INTEGRAL VARADHAN FORMULA FOR NON-LINEAR HEAT FLOW

SHIN-ICHI OHTA AND KOHEI SUZUKI

Abstract. We prove the integral Varadhan short-time formula for non-linear heat flow on measured Finsler manifolds. To the best of the authors' knowledge, this is the first result establishing a Varadhan-type formula for non-linear semigroups. We do not assume the reversibility of the metric, and the distance function can be asymmetric. Within this generality, we reveal that the probabilistic interpretation is well-suited for our formula; the probability that a particle starting from a set A can be found in another set B describes the distance from A to B. One side of the estimates (the upper bound of the probability) is also established in the nonsmooth setting of infinitesimally strictly convex metric measure spaces satisfying the local Sobolev-to-Lipschitz property.

1. Introduction

Aim/Motivation. The main aim of this article is to draw more attention to (geometric) analysis of *non-linear heat flow*. The linear theory had been developed intensively and extensively in connection with two powerful theories: *Dirichlet forms* related to probability theory and the Γ-*calculus* à la Bakry et al related to differential geometry as well as geometric analysis. A non-linear analogue to the Γ-calculus has been investigated on Finsler manifolds (of Ricci curvature bounded below in an appropriate way) by the first author and Sturm [Oht17a, Oht17b, Oht21, Oht22, OS14]. Then it is natural to expect a more general theory of non-linear heat semigroups as a non-linear counterpart to the theory of Dirichlet forms, however, there is surprisingly no result in such a direction. In this article, to motivate further studies of non-linear heat semigroups, we establish the integral Varadhan short-time formula for non-linear heat flow on Finsler manifolds. A large part of our discussion can be generalised to metric measure spaces under mild assumptions.

Background. On a complete Riemannian manifold (M, g) with the Riemannian distance d, let $p_t(x, y)$ be the heat kernel density, i.e., the fundamental solution to the heat equation $\partial_t u = \frac{1}{2}\Delta u$. From the probabilistic viewpoint, $p_t(x, y)$ is the density function of the transition probability of the Brownian motion in M. The fundamental *Varadhan short-time formula* [Var67] states that the short-time behaviour of $p_t(x, y)$ is governed by $\mathsf{d}(x, y)$ in the following way:

$$\lim_{t\downarrow 0} t \log p_t(x, y) = -\frac{1}{2}\mathsf{d}(x, y)^2 \ . \tag{1.1}$$

Date: March 6, 2024.

The first author is supported in part by JSPS Grant-in-Aid for Scientific Research (KAKENHI) 19H01786, 22H04942. The second author gratefully acknowledges funding by the Alexander von Humboldt Stiftung.

The formula (1.1), linking geometry, analysis, and probability, has been studied in various settings including complete connected Riemannian manifolds [Var67], Lipschitz manifolds [Nor97], degenerate diffusions on Euclidean spaces [CKS87], sub-Riemannian manifolds [Léa87a, Léa87b, BN18], and metric measure spaces satisfying the quasi Riemannian curvature-dimension condition [DS22b].

For spaces not admitting the heat kernel density $p_t(x, y)$, the formula (1.1) has been generalised as an integrated one as follows:

$$\lim_{t\downarrow 0} t \log \mathsf{P}_t(A, B) = -\frac{1}{2}\bar{\mathsf{d}}_{\mathsf{m}}(A, B)^2 \tag{1.2}$$

for $A, B \subset \mathsf{M}$ with $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$, where m is the reference measure on M,

$$\mathsf{P}_t(A, B) := \int_A \mathsf{T}_t \mathbf{1}_B \, \mathrm{d}\mathsf{m}$$

with the L^2-heat semigroup $(\mathsf{T}_t)_{t\geq 0}$, and $\bar{\mathsf{d}}_{\mathsf{m}}(A, B)$ is a suitable distance-like function. In the case of *linear* heat semigroups, (1.2) has been established in a general setting of local Dirichlet spaces [HR03, AH05, HM18], where $\bar{\mathsf{d}}_{\mathsf{m}}(A, B)$ is induced from a local Dirichlet form. If a local Dirichlet space admits a distance function in the domain of the Dirichlet form in a compatible way in the sense of the Rademacher-type property and the Sobolev-to-Lipschitz property, the set function $\bar{\mathsf{d}}_{\mathsf{m}}(A, B)$ is indeed realised as the distance between A and B; see [DS21] for details. We refer the readers to the following articles for particular spaces: the Wiener space and path/loop groups [Fan94, FZ99, AK01, AZ02, HR03], the configuration space [Zha01, DS22a], and the Wasserstein space [RS09].

Main results. By extending the argument in [HR03, AH05], we shall generalise (1.2) to *non-linear* heat flow $(\mathsf{T}_t)_{t\geq 0}$ on a Finsler manifold (M, F) equipped with a measure m. We do not assume the *reversibility* of F (i.e., $F(-v) \neq F(v)$ is allowed), thereby the distance function d can be asymmetric. In this case, $\bar{\mathsf{d}}_{\mathsf{m}}(A, B)$ is defined as

$$\bar{\mathsf{d}}_{\mathsf{m}}(A, B) := \sup_{f\in\mathbb{L}} \Big\{ \operatorname*{ess\,inf}_{x\in A} f(x) - \operatorname*{ess\,sup}_{y\in B} f(y) \Big\} \tag{1.3}$$

for measurable sets $A, B \subset \mathsf{M}$ with $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$, where

$$\mathbb{L} := \{ f \in H^1_{\mathrm{loc}}(\mathsf{M}) \cap L^\infty(\mathsf{M}) : F^*(-\mathbf{d}f) \leq 1 \text{ a.e.} \} \, . \tag{1.4}$$

We remark that $\bar{\mathsf{d}}_{\mathsf{m}}(A, B) < \infty$. The condition $F^*(-\mathbf{d}f) \leq 1$ roughly means that $-f$ is 1-Lipschitz, and one can regard that $\bar{\mathsf{d}}_{\mathsf{m}}(A, B)$ represents the distance *from A to B*. We refer to Subsection 2.1 for precise definitions and notations in Finsler geometry. We also set $\mathsf{d}(A, B) := \inf_{x\in A,\, y\in B} \mathsf{d}(x, y)$.

Theorem 1.1. *Let (M, F) be a C^∞-Finsler manifold equipped with a C^∞-measure m on M with $\mathsf{m}(\mathsf{M}) < \infty$. Assume that the uniform convexity and smoothness constants are finite. Then, for any measurable sets $A, B \subset \mathsf{M}$ with $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$, we have*

$$\lim_{t\downarrow 0} t \log \mathsf{P}_t(A, B) = -\frac{1}{2}\bar{\mathsf{d}}_{\mathsf{m}}(A, B)^2 \, . \tag{1.5}$$

Moreover, for any open sets $A, B \subset \mathsf{M}$ with $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$, we have

$$\lim_{t\downarrow 0} t \log \mathsf{P}_t(A, B) = -\frac{1}{2}\mathsf{d}(A, B)^2 \, .$$

To the best of our knowledge, Theorem 1.1 is the first result establishing the integral Varadhan formula for non-linear semigroups. It is unclear whether the pointwise Varadhan estimate (1.1) can be properly formulated in our setting, since there is no concept of heat kernel density $p_t(x,y)$ for non-linear heat semigroups.

