Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Stochastic Processes and their Applications 119 (2009) 3653–3670

stochastic processes and their applications

www.elsevier.com/locate/spa

Gradient estimates and Harnack inequalities on non-compact Riemannian manifolds[☆]

Marc Arnaudon[b], Anton Thalmaier[c], Feng-Yu Wang[a,d,*]

[a] *School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China*
[b] *Département de mathématiques, Université de Poitiers, Téléport 2 - BP 30179, F–86962 Futuroscope Chasseneuil Cedex, France*
[c] *Institute of Mathematics, University of Luxembourg, 162A, avenue de la Faïencerie, L–1511 Luxembourg, Luxembourg*
[d] *Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, Swansea, UK*

Received 1 October 2008; received in revised form 1 July 2009; accepted 1 July 2009
Available online 8 July 2009

Abstract

A gradient-entropy inequality is established for elliptic diffusion semigroups on arbitrary complete Riemannian manifolds. As applications, a global Harnack inequality with power and a heat kernel estimate are derived.

MSC: 58J65; 58J35; 60H30

Keywords: Harnack inequality; Heat equation; Gradient estimate; Diffusion semigroup

1. The main result

Let M be a non-compact complete connected Riemannian manifold, and P_t be the Dirichlet diffusion semigroup generated by $L = \Delta + \nabla V$ for some C^2 function V. We intend to establish reasonable gradient estimates and Harnack type inequalities for P_t. In case that $\mathrm{Ric} - \mathrm{Hess}_V$ is bounded below, a dimension-free Harnack inequality was established in [14] which, according

☆ Supported in part by WIMICS, NNSFC(10721091) and the 973-Project.
* Corresponding author at: School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China. Tel.: +86 010 5880 8811.
E-mail addresses: wangfy@bnu.edu.cn, F.Y.Wang@swansea.ac.uk (F.-Y. Wang).

0304-4149/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2009.07.001

to [15], is indeed equivalent to the corresponding curvature condition. See e.g. [2] for equivalent statements on heat kernel functional inequalities; see also [8,3,7] for a parabolic Harnack inequality using the dimension–curvature condition by shifting time, which goes back to the classical local parabolic Harnack inequality of Moser [9].

Recently, some sharp gradient estimates have been derived in [11,18] for the Dirichlet semigroup on relatively compact domains. More precisely, for $V = 0$ and a relatively compact open C^2 domain D, the Dirichlet heat semigroup P_t^D satisfies

$$|\nabla P_t^D f|(x) \le C(x,t) P_t^D f(x), \quad x \in D,\ t > 0, \tag{1.1}$$

for some locally bounded function $C\colon D\times]0,\infty[\to]0,\infty[$ and all $f \in \mathscr{B}_b^+$, the space of bounded non-negative measurable functions on M. Obviously, this implies the Harnack inequality

$$P_t^D f(x) \le \tilde{C}(x,y,t) P_t^D f(y), \quad t > 0,\ x, y \in D,\ f \in \mathscr{B}_b^+, \tag{1.2}$$

for some function $\tilde{C}\colon M^2\times]0,\infty[\to]0,\infty[$. The purpose of this paper is to establish inequalities analogous to (1.1) and (1.2) globally on the whole manifold M.

On the other hand however, both (1.1) and (1.2) are, in general, wrong for P_t in place of P_t^D. A simple counter-example is already the standard heat semigroup on $\mathbb{R}^d$. Hence, we turn to search for the following slightly weaker version of gradient estimate:

$$\begin{aligned} |\nabla P_t f(x)| &\le \delta \left[P_t(f \log f) - P_t f \log P_t f \right](x) + \frac{C(\delta,x)}{t\wedge 1} P_t f(x), \\ & x \in M,\ t > 0,\ \delta > 0,\ f \in \mathscr{B}_b^+, \end{aligned} \tag{1.3}$$

for some positive function $C\colon]0,\infty[\times M \to]0,\infty[$. When $\mathrm{Ric} - \mathrm{Hess}_V$ is bounded below, this kind of gradient estimate follows from [2, Proposition 2.6] but is new without curvature conditions. In particular, it implies the Harnack inequality with power introduced in [14] (see Theorem 1.2).

Theorem 1.1. *There exists a continuous positive function F on $]0,1]\times M$ such that*

$$\begin{aligned} |\nabla P_t f(x)| \le{} & \delta \left(P_t f \log f - P_t f \log P_t f \right)(x) \\ & + \left(F(\delta\wedge 1, x)\left(\frac{1}{\delta(t\wedge 1)} + 1 \right) + \frac{2\delta}{\mathrm{e}} \right) P_t f(x), \\ & \delta > 0,\ x \in M,\ t > 0,\ f \in \mathscr{B}_b^+. \end{aligned} \tag{1.4}$$

Theorem 1.2. *There exists a positive function $C \in C(]1,\infty[\times M^2)$ such that*

$$\begin{aligned} (P_t f(x))^\alpha &\le (P_t f^\alpha(y)) \exp\left\{ \frac{2(\alpha-1)}{\mathrm{e}} + \alpha C(\alpha,x,y)\left(\frac{\alpha\rho^2(x,y)}{(\alpha-1)(t\wedge 1)} + \rho(x,y) \right) \right\}, \\ & \alpha > 1,\ t > 0,\ x, y \in M,\ f \in \mathscr{B}_b^+, \end{aligned}$$

where ρ is the Riemannian distance on M. Consequently, for any $\delta > 2$ there exists a positive function $C_\delta \in C([0,\infty[\times M)$ such that the transition density $p_t(x,y)$ of P_t with respect to $\mu(\mathrm{d}x) := \mathrm{e}^{V(x)}\mathrm{d}x$, where $\mathrm{d}x$ is the volume measure, satisfies

$$p_t(x,y) \le \frac{\exp\left\{ -\rho(x,y)^2/(2\delta t) + C_\delta(t,x) + C_\delta(t,y) \right\}}{\sqrt{\mu(B(x,\sqrt{2t}))\mu(B(y,\sqrt{2t}))}}, \quad x, y \in M,\ t \in]0,1[.$$

Remark 1.1. According to the Varadhan asymptotic formula for short time behavior, one has $\lim_{t\to 0} 4t \log p_t(x, y) = -\rho(x, y)^2,\ x \neq y$. Hence, the above heat kernel upper bound is sharp for short time, as δ is allowed to approximate 2.

The paper is organized as follows: In Section 2 we provide a formula expressing P_t in terms of P_t^D and the joint distribution of (τ, X_τ), where X_t is the L-diffusion process and τ its hitting time to ∂D. Some necessary lemmas and technical results are collected. Proposition 2.5 is a refinement of a result in [18] to make the coefficient of $\rho(x, y)/t$ sharp and explicit. In Section 3 we use parallel coupling of diffusions together with Girsanov transformation to obtain a gradient estimate for Dirichlet heat semigroup. Finally, complete proofs of Theorems 1.1 and 1.2 are presented in Section 4.

To prove the indicated theorems, besides stochastic arguments, we make use of a local gradient estimate obtained in [11] for $V = 0$. For the convenience of the reader, we include a brief proof for the case with drift in the Appendix.

2. Some preparations

Let $X_s(x)$ be an L-diffusion process with starting point x and explosion time $\xi(x)$. For any bounded open C^2 domain $D \subset M$ such that $x \in D$, let $\tau(x)$ be the first hitting time of $X_s(x)$ at the boundary ∂D. We have

$$P_t f(x) = \mathbb{E}\left[f(X_t(x))\, 1_{\{t<\xi(x)\}}\right], \qquad P_t^D f(x) = \mathbb{E}\left[f(X_t(x))\, 1_{\{t<\tau(x)\}}\right].$$

Let $p_t^D(x, y)$ be the transition density of P_t^D with respect to μ.

We first provide a formula for the density $h_x(t, z)$ of $(\tau(x), X_{\tau(x)}(x))$ with respect to $\mathrm{d}t \otimes \nu(\mathrm{d}z)$, where ν is the measure on ∂D induced by $\mu(\mathrm{d}y) := \mathrm{e}^{V(y)}\mathrm{d}y$.

