $\gamma(\cdot)$ as the canonical filtration on $P_x(\bar{M})$ which is augmented to be complete and right continuous as usual.

It is well known that the distribution of $\{X_t^x\}_{0 \le t < \zeta(x)}$ and $\{U_t\}_{0 \le t < \zeta(x)}$ under $\mathbb{P}$ is the same as that of the canonical process $\{\gamma(t)\}_{0 \le t < \zeta(\gamma)}$ and its horizontal lift $\{U_t(\gamma)\}_{0 \le t < \zeta(\gamma)}$ under μ_x, where $\zeta(\gamma)$ denotes the explosion time of $\gamma(\cdot)$. Set

$$\tau_m(\gamma) = \tau_{D_m}(\gamma) := \inf\{s \ge 0 : \gamma(s) \notin D_m\}.$$

LEMMA 5.1. *For any $m \in \mathbb{N}$, there exists a stochastic process (vector field) $l_m : [0,1] \times P_x(\bar{M}) \to [0,1]$ such that*:

(1) $l_m(t, \gamma) = \begin{cases} 1, & t \le \tau_{m-1}(\gamma) \wedge 1, \\ 0, & t > \tau_m(\gamma). \end{cases}$

(2) Absolute continuity: *$l_m(t, \cdot)$ is adapted to the canonical filtration and $l_m(\cdot, \gamma)$ is absolutely continuous for μ_x-a.s. $\gamma \in P_x(\bar{M})$.*

(3) Local uniform moment estimates: *For every positive integer $k \in \mathbb{N}$, we have*

$$\sup_{x \in D_{m-1}} \int_{P_x(\bar{M})} \int_0^1 |l'_m(s, \gamma)|^k \, ds \, \mu_x(d\gamma) \le C_1(m, k) \tag{5.3}$$

for some positive constant $C_1(m, k)$ (which may depend on m and k).

PROOF. In the proof, the constant C (which may depend on m) will change in different lines. The main idea of the proof is inspired by the article of Thalmaier [67] and Thalmaier and Wang [68].

(1) Since for any $m \ge 1$, $D_m \subset D_{m+1} \uparrow M$, there exists a $m_0 \in \mathbb{N}$ such that

$$\begin{cases} x \in D_m, & \text{when } m \ge m_0, \\ x \notin D_m, & \text{when } 1 \le m < m_0. \end{cases}$$

When $x \notin D_m$, let $l_m(t,\gamma) \equiv 0$. In the following we will consider the case of $x \in D_m$ (which implies that $\tau_m(\gamma) > 0$) without loss of generality. Let $f_m : M \to \mathbb{R}_+$ be the function given by (5.2), we define a sequence of functions

$$T_m(t,\gamma) := \begin{cases} \int_0^t \frac{\mathrm{d}s}{[f_m(\gamma(s))]^2}, & t < \tau_m(\gamma), \\ \infty, & t \ge \tau_m(\gamma). \end{cases}$$

Then each $T_m(\cdot,\gamma)$ is an increasing right continuous function of t. For any $t \ge 0$, set

$$A_m(t,\gamma) := \inf\{s \ge 0 : T_m(s,\gamma) \ge t\}.$$

We may omit the parameter γ in the notation of $T_m(t,\gamma)$, $A_m(t,\gamma)$ for simplicity in the proof.

Since $\inf_{s\in[0,t]} f_m(\gamma(t)) > 0$ for $t < \tau_m(\gamma)$, then $T_m(t) < \infty$ for every $t < \tau_m$ and $T_m(\cdot)$ is strictly increasing and continuous in $[0,\tau_m)$ (with respect to the variable t). Therefore, $A_m(\cdot)$ is continuous on $[0, T_m(\tau_m))$ and $T_m(A_m(t)) = t$ for every $0 \le \tau_m < T_m(\tau_m)$. Furthermore, we have $T_m(\tau_m) = \infty$. To see this, we only need to observe that

$$\begin{aligned} f_m(\gamma(s)) = f_m(\gamma(s)) - f_m(\gamma(\tau_m)) &\le \frac{1}{2} \sup_{x\in D_m} |\nabla^2 f_m(x)| d(\gamma(s),\gamma(\tau_m))^2 \\ &\le C_m(\gamma)\sqrt{|s-\tau_m|}, \quad \forall s < \tau_m, \end{aligned}$$

where $C_m(\gamma)$ is a constant, and we applied the property that $d(\gamma(s),\gamma(\tau_m)) \le C_m(\gamma)|s - \tau_m|^{1/4}$ which is easy to prove by the Kolmogorov criterion. Combing the fact $T_m(\tau_m) = \infty$ with $T_m(t) < \infty$ for all $0 \le t < \tau_m$ immediately yields that $A_m(T_m(t)) = t$ for every $0 \le t \le \tau_m$ and $\tau_m > A_m(t)$ for every $0 \le t < \infty$.

Next, we use the truncation function $\phi : \mathbb{R} \to \mathbb{R}$ in (5.1) to define

$$l_m(t,\gamma) = \phi\left(\int_0^t \frac{\phi(T_m(s)-2)}{f_m^2(\gamma(s))}\,\mathrm{d}s\right) \tag{5.4}$$

which is clearly adapted to the canonical filtration. Suppose that $t \ge \tau_m > A_m(3)$, then

$$\begin{aligned} \int_0^t \frac{\phi(T_m(s)-2)}{f_m^2(\gamma(s))}\,\mathrm{d}s &\ge \int_0^t \mathbf{1}_{\{T_m(s)\le 3\}} f_m^{-2}(\gamma(s))\,\mathrm{d}s \\ &= \int_0^t \mathbf{1}_{\{s\le A_m(3)\}} f_m^{-2}(\gamma(s))\,\mathrm{d}s \\ &= T_m(A_m(3)) = 3, \end{aligned}$$

which implies $l_m(t,\gamma) = 0$ for $t \ge \tau_m$ by the definition of ϕ.

If $s \le \tau_{m-1}(\gamma)$, then $f_m(\gamma(s)) = 1$ and so $T_m(s) = s$. Consequently, $\phi(T_m(s)-2) = 1$ for every $s \le \tau_{m-1} \wedge 1$. Hence, we obtain

$$l_m(t,\gamma) = \phi(t \wedge 1) = 1, \quad \forall t \le \tau_{m-1} \wedge 1,$$

concluding the proof of part (1).

(2) Still by the expression of (5.4), we know the conclusion of part (2) holds.

(3) Now it only remains to verify the estimates (5.3). First,

$$\begin{aligned} |l_m'(t)| &= \left|\phi'\left(\int_0^t \frac{\phi(T_m(s)-2)}{f_m^2(\gamma(s))}\,\mathrm{d}s\right)\right| \frac{\phi(T_m(t)-2)}{f_m^2(\gamma(t))} \\ &\le \|\phi'\|_\infty f_m^{-2}(\gamma(t))\mathbf{1}_{\{\phi(T_m(t)-2)\ne 0\}} \le C f_m^{-2}(\gamma(t))\mathbf{1}_{\{T_m(t)\le 4\}}. \end{aligned}$$

Then, for every $k \in \mathbb{N}$,

$$
\begin{aligned}
\int_0^1 |l_m'(s)|^k \, \mathrm{d}s &\le C \int_0^1 f_m^{-2k}(\gamma(s)) \mathbf{1}_{\{T_m(s) \le 4\}} \, \mathrm{d}s \\
&\le C \int_0^1 f_m^{-2k+2}(\gamma(s)) \mathbf{1}_{\{s \le A_m(4)\}} \, \mathrm{d}T_m(s) \\
&= C \int_0^{4 \wedge T_m(1)} f_m^{-2k+2}(\gamma(A_m(r))) \, \mathrm{d}r \\
&\le C \int_0^4 f_m^{-2k+2}(\gamma(A_m(r))) \, \mathrm{d}r.
\end{aligned} \tag{5.5}
$$

Observe that the distribution of $X_\cdot^x$ under $\mathbb{P}$ is the same as that of $\gamma(\cdot)$ under μ_x,

$$
\begin{aligned}
&\sup_{x \in D_{m-1}} \int_{P_x(\bar{M})} \int_0^4 f_m^{-2k+2}(\gamma(A_m(s))) \, \mathrm{d}s \mu_x(\mathrm{d}\gamma) \\
&\quad = \sup_{x \in D_{m-1}} \mathbb{E}\left[\int_0^4 f_m^{-2k+2}(X^x_{A_m(s, X_\cdot^x)}) \, \mathrm{d}s \right].
\end{aligned} \tag{5.6}
$$

Let $S_{j,m}(\gamma) := \inf\{t > 0; f_m(\gamma(t)) \le \frac{1}{j}\}$. According to Itô's formula, we obtain for all $j, k \in \mathbb{N}$ and $x \in D_{m-1}$,

$$
\begin{aligned}
\mathbb{E}[f_m^{-k}(X_{A_m(t) \wedge S_{j,m}})] &= f_m^{-k}(x) + \frac{1}{2} \mathbb{E}\left[\int_0^{A_m(t) \wedge S_{j,m}} \Delta(f_m^{-k})(X_s) \, \mathrm{d}s \right] \\
&= 1 + \frac{1}{2} \mathbb{E}\left[\int_0^{A_m(t) \wedge S_{j,m}} (f_m^2 \Delta(f_m^{-k}))(X_{A_m(T_m(s))}) \, \mathrm{d}T_m(s) \right].
\end{aligned} \tag{5.7}
$$

We have applied the fact that $A_m(T_m(s)) = s$ for every $0 \le s < S_{j,m}$ and $f_m(x) = 1$ for all $x \in D_{m-1}$. Meanwhile, we have

$$
\begin{aligned}
f_m^2 \Delta(f_m^{-k}) &= k(k+1) f_m^{-k} |\nabla f_m|^2 - k f_m^{-k+1} \Delta f_m \\
&= k(k+1) f_m^{-k} |\phi'(\hat{d} - m + 2)|^2 |\nabla \hat{d}|^2 \\
&\quad - k f_m^{-k} (f_m \phi''(\hat{d} - m + 2) |\nabla \hat{d}|^2 + \phi'(\hat{d} - m + 2) f_m \Delta \hat{d}) \\
&\le k(k+1) f_m^{-k} \Big(\|\phi'\|_\infty + \|\phi''\|_\infty + \|\phi'\|_\infty \sup_{z \in D_m} |\Delta \hat{d}(z)| \Big) \\
&\le C f_m^{-k}.
\end{aligned}
$$

Putting this into (5.7), we arrive at

$$
\begin{aligned}
\mathbb{E}[f_m^{-k}(X_{A_m(t) \wedge S_{j,m}})] &\le 1 + C \mathbb{E}\left[\int_0^{A_m(t) \wedge S_{j,m}} f_m^{-k}(X_{A_m(T_m(s))}) \, \mathrm{d}T_m(s) \right] \\
&\le 1 + C \int_0^t \mathbb{E}[f_m^{-k}(X_{A_m(r) \wedge S_{j,m}})] \, \mathrm{d}r,
\end{aligned}
$$

where the last step follows from the procedure of change of variable $u = T_m(s)$ and the fact $A_m(t) \le t$.

Hence, by Grownwall's inequality we arrive at, for all $k, j \in \mathbb{N}$,

$$
\mathbb{E}[f_m^{-k}(X_{A_m(t) \wedge S_{j,m}})] \le C e^{Ct}.
$$

Then, letting $j \to \infty$ and observing that $A_m(t) \le \tau_m = \lim_{j \to \infty} S_{j,m}$, we obtain for all $k \in \mathbb{N}$,

$$
\mathbb{E}[f_m^{-k}(X_{A_m(t)})] \le C e^{Ct},
$$

combining this with (5.5) yields (5.3). This completes the proof for Lemma 5.1. □

6. Proof of the main estimates. In this section we shall apply the cut-off procedures, using the quantitative localised vector fields, introduced in Section 5, to obtain short time as well as asymptotic first and second order gradient estimates for the logarithmic heat kernel of a complete Riemannian manifold without imposing on it any curvature bounds.

Let $\{D_m\}_{m=1}^{\infty}$ and $\{f_m\}_{m=1}^{\infty}$ be the sequences of domains and functions constructed in Section 5. Recall that, for every m, $D_m = \{x \in M : f_m(x) > 0\}$ is a bounded connected open set. By Lemma A.1 from the Appendix, there exists a compact Riemannian manifold $\tilde{M}_m$ such that D_m is isometrically embedded into $\tilde{M}_m$ as an open set. We could and will view $D_m \subset \tilde{M}_m$ as an open subset of $\tilde{M}_m$. In particular, we have

$$d_{\tilde{M}_m}(x, y) = d(x, y), \quad \forall x, y \in B_o(2m-2), \tag{6.1}$$

where d and $d_{\tilde{M}_m}$ are the Riemannian distance function on M and $\tilde{M}_m$. We denote the heat kernel on M and $\tilde{M}_m$ by $p(t, x, y)$ and $p_{\tilde{M}_m}(t, x, y)$, respectively. For every $e \in \mathbb{R}^n$, we also let H_e^m denote the horizontal lift of ue on $TO(\tilde{M}_m)$.

Let us fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $\{B_t\}_{t\geq 0}$ be the standard $\mathbb{R}^n$-valued Brownian motion with $B_t = (B_t^1, \dots, B_t^n)$, and we denote by $\mathscr{F}_t$ the filtration generated by it. Now we fix an orthonormal basis $\{e_i\}_{i=1}^n$ of $\mathbb{R}^n$.