The asymmetry of the distance function d reveals the probabilistic nature of our Varadhan formula, which is not apparent in the symmetric setting. From the analytic (PDE) point of view, the semigroup $\mathsf{T}_t\mathbf{1}_B$ in $\mathsf{P}_t(A,B) = \int_A \mathsf{T}_t\mathbf{1}_B \,\mathrm{d}\mathsf{m}$ could be regarded as describing the heat propagation from B, thereby the appearance of the distance $\mathsf{d}(A,B)$ from A to B may be counter-intuitive. From the probabilistic viewpoint (which is dual to the PDE one), however, $\mathsf{d}(A,B)$ is natural since $\mathsf{T}_t\mathbf{1}_B(x)$ represents the probability that a Brownian motion starting from x lives in B at time t (in the Riemannian setting, to be precise; the existence of the Brownian motion is unknown in the Finsler case).

The above observation should be compared with the fact that heat flow is regarded as the gradient flow of the relative entropy in the L^2-Wasserstein space with respect to the reverse Finsler structure $\overline{F}(v) = F(-v)$ (see Remark 2.2). Since $\mathsf{d}(A,B)$ coincides with the distance from B to A with respect to $\overline{F}$, the analytic point of view seems consistent with $\overline{F}$.

The upper estimate in Theorem 1.1 is more flexible than the lower estimate, and can be generalised to the nonsmooth setting as follows, thanks to differential calculus developed in [AGS14, Gig18]. We remark that reversible Finsler manifolds (satisfying $F(-v) = F(v)$) also fall into this framework.

Theorem 1.2. *Let* $(\mathsf{X},\mathsf{d},\mathsf{m})$ *be an infinitesimally strictly convex metric measure space such that the Sobolev space* $W^{1,2}(\mathsf{X})$ *is reflexive. Then, for any measurable sets* $A, B \subset \mathsf{X}$ *with* $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$*, we have*

$$\lim_{t\downarrow 0} t \log \mathsf{P}_t(A,B) \le -\frac{1}{2}\bar{\mathsf{d}}_{\mathsf{m}}(A,B)^2 \ .$$

Furthermore, if the local Sobolev-to-Lipschitz property holds, then we have

$$\bar{\mathsf{d}}_{\mathsf{m}}(A,B) = \mathsf{d}_{\mathsf{m}}(A,B)$$

for any open sets $A, B \subset \mathsf{X}$ *with* $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$*.*

The proof strategy of Theorems 1.1, 1.2 follows the lines of [HR03, AH05]. In our non-linear setting, however, there are difficulties arising from the fact that the energy form is not bilinear, due to the lack of the Leibniz rule for the gradient operator. We, therefore, need to carefully avoid using the Leibniz rule and adapt the proofs of [HR03, AH05] to our setting. Another difficulty we encounter is the lack of symmetry: $\int_{\mathsf{M}} u_1 \Delta u_2 \,\mathrm{d}\mathsf{m} \neq \int_{\mathsf{M}} u_2 \Delta u_1 \,\mathrm{d}\mathsf{m}$, which is relevant to the proof of the lower bound of (1.5). We employ a linearised heat semigroup to sort it out in the Finsler case. This approach is, however, not applicable to the non-smooth setting in Theorem 1.2 as we do not know if the linearised semigroup exists in metric measure spaces.

The structure of the paper. After reviewing the basics of Finsler geometry in Section 2, we study the behaviour of the function $\bar{\mathsf{d}}_{\mathsf{m}}(A,B)$ in Section 3. Then, we prove the upper bound estimate in Theorem 1.1 as well as Theorem 1.2 in Section 4, and Section 5 is devoted to the proof of the lower bound estimate in Theorem 1.1.

2. Preliminaries

We first review the basics of Finsler geometry, and then introduce truncation functions as in [HR03] playing an essential role in the lower estimate in Section 5.

2.1. **Finsler manifolds.** We refer the readers to [Oht21] for a concise description of the following contents, and also to [GS01, OS09, OS14] for the behaviour of heat flow on Finsler manifolds.

Let M be a connected C^∞-manifold without boundary of dimension $n \geq 2$. Given local coordinates $(x^i)_{i=1}^n$ on an open set $U \subset \mathsf{M}$, we will denote by $(x^i, v^j)_{i,j=1}^n$ the fibre-wise linear coordinates of the tangent bundle TU given by

$$v = \sum_{j=1}^n v^j \frac{\partial}{\partial x^j}\Big|_x \in T_x\mathsf{M} \ , \qquad x \in U \ .$$

We say that a nonnegative function $F : T\mathsf{M} \to [0, \infty)$ is a C^∞-*Finsler structure* on M if the following three conditions hold:

(1) (*Regularity*) F is C^∞ on $T\mathsf{M} \setminus \{0\}$;
(2) (*Positive* 1*-homogeneity*) $F(cv) = cF(v)$ for every $v \in T\mathsf{M}$ and $c \geq 0$;
(3) (*Strong convexity*) For every $v \in T\mathsf{M} \setminus \{0\}$, the following $n \times n$ matrix is positive-definite:

$$\big(g_{ij}(v)\big)_{i,j=1}^n := \left(\frac{1}{2}\frac{\partial^2[F^2]}{\partial v^i \partial v^j}(v)\right)_{i,j=1}^n \ . \tag{2.1}$$

We call a pair (M, F) a C^∞-*Finsler manifold*. We stress that the 1-homogeneity is imposed only in the positive direction, thereby $F(-v) \neq F(v)$ is allowed. If $F(v) = F(-v)$ for all $v \in T\mathsf{M}$, then we say that (M, F) is *reversible*. The matrix $(g_{ij}(v))$ in (2.1) provides an inner product g_v of $T_x\mathsf{M}$ by

$$g_v\left(\sum_{i=1}^n a_i \frac{\partial}{\partial x^i}, \sum_{j=1}^n b_j \frac{\partial}{\partial x^j}\right) := \sum_{i,j=1}^n a_i b_j g_{ij}(v) \ .$$

We will also make use of their counterparts in the dual space $T_x^*\mathsf{M}$:

$$F^*(\alpha) := \sup_{v \in T_x\mathsf{M},\, F(v)=1} \alpha(v) \qquad \text{for } \alpha \in T_x^*\mathsf{M} \ ,$$

$$g_{ij}^*(\alpha) := \frac{1}{2}\frac{\partial^2[(F^*)^2]}{\partial \alpha_i \partial \alpha_j}(\alpha) \qquad \text{for } \alpha = \sum_{i=1}^n \alpha_i \,\mathrm{d}x^i \in T_x^*\mathsf{M} \setminus \{0\} \ ,$$

$$g_\alpha^*\left(\sum_{i=1}^n a_i \,\mathrm{d}x^i, \sum_{j=1}^n b_j \,\mathrm{d}x^j\right) := \sum_{i,j=1}^n a_i b_j g_{ij}^*(\alpha) \qquad \text{on } T_x^*\mathsf{M} \ . \tag{2.2}$$

We remark that, though $\alpha(v) \leq F^*(\alpha)F(v)$ holds by definition, $\alpha(v) \geq -F^*(\alpha)F(v)$ does not hold in general due to the irreversibility of F (for example, one cannot replace the LHS of (4.4) with its absolute value).