Lemma 2.1. *Let $K(z, x)$ be the Poisson kernel in D with respect to ν. Then*

$$h_x(t, z) = \int_D \left(-\partial_t p_t^D(x, y)\right) K(z, y)\, \mu(\mathrm{d}y). \tag{2.1}$$

Consequently, the density $s \mapsto \ell_x(s)$ of $\tau(x)$ satisfies the equation:

$$\ell_x(s) = \int_D \left(-\partial_t p_t^D(x, y)\right) \mu(\mathrm{d}y). \tag{2.2}$$

Proof. Every bounded continuous function $f : \partial D \to \mathbb{R}$ extends continuously to a function h on $\bar{D}$ which is harmonic in D and represented by

$$h(x) = \int_{\partial D} K(z, x) f(z)\, \nu(\mathrm{d}z).$$

Recall that $z \mapsto K(z, x)$ is the distribution density of $X_{\tau(x)}(x)$. Hence

$$\mathbb{E}[f(X_{\tau(x)}(x))] = h(x) = \int_{\partial D} K(z, x) f(z)\, \nu(\mathrm{d}z).$$

On the other hand, the identity

$$h(x) = \mathbb{E}[h(X_{t\wedge\tau(x)}(x))]$$

yields

$$\begin{aligned} h(x) &= \int_D p_t^D(x,y)h(y)\,\mu(\mathrm{d}y) + \int_{\partial D} \nu(\mathrm{d}z)\int_0^t h_x(s,z)f(z)\mathrm{d}s \\ &= \int_D p_t^D(x,y)\left(\int_{\partial D} K(z,y)f(z)\nu(\mathrm{d}z)\right)\mu(\mathrm{d}y) + \int_{\partial D}\nu(\mathrm{d}z)\int_0^t h_x(s,z)f(z)\mathrm{d}s \\ &= \int_{\partial D} f(z)\left(\int_D p_t^D(x,y)K(z,y)\,\mu(\mathrm{d}y) + \int_0^t h_x(s,z)\mathrm{d}s\right)\nu(\mathrm{d}z), \end{aligned}$$

which implies that

$$K(z,x) = \int_D p_t^D(x,y)K(z,y)\,\mu(\mathrm{d}y) + \int_0^t h_x(s,z)\mathrm{d}s. \tag{2.3}$$

Differentiating with respect to t gives

$$h_x(t,z) = -\partial_t \int_D p_t^D(x,y)K(z,y)\,\mu(\mathrm{d}y). \tag{2.4}$$

Since $\partial_t p_t^D(x,y)$ is bounded on $[\varepsilon,\varepsilon^{-1}]\times\bar{D}\times\bar{D}$ for any $\varepsilon\in]0,1[$, Eq. (2.1) follows by the dominated convergence.

Finally, Eq. (2.2) is obtained by integrating (2.1) with respect to $\nu(\mathrm{d}z)$. □

Lemma 2.2. *The following formula holds:*

$$\begin{aligned} P_t f(x) &= P_t^D f(x) + \int_{]0,t]\times\partial D} P_{t-s}f(z)h_x(s,z)\,\mathrm{d}s\,\nu(\mathrm{d}z) \\ &= P_t^D f(x) + \int_{]0,t]\times\partial D} P_{t-s}f(z)P_{s/2}^D h_\cdot(s/2,z)(x)\,\mathrm{d}s\,\nu(\mathrm{d}z). \end{aligned}$$

Proof. The first formula is standard due to the strong Markov property:

$$\begin{aligned} P_t f(x) &= \mathbb{E}\left[f(X_t(x))1_{\{t<\xi(x)\}}\right] = \mathbb{E}\left[f(X_t(x))1_{\{t<\tau(x)\}}\right] + \mathbb{E}\left[f(X_t(x))1_{\{\tau(x)<t<\xi(x)\}}\right] \\ &= P_t^D f(x) + \mathbb{E}\left[\mathbb{E}\left[f(X_t(x))1_{\{\tau(x)<t<\xi(x)\}}|(\tau(x),X_{\tau(x)}(x))\right]\right] \\ &= P_t^D f(x) + \int_{]0,t]\times\partial D} P_{t-s}f(z)h_x(s,z)\,\mathrm{d}s\;\nu(\mathrm{d}z). \end{aligned} \tag{2.5}$$

Next, since

$$\begin{aligned} \partial_s p_s^D(x,y) &= Lp_s^D(\cdot,y)(x) = LP_{s/2}^D p_{s/2}^D(\cdot,y)(x) \\ &= P_{s/2}^D(Lp_{s/2}^D(\cdot,y))(x) = P_{s/2}^D(\partial_u p_u^D(\cdot,y)|_{u=s/2})(x), \end{aligned}$$

it follows from (2.1) that

$$h_x(s,z) = P_{s/2}^D h_\cdot(s/2,z)(x). \tag{2.6}$$

This completes the proof. □

We remark that formula (2.6) can also be derived from the strong Markov property without invoking Eq. (2.1). Indeed, for any $u<s$ and any measurable set $A\subset\partial D$, the strong Markov

property implies that

$$\begin{aligned}\mathbb{P}\left\{\tau(x) > s,\ X_{\tau(x)}(x) \in A\right\} &= \mathbb{E}\left[\left(1_{\{u<\tau(x)\}}\right) \mathbb{P}\left\{\tau(x) > s,\ X_{\tau(x)}(x) \in A|\mathscr{F}_u\right\}\right] \\ &= \int_D p_u^D(x, y)\,\mathbb{P}\left\{\tau(y) > s-u,\ X_{\tau(y)}(y) \in A\right\} \mu(\mathrm{d}y),\end{aligned}$$

and thus,

$$h_x(s, z) = P_u^D h_\cdot(s-u, z)(x), \quad s > u > 0,\ x \in D,\ z \in \partial D.$$

Lemma 2.3. *Let D be a relatively compact open domain and $\rho_{\partial D}$ be the Riemannian distance to the boundary ∂D. Then there exists a constant $C > 0$ depending on D such that*

$$\mathbb{P}\{\tau(x) \le t\} \le C\mathrm{e}^{-\rho_{\partial D}^2(x)/16t}, \quad x \in D,\ t > 0.$$

Proof. For $x \in D$, let $R := \rho_{\partial D}(x)$ and ρ_x the Riemannian distance function to x. Since D is relatively compact, there exists a constant $c > 0$ such that $L\rho_x^2 \le c$ holds on D outside the cut-locus of x. Let $\gamma_t := \rho_x(X_t(x)),\ t \ge 0$. By Itô's formula, according to Kendall [6], there exists a one-dimensional Brownian motion b_t such that

$$\mathrm{d}\gamma_t^2 \le 2\sqrt{2}\gamma_t\,\mathrm{d}b_t + c\,\mathrm{d}t, \quad t \le \tau(x).$$

Thus, for fixed $t > 0$ and $\delta > 0$,

$$Z_s := \exp\left(\frac{\delta}{t}\gamma_s^2 - \frac{\delta}{t}cs - 4\frac{\delta^2}{t^2}\int_0^s \gamma_u^2 \mathrm{d}u\right), \quad s \le \tau(x)$$

is a supermartingale. Therefore,

$$\begin{aligned}\mathbb{P}\{\tau(x) \le t\} &= \mathbb{P}\left\{\max_{s\in[0,t]} \gamma_{s\wedge\tau(x)} \ge R\right\} \le \mathbb{P}\left\{\max_{s\in[0,t]} Z_{s\wedge\tau(x)} \ge \mathrm{e}^{\delta R^2/t-\delta c-4\delta^2R^2/t}\right\} \\ &\le \exp\left(c\delta - \frac{1}{t}(\delta R^2 - 4\delta^2 R^2)\right).\end{aligned}$$

The proof is completed by taking $\delta := 1/8$. □

Lemma 2.4. *On a measurable space $(E, \mathscr{F}, \tilde{\mu})$ satisfying $\tilde{\mu}(E) < \infty$, let $f \in L^1(\tilde{\mu})$ be non-negative with $\tilde{\mu}(f) > 0$. Then for every measurable function ψ such that $\psi f \in L^1(\tilde{\mu})$, there holds:*

$$\int_E \psi f\,\mathrm{d}\tilde{\mu} \le \int_E f \log\frac{f}{\tilde{\mu}(f)}\,\mathrm{d}\tilde{\mu} + \tilde{\mu}(f)\log\int_E \mathrm{e}^{\psi}\,\mathrm{d}\tilde{\mu}. \tag{2.7}$$

Proof. This is a direct consequence of [12] Lemma 6.45. We give a proof for completeness. Multiplying f by a positive constant, we can assume that $\tilde{\mu}(f) = 1$. If $\int_E \mathrm{e}^{\psi}\,\mathrm{d}\tilde{\mu} = \infty$, then (2.7) is clearly satisfied.

If $\int_E \mathrm{e}^{\psi}\,\mathrm{d}\tilde{\mu} < \infty$, then since $\int_E \mathrm{e}^{\psi}\,\mathrm{d}\tilde{\mu} \ge \int_{\{f>0\}} \mathrm{e}^{\psi}\,\mathrm{d}\tilde{\mu}$, we can assume that $f > 0$ everywhere. Now from the fact that $\mathrm{e}^{\psi}\frac{1}{f} \in L^1(f\tilde{\mu})$, we can apply Jensen's inequality to obtain

$$\log\left(\int_E \mathrm{e}^{\psi}\,\mathrm{d}\tilde{\mu}\right) = \log\left(\int_E \mathrm{e}^{\psi}\frac{1}{f}\,f\mathrm{d}\tilde{\mu}\right) \ge \int_E \log\left(\mathrm{e}^{\psi}\frac{1}{f}\right)\,f\mathrm{d}\tilde{\mu}$$

(note the right-hand-side belongs to $\mathbb{R} \cup \{-\infty\}$). To finish we remark that since $\psi f \in L^1(\tilde{\mu})$,

$$\int_E \log\left(e^{\psi}\frac{1}{f}\right) f\,d\tilde{\mu} = \int_E \psi f\,d\tilde{\mu} - \int_E f \log f\,d\tilde{\mu}. \quad \square$$

Finally, in order to obtain precise gradient estimate of the type (1.4), where the constant in front of $\rho(x, y)/t$ is explicit and sharp, we establish the following revision of [18, Theorem 2.1].