For $x \in D_m \subset \tilde{M}_m$ and U_0 a frame at x so that $U_0 \in O_x M = O_x \tilde{M}_m$, let U_t^m denote the solution to the following $O(\tilde{M}_m)$-valued stochastic differential equation:

$$\mathrm{d}U_t^m = \sum_{i=1}^{n} H_{e_i}^m(U_t^m) \circ \mathrm{d}B_t^i, \qquad U_0^m = U_0. \tag{6.2}$$

Set $X_t^{m,x} := \pi(U_t^m)$. This is a $\tilde{M}_m$-valued Brownian motion. Recall that $X_t^x := \pi(U_t)$, where U_t is the solution to (3.1) with the same driving Brownian motion B_t and the same initial value U_0 as in (6.2).

Throughout this section, for every $m, k \in \mathbb{N}$ with $k \geq m$, we define

$$\tau_m := \inf\{t > 0; X_t^x \notin D_m\}, \qquad \tau_m^k := \inf\{t > 0; X_t^{k,x} \notin D_m\}.$$

Note that for every $k > m$, $H_{e_i}^k = H_{e_i}$ on $\pi^{-1}(D_m)$. It is easy to verify that

$$\tau_m = \tau_m^k, \quad X_t^x = X_t^{k,x}, \quad \forall k \geq m > 1, 0 \leq t \leq \tau_m. \tag{6.3}$$

As before, the superscript x may be omitted from time to time when there is no risk of confusion. *The probability and the expectation for the functional generated by $X_\cdot^x$ or $X_\cdot^{m,x}$ (with respect to $\mathbb{P}$) are denoted by $\mathbb{P}_x$ and $\mathbb{E}_x$, respectively, in this section.*

If M is compact, then when m is large enough, we have $D_m = M$, and we can take $\tilde{M}_m = M$ (we do not have to apply Lemma A.1 when M is compact), then all the conclusions in this section automatically hold. Hence, in this section we always assume that M is noncompact.

We shall use the following estimates which are crucial for our proof.

LEMMA 6.1 ([7, 60, 70, 71]). *For any $x, y \in M$,*

$$\lim_{t\downarrow 0} t \log p(t, x, y) = -\frac{d(x, y)^2}{2}, \tag{6.4}$$

and the convergence is uniformly in (x, y) on $K \times K$ for any compact subset K.

Moreover, for every connected bounded open set $D \supseteq K$ with smooth boundary,

$$\lim_{t\downarrow 0} t \log \mathbb{P}_x(\tau_D < t) = -\frac{d(x, \partial D)^2}{2}, \quad \forall x \in K. \tag{6.5}$$

Here, $\tau_D := \inf\{t > 0; X_t \notin D\}$ is the first exit time from D and $d(x, \partial D) := \inf_{z\in\partial D} d(x, z)$. And the convergence is also uniform in x on K.

The asymptotic estimates (6.4) and (6.5) were first shown to hold for $\mathbb{R}^n$ in Varadhan [70, 71], extension to a complete Riemannian manifold was given in Molchanov [60]. In addition, Azencott [8] and [45] indicated that these statements may fail for an incomplete Riemannian manifold. We shall also use the following statement which follows readily from the small time asymptotics and the Gaussian heat kernel upper bounds.

LEMMA 6.2 ([8], [45], Lemma 2.2). *For any compact subset K of M and any positive number r, then there exists a positive number t_0 such that*

$$\sup_{t\in(0,t_0]}\sup_{d(z,y)\geq r, y\in K} p(t,z,y)\leq 1. \tag{6.6}$$

6.1. *Comparison theorem for functional integrals involving approximate heat kernels.* Let D_m denote the relatively compact subset, and let $l_m : [0,1]\times P_x(M)\to\mathbb{R}$ be the cut-off processes adapted to D_m, as constructed by Lemma 5.1. Let $p^{D_m}(t,x,y)$ denote the Dirichlet heat kernel on D_m. Let K be a compact set and $x,y\in K$ be such that $d(x,y)<d(x,\partial D^m)\vee d(y,\partial D^m)$. Then $p(t,x,y)$ and $p^{D_m}(t,x,y)$ are asymptotically the same for small t; see [8], Lemma 2.3 on page 156.

Below we give a quantitative estimate on p and p^{D_m} on a compact set $K\times K$ for sufficiently large m. By sufficiently large we mean that $m\geq m_0$ for a natural number m_0 and m_0 may depend on other data. In all the results below, it depends on the compact set K and the prescribed exponential factor $L>0$.

LEMMA 6.3. *Suppose that K is a compact subset of M and $L>1$ is a positive number. Then, for sufficiently large m, there exists a positive number $t_0=t_0(K,L,m)$ such that, for every $t\in(0,t_0]$,*

$$\begin{aligned}\sup_{x,y\in K}\bigl|p(t,x,y)-p^{D_m}(t,x,y)\bigr|&\leq e^{-\frac{2L}{t}},\\ \sup_{x,y\in K}\bigl|p_{\tilde{M}_m}(t,x,y)-p^{D_m}(t,x,y)\bigr|&\leq e^{-\frac{2L}{t}}.\end{aligned}\tag{6.7}$$

In particular, for every $t\in(0,t_0]$,

$$\sup_{x,y\in K}\bigl|p(t,x,y)-p_{\tilde{M}_m}(t,x,y)\bigr|\leq e^{-\frac{L}{t}}.\tag{6.8}$$

PROOF. The estimates in (6.7) could be found in Azencott [8], Section 4.2, and also in Bismut [12], Section III.a, and Hsu [43], The proof of Theorem 5.1.1. Here, we include a proof for the convenience of the reader. The technique and the intermediate estimates will be used later.

By the strong Markovian property,

$$P_tf(x)=\mathbb{E}_x\bigl[f(X_t)\mathbf{1}_{\{t\leq\tau_m\}}\bigr]+\mathbb{E}_x\bigl[\mathbb{E}_{X_{\tau_m}}\bigl[f(X_{t-\tau_m})\bigr]\mathbf{1}_{\{\tau_m<t<\zeta\}}\bigr],$$

and so, for any $x,y\in K$ and $t>0$,

$$p(t,x,y)=p^{D_m}(t,x,y)+\mathbb{E}_x\bigl[p(t-\tau_m,X_{\tau_m},y)\mathbf{1}_{\{\tau_m<t<\zeta\}}\bigr].\tag{6.9}$$

Since M is noncompact, given any number $L>1$, there exists a natural number m_0 such that

$$K\subset B_o(2m_0-2),\qquad d(K,\partial D_{m_0})\geq d\bigl(K,\partial B_o(2m_0-2)\bigr)>4L.$$

Then, according to (6.5) and (6.6), for every $m \ge m_0$, we could find a positive number $t_0(K, L, m)$ such that, for any $t \in (0, t_0]$,

$$\mathbb{P}_x(t > \tau_m) \le \exp\left(-\frac{d(x, \partial D_m)^2 - 1}{2t}\right) \le e^{-\frac{2L}{t}}, \quad \forall x \in K,$$

$$p(t, z, y) \le 1, \quad \text{for all } z \in \partial D_m, \text{ and } y \in K.$$

By these estimates we obtain that, for all $m \ge m_0$ and all $t \in (0, t_0]$,

$$\mathbb{E}_x\left[p(t - \tau_m, X_{\tau_m}, y)\mathbf{1}_{\{t > \tau_{D_m}\}}\right] \le \sup_{t\in(0,t_0)} \sup_{z\in\partial D_m, y\in K} p(t, z, y) \cdot \mathbb{P}_x(t > \tau_m) \le e^{-\frac{2L}{t}}.$$

Putting this into (6.9), we arrive at that for all $m \ge m_0$, all $x, y \in K$ and for all $t \in (0, t_0]$,

$$\left|p(t, x, y) - p^{D_m}(t, x, y)\right| \le e^{-\frac{2L}{t}}. \tag{6.10}$$

Note that, for every $m \ge m_0$, $D_m \subset \tilde{M}_m$ and $x \in K$,

$$d_{\tilde{M}_m}(x, \partial D_m) \ge d_{\tilde{M}_m}\big(x, \partial B_o(2m - 2)\big) = d\big(x, \partial B_o(2m - 2)\big),$$

which is due to (6.1). By the same argument for (6.10) and changing the constant t_0 if necessary, we could find a $t_0(K, L, m)$ such that, for all $m \ge m_0$,

$$\left|p_{\tilde{M}_m}(t, x, y) - p^{D_m}(t, x, y)\right| \le e^{-\frac{2L}{t}}, \quad x, y \in K, t \in (0, t_0].$$

This, together with (6.10), yields (6.7) and (6.8). □

LEMMA 6.4. *Suppose that K is a compact subset of M and $L > 1$ is a positive number*:

(1) *For m_0 sufficiently large and any $m > m_0$, there exists a $t_0(K, L, m)$ such that, for every $0 < s \le \frac{t}{2}$ and $0 < t \le t_0$, we have*

$$\sup_{x,y\in K} \sup_{z\in D_{m_0}} \left|\frac{p(t - s, x, z)}{p(t, x, y)} - \frac{p_{\tilde{M}_m}(t - s, x, z)}{p_{\tilde{M}_m}(t, x, y)}\right| \le 2e^{-\frac{4L}{t}}. \tag{6.11}$$

(2) *Suppose Υ_t is an $\mathscr{F}_t$ adapted process and for any $q > 0$ and $m \ge 1$ we set*

$$F_m^q(\Upsilon, X.) = \left(\int_0^s \Upsilon_r l_m'(r, X.)\,\mathrm{d}B_r\right)^q, \qquad F_m^q(\Upsilon, X_\cdot^m) = \left(\int_0^s \Upsilon_r l_m'(r, X_\cdot^m)\,\mathrm{d}B_r\right)^q.$$

We also assume that

$$\sup_{x\in K} \mathbb{E}_x\left[\int_0^{1\wedge\tau_m} |\Upsilon_s|^{2q}\,\mathrm{d}s\right] < \infty, \quad \forall m \ge 1 \tag{6.12}$$

for some $q \in \mathbb{N}$. Then, for every sufficiently large m (any m greater than some number $m_0(K, L)$), we can find a positive number $t_0(K, L, m)$ with the property that

$$\begin{aligned}\sup_{x,y\in K}&\left|\mathbb{E}_x\left[F_m^q(\Upsilon, X.)\frac{p(t - s, X_s, y)}{p(t, x, y)}\right] - \mathbb{E}_x\left[F_m^q(\Upsilon, X_\cdot^m)\frac{p_{\tilde{M}_m}(t - s, X_s^m, y)}{p_{\tilde{M}_m}(t, x, y)}\right]\right| \\ &\le C(m)e^{-\frac{L}{t}}\end{aligned} \tag{6.13}$$

for any $0 < t \le t_0$, $0 < s \le \frac{t}{2}$. Here, the positive constant $C(m)$ may depend on m and on $\alpha_m := \sup_{x\in K} \mathbb{E}_x[\int_0^1 |\Upsilon_r l_m'(r, X.)|^q\,\mathrm{d}r]$. (Note that $l_m'(r, X.) \ne 0$ only for $r < \tau_m = \tau_m^m$ so the quantity is well defined.)

PROOF. In the proof the constant C may represent different constants in different lines. Let $r_0 := \sup_{x,y\in K} d_M(x,y)$ denote the diameter of K. Since M is noncompact, we can choose a natural number $\tilde{m}_0$ (which may depend on K and L) such that

$$K \subset B_o(2\tilde{m}_0 - 2) \subset D_{\tilde{m}_0}$$

and for all $m > \tilde{m}_0$,

$$d\big(K, \partial B_o(2\tilde{m}_0 - 2)\big) = d_{\tilde{M}_m}\big(K, \partial B_o(2\tilde{m}_0 - 2)\big) > 4(L + r_0 + 1).$$

Also, by the heat kernel comparison (6.7) and (6.8), we can find a $m_0 > \tilde{m}_0$ so that for all $m > m_0$, there exists a constant $t_2(K, L, m) > 0$ such that

$$\big|p(t,z,y) - p_{\tilde{M}_m}(t,z,y)\big| \le e^{-\frac{4(L+r_0+1)^2}{t}}, \quad \forall t \in (0, t_2], z, y \in D_{m_0}. \tag{6.14}$$

According to the asymptotic relations (6.4) and (6.5), for every $m > m_0$ (taking m_0 larger as is necessary) we could find a constant $0 < t_1(K, L, m) \le t_2$ such that, for all $t \in (0, t_1]$,

$$p(t,z,y) \le e^{\frac{1}{t}}, \qquad p_{\tilde{M}_m}(t,z,y) \le e^{\frac{1}{t}}, \quad \forall z, y \in D_{m_0}, \tag{6.15}$$

$$p(t,z,y) \ge e^{-\frac{r_0^2+1}{t}}, \qquad p_{\tilde{M}_m}(t,z,y) \ge e^{-\frac{r_0^2+1}{t}}, \quad \forall z, y \in K, \tag{6.16}$$

$$\mathbb{P}_z(\tau_{m_0} < t) \le e^{-\frac{4(L+r_0+1)^2}{t}}, \quad \forall z \in K. \tag{6.17}$$

By the small time locally uniform heat kernel bound (6.6), for every $m > m_0$ there exists a number $0 < t_0(K, L, m) \le t_1$ such that, for all $t \in (0, t_0]$,

$$p(t, z_1, y) \vee p_{\tilde{M}_m}(t, z_2, y) \le 1, \quad \forall z_1 \in M \cap D_{m_0}^c, z_2 \in \tilde{M}_m \cap D_{m_0}^c, y \in K. \tag{6.18}$$

Therefore, for every $m > m_0$ and for every $0 < s \le \frac{t}{2}$, every $0 < t \le t_0$ and for all $x, y \in K$ and $z \in D_{m_0}$, we have

$$\begin{aligned} &\left|\frac{p(t-s,x,z)}{p(t,x,y)} - \frac{p_{\tilde{M}_m}(t-s,x,z)}{p_{\tilde{M}_m}(t,x,y)}\right| \\ &\quad \le \frac{p(t-s,x,z)|p_{\tilde{M}_m}(t,x,y) - p(t,x,y)| + p(t,x,y)|p_{\tilde{M}_m}(t-s,x,z) - p(t-s,x,z)|}{p(t,x,y)p_{\tilde{M}_m}(t,x,y)} \\ &\quad \le 2e^{\frac{2(1+r_0^2)}{t}} e^{\frac{2}{t}} e^{-\frac{4(L+r_0+1)^2}{t}} \le 2e^{-\frac{4L}{t}}. \end{aligned} \tag{6.19}$$

Here, the second step above is due to (6.14)–(6.16). Thus, we finish the proof of (1).