For $x, y \in \mathsf{M}$, we define

$$\mathsf{d}(x, y) := \inf_\eta \int_0^1 F\big(\dot\eta(t)\big)\,\mathrm{d}t \ ,$$

where the infimum is taken over all piecewise C^1-curves $\eta : [0,1] \to \mathsf{M}$ with $\eta(0) = x$ and $\eta(1) = y$. Then d provides an *asymmetric* distance function on M, namely the *triangle inequality*

$$\mathsf{d}(x,z) \le \mathsf{d}(x,y) + \mathsf{d}(y,z) \quad \text{for all } x,y,z \in \mathsf{M}$$

holds but $\mathsf{d}(y,x)$ may be different from $\mathsf{d}(x,y)$. Note that d is symmetric if and only if F is reversible. We define the *reversibility constant* of (M, F) as

$$\Lambda_F := \sup_{v \in T\mathsf{M}\setminus\{0\}} \frac{F(v)}{F(-v)} = \sup_{x,y\in\mathsf{M},\, x\neq y} \frac{\mathsf{d}(x,y)}{\mathsf{d}(y,x)} \ . \tag{2.3}$$

Observe that $\Lambda_F \in [1,\infty]$ in general, and $\Lambda_F = 1$ holds only in the reversible case.

We say that (M, F) is *forward complete* if any closed, forward bounded set $A \subset \mathsf{M}$ (i.e., $\sup_{y\in A} \mathsf{d}(x,y) < \infty$ for some or, equivalently, any $x \in \mathsf{M}$) is compact. The *backward completeness* is defined as the forward completeness of $(\mathsf{M}, \overline{F})$, where $\overline{F}$ is the *reverse* Finsler structure given by $\overline{F}(v) := F(-v)$. If $\Lambda_F < \infty$, then the forward and backward completenesses are mutually equivalent and we may simply call it the *completeness*.

Uniform convexity and smoothness. We define the *uniform convexity* and *smoothness constants* of (M, F) as

$$\mathsf{C}_F := \sup_{x\in\mathsf{M}} \sup_{v,w\in T_x\mathsf{M}\setminus\{0\}} \frac{F(w)^2}{g_v(w,w)} \ , \qquad \mathsf{S}_F := \sup_{x\in\mathsf{M}} \sup_{v,w\in T_x\mathsf{M}\setminus\{0\}} \frac{g_v(w,w)}{F(w)^2} \ , \tag{2.4}$$

respectively. Since g_v comes from the Hessian of F^2, C_F (resp. S_F) actually measures the convexity (resp. concavity) of F^2 in tangent spaces. We have $\mathsf{C}_F, \mathsf{S}_F \in [1,\infty]$ in general, and $\mathsf{C}_F = 1$ or $\mathsf{S}_F = 1$ holds only in Riemannian manifolds (see [Oht21, Proposition 1.6]). On a compact Finsler manifold, C_F and S_F are finite thanks to the smoothness and the strong convexity of F. We also remark that their dual expressions are given by

$$\mathsf{C}_F = \sup_{x\in\mathsf{M}} \sup_{\alpha,\beta\in T_x^*\mathsf{M}\setminus\{0\}} \frac{g_\alpha^*(\beta,\beta)}{F^*(\beta)^2} \ , \qquad \mathsf{S}_F = \sup_{x\in\mathsf{M}} \sup_{\alpha,\beta\in T_x^*\mathsf{M}\setminus\{0\}} \frac{F^*(\beta)^2}{g_\alpha^*(\beta,\beta)} \ . \tag{2.5}$$

We refer to [Oht09], [Oht21, §8.3.2] for more discussions on C_F and S_F. For example, the reversibility constant Λ_F in (2.3) can be bounded by C_F and S_F as

$$\Lambda_F \le \min\left\{\sqrt{\mathsf{C}_F}\ ,\ \sqrt{\mathsf{S}_F}\right\} \ .$$

Gradient vectors. For a differentiable function $f : \mathsf{M} \to \mathbb{R}$, its *gradient vector* at $x \in \mathsf{M}$ is defined to be the Legendre transform of the derivative of f:

$$\nabla f(x) := \mathcal{L}^*\big(\mathbf{d}f(x)\big) \in T_x\mathsf{M} \ .$$

Here the *Legendre transform* $\mathcal{L}^* : T_x^*\mathsf{M} \to T_x\mathsf{M}$ maps $\alpha \in T_x^*\mathsf{M}$ to the unique element $v \in T_x\mathsf{M}$ such that $F(v) = F^*(\alpha)$ and $\alpha(v) = F^*(\alpha)^2$. We remark that $(g_{ij}(\mathcal{L}^*(\alpha)))$ is the inverse matrix of $(g_{ij}^*(\alpha))$, provided $\alpha \neq 0$. In local coordinates, we can write down

$$\nabla f(x) = \sum_{i,j=1}^n g_{ij}^*\big(\mathbf{d}f(x)\big) \frac{\partial f}{\partial x^j}(x) \frac{\partial}{\partial x^i}\Big|_x$$

(when $\mathbf{d}f(x) \neq 0$; while $\nabla f(x) = 0$ if $\mathbf{d}f(x) = 0$). We say that f is 1-*Lipschitz* if

$$f(y) - f(x) \le \mathsf{d}(x,y) \quad \text{for all } x,y \in \mathsf{M} \ ,$$

which is equivalent to $F(\nabla f) \le 1$ when f is differentiable. For example, given $x_0 \in \mathsf{M}$, the functions $x \mapsto \mathsf{d}(x_0, x)$ and $x \mapsto -\mathsf{d}(x, x_0)$ are 1-Lipschitz by the triangle inequality.