Proposition 2.5. *Let D be a relatively compact open C^2 domain in M and K a compact subset of D. For any $\varepsilon > 0$, there exists a constant $C(\varepsilon) > 0$ such that*

$$|\nabla \log p_t^D(\cdot, y)(x)| \le \frac{C(\varepsilon)\log(1+t^{-1})}{\sqrt{t}} + \frac{(1+\varepsilon)\rho(x, y)}{2t},$$
$$t \in]0, 1[,\ x \in K,\ y \in D. \tag{2.8}$$

In addition, if D is convex, the above estimate holds for $\varepsilon = 0$ and some constant $C(0) > 0$.

Proof. Since $\delta := \min_K \rho_{\partial D} > 0$, it suffices to deal with the case where $0 < t \le 1 \wedge \delta$. To this end, we combine the argument in [18] with relevant results from [16,17]. Let $t \in (0, 1 \wedge \delta]$, $t_0 = t/2$ and $y \in D$ be fixed, and take

$$f(x, s) = p_{s+t_0}^D(x, y), \quad x \in D,\ s > 0.$$

(a) Applying Theorem A.1 of the Appendix to the cube

$$Q := B(x, \rho_{\partial D}(x)) \times [s - \rho_{\partial D}(x)^2/2, s] \subset D \times [-t_0, t_0], \quad s \le t_0,$$

we obtain

$$|\nabla \log f(x, s)| \le \frac{c_0}{\rho_{\partial D}(x)}\left(1 + \log\frac{A}{f(x, s)}\right), \quad s \le t_0, \tag{2.9}$$

where $A := \sup_Q f$ and $c_0 > 0$ is a constant depending on the dimension and curvature on D. By [7, Theorem 5.2],

$$A \le c_1 f\left(x, s + \rho_{\partial D}(x)^2\right), \quad s \in]0, 1],\ x \in D, \tag{2.10}$$

holds for some constant $c_1 > 0$ depending on D and L. Moreover, by the boundary Harnack inequality of [4] (which treats $Z = 0$ but generalizes easily to non-zero C^1 drift Z),

$$f\left(x, s + \rho_{\partial D}(x)^2\right) \le c_2 f(x, s), \quad s \in]0, 1],\ x \in D, \tag{2.11}$$

for some constant $c_2 > 0$ depending on D and L. Combining (2.9)–(2.11), there exists a constant $c > 0$ depending on D and L such that

$$|\nabla \log f(x, s)| \le \frac{c}{\sqrt{s}}, \quad x \in D,\ s \in]0, t_0] \text{ with } \rho_{\partial D}(x)^2 \le s. \tag{2.12}$$

(b) Let

$$\Omega = \left\{(x, s) : x \in D,\ s \in [0, t_0],\ \rho_{\partial D}(x)^2 \ge s\right\}$$

and $B = \sup_\Omega f$. Since $\partial_s f = Lf$, for any constant $b \ge 1$, we have

$$(L - \partial_s)\left(f \log\frac{bB}{f}\right) = -\frac{|\nabla f|^2}{f}.$$

Next, again by $\partial_s f = Lf$ and the Bochner–Weizenböck formula,

$$(L - \partial_s)\frac{|\nabla f|^2}{f} \geq -2k\frac{|\nabla f|^2}{f},$$

where $k \geq 0$ is such that $\mathrm{Ric} - \nabla Z \geq -k$ on D. Then the function

$$h := \frac{s|\nabla f|^2}{(1+2ks)f} - f\log\frac{bB}{f}$$

satisfies

$$(L - \partial_s)h \geq 0 \quad \text{on } D\times]0, \infty[. \tag{2.13}$$

Obviously $h(\cdot, 0) \leq 0$, and (2.12) yields $h(x, s) \leq 0$ for $s = \rho_{\partial D}(x)^2$ provided the constant b is large enough. Then the maximum principle and inequality (2.13) imply $h \leq 0$ on Ω. Thus,

$$|\nabla \log f(x, s)|^2 \leq (2k + s^{-1})\log\frac{bB}{f}, \qquad (x, s) \in \Omega. \tag{2.14}$$

(c) If D is convex, by [16, Theorem 2.1] with $\delta = \sqrt{t}$ and $t = 2t_0$, we obtain (note the generator therein is $\frac{1}{2}L$)

$$f(x, t_0) = p^D_{2t_0}(x, y) = p^D_{2t_0}(y, x) \geq c_1\varphi(y)\, t_0^{-d/2}\mathrm{e}^{-\rho(x,y)^2/8t_0}, \quad x \in K,\ y \in D$$

for some constant $c_1 > 0$, where $\varphi > 0$ is the first Dirichlet eigenfunction of L on D. On the other hand, the intrinsic ultracontractivity for P^D_t implies (see e.g. [10])

$$f(z, s) = p^D_{s+t_0}(z, y) \leq c_2\,\varphi(y)\, t_0^{-(d+2)/2}, \quad z, y \in D,\ s \leq t_0,$$

for some constant $c_2 > 0$ depending on D, K and L. Combining these estimates we obtain

$$\frac{B}{f(x, s)} \leq c_3\, t_0^{-1}\mathrm{e}^{\rho(x,y)^2/8t_0}, \quad x \in K,\ s \leq t_0,$$

for some constant $c_3 > 0$ depending on D, K and L. Hence by (2.14) for $s = t_0$ we get the existence of a constant $C > 0$ such that

$$|\nabla \log p^D_{2t_0}(\cdot, y)|^2 \leq (t_0^{-1} + 2k)\left(C + \log t_0^{-1} + \frac{\rho(x, y)^2}{8t_0}\right)$$

for all $y \in D$, $x \in K$ and $t_0 \in]0, 1[$ with $t_0 \leq \rho_{\partial D}(x)^2$. This completes the proof by noting that $t = 2t_0$.

(d) Finally, if D is not convex, then there exists a constant $\sigma > 0$ such that

$$\langle \nabla_X N, X\rangle \geq -\sigma|X|^2, \quad X \in T\partial D,$$

where N is the outward unit normal vector field of ∂D, and $T\partial D$ is the set of all vector fields tangent to ∂D. Let $\psi \in C^\infty(\bar{D})$ such that $\psi = 1$ for $\rho_{\partial D} \geq \varepsilon$, $1 \leq \psi \leq \mathrm{e}^{2\varepsilon\sigma}$ for $\rho_{\partial D} \leq \varepsilon$, and $N\log\psi|_{\partial D} \geq \sigma$. By Lemma 2.1 in [17], ∂D is convex under the metric $\tilde{g} := \psi^{-2}\langle\cdot,\cdot\rangle$. Let $\tilde{\Delta}$, $\tilde{\nabla}$ and $\tilde{\rho}$ be respectively the Laplacian, the gradient and the Riemannian distance induced by $\tilde{g}$. By Lemma 2.2 in [17],

$$L := \Delta + \nabla V = \psi^{-2}\left[\tilde{\Delta} + (d-2)\psi\nabla\psi\right] + \nabla V.$$

Since D is convex under $\tilde{g}$, as explained in the first paragraph in Section 2 of [17],

$$\tilde{g}(\tilde{\nabla}\tilde{\rho}(y,\cdot),\tilde{\nabla}\varphi)|_{\partial D}<0,$$

so that

$$\tilde{\sigma}(y):=\sup_D \tilde{g}(\tilde{\nabla}\tilde{\rho}(y,\cdot),\tilde{\nabla}\varphi)<\infty,\qquad y\in D.$$

Hence, repeating the proof of Theorem 2.1 in [16], but using $\tilde{\rho}$ and $\tilde{\nabla}$ in place of ρ and ∇ respectively, and taking into account that $\psi\to 1$ uniformly as $\varepsilon\to 0$, we obtain

$$\begin{aligned}p^D_{2t_0}(x,y)&\ge C_1(\varepsilon)\varphi(y)t_0^{-d/2}\mathrm{e}^{-C_2(\varepsilon)\tilde{\rho}(x,y)^2/8t_0}\\&\ge C_1(\varepsilon)\varphi(y)t_0^{-d/2}\mathrm{e}^{-C_2(\varepsilon)C_3(\varepsilon)\rho(x,y)^2/8t_0}\end{aligned}$$

for some constants $C_1(\varepsilon), C_2(\varepsilon), C_3(\varepsilon)>1$ with $C_2(\varepsilon), C_3(\varepsilon)\to 1$ as $\varepsilon\to 0$. Hence the proof is completed. □

3. Gradient estimate for Dirichlet heat semigroup using coupling of diffusion processes

Proposition 3.1. *Let D be a relatively compact C^2 domain in M. For every compact subset K of D, there exists a constant $C=C(K,D)>0$ such that for all $\delta>0$, $t>0$, $x_0\in K$ and for all bounded positive functions f on M,*

$$|\nabla P^D_t f(x_0)|\le \delta P^D_t\left(f\log\left(\frac{f}{P^D_t f(x_0)}\right)\right)(x_0)+C\left(\frac{1}{\delta(t\wedge 1)}+1\right)P^D_t f(x_0).\qquad(3.1)$$

Proof. We assume that $t\in]0,1[$, the other case will be treated at the very end of the proof.