For all $m > m_0$, let us split the terms as follows:

$$\begin{aligned} &\mathbb{E}_x\left[F_m^q(\Upsilon, X_\cdot)\frac{p(t-s,X_s,y)}{p(t,x,y)}\right] \\ &\quad = \mathbb{E}_x\left[F_m^q(\Upsilon, X_\cdot)\frac{p(t-s,X_s,y)}{p(t,x,y)}\mathbf{1}_{\{t\le\tau_{m_0}\}}\right] + \mathbb{E}_x\left[F_m^q(\Upsilon, X_\cdot)\frac{p(t-s,X_s,y)}{p(t,x,y)}\mathbf{1}_{\{t>\tau_{m_0}\}}\right] \\ &\quad =: I_1^m(s,t) + I_2^m(s,t). \end{aligned}$$

Since $l'_m(r, X_\cdot) \ne 0$ if only if $t < \tau_m$, then $l'_m(r, X^m_\cdot) = l'_m(r, X_\cdot)$, and we have

$$\begin{aligned} F_m^q(\Upsilon, X^m_\cdot) &= \left(\int_0^s \Upsilon_r l'_m(r, X^m_\cdot)\,\mathrm{d}B_r\right)^q = \left(\int_0^{s\wedge\tau_m} \Upsilon_r l'_m(r, X^m_\cdot)\,\mathrm{d}B_r\right)^q \\ &= \left(\int_0^s \Upsilon_r l'_m(r, X_\cdot)\,\mathrm{d}B_r\right)^q = F_m^q(\Upsilon, X_\cdot). \end{aligned}$$

Note also, $X_s^m = X_s$ for every $s \le \frac{t}{2} < \tau_m$. It holds that

$$\begin{aligned}
&\mathbb{E}_x\bigg[F_m^q(\Upsilon, X_\cdot^m)\frac{p_{\tilde M_m}(t-s, X_s^m, y)}{p_{\tilde M_m}(t,x,y)}\bigg]\\
&\quad= \mathbb{E}_x\bigg[F_m^q(\Upsilon, X_\cdot)\frac{p_{\tilde M_m}(t-s, X_s, y)}{p_{\tilde M_m}(t,x,y)}\mathbf{1}_{\{t\le\tau_{m_0}\}}\bigg]\\
&\qquad+ \mathbb{E}_x\bigg[F_m^q(\Upsilon, X_\cdot)\frac{p_{\tilde M_m}(t-s, X_s^m, y)}{p_{\tilde M_m}(t,x,y)}\mathbf{1}_{\{t>\tau_{m_0}\}}\bigg]\\
&\quad=: J_1^m(s,t) + J_2^m(s,t), \quad 0 < s < \frac{t}{2}.
\end{aligned}$$

Note that

$$\alpha_m = \sup_{x\in K}\mathbb{E}_x\bigg[\int_0^1 \big|\Upsilon_r l_m'(r, X_\cdot)\big|^q \,\mathrm{d}r\bigg] < \infty. \tag{6.20}$$

This follows from the moment estimates on l_m, (5.3), the assumption (6.12), and also

$$\alpha_m \le \sup_{x\in K}\mathbb{E}_x\bigg[\int_0^{1\wedge\tau_m} |\Upsilon_r|^{2q}\,\mathrm{d}r\bigg]^{1/2} \sup_{x\in K}\mathbb{E}_x\bigg[\int_0^{1\wedge\tau_m} \big|l_m'(r, X_\cdot)\big|^{2q}\,\mathrm{d}r\bigg]^{1/2}.$$

For all $m > m_0$, $x, y \in K$, $0 < s \le \frac{t}{2}$, and $0 < t \le t_0$, we may assume that $t_0 \le 2$,

$$\begin{aligned}
&\big|I_1^m(s,t) - J_1^m(s,t)\big|\\
&\quad\le \sup_{z\in D_{m_0}}\bigg|\frac{p(t-s,z,y)}{p(t,x,y)} - \frac{p_{\tilde M_m}(t-s,z,y)}{p_{\tilde M_m}(t,x,y)}\bigg|\mathbb{E}_x\bigg[\bigg|\int_0^s \Upsilon_r l_m'(r, X_\cdot)\,\mathrm{d}B_r\bigg|^q\bigg]\\
&\quad\le Ce^{-\frac{4L}{t}}\sup_{x\in K}\mathbb{E}_x\bigg[\int_0^{1\wedge\tau_m}\big|\Upsilon_r l_m'(r, X_\cdot)\big|^q\,\mathrm{d}r\bigg] = C\alpha_m e^{-\frac{4L}{t}}.
\end{aligned}$$

In the penultimate step we have applied Burkholder–Davies–Gundy inequality and (6.11). According to (6.15) and (6.18), we also have

$$\sup_{z\in M, y\in K} p(t,z,y) \le e^{\frac{1}{t}}, \quad \forall 0 < t \le t_0.$$

Combining this with (6.16)–(6.17), Cauchy–Schwarz inequality and Burkholder–Davies–Gundy inequality, we obtain that, for every $m > m_0$, $x, y \in K$, $0 < s \le \frac{t}{2}$ and $0 < t \le t_0$,

$$\begin{aligned}
&\big|I_2^m(s,t)\big|\\
&\quad\le Ce^{\frac{r_0^2+1}{t}}\sup_{r\in[\frac{t}{2},t], z\in M, y\in K} p(r,z,y)\mathbb{E}_x\bigg[\bigg|\int_0^s \Upsilon_r l_m'(r, X_\cdot)\,\mathrm{d}B_r\bigg|^{2q}\bigg]^{1/2}\mathbb{P}_x(\tau_{m_0} < t)^{1/2}\\
&\quad\le Ce^{\frac{r_0^2+1}{t}}e^{\frac{2}{t}}e^{-\frac{2(r_0+L+1)^2}{t}}\mathbb{E}_x\bigg[\int_0^{1\wedge\tau_m}\big|\Upsilon_r l_m'(r, X_\cdot)\big|^{2q}\,\mathrm{d}r\bigg]^{1/2} \le C\alpha_m e^{-\frac{2L}{t}}.
\end{aligned}$$

Here, in the last step we used (6.20). Similarly, we obtain that, for every $m > m_0$, $x, y \in K$,

$$\big|J_2^m(s,t)\big| \le C\alpha_m e^{-\frac{2L}{t}}, \quad 0 < s \le \frac{t}{2}, 0 < t \le t_0.$$

Combing the above estimates for I_1^m, I_2^m, J_1^m, J_2^m, we see that, for every $m > m_0$, $x, y \in K$, $0 < s \le \frac{t}{2}$ and $0 < t \le t_0$,

$$\left|\mathbb{E}_x\left[F_m^q(\Upsilon, X_\cdot)\frac{p(t-s, X_s, y)}{p(t,x,y)}\right] - \mathbb{E}_x\left[F_m^q(\Upsilon, X_\cdot^m)\frac{p_{\tilde{M}_m}(t-s, X_s^m, y)}{p_{\tilde{M}_m}(t,x,y)}\right]\right|$$
$$\le \left|I_1^m(s,t) - J_1^m(s,t)\right| + \left|I_2^m(s,t)\right| + \left|J_2^m(s,t)\right| \le C\alpha_m e^{-\frac{L}{t}},$$

which is (6.13), and we have finished the proof. □

REMARK 6.1. By the same arguments in the proof for (6.13), we could obtain the following under the conditions of Lemma 6.4. For sufficiently large m we could find a positive number $t_0(K, L, m)$ so that, for every $x, y \in K$, $0 < s \le \frac{t}{2}$ and $0 < t \le t_0$, the following estimates hold, replacing l_m' by l_m or $\mathrm{d}B_r$ by $\mathrm{d}r$:

$$\left|\mathbb{E}_x\left[\left(\int_0^s \Upsilon_r l_m(r, X_\cdot)\,\mathrm{d}B_r\right)^q\left(\frac{p(t-s, X_s, y)}{p(t,x,y)} - \frac{p_{\tilde{M}_m}(t-s, X_s^m, y)}{p_{\tilde{M}_m}(t,x,y)}\right)\right]\right| \le C(m)e^{-\frac{L}{t}}, \tag{6.21}$$

$$\left|\mathbb{E}_x\left[\left(\int_0^s \Upsilon_r l_m'(r, X_\cdot)\,\mathrm{d}r\right)^q\left(\frac{p(t-s, X_s, y)}{p(t,x,y)} - \frac{p_{\tilde{M}_m}(t-s, X_s^m, y)}{p_{\tilde{M}_m}(t,x,y)}\right)\right]\right| \le C(m)e^{-\frac{L}{t}}, \tag{6.22}$$

$$\left|\mathbb{E}_x\left[\left(\int_0^s \Upsilon_r l_m(r, X_\cdot)\,\mathrm{d}r\right)^q\left(\frac{p(t-s, X_s, y)}{p(t,x,y)} - \frac{p_{\tilde{M}_m}(t-s, X_s^m, y)}{p_{\tilde{M}_m}(t,x,y)}\right)\right]\right| \le C(m)e^{-\frac{L}{t}}. \tag{6.23}$$

These estimates will also be used in the proof for Proposition 6.6.

6.2. *Proof of the main theorem*: *Gradient estimates*.

LEMMA 6.5. *Let $t > 0$, $x \in M$ and $v \in T_xM$. Suppose that m is a natural number such that $x \in D_m$. Let $h \in L^{1,2}(\Omega; \mathbb{R}^n)$ be given by*

$$h(s) = \left(\frac{t-2s}{t}\right)^+ l_m(s, X_\cdot)U_0^{-1}v.$$

Then, for any $f \in C_b(M)$ we have

$$\langle \nabla P_t f(x), v\rangle_{T_xM} = -\mathbb{E}_x\left[f(X_t)\int_0^t \langle \Theta_s^h, \mathrm{d}B_s\rangle \mathbf{1}_{\{t<\zeta\}}\right], \tag{6.24}$$

where Θ_s^h defined by (3.2) with the h chosen above.

PROOF. When M is compact, (6.24) is just (4.11) established in Proposition 4.4. For general noncompact complete M, we will use the arguments based on truncation and approximation. For each $k > m$, let $\{U_t^k\}_{t\ge 0}$ be the horizontal Brownian motion on compact manifold $\tilde{M}_k$, as defined in (6.2) with $\pi(U_0^k) = x \in D_m$. Set $X_t^k = \pi(U_t^k)$ and $P_t^k f(x) = \mathbb{E}_x[f(X_t^k)]$. Let

$$h(s) = \left(\frac{t-2s}{t}\right)^+ l_m(s, X_\cdot)U_0^{-1}v. \tag{6.25}$$

According to (6.3), precisely $\tau_m = \tau_m^k$ and $h(s) \ne 0$ if and only if when $s \le \frac{t}{2} \wedge \tau_m$. So furthermore,

$$h(s) = \left(\frac{t-2s}{t}\right)^+ l_m(s, X_\cdot^k)\mathbf{1}_{\{s<\frac{t}{2}\wedge\tau_m\}}U_0^{-1}v, \quad \forall k > m,$$

$$h'(s) = \left(-\frac{2}{t}l_m(s, X_\cdot^k) + l_m'(s, X_\cdot^k)\left(\frac{t-2s}{t}\right)^+\right)\mathbf{1}_{\{s<\frac{t}{2}\wedge\tau_m\}}U_0^{-1}v, \quad \forall k > m,$$

which means that we can replace $X_\cdot$ by $X_\cdot^k$ in the expression of $h(s)$ and $h'(s)$. Let $\Theta_s^{h,k}$ be given by (3.2) with the manifold M replaced by $\tilde{M}_k$ (associated with $X_\cdot^k$). Therefore, by (3.2) we have the following expression:

$$\Theta_s^{h,k} = h'(s) + \mathrm{ric}_{U_s^k}^{\tilde{M}_k}\big(h(s)\big) = h'(s) + \mathrm{ric}_{U_s}\big(h(s)\big) = \Theta_s^h, \quad \forall k > m. \tag{6.26}$$