Remark 2.1 (Non-linearity of ∇)**.** As the differential $\mathbf{d}$ stems only from the differentiable structure of M, it is a linear operator and enjoys the chain and Leibniz rules. However, the Legendre transform $\mathcal{L}^*$ is non-linear ($\mathcal{L}^*(\alpha+\beta) \neq \mathcal{L}^*(\alpha) + \mathcal{L}^*(\beta)$) and irreversible ($\mathcal{L}^*(-\alpha) \neq \mathcal{L}^*(\alpha)$). Therefore, the gradient operator ∇ does not satisfy the Leibniz rule, and the chain rule holds only for non-decreasing functions:

$$\nabla(\varphi \circ f) = \mathcal{L}^*(\varphi' \circ f \cdot \mathbf{d}f) = \varphi' \circ f \cdot \nabla f \qquad \text{if } \varphi' \ge 0 \ ,$$

while we have $\nabla(\varphi \circ f) = -\varphi' \circ f \cdot \nabla(-f)$ if $\varphi' \le 0$. In the reversible case, the chain rule holds regardless of the sign of φ'.

Heat semigroup. Now, we fix a positive C^∞-measure m on M (in the sense that, in each local chart, $\mathsf{m} = \rho \, \mathrm{d}x^1 \cdots \mathrm{d}x^n$ for a positive C^∞-function ρ). Then the *divergence* of a differentiable vector field V on M with respect to m is defined in local coordinates as

$$\mathrm{div}_{\mathsf{m}} V := \sum_{i=1}^n \left(\frac{\partial V^i}{\partial x^i} + V^i \frac{\partial \psi}{\partial x^i} \right) ,$$

where $V = \sum_{i=1}^n V^i (\partial/\partial x^i)$ and we wrote $\mathsf{m} = e^\psi \, \mathrm{d}x^1 \cdots \mathrm{d}x^n$ in the coordinates. It can be generalised to measurable vector fields V in the distributional sense (against C^∞-functions of compact support) as

$$\int_{\mathsf{M}} \varphi \, \mathrm{div}_{\mathsf{m}} V \, \mathrm{d}\mathsf{m} = -\int_{\mathsf{M}} \mathbf{d}\varphi(V) \, \mathrm{d}\mathsf{m} \qquad \text{for all } \ \varphi \in C_c^\infty(\mathsf{M}) \ .$$

Then we define the *distributional Laplacian* $\Delta := \mathrm{div}_{\mathsf{m}} \circ \nabla$, i.e.,

$$\int_{\mathsf{M}} \varphi \Delta u \, \mathrm{d}\mathsf{m} := -\int_{\mathsf{M}} \mathbf{d}\varphi(\nabla u) \, \mathrm{d}\mathsf{m} \qquad \text{for all } \ \varphi \in C_c^\infty(\mathsf{M}) \ .$$

This Laplacian is again non-linear by the non-linearity of ∇. Moreover,

$$\int_{\mathsf{M}} u_1 \Delta u_2 \, \mathrm{d}\mathsf{m} \neq \int_{\mathsf{M}} u_2 \Delta u_1 \, \mathrm{d}\mathsf{m} \tag{2.6}$$

in general (in other words, $\mathbf{d}u_1(\nabla u_2) \neq \mathbf{d}u_2(\nabla u_1)$).

Define the *energy functional* $\mathcal{E}$ on $H^1_{\mathrm{loc}}(\mathsf{M})$ as

$$\mathcal{E}(u) := \frac{1}{2} \int_{\mathsf{M}} F(\nabla u)^2 \, \mathrm{d}\mathsf{m} = \frac{1}{2} \int_{\mathsf{M}} F^*(\mathbf{d}u)^2 \, \mathrm{d}\mathsf{m}$$

($H^1_{\mathrm{loc}}(\mathsf{M})$ is defined solely by the differentiable structure of M via local charts). The *Sobolev space* $H^1(\mathsf{M})$ is defined as the set of functions $u \in L^2(\mathsf{M}) \cap H^1_{\mathrm{loc}}(\mathsf{M})$ such that $\mathcal{E}(u) + \mathcal{E}(-u) < \infty$. Denote by $H^1_0(\mathsf{M})$ the closure of $C_c^\infty(\mathsf{M})$ with respect to the norm

$$\|u\|_{H^1} := \sqrt{\|u\|_{L^2}^2 + \mathcal{E}(u) + \mathcal{E}(-u)} \ .$$

Note that $(H^1(\mathsf{M}), \|\cdot\|_{H^1})$ is not a Hilbert space but a reflexive Banach space. If $\Lambda_F < \infty$ and (M, F) is complete, then we have $H^1_0(\mathsf{M}) = H^1(\mathsf{M})$ (see [Oht21, Lemma 11.4]).

The L^2*-heat semigroup* $u_t = \mathsf{T}_t f$ is defined as the solution to the Cauchy problem:

$$\partial_t u_t = \frac{1}{2} \Delta u_t \ , \qquad u_0 = f \ .$$

One can construct $(u_t)_{t\geq 0}$ as gradient flow of the energy $\mathcal{E}$ in $L^2(\mathsf{M})$, provided $\Lambda_F < \infty$. Precisely, we first construct $\mathsf{T}_t f$ for $f \in H_0^1(\mathsf{M})$ and then extend it to a contraction semigroup acting on $L^2(\mathsf{M})$ (see, e.g., [AGS08], [Oht21, §13.2]). Thanks to the regularizing effect as in [AGS08, Theorem 4.0.4], [AGS14, (4.26)], we have $\mathsf{T}_t L^2(\mathsf{M}) \subset H_0^1(\mathsf{M})$ for $t > 0$. We stress that the heat semigroup $(\mathsf{T}_t)_{t\geq 0}$ is non-linear. Indeed, $\mathsf{T}_t(f_1 + f_2) = \mathsf{T}_t f_1 + \mathsf{T}_t f_2$ does not hold, and $\mathsf{T}_t(cf) = c\mathsf{T}_t f$ holds only when $c \geq 0$ due to the irreversibility.

Remark 2.2. Besides the above interpretation of heat flow as gradient flow of the energy, one can also regard heat flow as gradient flow of the relative entropy in the L^2-Wasserstein space. In the Finsler case, however, we need to consider the Wasserstein space for the reverse Finsler structure $\overline{F}(v) = F(-v)$ (we refer to [OS09, OZ22] for details).

Linearised heat semigroups. In the last step of the lower estimate in Section 5, we will employ a linearisation of the heat semigroup to overcome a difficulty due to the asymmetry (2.6). See (5.10) in Lemma 5.7 below for the precise equation, here we recall a related result from [Oht21, §13.5].

Let $(u_t)_{t\geq 0}$ be a solution to the heat equation, and take a measurable one-parameter family $(V_t)_{t\geq 0}$ of nowhere vanishing vector fields such that $V_t(x) = \nabla u_t(x)$ for all x with $\mathbf{d}u_t(x) \neq 0$. Given $f \in H_0^1(\mathsf{M})$, there exists a (weak) solution $(f_t)_{t\geq 0} \subset H_0^1(\mathsf{M})$ to the *linearised heat equation*

$$\partial_t f_t = \frac{1}{2}\Delta^{\nabla u_t} f_t \;, \qquad f_0 = f \;, \tag{2.7}$$

where $\Delta^{\nabla u_t} := \mathrm{div}_{\mathsf{m}} \circ \nabla^{\nabla u_t}$ is the *linearised Laplacian* defined with

$$\nabla^{\nabla u_t} h := \sum_{i,j=1}^{n} g_{ij}^*\big((\mathcal{L}^*)^{-1}(V_t)\big) \frac{\partial h}{\partial x^j} \frac{\partial}{\partial x^i} \;. \tag{2.8}$$

Note that we suppressed the dependence on the choice of V_t for simplicity. We also remark that $\Delta^{\nabla u_t} u_t = \Delta u_t$. Linearised heat semigroups play an essential role in geometric analysis on Finsler manifolds, including gradient estimates [OS14] as well as functional and geometric inequalities [Oht17b, Oht22].