We write $\nabla V=Z$ so that $L=\Delta+Z$. Since P^D_t only depends on the Riemannian metric and the vector field Z on the domain D, by modifying the metric and Z outside of D we may assume that $\mathrm{Ric}-\nabla Z$ is bounded below (see e.g. [13]); that is,

$$\mathrm{Ric}-\nabla Z\ge -\kappa\qquad(3.2)$$

for some constant $\kappa\ge 0$.

Fix $x_0\in K$. Let f be a positive bounded function on M and X_s a diffusion with generator L, starting at x_0. For fixed $t\le 1$, let

$$v=\frac{\nabla P^D_t f(x_0)}{|\nabla P^D_t f(x_0)|}$$

and denote by $u\mapsto\varphi(u)$ the geodesics in M satisfying $\dot{\varphi}(0)=v$. Then

$$\left.\frac{\mathrm{d}}{\mathrm{d}u}\right|_{u=0}P^D_t f(\varphi(u))=\left|\nabla P^D_t f(x_0)\right|.$$

To formulate the coupling used in [1], we introduce some notations.

If Y is a semimartingale in M, we denote by $\mathrm{d}Y$ its Itô differential and by $\mathrm{d}_m Y$ the martingale part of $\mathrm{d}Y$: in local coordinates,

$$\mathrm{d}Y=\left(\mathrm{d}Y^i+\frac{1}{2}\Gamma^i_{jk}(Y)\,\mathrm{d}\langle Y^j,Y^k\rangle\right)\frac{\partial}{\partial x^i}$$

where Γ^i_{jk} are the Christoffel symbols of the Levi–Civita connection; if $\mathrm{d}Y^i = \mathrm{d}M^i + \mathrm{d}A^i$ where M^i is a local martingale and A^i a finite variation process, then

$$\mathrm{d}_m Y = \mathrm{d}M^i \frac{\partial}{\partial x^i}.$$

Alternatively, if $Q(Y)\colon T_{Y_0}M \to T_{Y_\cdot}M$ is the parallel translation along Y, then

$$\mathrm{d}Y_t = Q(Y)_t\,\mathrm{d}\left(\int_0^\cdot Q(Y)_s^{-1} \circ \mathrm{d}Y_s\right)_t$$

and

$$\mathrm{d}_m Y_t = Q(Y)_t\,\mathrm{d}N_t$$

where N_t is the martingale part of the Stratonovich integral $\int_0^t Q(Y)_s^{-1} \circ \mathrm{d}Y_s$.

For $x, y \in M$, and y not in the cut-locus of x, let

$$I(x,y) = \sum_{i=1}^{d-1} \int_0^{\rho(x,y)} \left(|\nabla_{\dot{e}(x,y)} J_i|^2 + \left\langle R(\dot{e}(x,y), J_i)J_i + \nabla_{\dot{e}(x,y)} Z, \dot{e}(x,y) \right\rangle \right)_s \mathrm{d}s \qquad (3.3)$$

where $\dot{e}(x,y)$ is the tangent vector of the unit speed minimal geodesic $e(x,y)$ and $(J_i)_{i=1}^d$ are Jacobi fields along $e(x,y)$ which together with $\dot{e}(x,y)$ constitute an orthonormal basis of the tangent space at x and y:

$$J_i(\rho(x,y)) = P_{x,y} J_i(0), \quad i = 1, \ldots, d-1;$$

here $P_{x,y}\colon T_xM \to T_yM$ is the parallel translation along the geodesic $e(x,y)$.

Let $c \in]0,1[$. For $h > 0$ but smaller than the injectivity radius of D, and $t > 0$, let X^h be the semimartingale satisfying $X_0^h = \varphi(h)$ and

$$\mathrm{d}X_s^h = P_{X_s, X_s^h} \mathrm{d}_m X_s + Z(X_s^h)\,\mathrm{d}s + \xi_s^h \mathrm{d}s, \qquad (3.4)$$

where

$$\xi_s^h := \left(\frac{h}{ct} + \kappa h\right) n(X_s^h, X_s)$$

with $n(X_s^h, X_s)$ the derivative at time 0 of the unit speed geodesic from X_s^h to X_s, and $P_{X_s, X_s^h}\colon T_{X_s}M \to T_{X_s^h}M$ the parallel transport along the minimal geodesic from X_s to X_s^h. By convention, we put $n(x,x) = 0$ and $P_{x,x} = \mathrm{Id}$ for all $x \in M$.

By the second variational formula and (3.2) (cf. [1]), we have

$$\mathrm{d}\rho(X_s, X_s^h) \le \left\{ I(X_s, X_s^h) - \frac{h}{ct} - \kappa h \right\} \mathrm{d}s \le -\frac{h}{ct}\,\mathrm{d}s, \quad s \le T_h,$$

where $T_h := \inf\{s \ge 0 : X_s = X_s^h\}$. Thus, (X_s, X_s^h) never reaches the cut-locus. In particular, $T_h \le ct$ and

$$X_s = X_s^h, \quad s \ge ct. \qquad (3.5)$$

Moreover, we have $\rho(X_s, X_s^h) \le h$ and

$$|\xi_s^h|^2 \le h^2 \left(\kappa + \frac{1}{ct}\right)^2. \qquad (3.6)$$

We want to compensate the additional drift of X^h by a change of probability. To this end, let

$$M_s^h = -\int_0^{s\wedge ct} \left\langle \xi_r^h, P_{X_r, X_r^h}\, \mathrm{d}_m X_r \right\rangle,$$

and

$$R_s^h = \exp\left(M_s^h - \frac{1}{2}[M^h]_s\right).$$

Clearly R^h is a martingale, and under $\mathbb{Q}^h = R^h \cdot \mathbb{P}$, the process X^h is a diffusion with generator L.

Letting $\tau(x_0)$ (resp. τ^h) be the hitting time of ∂D by X (resp. by X^h), we have

$$1_{\{t<\tau^h\}} \le 1_{\{t<\tau(x_0)\}} + 1_{\{\tau(x_0)\le t<\tau^h\}}.$$

But, since $X_s^h = X_s$ for $s \ge ct$, we obtain

$$1_{\{\tau(x_0)\le t<\tau^h\}} = 1_{\{\tau(x_0)\le ct\}} 1_{\{t<\tau^h\}}.$$

Consequently,

$$\begin{aligned}\frac{1}{h}\left(P_t^D f(\varphi(h)) - P_t^D f(x_0)\right) &= \frac{1}{h}\,\mathbb{E}\left[f(X_t^h) R_t^h 1_{\{t<\tau^h\}} - f(X_t(0)) 1_{\{t<\tau(x_0)\}}\right] \\ &\le \frac{1}{h}\,\mathbb{E}\left[f(X_t^h) R_t^h 1_{\{t<\tau(x_0)\}} - f(X_t(0)) 1_{\{t<\tau(x_0)\}}\right] + \frac{1}{h}\,\mathbb{E}\left[f(X_t^h) R_t^h 1_{\{\tau(x_0)\le ct\}} 1_{\{t<\tau^h\}}\right],\end{aligned}$$

and since $X_t^h = X_t$ this yields

$$\begin{aligned}\frac{1}{h}\left(P_t^D f(\varphi(h)) - P_t^D f(x_0)\right) &\le \mathbb{E}\left[f(X_t) 1_{\{t<\tau(x_0)\}} \frac{1}{h}(R_t^h - 1)\right] \\ &\quad + \frac{1}{h}\,\mathbb{E}\left[f(X_t^h) R_t^h 1_{\{\tau(x_0)\le ct\}} 1_{\{t<\tau^h\}}\right]. \end{aligned} \tag{3.7}$$