Here, both sides of (6.26) vanish for $s > \tau_m$. Meanwhile, we have used the fact that $U_s = U_s^k$ when $s < \tau_m$ and $\mathrm{Ric}_z^{\tilde{M}_k} = \mathrm{Ric}_z$ for every $z \in D_m$ and

$$\mathrm{ric}_{U_s^k}^{\tilde{M}_k}\big(h(s)\big) = \mathrm{ric}_{U_s^k}^{\tilde{M}_k}\big(h(s)\big)\mathbf{1}_{\{s<\tau_m\}} = \mathrm{ric}_{U_s}\big(h(s)\big), \quad \forall k > m.$$

Moreover, we observe that, for any compact set $K \subset D_m$ and $q > 0$,

$$\begin{aligned} \sup_{x\in K} \mathbb{E}_x\bigg[\int_0^{1\wedge\tau_m} \big|\mathrm{ric}_{U_s}\big(U_0^{-1}v\big)\big|^q \,\mathrm{d}s\bigg] &\le |v|^q \sup_{x\in K} \mathbb{E}_x\bigg[\int_0^{1\wedge\tau_m} |\,\mathrm{ric}_{U_s}\mathbf{1}_{\{s<\tau_m\}}|^q \,\mathrm{d}s\bigg] \\ &\le |v|^q \sup_{z\in D_m} \|\mathrm{Ric}_z\|^q < \infty. \end{aligned} \tag{6.27}$$

Combining this with (5.3) and the fact that $h(s) \neq 0$ only if $s \le t \wedge \tau_m = t \wedge \tau_m^k$ yields immediately that

$$\sup_{x\in D_m, v\in T_xM, |v|=1} \mathbb{E}_x\bigg[\int_0^t |\Theta_s^h|^2 \,\mathrm{d}s\bigg] < \infty, \quad \forall t > 0. \tag{6.28}$$

Thus, applying Proposition 4.4 to $P_t^k f$ (note that $\tilde{M}_k$ is compact) and using (6.26), we obtain that, for all $v \in T_xM$,

$$\langle \nabla P_t^k f(x), v\rangle_{T_xM} = -\mathbb{E}_x\bigg[f\big(X_t^k\big)\int_0^t \langle\Theta_s^{h,k}, \mathrm{d}B_s\rangle\bigg] = -\mathbb{E}_x\bigg[f\big(X_t^k\big)\int_0^t \langle\Theta_s^h, \mathrm{d}B_s\rangle\bigg]. \tag{6.29}$$

For any function $\psi \in C_c^\infty(M)$ and vector field $V \in C_c^\infty(M; TM)$ with supports in D_m satisfying that $|V(x)| \le 1$ for all $x \in D_m$, we can use ∇ and $\mathrm{d}x$ for the the gradient operator and the Riemannian volume measure on both manifolds M and $\tilde{M}_k$, so we have

$$\begin{aligned} &\int_M \mathbb{E}_x\bigg[f\big(X_t^k\big)\int_0^t \langle\Theta_s^{h(x)}, \mathrm{d}B_s\rangle\bigg]\psi(x)\,\mathrm{d}x \\ &= \int_M \langle\nabla P_t^k f(x), V(x)\rangle_{T_xM}\psi(x)\,\mathrm{d}x \\ &= \int_{\tilde{M}_k} \langle\nabla P_t^k f(x), V(x)\rangle_{T_xM}\psi(x)\,\mathrm{d}x \\ &= -\int_{\tilde{M}_k} \mathbb{E}_x\big[f\big(X_t^k\big)\big]\,\mathrm{div}(V\psi)(x)\,\mathrm{d}x = -\int_M \mathbb{E}_x\big[f\big(X_t^k\big)\big]\,\mathrm{div}(V\psi)(x)\,\mathrm{d}x. \end{aligned} \tag{6.30}$$

Here, $h(x)$ is defined by (6.25) with $v = V(x)$.

Meanwhile, note that $X_t = X_t^k$ if $t < \tau_k$, for every $x \in D_m$ it holds

$$\begin{aligned} &\lim_{k\to\infty}\bigg|\mathbb{E}_x\bigg[f\big(X_t^k\big)\int_0^t \langle\Theta_s^{h(x)}, \mathrm{d}B_s\rangle\bigg] - \mathbb{E}_x\bigg[f(X_t)\int_0^t \langle\Theta_s^{h(x)}, \mathrm{d}B_s\rangle\mathbf{1}_{\{t<\zeta\}}\bigg]\bigg| \\ &\le \lim_{k\to\infty} \mathbb{E}_x\bigg[\big|f\big(X_t^k\big) - f(X_t)\mathbf{1}_{\{t<\zeta\}}\big|\bigg|\int_0^t \langle\Theta_s^{h(x)}, \mathrm{d}B_s\rangle\bigg|\bigg] \\ &\le \lim_{k\to\infty} \sqrt{\mathbb{E}_x\big[\big|f\big(X_t^k\big) - f(X_t)\mathbf{1}_{\{t<\zeta\}}\big|^2\big]}\sqrt{\mathbb{E}_x\bigg[\bigg|\int_0^t \langle\Theta_s^{h(x)}, \mathrm{d}B_s\rangle\bigg|^2\bigg]} \\ &\le \lim_{k\to\infty} \sqrt{2}C\|f\|_\infty\sqrt{\mathbb{P}_x(\tau_k \le t < \zeta)} = 0, \end{aligned}$$

where

$$C := \sup_{x\in D_m, v\in T_xM, |v|=1} \mathbb{E}_x\left[\int_0^t |\Theta_s^h|^2\,\mathrm{d}s\right]$$

is finite for every $t>0$ which is due to (6.28). With this we may take $k\to\infty$ in (6.30), then

$$\lim_{k\to\infty}\mathbb{E}_x\left[f(X_t^k)\int_0^t\langle\Theta_s^{h(x)},\mathrm{d}B_s\rangle\right]=\mathbb{E}_x\left[f(X_t)\int_0^t\langle\Theta_s^{h(x)},\mathrm{d}B_s\rangle\mathbf{1}_{\{t<\zeta\}}\right]$$

and, consequently,

$$\int_{D_m}\mathbb{E}_x\left[f(X_t)\int_0^t\langle\Theta_s^{h(x)},\mathrm{d}B_s\rangle\mathbf{1}_{\{t<\zeta\}}\right]\psi(x)\,\mathrm{d}x=\int_{D_m}\mathbb{E}_x\left[f(X_t)\mathbf{1}_{\{t<\zeta\}}\right]\operatorname{div}(V\psi)(x)\,\mathrm{d}x.$$

Since m is arbitrary, so it follows that for all test vector fields $V\in C_c^\infty(M;TM)$ and test functions $\psi\in C_c^\infty(M)$,

$$\int_M\mathbb{E}_x\left[f(X_t)\int_0^t\langle\Theta_s^{h(x)},\mathrm{d}B_s\rangle\mathbf{1}_{\{t<\zeta\}}\right]\psi(x)\,\mathrm{d}x=\int_M P_tf(x)\operatorname{div}(V\psi)(x)\,\mathrm{d}x,$$

which means that the weak (distributional) gradient $\nabla P_t f$ exists

$$\langle\nabla P_tf(x),V(x)\rangle_{T_xM}=\mathbb{E}_x\left[f(X_t)\int_0^t\langle\Theta_s^{h(x)},\mathrm{d}B_s\rangle\mathbf{1}_{\{t<\zeta\}}\right],\quad x\in M.$$

According to the same arguments in the proof of Lemma A.2 in the Appendix, the functional $x\mapsto\mathbb{E}_x\left[f(X_t)\ \int_0^t\langle\Theta_s^{h(x)},\mathrm{d}B_s\rangle\mathbf{1}_{\{t<\zeta\}}\right]$ is continuous. So we have verified that the distributional derivative $\nabla P_t f$ exists and is continuous, then $\nabla P_t f$ is the classical gradient and expression (6.24) holds. □

Now we present an estimate for the difference between the gradients of logarithmic heat kernels. Note that, for every $x,y\in K\subset B_o(2m-2)\subset D_m$, we could view $\nabla_x\log p(t,x,y)$ and $\nabla_x\log p_{\tilde M_m}(t,x,y)$ as vectors in T_xM so that

$$\left|\nabla_x\log p(t,x,y)-\nabla_x\log p_{\tilde M_m}(t,x,y)\right|_{T_xM}$$

is well defined.

Let ∂ be the cemetery point. We make the convention that $p(t,\partial,y)=0$ for all t.

PROPOSITION 6.6. *Suppose that K is a compact subset of M and $L>1$ is a positive number. Then, for every suficiently large m, we could find a number $t_0(K,L,m)$, depending on K, L, m, such that, for every $0<t\le t_0$,*

$$\sup_{x,y\in K}\left|\nabla_x\log p(t,x,y)-\nabla_x\log p_{\tilde M_m}(t,x,y)\right|_{T_xM}\le C(m)e^{-\frac{L}{t}},\tag{6.31}$$

where $C(m)$ is a positive constant which may depend on m.

PROOF. Let us fix points $x,y\in K$ and a unit vector $v\in T_xM$. Let m be a natural number such that $B_o(2m-2)\supset K$. Let $t>0$ be fixed.

By (6.24), where $\Theta^h=h'(t)+\frac12\operatorname{ric}_{U_t}(h(t))$, we have, for every $f\in C_c(M)$,

$$\begin{aligned}\langle\nabla P_tf(x),v\rangle_{T_xM}&=\frac{2}{t}\mathbb{E}_x\left[\int_0^{\frac t2}\langle l_m(s)U_0^{-1}v,\mathrm{d}B_s\rangle f(X_t)\mathbf{1}_{\{t<\zeta\}}\right]\\&\quad-\mathbb{E}_x\left[\int_0^{\frac t2}\left\langle\left(\frac{t-2s}{t}\right)l_m'(s)U_0^{-1}v,\mathrm{d}B_s\right\rangle f(X_t)\mathbf{1}_{\{t<\zeta\}}\right]\\&\quad-\frac12\mathbb{E}_x\left[\int_0^{\frac t2}\left\langle\operatorname{ric}_{U_t}\left(\left(\frac{t-2s}{t}\right)l_m(s)U_0^{-1}v\right),\mathrm{d}B_s\right\rangle f(X_t)\mathbf{1}_{\{t<\zeta\}}\right].\end{aligned}$$

Since f has compact support, the indicator function $\mathbf{1}_{\{t<\zeta\}}$ can be removed. Taking the conditional expectation on $\sigma(X_t)$, we obtain, for all $x \in M$ and almost everywhere $y \in M$ (with respect to volume measure on M),

$$
\begin{aligned}
&t\langle \nabla_x \log p(t,x,y), v\rangle_{T_xM} \\
&\quad = t\frac{\langle \nabla_x p(t,x,y), v\rangle_{T_xM}}{p(t,x,y)} = 2\mathbb{E}_x\left[\mathbf{1}_{\{t<\zeta\}} \int_0^{\frac{t}{2}} \langle l_m(s) U_0^{-1} v, \mathrm{d}B_s\rangle \Big| X_t = y\right] \\
&\qquad - t\mathbb{E}_x\left[\mathbf{1}_{\{t<\zeta\}} \int_0^{\frac{t}{2}} \left\langle \left(\frac{t-2s}{t}\right) l_m'(s) U_0^{-1} v, \mathrm{d}B_s\right\rangle \Big| X_t = y\right] \\
&\qquad - \frac{t}{2}\mathbb{E}_x\left[\mathbf{1}_{\{t<\zeta\}} \int_0^{\frac{t}{2}} \left\langle \mathrm{ric}_{U_t}\left(\frac{t-2s}{t}\right) l_m(s) U_0^{-1} v, \mathrm{d}B_s\right\rangle \Big| X_t = y\right] \\
&\quad = \mathbb{E}_x\left[\mathbf{1}_{\{t<\zeta\}} \int_0^{\frac{t}{2}} g_m(s)\langle U_0^{-1} v, \mathrm{d}B_s\rangle \Big| X_t = y\right],
\end{aligned}
\tag{6.32}
$$

where

$$
g_m(s) := 2l_m(s) - t\left(\frac{t-2s}{t}\right) l_m'(s) - \frac{t}{2}\mathrm{ric}_{U_t}\left(\frac{t-2s}{t}\right) l_m(s).
$$

Thus,

$$
\begin{aligned}
t\langle \nabla_x \log p(t,x,y), v\rangle_{T_xM} &= \mathbb{E}_x\left[\mathbf{1}_{\{t<\zeta\}} \int_0^{\frac{t}{2}} g_m(s)\langle U_0^{-1} v, \mathrm{d}B_s\rangle \Big| X_t = y\right] \\
&= \mathbb{E}_x\left[\int_0^{\frac{t}{2}} g_m(s)\langle U_0^{-1} v, \mathrm{d}B_s\rangle \frac{p(\frac{t}{2}, X_{\frac{t}{2}}, y)}{p(t,x,y)} \mathbf{1}_{\{\frac{t}{2}<\zeta\}}\right] \\
&= \mathbb{E}_x\left[\int_0^{\frac{t}{2}} g_m(s)\langle U_0^{-1} v, \mathrm{d}B_s\rangle \frac{p(\frac{t}{2}, X_{\frac{t}{2}}, y)}{p(t,x,y)}\right].
\end{aligned}
\tag{6.33}
$$

We have used the property that for $p(\frac{t}{2}, X_{\frac{t}{2}}, y) = 0$ whenever $\frac{t}{2} \geq \zeta(x)$.