2.2. **Truncation functions.** We shall introduce some useful functions as in [HR03, §2.1]. Let $\zeta : [0, \infty) \to [0, \infty)$ be a bounded concave C^3-function satisfying:

(1) $\zeta(t) = t$ for $t \in [0, 1]$ and $0 < \zeta'(t) \leq 1$ for all $t \geq 0$;
(2) There is a positive constant C such that $0 \leq -\zeta''(t) \leq C\zeta'(t)$ for all $t \geq 0$.

For instance, any C^∞-function ζ such that $\zeta(0) = 0$ and ζ' is non-increasing with

$$\zeta'(t) = \begin{cases} 1 & 0 \leq t \leq 1 \;, \\ e^{-t} & t \geq 2 \end{cases}$$

satisfies the required properties. By these conditions, the monotone limit $\lim_{t\to\infty} \zeta(t) = L$ exists. Define $\phi^K(t) := K\zeta(t/K)$ for $K > 0$. For notational simplicity, we do not write K explicitly when no confusion could occur. We also set

$$\Phi(t) := \int_0^t \phi'(s)^2 \,\mathrm{d}s \;, \qquad \Psi(t) := t\phi'(t)^2 \;.$$

Then we immediately have the following estimates:

$$0 < \phi'(t) \le 1\ , \quad 0 \le -\phi''(t) \le \frac{C}{K}\phi'(t)\ , \quad \Phi(t) = \Psi(t) = t \text{ on } [0, K]\ , \tag{2.9}$$
$$0 \le \Psi(t) \le \Phi(t) \le \int_0^t \phi'(s)\,\mathrm{d}s = \phi(t) \le KL\ .$$

3. Maximal functions

In this section, we assume $\Lambda_F < \infty$ and the completeness of (M, F). To begin the proof of Theorem 1.1, fix $x_0 \in \mathsf{M}$ and let $(B_{r_k}(x_0))_{k\in\mathbb{N}}$ and $(\chi_k)_{k\in\mathbb{N}}$ be sequences of open forward balls (i.e., $B_r(x) := \{y \in \mathsf{M} : \mathsf{d}(x, y) < r\}$) and functions such that $\lim_{k\to\infty} r_k = \infty$, $0 \le \chi_k \le 1$, $\chi_k \equiv 1$ on $B_{r_k}(x_0)$, $\chi_k \equiv 0$ on $\mathsf{M} \setminus B_{r_{k+1}}(x_0)$, and $-\chi_k$ is 1-Lipschitz. Note that, in particular, χ_k is compactly supported. Throughout this article, for simplicity, we omit an arbitrarily fixed centre x_0 and write $B_k := B_{r_k}(x_0)$.

We next introduce a distance-like function $\bar{\mathsf{d}}_B$ for a measurable set $B \subset \mathsf{M}$. We will always assume $0 < \mathsf{m}(A), \mathsf{m}(B) < \infty$. For $R \ge 0$, define

$$\mathbb{L}_{B,R} := \{f \in \mathbb{L} : f = 0 \text{ on } B \text{ and } 0 \le f \le R \text{ a.e.}\}\ . \tag{3.1}$$

Recall the definitions (1.3), (1.4) of $\bar{\mathsf{d}}_{\mathsf{m}}(A, B)$ and $\mathbb{L}$ given in the introduction. We also set

$$\mathsf{d}_{\mathsf{m}}(A, B) := \operatorname*{ess\,inf}_{x\in A} \inf_{y\in B} \mathsf{d}(x, y)\ .$$

Proposition 3.1. *For any measurable set $B \subset \mathsf{M}$, there exists a unique $[0, \infty]$-valued measurable function $\bar{\mathsf{d}}_B$ such that, for every $R > 0$, the function $\bar{\mathsf{d}}_B \wedge R$ is the maximal element of $\mathbb{L}_{B,R}$. Precisely, $\bar{\mathsf{d}}_B \wedge R \in \mathbb{L}_{B,R}$ and $f \le \bar{\mathsf{d}}_B \wedge R$ holds a.e. for any $f \in \mathbb{L}_{B,R}$. Furthermore, for any measurable set $A \subset \mathsf{M}$, we have*

$$\bar{\mathsf{d}}_{\mathsf{m}}(A, B) = \operatorname*{ess\,inf}_{x\in A} \bar{\mathsf{d}}_B(x)\ .$$

Proof. We can follow the lines of [AH05, Proposition 3.11] by replacing E_k, $\mathbb{D}$, $\mathbb{D}_{A,M}$ and $\mathbb{D}_0$ there by B_k, $H^1_0(\mathsf{M}) = H^1(\mathsf{M})$, $\mathbb{L}_{B,R}$ and $\mathbb{L}$. ■

Remark 3.2. The requirement $F^*(-\mathbf{d}f) \le 1$ in (1.4) means that $\bar{\mathsf{d}}_B$ is regarded as the distance "to" the set B rather than the distance "from" B. We need to distinguish them in the present situation where the distance function d is asymmetric.

Set $\mathsf{d}_B(x) := \inf_{y\in B} \mathsf{d}(x, y)$ for $x \in \mathsf{M}$. Note that $-\mathsf{d}_B$ is 1-Lipschitz by the triangle inequality, and $F^*(-\mathbf{d}\mathsf{d}_B) \le 1$ a.e. We compare d_B with $\bar{\mathsf{d}}_B$ obtained in Proposition 3.1.

Lemma 3.3. *For any measurable set $B \subset \mathsf{M}$, we have $\mathsf{d}_B \le \bar{\mathsf{d}}_B$ a.e. In particular, for any measurable sets $A, B \subset \mathsf{M}$, we have*

$$\mathsf{d}_{\mathsf{m}}(A, B) \le \bar{\mathsf{d}}_{\mathsf{m}}(A, B)\ .$$

Proof. Since $\mathsf{d}_B \wedge R \in \mathbb{L}_{B,R}$ for every $R > 0$, we have $\mathsf{d}_B \wedge R \le \bar{\mathsf{d}}_B \wedge R$ by the maximality of $\bar{\mathsf{d}}_B \wedge R$ in $\mathbb{L}_{B,R}$. Letting $R \to \infty$, we obtain $\mathsf{d}_B \le \bar{\mathsf{d}}_B$ a.e. The latter assertion then follows from the definition of $\mathsf{d}_{\mathsf{m}}(A, B)$ and Proposition 3.1 as

$$\mathsf{d}_{\mathsf{m}}(A, B) = \operatorname*{ess\,inf}_{x\in A} \mathsf{d}_B(x) \le \operatorname*{ess\,inf}_{x\in A} \bar{\mathsf{d}}_B(x) = \bar{\mathsf{d}}_{\mathsf{m}}(A, B)\ .$$

■

For open sets, these distance-like functions actually coincide with the usual distance $\mathsf{d}(A,B)=\inf_{x\in A,\,y\in B}\mathsf{d}(x,y)=\inf_{x\in A}\mathsf{d}_B(x)$.