The left hand side converges to the quantity to be evaluated as h goes to 0. Hence, it is enough to find appropriate lim sup's for the two terms of the right hand side. We begin with the first term. Letting

$$Y_s^h = \left| M_s^h - \frac{1}{2}[M^h]_s \right|$$

and noting that $\langle n(X_r^h, X_r), P_{X_r, X_r^h} \mathrm{d}_m X_r\rangle = \sqrt{2}\,\mathrm{d}b_r$ up to the coupling time T_h for some one-dimensional Brownian motion b_r, we have

$$\begin{aligned}R_t^h &= \exp\left(M_t^h - \frac{1}{2}[M^h]_t\right) \le 1 + M_t^h - \frac{1}{2}[M^h]_t + (Y_t^h)^2 \exp(Y_t^h) \\ &= 1 + M_t^h - \int_0^t |\xi_s^h|^2 \mathrm{d}s + (Y_t^h)^2 \exp(Y_t^h).\end{aligned}$$

From the assumptions, $\exp(Y_t^h)$ and Y_t^h/h have all their moments bounded, uniformly in $h > 0$. Consequently, since f is bounded,

$$\limsup_{h\to 0} \mathbb{E}\left[f(X_t) 1_{\{t<\tau(x_0)\}} \frac{1}{h}\left(\int_0^t |\xi_r^h|^2\,\mathrm{d}r + (Y_t^h)^2 \exp(Y_t^h)\right)\right] = 0,$$

which implies

$$\limsup_{h\to 0}\mathbb{E}\left[f(X_t)1_{\{t<\tau(x_0)\}}\frac{1}{h}(R_t^h-1)\right]$$
$$\le \limsup_{h\to 0}\mathbb{E}\left[f(X_t)1_{\{t<\tau(x_0)\}}\frac{1}{h}\int_0^s\left\langle \xi_r^h, P_{X_r,X_r^h}\,\mathrm{d}_m X_r\right\rangle\right].$$

Using Lemma 2.4 and estimate (3.6), we have for $\delta>0$

$$\mathbb{E}\left[f(X_t)1_{\{t<\tau(x_0)\}}\frac{1}{h}\int_0^s\left\langle \xi_r^h, P_{X_r,X_r^h}\mathrm{d}_m X_r\right\rangle\right]\le \delta P_t^D\left(f\log\left(\frac{f}{P_t^D f(x_0)}\right)\right)(x_0)$$
$$+\delta P_t^D f(x_0)\log\mathbb{E}\left[1_{\{t<\tau(x_0)\}}\exp\left(\frac{1}{\delta h}\int_0^{ct}\left\langle \xi_s^h, P_{X_s,X_s^h}\mathrm{d}_m X_s\right\rangle\right)\right]$$
$$\le \delta P_t^D\left(f\log\left(\frac{f}{P_t^D f(x_0)}\right)\right)(x_0)+\delta P_t^D f(x_0)\log\mathbb{E}\left[\exp\left(\frac{1}{\delta^2 h^2}\int_0^{ct}\left|\xi_s^h\right|^2\mathrm{d}s\right)\right]$$
$$\le \delta P_t^D\left(f\log\left(\frac{f}{P_t^D f(x_0)}\right)\right)(x_0)+\delta P_t^D f(x_0)\frac{ct}{\delta^2}\left(\frac{1}{c^2t^2}+\kappa^2\right)$$
$$\le \delta P_t^D\left(f\log\left(\frac{f}{P_t^D f(x_0)}\right)\right)(x_0)+\frac{C'}{c\delta t}P_t^D f(x_0),$$

where $C'=1+(c\kappa)^2$ (recall that $t\le 1$). Since the last expression is independent of h, this proves that

$$\limsup_{h\to 0}\mathbb{E}\left[f(X_t)1_{\{t<\tau(x_0)\}}\frac{1}{h}(R_t^h-1)\right]$$
$$\le \delta P_t^D\left(f\log\left(\frac{f}{P_t^D f(x_0)}\right)\right)(x_0)+\frac{C'}{c\delta t}P_t^D f(x_0). \tag{3.8}$$

We are now going to estimate lim sup of the second term in Eq. (3.7). By the strong Markov property, we have

$$\mathbb{E}\left[f(X_t^h)R_t^h 1_{\{\tau(x_0)\le ct\}}1_{\{t<\tau^h\}}\right]=\mathbb{E}_{\mathbb{Q}^h}\left[P_{t-ct}^D f(X_{ct}^h)1_{\{\tau(x_0)\le ct<\tau^h\}}\right]$$
$$\le \|P_{t-ct}^D f\|_\infty\,\mathbb{Q}^h\left\{\tau(x_0)\le ct<\tau^h\right\}. \tag{3.9}$$

Since $\rho(X_s^h, X_s)\le h\frac{ct-s}{ct}$ for $s\in[0,ct]$, we have on $\{\tau(x_0)\le ct<\tau^h\}$:

$$\rho_{\partial D}(X_{\tau(x_0)}^h)\le h\frac{ct-\tau(x_0)}{ct}.$$

For $s\in[0,\tau^h-\tau(x_0)]$, define

$$Y_s'=\rho(X_{\tau(x_0)+s}^h,\partial D),$$

and for fixed small $\varepsilon>0$ (but $\varepsilon>h$), let $S'=\inf\{s\ge 0,\ Y_s'=\varepsilon \text{ or } Y_s'=0\}$. Since under $\mathbb{Q}^h$ the process X_s^h is generated by L, the drift of $\rho(X_s^h,\partial D)$ is $L\rho(\cdot,\partial D)$ which is bounded in a neighborhood of ∂D. Thus, for a sufficiently small $\varepsilon>0$, there exists a $\mathbb{Q}^h$-Brownian motion β started at 0, and a constant $N>0$ such that

$$Y_s:=h\frac{ct-\tau(x_0)}{ct}+\sqrt{2}\beta_s+Ns\ge Y_s',\quad s\in[0,S'].$$

Let

$$S = \inf\{u \geq 0,\ Y_u = \varepsilon \text{ or } Y_u = 0\}.$$

Taking into account that on $\{\tau(x_0) = u\}$,

$$\{Y'_{S'} = \varepsilon\} \cup \{S' > ct - u\} \subset \{Y_S = \varepsilon\} \cup \{S > ct - u\},$$

we have for $u \in [0, ct]$,

$$\begin{aligned}\mathbb{Q}^h\left\{ct < \tau^h | \tau(x_0) = u\right\} &\leq \mathbb{Q}^h\{Y_{S'} = \varepsilon | \tau(x_0) = u\} + \mathbb{Q}^h\left\{S' \geq ct - u | \tau(x_0) = u\right\}\\ &\leq \mathbb{Q}^h\{Y_S = \varepsilon | \tau(x_0) = u\} + \mathbb{Q}^h\{S \geq ct - u | \tau(x_0) = u\}\\ &\leq \mathbb{Q}^h\{Y_S = \varepsilon | \tau(x_0) = u\} + \frac{1}{ct - u}\mathbb{E}_{\mathbb{Q}^h}[S | \tau(x_0) = u].\end{aligned}$$

Now using the fact that e^{-NY_s} is a martingale and $Y_s^2 - 2s$ a submartingale, we get

$$\mathbb{Q}^h\{Y_S = \varepsilon | \tau(x_0) = u\} = \frac{1 - \mathrm{e}^{-Nh\frac{ct-u}{ct}}}{1 - \mathrm{e}^{-N\varepsilon}} \leq C_1 h$$

and

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}^h}[S | \tau(x_0) = u] &\leq \mathbb{E}_{\mathbb{Q}^h}\left[Y_S^2 | \tau(x_0) = u\right]\\ &\leq \varepsilon^2\, \mathbb{Q}^h\{Y_S = \varepsilon | \tau(x_0) = u\}\\ &= \varepsilon^2 \frac{1 - \mathrm{e}^{-Nh\frac{ct-u}{ct}}}{1 - \mathrm{e}^{-N\varepsilon}} \leq C_2\, \frac{h(ct - u)}{ct}\end{aligned}$$

for some constants $C_1, C_2 > 0$. Thus,

$$\begin{aligned}\mathbb{Q}^h\left\{ct < \tau^h | \tau(x_0) = u\right\} &\leq C_1 h + \frac{1}{ct - u}\, C_2\, \frac{h(ct - u)}{ct}\\ &\leq C_1 h + C_3 \frac{h}{ct} \leq C_4 \frac{h}{t}\end{aligned}$$

for some constants $C_3, C_4 > 0$ (recall that $t \leq 1$). Denoting by ℓ^h the density of $\tau(x_0)$ under $\mathbb{Q}^h$, this implies

$$\begin{aligned}\mathbb{Q}^h\left\{\tau(x_0) \leq ct < \tau^h\right\} &= \int_0^{ct} \ell^h(u)\, \mathbb{Q}^h\{ct < \tau^h | \sigma^h = u\}\, \mathrm{d}u\\ &\leq C_4 \frac{h}{t} \int_0^{ct} \ell^h(u)\, \mathrm{d}u\\ &= C_4 \frac{h}{t}\, \mathbb{Q}^h\{\tau(x_0) \leq ct\}.\end{aligned}$$