Based on the heat kernel estimates in the previous lemmas, by the proof of Lemma A.2 we know immediately

$$
x \mapsto \mathbb{E}_x\left[\int_0^{\frac{t}{2}} g_m(s)\langle U_0^{-1} v, \mathrm{d}B_s\rangle \frac{p(\frac{t}{2}, X_{\frac{t}{2}}, y)}{p(t,x,y)}\right]
$$

is continuous. So the expression above is true for all $x, y \in M$.

Since $l_m'(s, X_\cdot^m) = l_m'(s, X_\cdot)$ and $l_m(s, X_\cdot^m) = l_m(s, X_\cdot)$ for $s < \tau_m$ and $X_s = X_s^m$, applying the same arguments above to $\tilde{M}_m$, we have

$$
\langle t\nabla \log p_{\tilde{M}_m}(t,x,y), v\rangle_{T_xM} = \mathbb{E}_x\left[\int_0^{\frac{t}{2}} g_m(s)\langle U_0^{-1} v, \mathrm{d}B_s\rangle \frac{p_{\tilde{M}_m}(\frac{t}{2}, X_{\frac{t}{2}}^m, y)}{p_{\tilde{M}_m}(t,x,y)}\right].
\tag{6.34}
$$

To apply Lemma 6.4, it remains to make moment estimates for $\int_0^t g(s)\langle v, U_0\,\mathrm{d}B_s\rangle$. For any $m \in \mathbb{N}$ large enough and $q > 0$, (6.27) implies that condition (6.12) in Lemma 6.4 holds for the process $\Upsilon_t = \mathrm{ric}_{U_t}$ and we could apply (6.21) and (6.13) to conclude the estimates. □

We are now in a position to proceed to prove the gradient estimates for $\log p(t,x,y)$.

THEOREM 6.7. *The following statements hold*:

(1) *Suppose* $x, y \in M$ *and* $x \notin \mathrm{Cut}_M(y)$, *then*

$$\lim_{t\downarrow 0} t\nabla_x \log p(t,x,y) = -\nabla_x\left(\frac{d^2(x,y)}{2}\right). \tag{6.35}$$

Here, the convergence is uniformly in x *on any compact subset of* $M \setminus \mathrm{Cut}_M(y)$.

(2) *Let* K *be a compact subset of* M. *Then there exists a positive constant* $C(K)$, *which may depend on* K, *such that*

$$\left|\nabla_x \log p(t,x,y)\right|_{T_xM} \le C(K)\left(\frac{d(x,y)}{t} + \frac{1}{\sqrt{t}}\right), \quad x, y \in K, t \in (0,1]. \tag{6.36}$$

PROOF. In the proof the constant C (which depends on $\tilde{K}$ or K) may change from line to line. For every $m \in \mathbb{N}$ with $K \subset B_o(2m-2) \subset D_m$, we have

$$\begin{aligned} &t\nabla_x \log p(t,x,y) \\ &\quad = t\nabla_x \log p_{\tilde{M}_m}(t,x,y) + \big(t\nabla_x \log p(t,x,y) - t\nabla_x \log p_{\tilde{M}_m}(t,x,y)\big). \end{aligned} \tag{6.37}$$

For each compact set $\tilde{K} \subset M \setminus \mathrm{Cut}_M(y)$, by (6.31) we could choose a $m_0 \in \mathbb{N}$ large enough such that $\tilde{K} \subset B_o(2m_0 - 2) \subset D_{m_0}$ and

$$\lim_{t\downarrow 0} \sup_{x\in\tilde{K}} \left|t\nabla_x \log p(t,x,y) - t\nabla_x \log p_{\tilde{M}_{m_0}}(t,x,y)\right|_{T_xM} = 0.$$

At the same time, since $\tilde{M}_{m_0}$ is compact and $\tilde{K}$ is outside of the cut locus $\mathrm{Cut}_{\tilde{M}_m}(y)$, we have

$$\begin{aligned} &\lim_{t\downarrow 0} \sup_{x\in\tilde{K}} \left|t\nabla_x \log p_{\tilde{M}_{m_0}}(t,x,y) + \nabla_x\left(\frac{d^2(x,y)}{2}\right)\right|_{T_xM} \\ &\quad = \lim_{t\downarrow 0} \sup_{x\in\tilde{K}} \left|t\nabla_x \log p_{\tilde{M}_{m_0}}(t,x,y) + \nabla_x\left(\frac{d^2_{\tilde{M}_{m_0}}(x,y)}{2}\right)\right|_{T_x\tilde{M}_{m_0}} = 0. \end{aligned}$$

In the first step, we used that $d_{\tilde{M}_{m_0}}(x,y) = d(x,y)$ for $x, y \in \tilde{K}$, while the second step is due to Corollary 2.29 from Malliavin and Stroock [59] (see also Bismut [12] and Norris [62]). Plugging this into (6.37) with $m = m_0$, then we have shown (6.35).

Given a compact set $K \subset M$ and a constant $L > 1$, based on (6.31), there exists a sufficiently large natural number m_0 such that $K \subset B_o(2m_0 - 2) \subset D_{m_0}$ and $t_0 \in (0,1)$ such that

$$\sup_{x,y\in K} \left|\nabla_x \log p(t,x,y) - \nabla_x \log p_{\tilde{M}_{m_0}}(t,x,y)\right|_{T_xM} \le Ce^{-\frac{L}{t}}, \quad \forall t \in (0, t_0]. \tag{6.38}$$

Since $\tilde{M}_{m_0}$ is compact, we can apply Hsu [43], Theorem 5.5.3, or Sheu [64] to show that, for all $x, y \in K$ and $t \in (0,1]$,

$$\begin{aligned} \left|\nabla_x \log p_{\tilde{M}_{m_0}}(t,x,y)\right|_{T_xM} &\le C(K)\left(\frac{d_{\tilde{M}_{m_0}}(x,y)}{t} + \frac{1}{\sqrt{t}}\right) \\ &= C(K)\left(\frac{d(x,y)}{t} + \frac{1}{\sqrt{t}}\right). \end{aligned} \tag{6.39}$$

Combining (6.38) and (6.39) into (6.37) with $m = m_0$, we immediately find (6.36) holds for all $t \in (0, t_0]$.

Also note that, for all $x, y \in K$ and for all $t \in [t_0, 1]$,

$$\left|\nabla_x \log p(t,x,y)\right|_{T_xM} \le C(K, t_0) \le C\left(\frac{d(x,y)}{t} + \frac{1}{\sqrt{t}}\right).$$

By now we have completed the proof of (6.36). $\square$

REMARK.

(1) The gradient estimate (6.36) was proved in [43, 64, 66] for a complete manifold with Ricci curvature bounded from below by a constant C_0. In that case the constant $C(K)$ in (6.36) in uniform and only depends on C_0, see also [49] for the case of the estimates for heat kernel associated with the Witten Laplacian operator.

(2) By carefully tracking the proof, we know the constant $C(K)$ from (6.36) depends only on $C_1(m_0)$, $\inf_{x\in D_{m_0}} \|\mathrm{Ric}_x\|$ and $\sup_{x\in D_{m_0}} \mathbb{E}_x \int_0^1 |l'_{m_0}(s)|^2\,\mathrm{d}s$, where $C_1(m_0)$ is the positive constant such that

$$\begin{aligned}\big|\nabla_x \log p_{\tilde{M}_{m_0}}(t,x,y)\big|_{T_xM} &\le C_1(m_0)\bigg(\frac{d_{\tilde{M}_{m_0}}(x,y)}{t}+\frac{1}{\sqrt{t}}\bigg)\\ &= C_1(m_0)\bigg(\frac{d(x,y)}{t}+\frac{1}{\sqrt{t}}\bigg).\end{aligned}$$

6.3. *Proof of Theorem* 3.1 *and the main theorem*: *Hessian estimates*. Now we can prove the claim for the second order gradient of logarithmic heat kernel. In Proposition 4.4 we have established a second order gradient formula for $P_t f$ on a compact manifold. In its proof we exchanged the differential and the integral operators several times which may not hold if M is not compact. So it is not trivial to extend Proposition 4.4 to a noncompact manifold.

To prove Theorem 3.1, we begin with comparing the terms in $\nabla^2 P_t^k$ and $\nabla^2 P_t$.

LEMMA 6.8. *Given a point $x\in M$ and a vector $v\in T_xM$, suppose that m is sufficiently large so $x\in D_m$ and $k>m$. Let $\{U_t^k\}_{t\ge 0}$ be the horizontal Brownian motion on $\tilde{M}_k$, as defined in* (6.2). *Set $X_t^k=\pi(U_t^k)$ and $P_t^k f(x)=\mathbb{E}_x[f(X_t^k)]$. Let $h(s)=(\frac{t-2s}{t})^+ l_m(s,X_\cdot)U_0^{-1}v$, and define*

$$I\Big(\frac{t}{2},X_\cdot,v\Big):=\bigg(\int_0^{\frac{t}{2}}\langle\Theta_s^h,\mathrm{d}B_s\rangle\bigg)^2-\int_0^{\frac{t}{2}}\langle\Lambda_s^h,\mathrm{d}B_s\rangle-\int_0^{\frac{t}{2}}|\Theta_s^h|^2\,\mathrm{d}s. \tag{6.40}$$

Let $I(\frac{t}{2},X_\cdot^k,v)$ be defined with the corresponding terms in $\tilde{M}_k$. Then we have

$$I\Big(\frac{t}{2},X_\cdot^k,v\Big)=I\Big(\frac{t}{2},X_\cdot,v\Big)=\bigg(\int_0^t\langle\Theta_s^h,\mathrm{d}B_s\rangle\bigg)^2-\int_0^t\langle\Lambda_s^h,\mathrm{d}B_s\rangle-\int_0^t|\Theta_s^h|^2\,\mathrm{d}s. \tag{6.41}$$

Furthermore, it holds that

$$\sup_{x\in D_m, v\in T_xM, |v|=1}\mathbb{E}_x\bigg[\bigg|I\Big(\frac{t}{2},X_\cdot,v\Big)\bigg|^2\bigg]<\infty,\quad \forall t>0. \tag{6.42}$$

PROOF. Let $\Theta_s^{h,k}$, $\Gamma_s^{h,k}$, $\Lambda_s^{h,k}$ be the corresponding terms of Θ_s^h, Γ_s^h, Λ_s^h defined on $\tilde{M}_k$. By (6.26) we have

$$\Theta_s^{h,k}=h'(s)+\mathrm{ric}_{U_s^k}^{\tilde{M}_k}(h(s))=h'(s)+\mathrm{ric}_{U_s}(h(s))=\Theta_s^h,\quad \forall k>m. \tag{6.43}$$

Still based on (3.2), (3.3) and the same arguments for (6.43), we can obtain that

$$\Gamma_s^{h,k}=\Gamma_s^h,\qquad \Lambda_s^{h,k}=\Lambda_s^h,\quad \forall k>m.$$

Therefore, the term $I(\frac{t}{2},X_\cdot^k,v)$ in (6.40) is independent of k, and the required identity (6.41) holds. Finally, (6.42) immediately follows from the moment estimates (5.3) for l'_m and the same arguments for (6.27). □

Proof of Theorem 3.1. The idea of the proof is similar to that of Lemma 6.5. For convenience of the reader, here we provide a detailed proof. Let $m_0 \in \mathbb{N}$ satisfy that $x \in D_{m_0+1}$, then for every $k > m > m_0$ it holds that $B_o(2m-2) \subset D_m \subset D_k$. Let $h(s) = (\frac{t-2s}{t})^+ l_m(s, X_\cdot) U_0^{-1} v = (\frac{t-2s}{t})^+ l_m(s, X_\cdot^k) U_0^{-1} v$. We can apply (4.12) in Proposition 4.4 to the compact manifold $\tilde{M}_k$ to obtain that, for every $k > m$,

$$\begin{aligned}
&\langle \nabla^2 P_t^k f(x), v \otimes v\rangle_{T_xM\otimes T_xM} \\
&= \mathbb{E}_x\Big[f(X_t^k)\Big(\Big(\int_0^t \langle \Theta_s^{h,k}, \mathrm{d}B_s\rangle\Big)^2 - \int_0^t \langle \Lambda_s^{h,k}, \mathrm{d}B_s\rangle - \int_0^t |\Theta_s^{h,k}|^2\,\mathrm{d}s\Big)\Big] \\
&= \mathbb{E}_x\Big[f(X_t^k) I\Big(\frac{t}{2}, X_\cdot^k, v\Big)\Big] = \mathbb{E}_x\Big[f(X_t^k) I\Big(\frac{t}{2}, X_\cdot, v\Big)\Big],
\end{aligned} \tag{6.44}$$

where the process $\Theta_s^{h,k}$, $\Lambda_s^{h,k}$ are defined by (3.2), (3.3) on $\tilde{M}_k$, and in the last step we have applied (6.41).