Lemma 3.4. *For any open set $B\subset\mathsf{M}$, we have $\bar{\mathsf{d}}_B=\mathsf{d}_B$ a.e. In particular, for any open sets $A,B\subset\mathsf{M}$, we have*

$$\bar{\mathsf{d}}_{\mathsf{m}}(A,B)=\mathsf{d}_{\mathsf{m}}(A,B)=\mathsf{d}(A,B)\ .$$

Proof. Since $\bar{\mathsf{d}}_B\wedge R\in\mathbb{L}_{B,R}$ and B is open, from the proposition below, its continuous version f satisfies $f\equiv 0$ on B, and $-f$ is 1-Lipschitz. Hence, for all $y\in B$ and a.e. $x\in\mathsf{M}$, we have

$$\bar{\mathsf{d}}_B(x)\wedge R=f(x)=f(x)-f(y)\le\mathsf{d}(x,y)\ .$$

Letting $R\to\infty$ and taking the infimum in $y\in B$, we conclude $\bar{\mathsf{d}}_B\le\mathsf{d}_B$ a.e. Combining this with Lemma 3.3, we deduce the former assertion. Then $\bar{\mathsf{d}}_{\mathsf{m}}(A,B)=\mathsf{d}_{\mathsf{m}}(A,B)$ follows from the definition of $\mathsf{d}_{\mathsf{m}}(A,B)$ and Proposition 3.1, while $\mathsf{d}_{\mathsf{m}}(A,B)=\mathsf{d}(A,B)$ is immediate since A is open and d_B is continuous. ∎

Though the next proposition (called the *local Sobolev-to-Lipschitz property*) should be a known fact, we give a short proof for completeness.

Proposition 3.5. *For any $f\in\mathbb{L}$, there exists a bounded 1-Lipschitz function $-\tilde{f}$ such that $\tilde{f}=f$ a.e.*

Proof. By multiplying with a smooth cut-off function in each local chart, one can reduce the existence of a (locally) Lipschitz function $\tilde{f}$ such that $\tilde{f}=f$ a.e. to the Euclidean case. The Euclidean case can be seen, e.g., in [EG15, Theorem 4.5]. Then $F^*(-\mathbf{d}\tilde{f})=F^*(-\mathbf{d}f)\le 1$ a.e. yields that $-\tilde{f}$ is 1-Lipschitz. ∎

4. Upper estimate

For measurable sets $A,B\subset\mathsf{M}$, recall that we define

$$\mathsf{P}_t(A,B):=\int_A \mathsf{T}_t\mathbf{1}_B\,\mathrm{d}\mathsf{m}\ .$$

In this section, we establish the upper estimate of (1.5). Our proof is essentially along the lines of [AH05, Theorem 4.1] or [HR03, Theorem 2.8], where the former is a generalisation of the latter to admit infinite total mass. The upper estimate is less demanding than the lower estimate. In fact, after establishing the upper estimate in the Finsler setting, we will see that it also holds true in metric measure spaces under mild assumptions in Subsection 4.2.

4.1. **Finsler case.** Let $(\mathsf{M},F,\mathsf{m})$ be a complete Finsler manifold such that $\Lambda_F<\infty$ equipped with a measure m.

Proposition 4.1. *For any measurable sets $A,B\subset\mathsf{M}$ with $0<\mathsf{m}(A),\mathsf{m}(B)<\infty$ and $t>0$, we have*

$$\mathsf{P}_t(A,B)\le\sqrt{\mathsf{m}(A)}\sqrt{\mathsf{m}(B)}\exp\left(-\frac{\bar{\mathsf{d}}_{\mathsf{m}}(A,B)^2}{2t}\right)\ .$$

In particular,

$$\limsup_{t\downarrow 0} t\log \mathsf{P}_t(A,B) \le -\frac{\bar{\mathsf{d}}_{\mathsf{m}}(A,B)^2}{2} \ .$$

Proof. Given $R>0$, take $f\in \mathbb{L}_{B,R}$ and set $f_k := f\chi_k \in H^1_0(\mathsf{M})$ for $k\in\mathbb{N}$, with χ_k chosen in Section 3. We also set $u_t := \mathsf{T}_t \mathbf{1}_B \in H^1_0(\mathsf{M})$ and $u_{t,k} := u_t\chi_k \in H^1_0(\mathsf{M})$. Note that $0\le u_t\le 1$ a.e. (see, e.g., [Oht21, Lemma 13.13]). For $\alpha\ge 0$ (chosen later), we consider the function

$$\xi(t) := \int_{\mathsf{M}} (e^{\alpha f}u_t)^2 \,\mathsf{dm} \ .$$

By the heat equation and the integration by parts, we have

$$\begin{aligned}\xi'(t) &= \int_{\mathsf{M}} e^{2\alpha f} u_t \Delta u_t \,\mathsf{dm} = \lim_{k\to\infty}\int_{\mathsf{M}} e^{2\alpha f_k} u_{t,k}\Delta u_t \,\mathsf{dm} \\ &= -\lim_{k\to\infty}\int_{\mathsf{M}} \mathbf{d}(e^{2\alpha f_k}u_{t,k})(\nabla u_t)\,\mathsf{dm} \ .\end{aligned} \tag{4.1}$$

Let us now see that

$$\lim_{k\to\infty}\int_{\mathsf{M}} \mathbf{d}(e^{2\alpha f_k}u_{t,k})(\nabla u_t)\,\mathsf{dm} = \lim_{k\to\infty}\int_{\mathsf{M}} \mathbf{d}(e^{2\alpha f_k}u_{t,k})(\nabla u_{t,k})\,\mathsf{dm} \ . \tag{4.2}$$

Note first that

$$\begin{aligned}&\left|\int_{\mathsf{M}} \mathbf{d}(e^{2\alpha f_k}u_{t,k})(\nabla u_t)\,\mathsf{dm} - \int_{\mathsf{M}} \mathbf{d}(e^{2\alpha f_k}u_{t,k})(\nabla u_{t,k})\,\mathsf{dm}\right| \\ &\le \int_{\mathsf{M}} F^*\big(\mathbf{d}(e^{2\alpha f_k}u_{t,k})\big)F(\nabla u_t - \nabla u_{t,k})\,\mathsf{dm} \\ &\le \sqrt{2\mathcal{E}(e^{2\alpha f_k}u_{t,k})}\cdot \|F(\nabla u_t-\nabla u_{t,k})\|_{L^2} \ .\end{aligned}$$