In terms of $D^{-h} = \{x \in D,\ \rho_{\partial D}(x) > h\}$ and $\sigma^h = \inf\{s > 0,\ X_s^h \in \partial D^{-h}\}$, we have $\sigma^h \leq \tau(x_0)$ a.s. Hence, by Lemma 2.3,

$$\mathbb{Q}^h\{\tau(x_0) \leq ct\} \leq \mathbb{Q}^h\left\{\sigma^h \leq ct\right\} \leq C \exp\left\{-\frac{\rho_{\partial D^{-h}}(\varphi(h))}{16ct}\right\},$$

where we used that X_s^h is generated by L under $\mathbb{Q}^h$. This implies

$$\mathbb{Q}^h\left\{\tau(x_0) \leq ct < \tau^h\right\} \leq C_5 \frac{h}{t} \exp\left\{-\frac{\rho_{\partial D^{-h}}(\varphi(h))}{16ct}\right\}. \tag{3.10}$$

Since $\frac{1}{h}\left(P_t^D(\varphi(h)) - P_t^D(x_0)\right)$ converges to $|\nabla P_t^D f(x_0)|$, we obtain from (3.7)–(3.10),

$$|\nabla P_t^D f(x_0)| \le \delta P_t^D \left(f \log \left(\frac{f}{P_t^D f(x_0)} \right) \right)(x_0) + \frac{C'}{c\delta t} P_t^D f(x_0) + C_5 \, \|P_{t-ct}^D f\|_\infty \frac{1}{t} \exp\left\{ -\frac{\rho_{\partial D}(x_0)}{16ct} \right\}. \tag{3.11}$$

Finally, as explained in steps (c) and (d) of the proof of Proposition 2.5, for any compact set $K \subset D$, there exists a constant $C(K, D) > 0$ such that

$$\|P_{t-ct}^D f\|_\infty \le e^{C(K,D)/t} P_t^D f(x_0), \quad c \in [0, 1/2], \; x_0 \in K, \; t \in]0, 1].$$

Combining this with (3.11), we arrive at

$$|\nabla P_t^D f(x_0)| \le \delta P_t^D \left(f \log \left(\frac{f}{P_t^D f(x_0)} \right) \right)(x_0) + \frac{C'}{c\delta t} P_t^D f(x_0) + C_5 \frac{1}{t} \exp\left\{ -\frac{\rho_{\partial D}(x_0)}{16ct} \right\} \exp\left\{ \frac{C(K, D)}{t} \right\} P_t^D f(x_0). \tag{3.12}$$

Finally, choosing c such that

$$0 < c < \frac{1}{2} \wedge \frac{\mathrm{dist}(K, \partial D)}{16C(K, D)},$$

we get for some constant $C > 0$,

$$|\nabla P_t^D f(x_0)| \le \delta P_t^D \left(f \log \left(\frac{f}{P_t^D f(x_0)} \right) \right)(x_0) + C \left(\frac{1}{\delta t} + 1 \right) P_t^D f(x_0), \quad x_0 \in K, \; \delta > 0, \tag{3.13}$$

which implies the desired inequality.

To finish we consider the case $t > 1$. From the semigroup property, we have $P_t^D f = P_1^D(P_{t-1}^D f)$. So letting $g = P_{t-1}^D f$ and applying (3.13) to g at time 1, we obtain

$$|\nabla P_t^D f(x_0)| \le \delta P_1^D \left(g \log \left(\frac{g}{P_1^D g(x_0)} \right) \right)(x_0) + C \left(\frac{1}{\delta} + 1 \right) P_1^D g(x_0).$$

Now using $P_1^D g = P_t^D f$, we get

$$|\nabla P_t^D f(x_0)| \le \delta P_1^D (g \log g)(x_0) - P_t^D f(x_0) \log P_t^D f(x_0) + C \left(\frac{1}{\delta} + 1 \right) P_t^D f(x_0).$$

Letting $\varphi(x) = x \log x$, we have for $z \in D$

$$\begin{aligned} g \log g(z) &= \varphi\left(\mathbb{E}\left[f(X_{t-1}(z)) 1_{\{t-1<\tau(z)\}} \right] \right) \\ &\le \mathbb{E}\left[\varphi\left(f(X_{t-1}(z)) 1_{\{t-1<\tau(z)\}} \right) \right] \\ &= \mathbb{E}\left[\varphi(f)(X_{t-1}(z)) 1_{\{t-1<\tau(z)\}} \right] \\ &= P_{t-1}^D (f \log f)(z), \end{aligned}$$

where we successively used the convexity of φ and the fact that $\varphi(0) = 0$. This implies

$$|\nabla P_t^D f(x_0)| \le \delta P_t^D \left(f \log \left(\frac{f}{P_t^D f(x_0)} \right) \right)(x_0) + C \left(\frac{1}{\delta} + 1 \right) P_t^D f(x_0),$$

which is the desired inequality for $t > 1$. □

4. Proof of Theorems 1.1 and 1.2

Proof of Theorem 1.1. We assume that $t \in]0, 1[$ and refer to the end of the proof of Proposition 3.1 for the case $t > 1$. Fixing $\delta > 0$ and $x_0 \in M$, we take $R = 160/(\delta \wedge 1)$. Let D be a relatively compact open domain with C^2 boundary containing $B(x_0, 2R)$ and contained in $B(x_0, 2R + \varepsilon)$ for some small $\varepsilon > 0$. By the countable compactness of M, it suffices to prove that there exists a constant $C = C(D)$ such that (1.4) holds on $B(x_0, R)$ with C in place of $F(\delta \wedge 1, x_0)$. We now fix $x \in B(x_0, R)$, $t \in]0, 1]$ and $f \in \mathscr{B}_b^+$. Without loss of generality, we may and will assume that $P_t f(x) = 1$.

(a) Let $P_s(x, \mathrm{d}y)$ be the transition kernel of the L-diffusion process, and for $x \in D$, $z \in M$, let

$$\nu_s(x, \mathrm{d}z) = \int_{\partial D} h_x(s/2, y)\, P_{t-s}(y, \mathrm{d}z)\, \nu(\mathrm{d}y),$$

where ν is the measure on ∂D induced by $\mu(\mathrm{d}y) = \mathrm{e}^{V(y)}\mathrm{d}y$. By Lemma 2.2 we have

$$P_t f(x) = P_t^D f(x) + \int_{]0,t]\times D\times M} p_{s/2}^D(x, y)\, f(z)\, \mathrm{d}s\mu(\mathrm{d}y)\nu_s(y, \mathrm{d}z).$$

Then

$$\begin{aligned}|\nabla P_t f(x)| \le\ & |\nabla P_t^D f(x)| \\ & + \int_{]0,t]\times D\times M} |\nabla \log p_{s/2}^D(\cdot, y)(x)|\, p_{s/2}^D(x, y) f(z)\, \mathrm{d}s\mu(\mathrm{d}y)\nu_s(y, \mathrm{d}z) \\ =:\ & I_1 + I_2. \end{aligned} \tag{4.1}$$

(b) By Proposition 3.1, we have

$$I_1 \le \delta P_t^D(f \log f)(x) + \frac{\delta}{\mathrm{e}} + C\left(\frac{1}{\delta t} + 1\right), \quad x \in B(x_0, R),\ t \in]0, 1[,\ \delta > 0 \tag{4.2}$$

for some $C = C(D) > 0$.