According to (6.40) and integration by parts formula (on compact manifold $\tilde{M}_k$), for any $\psi \in C_c^\infty(M)$, $V \in C_c^\infty(M; TM)$ with $\operatorname{supp}\psi \cup \operatorname{supp} V \subset D_m$ we have

$$\begin{aligned}
&\int_M \mathbb{E}_x\Big[f(X_t^k) I\Big(\frac{t}{2}, X_\cdot, V(x)\Big)\Big]\psi(x)\,\mathrm{d}x \\
&= \int_M \langle \nabla^2 P_t^k f(x), V(x)\otimes V(x)\rangle_{T_xM\otimes T_xM}\psi(x)\,\mathrm{d}x \\
&= \int_{\tilde{M}_k} \langle \nabla^2 P_t^k f(x), V(x)\otimes V(x)\rangle_{T_xM\otimes T_xM}\psi(x)\,\mathrm{d}x \\
&= \int_{\tilde{M}_k} \mathbb{E}_x[f(X_t^k)]\Psi(\psi, V)(x)\,\mathrm{d}x = \int_M \mathbb{E}_x[f(X_t^k)]\Psi(\psi, V)(x)\,\mathrm{d}x.
\end{aligned} \tag{6.45}$$

Here, we denote the gradient operator and Riemannian volume measure on both M and $\tilde{M}_k$ by ∇ and $\mathrm{d}x$, and we set

$$\begin{aligned}
\Psi(\psi, V)(x) &:= \operatorname{div}\big(\operatorname{div}(V\psi)V\big)(x) + \operatorname{div}(\psi\nabla_V V)(x) \\
&= \psi(x)\big(\operatorname{div}(\nabla_V V) + (\operatorname{div} V)^2 + \langle V, \nabla \operatorname{div} V\rangle_{T_xM}\big)(x) \\
&\quad + 2\langle \nabla\psi, \nabla_V V + (\operatorname{div} V)V\rangle_{T_xM}(x) + \langle \nabla^2\psi(x), V(x)\otimes V(x)\rangle_{T_xM\otimes T_xM}.
\end{aligned}$$

The second and last step above follow from the properties that Riemannian volume measure $\mathrm{d}x$, and the second order gradient operator ∇^2 on M are the same as that on $\tilde{M}_k$, when they are restricted on D_m, the third equality is due to the integration by parts formula. Meanwhile, note that $X_t = X_t^k$ if $t < \tau_k$, for every $x \in D_m$ it holds

$$\begin{aligned}
&\lim_{k\to\infty}\Big|\mathbb{E}_x\Big[f(X_t^k) I\Big(\frac{t}{2}, X_\cdot, V(x)\Big)\Big] - \mathbb{E}_x\Big[f(X_t) I\Big(\frac{t}{2}, X_\cdot, V(x)\Big)\mathbf{1}_{\{t<\zeta\}}\Big]\Big| \\
&\le \lim_{k\to\infty} \mathbb{E}_x\Big[\big|f(X_t^k) - f(X_t)\mathbf{1}_{\{t<\zeta\}}\big|\Big|I\Big(\frac{t}{2}, X_\cdot, V(x)\Big)\Big|\Big] \\
&\le \lim_{k\to\infty} \sqrt{\mathbb{E}_x\big[|f(X_t^k) - f(X_t)\mathbf{1}_{\{t<\zeta\}}|^2\big]}\sqrt{\mathbb{E}_x\Big[\Big|I\Big(\frac{t}{2}, X_\cdot, V(x)\Big)\Big|^2\Big]} \\
&\le \lim_{k\to\infty} \sqrt{2}C\|f\|_\infty \sqrt{\mathbb{P}_x(\tau_k \le t < \zeta)} = 0,
\end{aligned}$$

where the last inequality is due to (6.42).

Putting this into (6.45), letting $k \to \infty$, we see that, for every $\psi \in C_c^\infty(M)$ and $V \in C^\infty(M; TM)$ with $\operatorname{supp}\psi \cup \operatorname{supp} V \subset D_m$,

$$\int_{D_m} \mathbb{E}_x\left[f(X_t)I\left(\frac{t}{2}, X_\cdot, V(x)\right)\mathbf{1}_{\{t<\zeta\}}\right]\psi(x)\,\mathrm{d}x = \int_{D_m} \mathbb{E}_x\left[f(X_t)\mathbf{1}_{\{t<\zeta\}}\right]\Psi(\psi, V)(x)\,\mathrm{d}x,$$

which implies the weak (distributional) second order gradient $\nabla^2 P_t f$ exists on D_m and

$$\langle \nabla^2 P_t f(x), v \otimes v\rangle_{T_xM\otimes T_xM} = \mathbb{E}_x\left[f(X_t)I\left(\frac{t}{2}, X_\cdot, v\right)\mathbf{1}_{\{t<\zeta\}}\right], \quad x \in D_m, v \in T_xM. \tag{6.46}$$

As shown by Lemma A.2 in the Appendix, the functional $x \mapsto \mathbb{E}_x[f(X_t)I(\frac{t}{2}, X_\cdot, V(x))\mathbf{1}_{\{t<\zeta\}}]$ is continuous. Now the distributional derivative $\nabla^2 P_t f$ exists and is continuous, then $\nabla^2 P_t f$ is the classical second order gradient on D_m, and expression (3.4) holds.

PROPOSITION 6.9. *Suppose that K is a compact subset of M and $L > 1$ is a positive constant. Then, for any sufficiently large m, we could find a $t_0(K, L, m)$ such that, for any $t \in (0, t_0]$,*

$$\sup_{x,y\in K} \left|t\nabla_x^2 \log p(t, x, y) - t\nabla_x^2 \log p_{\tilde{M}_m}(t, x, y)\right|_{T_xM\otimes T_xM} \le C(m)e^{-\frac{L}{t}}, \tag{6.47}$$

where $C(m)$ is a positive constant which may depend on m.

PROOF. Let us fix $x, y \in K$ and a unit vector $v \in T_xM$ with $|v| = 1$. Suppose that $m \in \mathbb{N}$ such that $K \subset B_o(2m-2) \subset D_m$. Then by (3.4) we have

$$\langle \nabla^2 P_t f(x), v \otimes v\rangle_{T_xM\otimes T_xM} = \mathbb{E}_x\left[f(X_t)I\left(\frac{t}{2}, X_\cdot, v\right)\mathbf{1}_{\{t<\zeta\}}\right],$$

where $I(\frac{t}{2}, X_\cdot, v)$ is defined by (6.40) with $h(s) := (\frac{t-2s}{t})^+ l_m(s, X_\cdot)U_0^{-1}v$.

By this representation and following the same arguments of (6.33) and (6.34), we obtain

$$\frac{\langle \nabla_x^2 p(t, x, y), v \otimes v\rangle_{T_xM\otimes T_xM}}{p(t, x, y)} = \mathbb{E}_x\left[I\left(\frac{t}{2}, X_\cdot, v\right)\frac{p(\frac{t}{2}, X_{\frac{t}{2}}, y)}{p(t, x, y)}\right],$$

$$\frac{\langle \nabla_x^2 p_{\tilde{M}_m}(t, x, y), v \otimes v\rangle_{T_xM\otimes T_xM}}{p_{\tilde{M}_m}(t, x, y)} = \mathbb{E}_x\left[I\left(\frac{t}{2}, X_\cdot, v\right)\frac{p_{\tilde{M}_m}(\frac{t}{2}, X^m_{\frac{t}{2}}, y)}{p_{\tilde{M}_m}(t, x, y)}\right].$$

Based on above expression and following the same arguments in the proof of Proposition 6.6 (especially applying (6.13) and (6.21)–(6.23)), we could find a $m_0(K, L) \in \mathbb{N}$ such that, for all $m \ge m_0$, there exists a $t_0(K, L, m) > 0$ such that

$$\sup_{x,y\in K}\left|\frac{\nabla_x^2 p(t, x, y)}{p(t, x, y)} - \frac{\nabla_x^2 p_{\tilde{M}_m}(t, x, y)}{p_{\tilde{M}_m}(t, x, y)}\right|_{T_xM\otimes T_xM} \le C(m)e^{-\frac{L}{t}}, \quad t \in (0, t_0]. \tag{6.48}$$

Meanwhile, we have

$$\begin{aligned}&\langle \nabla_x^2 \log p(t, x, y), v \otimes v\rangle_{T_xM\otimes T_xM}\\ &\quad = \frac{\langle \nabla_x^2 p(t, x, y), v \otimes v\rangle_{T_xM\otimes T_xM}}{p(t, x, y)} - \left(\frac{\langle \nabla_x p(t, x, y), v\rangle_{T_xM}}{p(t, x, y)}\right)^2,\end{aligned}$$

and the similar expression holds for $\langle \nabla_x^2 \log p_{\tilde{M}_m}(t, x, y), v \otimes v\rangle_{T_xM\otimes T_xM}$. Together with (6.31) and (6.48), this yields (6.47) and concludes the proof. □

With (6.47) we are in the position to prove the second part of the main theorem on the short time and asymptotic second order gradient estimates.

THEOREM 6.10. *The following statements hold*:

(1) *Suppose* $y \in M$ *and* $\tilde{K} \subset M \setminus \mathrm{Cut}_M(y)$ *is a compact set*, *then*

$$\lim_{t \downarrow 0} \sup_{x \in \tilde{K}} \left| t \nabla_x^2 \log p(t, x, y) + \nabla_x^2 \left(\frac{d^2(x, y)}{2} \right) \right|_{T_x M \otimes T_x M} = 0. \tag{6.49}$$

(2) *For any* $y \in M$ *and* $\delta < i(y)$ *there exist positive constants* t_0 *and* C_1 *such that*

$$\left| t \nabla_x^2 \log p(t, x, y) + \mathbf{I}_{T_x M} \right|_{T_x M \otimes T_x M} \le C_1 \big(d(x, y) + \sqrt{t} \big), \quad x \in B_y(\delta), t \in (0, t_0], \tag{6.50}$$

where $\mathbf{I}_{T_x M}$ *is the identical map on* $T_x M$.

(3) *Suppose* $K \subset M$ *is a compact subset of* M, *then there exists a positive constant* $C_2(K)$ *such that*

$$\left| \nabla_x^2 \log p(t, x, y) \right|_{T_x M \otimes T_x M} \le C_2 \left(\frac{d^2(x, y)}{t^2} + \frac{1}{t} \right), \quad x, y \in K, t \in (0, 1]. \tag{6.51}$$

PROOF. By Malliavin and Stroock [59], Corollary 2.29, Gong and Ma [36], Theorem 3.1, and Stroock [66] (or Sheu [64]), we know (6.49)–(6.51) hold when M is compact. Then, using the estimates (6.47) and following the same procedure as in the proof of Theorem 6.7, we can verify that (6.49)–(6.51) hold for any complete Riemannian manifold. □

APPENDIX: APPROXIMATION PROCEDURE

Let (M, g) and $D_m \subset M$ be the same terms in Section 5.

LEMMA A.1. *For every* $m \in \mathbb{Z}_+$, *there exists a* (*smooth*) *compact Riemannian manifold* $(\tilde{M}_m, \tilde{g}_m)$ *such that* (D_m, g) *is isometrically embedded into* $(\tilde{M}_m, \tilde{g}_m)$ *as an open set. In particular, if* $y, x \in D_m$ *and* $x \notin \mathrm{cut}_y(M)$, *then* $x \notin \mathrm{cut}_y(\tilde{M}_m)$.

PROOF. Let $G_m = D_{m+1}$, recall that ∂G_m is a connected smooth $n-1$-dimensional submanifold of M. Hence, $\overline{G_m}$ is an n-dimensional manifold with smooth boundary; then there exist an atlas of local charts $\{(V_i, \psi_i)\}_{i=1}^N$ of $\overline{G_m}$ such that:

(1) $\bigcup_{i=1}^N V_i = \overline{G_m}$;

(2) For $i = 1, \dots, N_1 \le N$, these are charts for the interior. So $V_i \cap \partial G_m = \varnothing$, and $\psi_i : V_i \to \mathbf{B}^n := \{z \in \mathbb{R}^n; |z| < 1\}$ is a smooth diffeomorphsim for all $1 \le i \le N_1$;

(3) For all $i > N_1$, $V_i \cap \partial G_m \neq \varnothing$,

$$\psi_i : V_i \to \mathbf{B}^{n,+} := \{ z = (z_1, \dots, z_n) \in \mathbb{R}^n; |z| < 1, z_1 \ge 0 \}$$

is a smooth diffeomorphsim and $\psi_i(V_i \cap \partial G_m) = \partial \mathbf{B}^{n,+}$.

By the Whitney embedding theorem, we could embed M into a (ambient) Euclidean space $\mathbb{R}^p$. Let $\hat{G}_m$ be an identical copy of G_m in $\mathbb{R}^p$, endowed with the local charts $\{(\hat{V}_i, \hat{\psi}_i)\}_{i=1}^N$ (which is also an identical copy of $\{(V_i, \psi_i)\}_{i=1}^N$). We define $h : \partial G_m \to \partial \hat{G}_m$ by $h(x) := \hat{\psi}_i^{-1}(\psi_i(x))$, if $x \in V_i \cap \partial G_m$, h is well defined and is a smooth diffeomorphism.