We find from $u_t\in H^1_0(\mathsf{M})$,

$$F^*(-\mathbf{d}f_k) \le F^*(-f\mathbf{d}\chi_k) + F^*(-\chi_k\mathbf{d}f) \le R+1 \quad \text{a.e.}$$

and $\Lambda_F<\infty$ that $\mathcal{E}(e^{2\alpha f_k}u_{t,k})$ is bounded above uniformly in k. Moreover, since $\mathbf{d}u_t = \mathbf{d}u_{t,k}$ on B_k, we observe

$$\begin{aligned}\|F(\nabla u_t - \nabla u_{t,k})\|^2_{L^2} &\le \int_{\mathsf{M}\setminus B_k} \big(F(\nabla u_t) + \Lambda_F F(\nabla u_{t,k})\big)^2\,\mathsf{dm} \\ &\le \int_{\mathsf{M}\setminus B_k} \Big(F(\nabla u_t) + \Lambda_F\big(u_t F^*(\mathbf{d}\chi_k) + \chi_k F^*(\mathbf{d}u_t)\big)\Big)^2\,\mathsf{dm} \\ &\le \int_{\mathsf{M}\setminus B_k} \big(\Lambda_F^2 u_t + (\Lambda_F+1)F(\nabla u_t)\big)^2\,\mathsf{dm} \xrightarrow{k\to\infty} 0 \ .\end{aligned}$$

Therefore, (4.2) has been shown and the RHS of (4.1) can be expanded as

$$\begin{aligned}&-\lim_{k\to\infty}\int_{\mathsf{M}} \mathbf{d}(e^{2\alpha f_k}u_{t,k})(\nabla u_{t,k})\,\mathsf{dm} \\ &= -\lim_{k\to\infty}\int_{\mathsf{M}} e^{2\alpha f_k}\big(\mathbf{d}u_{t,k}(\nabla u_{t,k}) + 2\alpha u_{t,k}\,\mathbf{d}f_k(\nabla u_{t,k})\big)\,\mathsf{dm} \ .\end{aligned} \tag{4.3}$$

On B_k, it follows from $\chi_k\equiv 1$ and $f\in\mathbb{L}_{B,R}$ that $F^*(-\mathbf{d}f_k) = F^*(-\mathbf{d}f)\le 1$, thereby,

$$-\mathbf{d}f_k(\nabla u_{t,k}) \le F^*(-\mathbf{d}f_k)F(\nabla u_{t,k}) \le F(\nabla u_{t,k}) \ . \tag{4.4}$$

Combining (4.1)–(4.4), we obtain

$$\begin{aligned}\xi'(t) &= -\lim_{k\to\infty}\int_{\mathsf{M}} e^{2\alpha f_k}\big(\mathbf{d}u_{t,k}(\nabla u_{t,k}) + 2\alpha u_{t,k}\,\mathbf{d}f_k(\nabla u_{t,k})\big)\,\mathrm{d}\mathsf{m}\\ &\le -\lim_{k\to\infty}\int_{\mathsf{M}} e^{2\alpha f_k}\big(F(\nabla u_{t,k})^2 - 2\alpha u_{t,k}F(\nabla u_{t,k})\big)\,\mathrm{d}\mathsf{m}\\ &\le \lim_{k\to\infty}\int_{\mathsf{M}} e^{2\alpha f_k}(\alpha u_{t,k})^2\,\mathrm{d}\mathsf{m}\\ &= \alpha^2\int_{\mathsf{M}} e^{2\alpha f}u_t^2\,\mathrm{d}\mathsf{m} = \alpha^2\xi(t)\ ,\end{aligned}$$

which yields

$$\xi(t) \le \xi(0)e^{\alpha^2 t}\ .$$

Now, we take $f = \bar{\mathsf{d}}_B \wedge \bar{\mathsf{d}}_{\mathsf{m}}(A,B) \in \mathbb{L}_{B,\bar{\mathsf{d}}_{\mathsf{m}}(A,B)}$. Since $\bar{\mathsf{d}}_B = 0$ a.e. on B and $\bar{\mathsf{d}}_B \ge \bar{\mathsf{d}}_{\mathsf{m}}(A,B)$ a.e. on A by Proposition 3.1, we have

$$\begin{aligned}\mathsf{P}_t(A,B) &= \int_{\mathsf{M}} u_t\mathbf{1}_A\,\mathrm{d}\mathsf{m} \le \|e^{\alpha f}u_t\|_{L^2}\|e^{-\alpha f}\mathbf{1}_A\|_{L^2} \le \sqrt{\xi(0)}e^{\alpha^2 t/2}\sqrt{\mathsf{m}(A)}e^{-\alpha\bar{\mathsf{d}}_{\mathsf{m}}(A,B)}\\ &= \sqrt{\mathsf{m}(A)}\sqrt{\mathsf{m}(B)}\exp\left(\frac{\alpha^2 t}{2} - \alpha\bar{\mathsf{d}}_{\mathsf{m}}(A,B)\right)\ .\end{aligned}$$

Choosing the optimal value $\alpha = \bar{\mathsf{d}}_{\mathsf{m}}(A,B)/t$, we conclude

$$\mathsf{P}_t(A,B) \le \sqrt{\mathsf{m}(A)}\sqrt{\mathsf{m}(B)}\exp\left(-\frac{\bar{\mathsf{d}}_{\mathsf{m}}(A,B)^2}{2t}\right)\ .$$

■

Set $\Phi_t := \Phi(-t\log \mathsf{T}_t\mathbf{1}_B)$ for $t>0$ and Φ as in Subsection 2.2. Since Φ is bounded, for any finite measure ν mutually absolutely continuous with m, $(\Phi_t)_{t>0}$ is uniformly bounded and hence weakly relatively compact in $L^2(\nu)$. The following corollary will play a role in the lower estimate.

Corollary 4.2. *For any measurable set $B \subset \mathsf{M}$ with $0 < \mathsf{m}(B) < \infty$ and any weak $L^2(\nu)$-limit Φ_0 of $(\Phi_t)_{t>0}$ as $t \to 0$, we have*

$$\Phi_0 \ge \Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right) \quad a.e.$$

Proof. We can follow the lines of [HR03, Lemma 2.9] thanks to Propositions 3.1, 4.1. ■

4.2. **Nonsmooth case.** In this subsection, we briefly explain that the upper estimate (Proposition 4.1) can be generalised to the nonsmooth setting of metric measure spaces by utilising the differential calculus developed in [AGS14, Gig18].