(c) By Proposition 2.5 with $\varepsilon = 1$, we have

$$I_2 \le \int_{]0,t]\times M\times D} \left[\frac{C \log(\mathrm{e} + s^{-1})}{\sqrt{s}} + \frac{2\rho(x, y)}{s}\right] p_{s/2}^D(x, y)\, f(z)\, \mathrm{d}s\nu_s(y, \mathrm{d}z)\mu(\mathrm{d}y) \tag{4.3}$$

for some $C = C(D) > 0$ and all $t \in]0, 1]$. Applying Lemma 2.4 to the measure $\tilde{\mu} := p_{s/2}^D(x, y)\, \mathrm{d}s\, \nu_s(y, \mathrm{d}z)\mu(\mathrm{d}y)$ on $E :=]0, t] \times M \times D$ so that

$$\tilde{\mu}(E) = \mathbb{P}(\tau(x) \le t < \xi(x)) \le 1,$$

we obtain

$$\begin{aligned} I_2 \le\ & \delta\, \mathbb{E}\left[(f \log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] + \frac{\delta}{\mathrm{e}} + \delta\mathbb{E}\left[f(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] \\ & \times \log \int_{]0,t]\times M\times D} \exp\left\{\frac{C \log(\mathrm{e} + s^{-1})}{\delta\sqrt{s}} + \frac{2\rho(x, y)}{s\delta}\right\} \mathrm{d}s\, p_{s/2}^D(x, y)\nu_s(y, \mathrm{d}z)\, \mu(\mathrm{d}y) \\ \le\ & \delta\mathbb{E}\left[(f \log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] + \frac{\delta}{\mathrm{e}} + \delta\mathbb{E}\left[f(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] \\ & \times \log \int_{]0,t]\times M\times D} \exp\left\{\frac{A}{\delta} + \frac{9R}{s\delta}\right\} \mathrm{d}s\, p_{s/2}^D(x, y)\nu_s(y, \mathrm{d}z)\, \mu(\mathrm{d}y), \end{aligned} \tag{4.4}$$

where

$$A := \sup_{r>0} \left\{C\sqrt{r}\log(\mathrm{e}+r) - r\right\} < \infty.$$

We get

$$\begin{aligned}
I_2 &\le \delta\mathbb{E}\left[(f\log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] + \frac{\delta}{\mathrm{e}} \\
&\quad + \delta\mathbb{E}\left[f(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right]\left(\log\mathbb{E}\left[\exp\left(9R/\delta\tau(x)\right)\right] + \frac{A}{\delta}\right) \\
&\le \delta\mathbb{E}\left[(f\log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] + \frac{\delta}{\mathrm{e}} + \delta\log\mathbb{E}\left[\exp\left(9R/\delta\tau(x)\right)\right] + A \\
&\le \delta\mathbb{E}\left[(f\log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] + \delta\log\mathbb{E}\left[\exp\left(\frac{9R}{(\delta\wedge 1)\tau(x)}\right)^{\frac{\delta\wedge 1}{\delta}}\right] + A + \frac{\delta}{\mathrm{e}} \\
&= \delta\mathbb{E}\left[(f\log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] \\
&\quad + (\delta\wedge 1)\log\mathbb{E}\left[\exp\left(\frac{9R}{(\delta\wedge 1)\tau(x)}\right)\right] + A + \frac{\delta}{\mathrm{e}}.
\end{aligned} \tag{4.5}$$

By Lemma 2.3 and noting that $\rho_\partial(x) \ge R$, we have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\frac{9R}{(\delta\wedge 1)\tau(x)}\right)\right] \le 1 + \mathbb{E}\left[\frac{9R}{(\delta\wedge 1)\tau(x)}\exp\left(\frac{9R}{(\delta\wedge 1)\tau(x)}\right)\right] \\
&= 1 + \int_0^\infty \frac{9Rs}{(\delta\wedge 1)}\exp\left(\frac{9Rs}{(\delta\wedge 1)}\right)\frac{\mathrm{d}}{\mathrm{d}s}\left(-\mathbb{P}\{\tau(x)\le s^{-1}\}\right)\mathrm{d}s \\
&= 1 + \frac{9R}{(\delta\wedge 1)}\int_0^\infty\left(\frac{9R}{(\delta\wedge 1)}s + 1\right)\exp\left(\frac{9Rs}{(\delta\wedge 1)}\right)\mathbb{P}\{\tau(x)\le s^{-1}\}\,\mathrm{d}s \\
&\le 1 + \frac{9R}{(\delta\wedge 1)}\int_0^\infty\left(\frac{9R}{(\delta\wedge 1)}s + 1\right)\exp\left(\frac{9Rs}{(\delta\wedge 1)}\right)\exp\left(\frac{-R^2 s}{16}\right)\mathrm{d}s \\
&= 1 + \frac{9R}{(\delta\wedge 1)}\int_0^\infty\left(\frac{9R}{(\delta\wedge 1)}s + 1\right)\exp\left(\frac{-Rs}{(\delta\wedge 1)}\right)\mathrm{d}s \\
&= 1 + 9\int_0^\infty (9u+1)\exp(-u)\,\mathrm{d}u =: A',
\end{aligned}$$

since $R = 160/(\delta\wedge 1)$. This along with (4.5) yields

$$I_2 \le \delta\,\mathbb{E}\left[(f\log f)(X_t(x))1_{\{\tau(x)\le t<\xi(x)\}}\right] + \log A' + A + \frac{\delta}{\mathrm{e}}. \tag{4.6}$$

The proof is completed by combining (4.6) with (4.1), (4.2) and (4.4). □

Proof of Theorem 1.2. By Theorem 1.1,

$$\begin{aligned}
|\nabla P_t f(x)| &\le \delta\left(P_t(f\log f)(x) - (P_t f)(x)\log P_t f(x)\right) \\
&\quad + \left(F(\delta\wedge 1, x)\left(\frac{1}{\delta(t\wedge 1)} + 1\right) + \frac{2\delta}{e}\right)P_t f(x), \quad \delta > 0,\ x \in M.
\end{aligned} \tag{4.7}$$

For $\alpha > 1$ and $x \neq y$, let $\beta(s) = 1 + s(\alpha - 1)$ and let $\gamma: [0, 1] \to M$ be the minimal geodesic from x to y. Then $|\dot\gamma| = \rho(x, y)$. Applying (4.7) with $\delta = \frac{\alpha-1}{\alpha\rho(x,y)}$, we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\log(P_t f^{\beta(s)})^{\alpha/\beta(s)}(\gamma_s) &= \frac{\alpha(\alpha-1)}{\beta(s)^2}\,\frac{P_t(f^{\beta(s)}\log f^{\beta(s)}) - (P_t f^{\beta(s)})\log P_t f^{\beta(s)}}{P_t f^{\beta(s)}}(\gamma_s)\\
&\quad + \frac{\alpha}{\beta(s)}\,\frac{\langle \nabla P_t f^{\beta(s)}, \dot\gamma_s\rangle}{P_t f^{\beta(s)}}(\gamma_s)\\
&\ge \frac{\alpha\rho(x,y)}{\beta(s) P_t f^{\beta(s)}(\gamma_s)}\left\{\frac{\alpha-1}{\alpha\rho(x,y)}\left(P_t(f^{\beta(s)}\log f^{\beta(s)}) - (P_t f^{\beta(s)})\log P_t f^{\beta(s)}\right)(\gamma_s)\right.\\
&\quad \left. - |\nabla P_t f^{\beta(s)}(\gamma_s)|\right\}\\
&\ge -F\left(\frac{\alpha-1}{\alpha\rho(x,y)}\wedge 1, \gamma_s\right)\left(\frac{\alpha^2\rho^2(x,y)}{\beta(s)(\alpha-1)(t\wedge 1)} + \frac{\alpha\rho(x,y)}{\beta(s)}\right) - \frac{2(\alpha-1)}{\mathrm{e}\beta(s)}\\
&\ge -C(\alpha,x,y)\left(\frac{\alpha\rho^2(x,y)}{(\alpha-1)(t\wedge 1)} + \rho(x,y)\right) - \frac{2(\alpha-1)}{\mathrm{e}}
\end{aligned}$$

where $C(\alpha, x, y) := \sup_{s\in[0,1]} \frac{1}{\alpha} F\left(\frac{\alpha-1}{\alpha\rho(x,y)} \wedge 1, \gamma_s\right)$. This implies the desired Harnack inequality.

Next, for fixed $\alpha \in]1, 2[$, let

$$K(\alpha, t, x) = \sup\left\{C(\alpha, x, y) : y \in B(x, \sqrt{2t})\right\}, \quad t > 0,\ x \in M.$$

Note $K(\alpha, t, x)$ is finite and continuous in $(\alpha, t, x) \in]1, 2[\times]0, 1[\times M$. Let $p := 2/\alpha$. For fixed $t \in]0, 1[$, the Harnack inequality gives for $y \in B(x, \sqrt{2t})$,

$$(P_t f(x))^2 \le (P_t f^\alpha(y))^p \exp\left\{\frac{2(2-p)}{\mathrm{e}} + 2K(\alpha, t, x)\left(\frac{2\alpha}{\alpha-1} + \sqrt{2t}\right)\right\}.$$

Then, choosing $T > t$ such that $q := p/2(p-1) < T/t$,

$$\begin{aligned}
&\mu\left(B(x,\sqrt{2t})\right)\exp\left\{-\frac{2(2-p)}{\mathrm{e}} - 2K(\alpha,t,x)\left(\frac{2\alpha}{\alpha-1}+\sqrt{2t}\right) - \frac{t}{T-qt}\right\}(P_t f(x))^2\\
&\quad \le \int_{B(x,\sqrt{2t})} (P_t f^\alpha(y))^p \exp\left\{-\frac{\rho(x,y)^2}{2(T-qt)}\right\}\mu(\mathrm{d}y).
\end{aligned}$$

Similarly to the proof of [1, Corollary 3], we obtain that for any $\delta > 2$, choosing $\alpha = \frac{2\delta}{2+\delta} \in]1, 2[$ such that $\delta > \frac{2}{2-\alpha} = \frac{p}{p-1} > 2$, there is a constant $c(\delta) > 0$ such that the following estimate holds:

$$\begin{aligned}
E_\delta(x,t) &:= \int_M p_t(x,y)^2 \exp\left\{\frac{\rho(x,y)^2}{\delta t}\right\}\mu(\mathrm{d}y)\\
&\le \frac{\exp\left\{c(\delta)K(\alpha,t,x)(1+\sqrt{2t})\right\}}{\mu(B(x,\sqrt{2t}))}, \quad t > 0,\ x \in M.
\end{aligned}$$

By [5, Eq. (3.4)], this implies the desired heat kernel upper bound for $C_\delta(t, x) := c(\delta)K(\alpha, t, x)(1 + \sqrt{2t})$. □

Appendix

The aim of the Appendix is to explain that the arguments in Souplet–Zhang [11] and Zhang [18] for gradient estimates of solutions to heat equations work as well in the case with drift.