We glue the boundary of G_m and $\tilde{G}_m$ together to get $\tilde{M}_m := (G_m \sqcup \hat{G}_m)/\sim$, where $\sim$ is an equivalent relation such that $x \sim y$ if and only if $h(x) = y$, $x \in \partial G_m$, $y \in \partial \hat{G}_m$. Then $\tilde{M}_m$ is a smooth compact manifold without boundary. In fact, $\{(U_i, \phi_i)\}_{i=1}^{N+N_1} = \{(V_i, \psi_i)\}_{i=1}^{N_1} \cup \{(\hat{V}_i, \hat{\psi}_i)\}_{i=1}^{N_1} \cup \{(\tilde{V}_i, \tilde{\psi}_i)\}_{i=N_1+1}^N$ is a local charts of $\tilde{M}_m$. Here, $\tilde{V}_i = (V_i \sqcup \hat{V}_i)/\sim$ for every $N_1 < i \le N$ and

$$\tilde{\psi}_i(x) = \begin{cases} \psi_i(x), & \text{if } x \in V_i, \\ \mathbf{S}(\hat{\psi}_i(x)), & \text{if } x \in \hat{V}_i, \end{cases}$$

where $\mathbf{S}: \mathbb{R}^n \to \mathbb{R}^n$ is a map such that $\mathbf{S}x = (-x_1, x_2, \dots, x_n)$ for all $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$. It is easy to see $\tilde{\psi}_i : \tilde{V}_i \to \mathbf{B}^n$, $N_1 < i \le N$ is a smooth diffeomorphsim, and the transition map between different local charts on $\{(U_i, \phi_i)\}_{i=1}^{N+N_1}$ is smooth.

We construct a smooth Riemannian metric $\tilde{g}_m$ on $\tilde{M}_m$ to ensure that $\tilde{g}_m(z) = g(z)$ for every $z \in D_m$. For the open set $D_m \subset G_m \subset \tilde{M}_m$, by the standard procedure (via the finite local charts) we could construct a function $\chi_m : \tilde{M}_m \to [0, 1]$ such that $\operatorname{supp} \chi_m \subset G_m$ and $\chi_m(x) = 1$ for every $x \in D_m$. Note that G_m could also be viewed as an open subset of $\tilde{M}_m$, so $\hat{g}_m(x) := g(x)\chi_m(x)$, $x \in \tilde{M}_m$ is well defined on $\tilde{M}_m$. Fixing a smooth Riemannian metric g_m^0 on $\tilde{M}_m$, which exists, we set

$$\tilde{g}_m(x) := g(x)\chi_m(x) + g_m^0(x)\big(1 - \chi_m(x)\big), \quad x \in \tilde{M}_m.$$

It is easy to see $\tilde{g}_m$ is a smooth Riemannian metric on $\tilde{M}_m$ and $\tilde{g}_m(x) = g(x)$ for each $x \in D_m$. By now we have completed the proof. □

Let $I(t, X_\cdot, v)$ be as defined in (6.40).

LEMMA A.2. *For every fixed $f \in C_c^\infty(M)$, $V \in C^\infty(M; TM)$ with compact supports and $t > 0$, the function*

$$F(x) := \mathbb{E}\bigg[f(X_t^x) I\bigg(\frac{t}{2}, X_\cdot^x, V(x)\bigg)\mathbf{1}_{\{t<\zeta(x)\}}\bigg], \quad x \in M$$

is continuous.

PROOF. Let $\zeta(x)$ denote the explosion time of the solution X_t^x to (3.1) with the initial value x. Let U be a frame at x. Then the explosion time of the horizontal Brownian motion agree with $\xi(x)$ almost surely. So we use ξ for the explosion time of both. Furthermore, by Elworthy [28] there exist a maximal solution flow $\{U_t(\cdot, \omega)\}_{0 \le t < \zeta(\cdot,\omega)}$ to (3.1) such that $U_t(u, \omega)$ is the solution of (3.1) with initial value $u \in OM$, and there is a null set Δ such that, for all $\omega \notin \Delta$:

(1) For each $t > 0$, set $\Xi_t(\omega) := \{u \in OM : t < \zeta(u, \omega)\}$, Then Ξ_t is open in OM (i.e., $\zeta(\cdot, \omega) : OM \to \mathbb{R}_+$ is lower semicontinuous), and $U_t(\cdot, \omega) : \Xi_t(\omega) \to OM$ is a C^∞ diffeomorphism onto its image.

(2) For each fixed $u \in OM$ with $\pi(u) = x$, there exists a null set $\Delta(u)$, depending on u, such that $\zeta(u, \omega) = \zeta(X_\cdot^x)$ for each $\omega \notin \Delta(u) \cup \Delta$.

Fix a point $x_0 \in M$. For each sequence $\{x_k\}_{k=1}^\infty$ with $\lim_{k\to\infty} x_k = x_0$, we take a sequence $\{u_k\}_{k=1}^\infty$ and U_0 in OM such that $\pi(u_k) = x_k$, $\pi(U_0) = x_0$ and $\lim_{k\to\infty} u_k = u_0$ in OM. Set $\tilde{\Delta} := (\bigcup_{k=0}^\infty \Delta(u_k)) \cup \Delta$. For each k and $\omega \notin \tilde{\Delta}$, $\zeta(U_k, \omega) = \zeta(x_k, \omega)$. By the lower semicontinuity of ζ, $\zeta(x_0) \le \liminf_{k\to\infty} \zeta(x_k)$, hence, $u_k \in \Xi_t(\omega)$ for each $t < \zeta(x_0)$ when k is large enough. By the property (1) above, we have immediately

$$\lim_{k\to\infty} U_t(u_k, \omega)\mathbf{1}_{\{t<\zeta(x_k)\}} = U_t(u_0, \omega)\mathbf{1}_{\{t<\zeta(x_0)\}}, \quad \omega \notin \tilde{\Delta}, t > 0.$$

Combing this with the definition $\Theta(s, X_\cdot, v)$ and the expression (5.4) of l_m, we see that

$$\lim_{k\to\infty} \Theta\big(s, X_\cdot^{x_k}, V(x_k)\big) = \Theta\big(s, X_\cdot^{x_0}, V(x_0)\big), \quad s > 0. \tag{A.1}$$

Let $h(s, X_\cdot, V(x)) := (\frac{t-2s}{t} \vee 0) \cdot l_m(s, X_\cdot) \cdot u_0(x)^{-1} V(x)$, where $u_0(\cdot)$ is a smooth section of OM with $\pi(u_0(x)) = x$. We only need to demonstrate the proof for one of the term in

$I(t, X^x_\cdot, v)$, for this we set

$$w(x) := \mathbb{E}\Big[f(X_t^x)\Big(\int_0^{\frac{t}{2}} \langle \Theta(s, X^x_\cdot, V(x)), \mathrm{d}B_s\rangle\Big)\mathbf{1}_{\{t<\zeta(x)\}}\Big]$$
$$=: \mathbb{E}\Big[f(X_t^x)\Big(\int_0^{\frac{t}{2}} \Big\langle\Big(h'(s) + \frac{1}{2}\mathrm{ric}_{U_s} h(s)\Big), \mathrm{d}B_s\Big\rangle\Big)\mathbf{1}_{\{t<\zeta(x)\}}\Big].$$

For simplicity, we only prove the continuity for the function $x \to w(x)$, the continuity property for the other terms in $F(x)$ could be verified similarly.

According to (5.3), we obtain

$$\sup_{k>0}\mathbb{E}\Big[\Big|\int_0^{\frac{t}{2}} \langle \Theta(s, X^{x_k}_\cdot, V(x_k)), \mathrm{d}B_s\rangle\Big|^4\Big] < \infty.$$

Based on this and (A.1), we have

$$\lim_{k\to\infty}\mathbb{E}\Big[\Big|\int_0^{\frac{t}{2}} \langle \Theta(s, X^{x_k}_\cdot, V(x_k)), \mathrm{d}B_s\rangle - \int_0^{\frac{t}{2}} \langle \Theta(s, X^{x_0}_\cdot, V(x_0)), \mathrm{d}B_s\rangle\Big|^2\Big] = 0.$$

Similarly from (A.1), we arrive at

$$\lim_{k\to\infty}\mathbb{E}\big[\big|f(X_t^{x_k})\mathbf{1}_{\{t<\zeta(x_k)\}} - f(X_t^{x_0})\mathbf{1}_{\{t<\zeta(x_0)\}}\big|^2\big] = 0.$$

Therefore, by Cauchy–Schwarz inequality

$$\begin{aligned}
&\lim_{k\to\infty}\big|w(x_k) - w(x_0)\big|^2 \\
&\quad\le 2\|f\|_\infty^2 \lim_{k\to\infty}\mathbb{E}\Big[\Big|\int_0^{\frac{t}{2}} \langle \Theta(s, X^{x_k}_\cdot, V(x_k)), \mathrm{d}B_s\rangle - \int_0^{\frac{t}{2}} \langle \Theta(s, X^{x_0}_\cdot, V(x_0)), \mathrm{d}B_s\rangle\Big|^2\Big] \\
&\qquad + 2\sup_{k>0}\mathbb{E}\Big[\Big|\int_0^{\frac{t}{2}} \langle \Theta(s, X^{x_k}_\cdot, V(x_k)), \mathrm{d}B_s\rangle\Big|^4\Big] \\
&\qquad \cdot \lim_{k\to\infty}\mathbb{E}\big[\big|f(X_t^{x_k})\mathbf{1}_{\{t<\zeta(x_k)\}} - f(X_t^{x_0})\mathbf{1}_{\{t<\zeta(x_0)\}}\big|^2\big] \\
&\quad = 0.
\end{aligned}$$

Since $\{x_k\}_{k=1}^\infty$ is arbitrarily chosen, $w(\cdot)$ is continuous at $x_0 \in M$. Again, x_0 is arbitrary, so $w(\cdot)$ is continuous on M. This completes the proof for the lemma. □

Acknowledgments. We would like to thank Christian Bär and Robert Neel for helpful comments and the referees for their valuable comments and suggestions.

Funding. The research of Xin Chen is supported by the National Natural Science Foundation of China (No. 12122111). The research of Xue-Mei Li is supported by EPSRC (Nos. EP/E058124/1, S023925/1 and EP/V026100/1). The research of Bo Wu is supported by the National Natural Science Foundation of China (No. 12071085).

REFERENCES

[1] AIDA, S. (2000). Logarithmic derivatives of heat kernels and logarithmic Sobolev inequalities with unbounded diffusion coefficients on loop spaces. *J. Funct. Anal.* **174** 430–477. MR1768982 https://doi.org/10.1006/jfan.2000.3592

[2] AIDA, S. (2004). Precise Gaussian estimates of heat kernels on asymptotically flat Riemannian manifolds with poles. In *Recent Developments in Stochastic Analysis and Related Topics* 1–19. World Sci. Publ., Hackensack, NJ. MR2200501

[3] AIDA, S. (2015). Asymptotics of spectral gaps on loop spaces over a class of Riemannian manifolds. *J. Funct. Anal.* **269** 3714–3764. MR3418070 https://doi.org/10.1016/j.jfa.2015.09.023

[4] AIRAULT, H. and MALLIAVIN, P. (1992). Integration on loop groups. II. Heat equation for the Wiener measure. *J. Funct. Anal.* **104** 71–109. MR1152460 https://doi.org/10.1016/0022-1236(92)90091-V

[5] AIRAULT, H., MALLIAVIN, P. and REN, J. (1999). Smoothness of stopping times of diffusion processes. *J. Math. Pures Appl.* (9) **78** 1069–1091. MR1732053 https://doi.org/10.1016/S0021-7824(99)00026-4

[6] ARNAUDON, M., PLANK, H. and THALMAIER, A. (2003). A Bismut type formula for the Hessian of heat semigroups. *C. R. Math. Acad. Sci. Paris* **336** 661–666. MR1988128 https://doi.org/10.1016/S1631-073X(03)00123-7

[7] AZENCOTT, R. (1980). Grandes déviations et applications. In *Eighth Saint Flour Probability Summer School*—1978 (*Saint Flour*, 1978). *Lecture Notes in Math.* **774** 1–176. Springer, Berlin. MR0590626

[8] AZENCOTT, R., ed. (1981). Géodésiques et diffusions en temps petit, sḿinaire de probabilitś universitd́e Paris VII, Asterisque 84–85, Ch. 5 by L. Elie and Ch. IX by C. Bellaiche. Math. Soc. France.

[9] BALL, K., BARTHE, F., BEDNORZ, W., OLESZKIEWICZ, K. and WOLFF, P. (2013). L^1-smoothing for the Ornstein–Uhlenbeck semigroup. *Mathematika* **59** 160–168. MR3028177 https://doi.org/10.1112/S0025579312001064

[10] BEN AROUS, G. (1988). Développement asymptotique du noyau de la chaleur hypoelliptique hors du cut-locus. *Ann. Sci. Éc. Norm. Supér.* (4) **21** 307–331. MR0974408

[11] BEN AROUS, G. and LÉANDRE, R. (1991). Décroissance exponentielle du noyau de la chaleur sur la diagonale. I. *Probab. Theory Related Fields* **90** 175–202. MR1128069 https://doi.org/10.1007/BF01192161

[12] BISMUT, J.-M. (1984). *Large Deviations and the Malliavin Calculus*. *Progress in Mathematics* **45**. Birkhäuser, Inc., Boston, MA. MR0755001

[13] CHEEGER, J., GROMOV, M. and TAYLOR, M. (1982). Finite propagation speed, kernel estimates for functions of the Laplace operator, and the geometry of complete Riemannian manifolds. *J. Differential Geom.* **17** 15–53. MR0658471

[14] CHEEGER, J. and YAU, S. T. (1981). A lower bound for the heat kernel. *Comm. Pure Appl. Math.* **34** 465–480. MR0615626 https://doi.org/10.1002/cpa.3160340404

[15] CHEN, X., LI, X.-M. and WU, B. Stochastic analysis on loop space over general Riemannian manifold. Preprint.

[16] CHEN, X., LI, X.-M. and WU, B. (2010). A Poincaré inequality on loop spaces. *J. Funct. Anal.* **259** 1421–1442. MR2659766 https://doi.org/10.1016/j.jfa.2010.05.006

[17] CHEN, X., LI, X.-M. and WU, B. (2011). A concrete estimate for the weak Poincaré inequality on loop space. *Probab. Theory Related Fields* **151** 559–590. MR2851693 https://doi.org/10.1007/s00440-010-0308-5

[18] CHEN, X. and WU, B. (2014). Functional inequality on path space over a non-compact Riemannian manifold. *J. Funct. Anal.* **266** 6753–6779. MR3198853 https://doi.org/10.1016/j.jfa.2014.03.017

[19] CHENG, S. Y., LI, P. and YAU, S. T. (1981). On the upper estimate of the heat kernel of a complete Riemannian manifold. *Amer. J. Math.* **103** 1021–1063. MR0630777 https://doi.org/10.2307/2374257

[20] DAVIES, E. B. (1989). *Heat Kernels and Spectral Theory*. *Cambridge Tracts in Mathematics* **92**. Cambridge Univ. Press, Cambridge. MR0990239 https://doi.org/10.1017/CBO9780511566158

[21] DRIVER, B. K. (1992). A Cameron–Martin type quasi-invariance theorem for Brownian motion on a compact Riemannian manifold. *J. Funct. Anal.* **110** 272–376. MR1194990 https://doi.org/10.1016/0022-1236(92)90035-H

[22] DRIVER, B. K. (1994). A Cameron–Martin type quasi-invariance theorem for pinned Brownian motion on a compact Riemannian manifold. *Trans. Amer. Math. Soc.* **342** 375–395. MR1154540 https://doi.org/10.2307/2154699

[23] EBERLE, A. (2001). Local Poincaré inequalities on loop spaces. *C. R. Acad. Sci. Paris Sér. I Math.* **333** 1023–1028. MR1872466 https://doi.org/10.1016/S0764-4442(01)02174-7

[24] EBERLE, A. (2002). Absence of spectral gaps on a class of loop spaces. *J. Math. Pures Appl.* (9) **81** 915–955. MR1946909 https://doi.org/10.1016/S0021-7824(02)01260-6

[25] EELLS, J. and ELWORTHY, K. D. (1971). On Fredholm manifolds. In *Actes du Congrès International des Mathématiciens* (*Nice*, 1970), *Tome* 2 215–219. Gauthier-Villars, Paris. MR0433487

[26] ELDAN, R. and LEE, J. R. (2015). Talagrand's convolution conjecture on Gaussian space. In 2015 *IEEE* 56*th Annual Symposium on Foundations of Computer Science—FOCS* 2015 1395–1408. IEEE Computer Soc., Los Alamitos, CA. MR3473377

[27] ELWORTHY, K. D. (1978). Stochastic dynamical systems and their flows. In *Stochastic Analysis* (*Proc. Internat. Conf., Northwestern Univ., Evanston, Ill.*, 1978) 79–95. Academic Press, New York. MR0517235

[28] ELWORTHY, K. D. (1982). *Stochastic Differential Equations on Manifolds*. *London Mathematical Society Lecture Note Series* **70**. Cambridge Univ. Press, Cambridge. MR0675100

[29] ELWORTHY, K. D., LE JAN, Y. and LI, X.-M. (1999). *On the Geometry of Diffusion Operators and Stochastic Flows*. *Lecture Notes in Math*. **1720**. Springer, Berlin. MR1735806 https://doi.org/10.1007/BFb0103064

[30] ELWORTHY, K. D. and LI, X.-M. (1994). Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** 252–286. MR1297021 https://doi.org/10.1006/jfan.1994.1124

[31] ELWORTHY, K. D. and LI, X.-M. (1994). Differentiation of heat semigroups and applications. In *Probability Theory and Mathematical Statistics* (*Vilnius*, 1993) 239–251. TEV, Vilnius. MR1649579

[32] ELWORTHY, K. D. and LI, X.-M. (1996). A class of integration by parts formulae in stochastic analysis. I. In *Itô's Stochastic Calculus and Probability Theory* 15–30. Springer, Tokyo. MR1439515

[33] ENGOULATOV, A. (2006). A universal bound on the gradient of logarithm of the heat kernel for manifolds with bounded Ricci curvature. *J. Funct. Anal.* **238** 518–529. MR2253731 https://doi.org/10.1016/j.jfa.2006.02.013

[34] FANG, S. and MALLIAVIN, P. (1993). Stochastic analysis on the path space of a Riemannian manifold. I. Markovian stochastic calculus. *J. Funct. Anal.* **118** 249–274. MR1245604 https://doi.org/10.1006/jfan.1993.1145

[35] GONG, F., RÖCKNER, M. and WU, L. (2001). Poincaré inequality for weighted first order Sobolev spaces on loop spaces. *J. Funct. Anal.* **185** 527–563. MR1856276 https://doi.org/10.1006/jfan.2001.3775

[36] GONG, F.-Z. and MA, Z.-M. (1998). The log-Sobolev inequality on loop space over a compact Riemannian manifold. *J. Funct. Anal.* **157** 599–623. MR1638269 https://doi.org/10.1006/jfan.1998.3263

[37] GOZLAN, N., LI, X.-M., MADIMAN, M., ROBERTO, C. and SAMSON, P. M. (2021). Log-Hessian and deviation bounds for Markov semi-groups, and regularization effect in $\mathbb{L}^1$. *Potential Anal.*. Available at arXiv:1907.10896. https://doi.org/10.1007/s11118-021-09934-z

[38] GREENE, R. E. and WU, H. (1979). *Function Theory on Manifolds Which Possess a Pole*. *Lecture Notes in Math*. **699**. Springer, Berlin. MR0521983

[39] GROSS, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083. MR0420249 https://doi.org/10.2307/2373688

[40] GROSS, L. (1991). Logarithmic Sobolev inequalities on loop groups. *J. Funct. Anal.* **102** 268–313. MR1140628 https://doi.org/10.1016/0022-1236(91)90123-M

[41] HAMILTON, R. S. (1993). A matrix Harnack estimate for the heat equation. *Comm. Anal. Geom.* **1** 113–126. MR1230276 https://doi.org/10.4310/CAG.1993.v1.n1.a6

[42] HSU, E. P. (1999). Estimates of derivatives of the heat kernel on a compact Riemannian manifold. *Proc. Amer. Math. Soc.* **127** 3739–3744. MR1618694 https://doi.org/10.1090/S0002-9939-99-04967-9

[43] HSU, E. P. (2002). *Stochastic Analysis on Manifolds*. *Graduate Studies in Mathematics* **38**. Amer. Math. Soc., Providence, RI. MR1882015 https://doi.org/10.1090/gsm/038

[44] HSU, E. P. and OUYANG, C. (2009). Quasi-invariance of the Wiener measure on the path space over a complete Riemannian manifold. *J. Funct. Anal.* **257** 1379–1395. MR2541273 https://doi.org/10.1016/j.jfa.2009.05.017

[45] HSU, P. (1990). Heat kernel on noncomplete manifolds. *Indiana Univ. Math. J.* **39** 431–442. MR1089046 https://doi.org/10.1512/iumj.1990.39.39023

[46] KOTSCHWAR, B. L. (2007). Hamilton's gradient estimate for the heat kernel on complete manifolds. *Proc. Amer. Math. Soc.* **135** 3013–3019. MR2317980 https://doi.org/10.1090/S0002-9939-07-08837-5

[47] LEHEC, J. (2016). Regularization in L_1 for the Ornstein–Uhlenbeck semigroup. *Ann. Fac. Sci. Toulouse Math.* (6) **25** 191–204. MR3485296 https://doi.org/10.5802/afst.1492

[48] LI, P. and YAU, S.-T. (1986). On the parabolic kernel of the Schrödinger operator. *Acta Math.* **156** 153–201. MR0834612 https://doi.org/10.1007/BF02399203

[49] LI, X.-D. (2016). Hamilton's Harnack inequality and the W-entropy formula on complete Riemannian manifolds. *Stochastic Process. Appl.* **126** 1264–1283. MR3461198 https://doi.org/10.1016/j.spa.2015.11.002

[50] LI, X.-M. (1992). Stochastic differential equations on noncompact manifolds. Thesis, Univ. Warwick.

[51] LI, X.-M. (1994). Strong p-completeness of stochastic differential equations and the existence of smooth flows on noncompact manifolds. *Probab. Theory Related Fields* **100** 485–511. MR1305784 https://doi.org/10.1007/BF01268991

[52] LI, X.-M. (2018). Generalised Brownian bridges: Examples. *Markov Process. Related Fields* **24** 151–163. MR3793947

[53] LI, X.-M. (2018). Doubly damped stochastic parallel translations and Hessian formulas. In *Stochastic Partial Differential Equations and Related Fields*. *Springer Proc. Math. Stat.* **229** 345–357. Springer, Cham. MR3828180 https://doi.org/10.1007/978-3-319-74929-7_2

[54] LI, X.-M. (2021). Hessian formulas and estimates for parabolic Schrödinger operators. *J. Stoch. Anal.* **2** Art. 7, 53 pp. MR4304478

[55] LI, X.-M. and SCHEUTZOW, M. (2011). Lack of strong completeness for stochastic flows. *Ann. Probab.* **39** 1407–1421. MR2857244 https://doi.org/10.1214/10-AOP585

[56] LI, X.-M. and THOMPSON, J. (2018). First order Feynman–Kac formula. *Stochastic Process. Appl.* **128** 3006–3029. MR3834851 https://doi.org/10.1016/j.spa.2017.10.010

[57] MALLIAVIN, P. (1978). *Géométrie Différentielle Stochastique*. *Séminaire de Mathématiques Supérieures* [*Seminar on Higher Mathematics*] **64**. Presses de l'Université de Montréal, Montreal, Que. Notes prepared by Danièle Dehen and Dominique Michel. MR0540035

[58] MALLIAVIN, P. (1999). The canonic diffusion above the diffeomorphism group of the circle. *C. R. Acad. Sci. Paris Sér. I Math.* **329** 325–329. MR1713340 https://doi.org/10.1016/S0764-4442(00)88575-4

[59] MALLIAVIN, P. and STROOCK, D. W. (1996). Short time behavior of the heat kernel and its logarithmic derivatives. *J. Differential Geom.* **44** 550–570. MR1431005

[60] MOLČANOV, S. A. (1975). Diffusion processes, and Riemannian geometry. *Uspekhi Mat. Nauk* **30** 3–59. MR0413289

[61] NEEL, R. (2007). The small-time asymptotics of the heat kernel at the cut locus. *Comm. Anal. Geom.* **15** 845–890. MR2395259

[62] NORRIS, J. R. (1993). Path integral formulae for heat kernels and their derivatives. *Probab. Theory Related Fields* **94** 525–541. MR1201558 https://doi.org/10.1007/BF01192562

[63] SETTI, A. G. (1993). Eigenvalue estimates for the Laplacian with lower order terms on a compact Riemannian manifold. In *Differential Geometry*: *Riemannian Geometry* (*Los Angeles*, *CA*, 1990). *Proc. Sympos. Pure Math.* **54** 521–527. Amer. Math. Soc., Providence, RI. MR1216643

[64] SHEU, S. J. (1991). Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Probab.* **19** 538–561. MR1106275

[65] SOUPLET, P. and ZHANG, Q. S. (2006). Sharp gradient estimate and Yau's Liouville theorem for the heat equation on noncompact manifolds. *Bull. Lond. Math. Soc.* **38** 1045–1053. MR2285258 https://doi.org/10.1112/S0024609306018947

[66] STROOCK, D. W. (1996). An estimate on the Hessian of the heat kernel. In *Itô's Stochastic Calculus and Probability Theory* 355–371. Springer, Tokyo. MR1439536

[67] THALMAIER, A. (1997). On the differentiation of heat semigroups and Poisson integrals. *Stoch. Stoch. Rep.* **61** 297–321. MR1488139 https://doi.org/10.1080/17442509708834123

[68] THALMAIER, A. and WANG, F.-Y. (1998). Gradient estimates for harmonic functions on regular domains in Riemannian manifolds. *J. Funct. Anal.* **155** 109–124. MR1622800 https://doi.org/10.1006/jfan.1997.3220

[69] THOMPSON, J. (2019). Derivatives of Feynman–Kac semigroups. *J. Theoret. Probab.* **32** 950–973. MR3959634 https://doi.org/10.1007/s10959-018-0824-2

[70] VARADHAN, S. R. S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** 261–286. MR0203230 https://doi.org/10.1002/cpa.3160190303

[71] VARADHAN, S. R. S. (1967). Diffusion processes in a small time interval. *Comm. Pure Appl. Math.* **20** 659–685. MR0217881 https://doi.org/10.1002/cpa.3160200404

[72] VAROPOULOS, N. T. (1989). Small time Gaussian estimates of heat diffusion kernels. I. The semigroup technique. *Bull. Sci. Math.* **113** 253–277. MR1016211

[73] VAROPOULOS, N. T. (1990). Small time Gaussian estimates of heat diffusion kernels. II. The theory of large deviations. *J. Funct. Anal.* **93** 1–33. MR1070036 https://doi.org/10.1016/0022-1236(90)90136-9