Theorem A.1. *Let $L = \Delta + Z$ for a C^1 vector field Z. Fix $x_0 \in M$ and $R,\ T,\ t_0 > 0$ such that $B(x_0, R) \subset M$. Assume that*

$$Ric - \nabla Z \geq -K \tag{A.1}$$

on $B(x_0, R)$. There exists a constant c depending only on d, the dimension of the manifold, such that for any positive solution u of

$$\partial_t u = Lu \tag{A.2}$$

on $Q_{R,T} := B(x_0, R) \times [t_0 - T, t_0]$, the estimate

$$|\nabla \log u| \leq c \left(\frac{1}{R} + T^{-1/2} + \sqrt{K} \right) \left(1 + \log \frac{\sup_{Q_{R,T}} u}{u} \right)$$

holds on $Q_{R/2,T/2}$.

Proof. Without loss of generality, let $N := \sup_{Q_{T,R}} u = 1$; otherwise replace u by u/N. Let $f = \log u$ and $\omega = \frac{|\nabla f|^2}{(1-f)^2}$. By (A.2) we have

$$Lf + |\nabla f|^2 - \partial_t f = 0$$

so that

$$\begin{aligned}
\partial_t \omega &= \frac{2\langle \nabla f, \nabla \partial_t f \rangle}{(1-f)^2} + \frac{2\,|\nabla f|^2 \partial_t f}{(1-f)^3} \\
&= \frac{2\langle \nabla f, \nabla (Lf + |\nabla f|^2) \rangle}{(1-f)^2} + \frac{2\,|\nabla f|^2 (Lf + |\nabla f|^2)}{(1-f)^3} \\
&= \frac{2\langle \nabla f, \nabla (\Delta f + |\nabla f|^2) \rangle}{(1-f)^2} + \frac{2\,|\nabla f|^2 (\Delta f + |\nabla f|^2)}{(1-f)^3} \\
&\quad + \frac{2\langle \nabla_{\nabla f} Z, \nabla f \rangle + 2\,\mathrm{Hess}_f(\nabla f, Z)}{(1-f)^2} + \frac{2\,|\nabla f|^2 \langle Z, \nabla f \rangle}{(1-f)^3}.
\end{aligned} \tag{A.3}$$

Moreover,

$$\begin{aligned}
L\omega &= \Delta\omega + \frac{\langle Z, \nabla |f|^2 \rangle}{(1-f)^2} + \frac{2|\nabla f|^2 \langle Z, \nabla f \rangle}{(1-f)^3} \\
&= \Delta\omega + \frac{2\,\mathrm{Hess}_f(\nabla f, Z)}{(1-f)^2} + \frac{2\,|\nabla f|^2 \langle Z, \nabla f \rangle}{(1-f)^3}.
\end{aligned} \tag{A.4}$$

Finally, by the proof of [11, (2.9)] with $-k$ replaced by $\mathrm{Ric}(\nabla f, \nabla f)/|\nabla f|^2$, we obtain

$$\Delta\omega - \left\{ \frac{2\,\langle \nabla f, \nabla (\Delta f + |\nabla f|^2) \rangle}{(1-f)^2} + \frac{2\,|\nabla f|^2 (\Delta f + |\nabla f|^2)}{(1-f)^3} \right\}$$

$$\geq \frac{2f}{1-f}\langle \nabla f, \nabla \omega \rangle + 2(1-f)\omega^2 + \frac{2\omega \operatorname{Ric}(\nabla f, \nabla f)}{|\nabla f|^2}. \tag{A.5}$$

Combining (A.1) and (A.3)–(A.5), we arrive at

$$L\omega - \partial_t \omega \geq \frac{2f}{1-f}\langle \nabla f, \nabla \omega \rangle + 2(1-f)\omega^2 - 2K\omega.$$

This implies the desired estimate by the Li-Yau cut-off argument as in [11]; the only difference is, using the notation in [11], in the calculation of $-(\Delta\psi)\omega$ after Eq. 2.13 in [11]. By (A.1) and the generalized Laplacian comparison theorem (see [3, Theorem 4.2]), we have

$$Lr \leq \sqrt{Kd}\coth\left(\sqrt{K/d}\, r\right) \leq \frac{d}{r} + \sqrt{Kd},$$

and then

$$-(L\psi)\omega = -(\partial_r^2 \psi + (\partial_r \psi) Lr)\omega \leq \left(|\partial_r \psi|^2 + |\partial_r \psi|\frac{d}{r} + \sqrt{Kd}\,|\partial_r \psi|\right)\omega.$$

The remainder of the proof is the same as in the proof of [11, Theorem 1.1], using $L\psi$ in place of $\Delta\psi$.

References

[1] M. Arnaudon, A. Thalmaier, F.-Y. Wang, Harnack inequality and heat kernel estimates on manifolds with curvature unbounded below, Bull. Sci. Math. 130 (2006) 223–233.
[2] D. Bakry, On Sobolev and logarithmic Sobolev inequalities for Markov semigroups, in: K.D. Elworthy, S. Kusuoka, I. Shigekawa (Eds.), New Trends in Stochastic Analysis, World Sci. Publ., River Edge, NJ, 1997, pp. 43–75.
[3] D. Bakry, Z. Qian, Harnack inequalities on a manifold with positive or negative Ricci curvature, Rev. Math. Iberoamericana 15 (1999) 143–179.
[4] E.B. Fabes, N. Garofalo, S. Salsa, A backward Harnack inequality and Fatou theorem for nonnegative solutions of parabolic equations, Illinois J. Math. 30 (1986) 536–565.
[5] A. Grigor'yan, Gaussian upper bounds for the heat kernel on arbitrary manifolds, J. Differential Geom. 45 (1997) 33–52.
[6] W.S. Kendall, The radial part of Brownian motion on a manifold: A semimartingale property, Ann. Probab. 15 (1987) 1491–1500.
[7] X.-D. Li, Liouville theorems for symmetric diffusion operators on complete Riemannian manifolds, J. Math. Pures Appl. 84 (2005) 1295–1361.
[8] P. Li, S.-T. Yau, On the parabolic kernel of the Schrödinger operator, Acta Math. 156 (1986) 153–201.
[9] J. Moser, A Harnack inequality for parabolic differential equations, Comm. Pure Appl. Math. 17 (1964) 101–134; Correction: Comm. Pure Appl. Math. 20 (1967) 231–236.
[10] E.M. Ouhabaz, F.-Y. Wang, Sharp estimates for intrinsic ultracontractivity on $C^{1,\alpha}$-domains, Manu. Math. 122 (2007) 229–244.
[11] P. Souplet, Qi S. Zhang, Sharp gradient estimate and Yau's Liouville theorem for the heat equation on noncompact manifolds, Bull. London Math. Soc. 38 (2006) 1045–1053.
[12] D.W. Stroock, An Introduction to the Analysis of Paths on a Riemannian Manifold, in: Mathematical Surveys and Monographs, vol. 74, American Mathematical Society, 1991.
[13] A. Thalmaier, F.-Y. Wang, Gradient estimates for harmonic functions on regular domains in Riemannian manifolds, J. Funct. Anal. 155 (1998) 109–124.
[14] F.-Y. Wang, Logarithmic Sobolev inequalities on noncompact Riemannian manifolds, Probab. Theory Related Fields 109 (1997) 417–424.
[15] F.-Y. Wang, Equivalence of dimension-free Harnack inequality and curvature condition, Integral Equ. Oper. Theory 48 (2004) 547–552.
[16] F.-Y. Wang, Estimates of Dirichlet heat kernels, Stoch. Proc. Appl. 74 (1998) 217–234.
[17] F.-Y. Wang, Estimates of the first Neumann eigenvalue and the log-Sobolev constant on non-convex manifolds, Math. Nachr. 280 (2007) 1431–1439.
[18] Qi S. Zhang, Some gradient estimates for the heat equation on domains and for an equation by Perelman, Int. Math. Res. Not. (2006), pp. 39 (Article ID 92314).