Setting. Let (X,d) be a complete separable metric space (with a usual symmetric distance function), and m be a fully supported Borel measure on X such that $\mathsf{m}(B) < \infty$ for every bounded m-measurable set $B \subset \mathsf{X}$. A Borel probability measure π on the set $C([0,1],\mathsf{X})$ of continuous curves $\gamma : [0,1] \to \mathsf{X}$ is called a *test plan* if there is a constant $C > 0$ such that $(e_t)_{\#}\pi \le C\mathsf{m}$ for all $t \in [0,1]$ and

$$\int_{C([0,1],\mathsf{X})}\int_0^1 |\dot\gamma_t|^2\,\mathrm{d}t\,\pi(\mathrm{d}\gamma) < \infty\ ,$$

where $e_t(\gamma) := \gamma_t$ is the evaluation map, $(e_t)_\# \pi$ denotes the push-forward of π by e_t, and $|\dot{\gamma}_t| := \lim_{s\to t} \mathsf{d}(\gamma_s, \gamma_t)/|s-t|$ is the metric speed.

The *Sobolev class* $S^2(\mathsf{X})$ consists of measurable functions f such that there is a nonnegative function $g \in L^2(\mathsf{X})$ satisfying

$$\int_{C([0,1],\mathsf{X})} |f(\gamma_1) - f(\gamma_0)| \, \pi(\mathrm{d}\gamma) \le \int_{C([0,1],\mathsf{X})} \int_0^1 g(\gamma_t) |\dot{\gamma}_t| \, \mathrm{d}t \, \pi(\mathrm{d}\gamma)$$

for all test plans π. The minimal function g is called the *minimal weak upper gradient* and denoted by $|\mathrm{D}f|$. We define the *Sobolev space* $W^{1,2}(\mathsf{X}) := S^2(\mathsf{X}) \cap L^2(\mathsf{X})$ equipped with the norm

$$\|f\|_{W^{1,2}} := \sqrt{\|f\|_{L^2}^2 + \big\||\mathrm{D}f|\big\|_{L^2}^2} \, .$$

(Denoting the Sobolev space by $W^{1,2}(\mathsf{X})$ follows the notation in [AGS14, Gig18], while in the Finsler setting we use $H^1(\mathsf{M}), H^1_0(\mathsf{M})$ as in [Oht21].) We remark that $W^{1,2}(\mathsf{X})$ is not necessarily separable, reflexive, nor a Hilbert space in this generality (see [ACD15], [Gig18, Theorem 2.1.5]).

Heat semigroup. The *Cheeger energy* $\mathsf{Ch} : L^2(\mathsf{X}) \to [0, \infty]$ is defined as

$$\mathsf{Ch}(f) := \frac{1}{2} \int_{\mathsf{X}} |\mathrm{D}f|^2 \, \mathrm{d}\mathsf{m}$$

for $f \in W^{1,2}(\mathsf{X})$, and $\mathsf{Ch}(f) := \infty$ otherwise. Note that Ch is convex, lower semi-continuous and the domain $W^{1,2}(\mathsf{X})$ is dense in $L^2(\mathsf{X})$. For $f \in W^{1,2}(\mathsf{X})$ such that the sub-differential $\partial^- \mathsf{Ch}(f) \subset L^2(\mathsf{X})$ is nonempty, $\Delta f \in L^2(\mathsf{X})$ is defined as $\Delta f := -h$, where h is the element of minimal L^2-norm in $\partial^- \mathsf{Ch}(f)$. Note that this Laplacian is not necessarily linear.

Thanks to the theory of gradient flows for convex functions on Hilbert spaces (we refer to [AGS08]), we have the *heat semigroup* $(\mathsf{T}_t)_{t \ge 0}$ of continuous operators from $L^2(\mathsf{X})$ to itself such that, for every $f \in L^2(\mathsf{X})$, $t \mapsto \mathsf{T}_t f \in L^2(\mathsf{X})$ is continuous on $[0, \infty)$, absolutely continuous on $(0, \infty)$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathsf{T}_t f = \frac{1}{2} \Delta \mathsf{T}_t f \quad \text{a.e. } t > 0 \, .$$

Differentials and gradients. According to [Gig18, Definition 2.2.1], there exists a dual pair of Banach spaces called the *tangent module* and the *cotangent module*:

$$\big(L^2(T\mathsf{X}), \|\cdot\|_{L^2(T\mathsf{X})}\big) \, , \qquad \big(L^2(T^*\mathsf{X}), \|\cdot\|_{L^2(T^*\mathsf{X})}\big) \, .$$

These spaces have $L^2(\mathsf{X})$-normed module structures, namely they are $L^\infty(\mathsf{X})$-premodules and there exist maps $|\cdot| : L^2(T\mathsf{X}) \to L^2(\mathsf{X})$, $|\cdot|_* : L^2(T^*\mathsf{X}) \to L^2(\mathsf{X})$ such that

$$\big\||v|\big\|_{L^2} = \|v\|_{L^2(T\mathsf{X})} \text{ for } v \in L^2(T\mathsf{X}) \, , \qquad \big\||\omega|_*\big\|_{L^2} = \|\omega\|_{L^2(T^*\mathsf{X})} \text{ for } \omega \in L^2(T^*\mathsf{X}) \, .$$

There exist a linear operator $\mathbf{d} : S^2(\mathsf{X}) \to L^2(T^*\mathsf{X})$ and a (not necessarily linear) multi-valued operator $\mathrm{Grad} : S^2(\mathsf{X}) \to L^2(T\mathsf{X})$ satisfying

$$\mathbf{d}f(\mathrm{Grad}f) = |\mathrm{Grad}f|^2 = |\mathbf{d}f|_*^2 = |\mathrm{D}f|^2$$

for $f \in S^2(\mathsf{X})$.

We say that $(\mathsf{X},\mathsf{d},\mathsf{m})$ is *infinitesimally strictly convex* if $\mathrm{Grad}(f)$ consists of exactly one element, which is denoted by ∇f (see, e.g., [Gig18, Definition 2.3.9]). In this case, we have

$$\mathsf{Ch}(f)=\frac{1}{2}\int_{\mathsf{X}}\mathbf{d}f(\nabla f)\,\mathrm{d}\mathsf{m},\qquad f\in W^{1,2}(\mathsf{X})\ .$$

The *differential* operator $\mathbf{d}$ enjoys the locality as well as the Leibniz and chain rules in an appropriate sense. On the other hand, the *gradient* operator ∇ satisfies the locality and the chain rule, whereas the Leibniz rule does not hold.

Upper estimate. Define $W^{1,2}_{\mathrm{loc}}(\mathsf{X})$ as the set of functions f admitting $(f_k)_{k\in\mathbb{N}}\subset W^{1,2}(\mathsf{X})$ such that $f=f_k$ on B_k (with B_k as in Section 3; $W^{1,2}_{\mathrm{loc}}(\mathsf{X})$ does not depend on the choice of B_k). Then we define

$$\mathbb{L}:=\{f\in W^{1,2}_{\mathrm{loc}}(\mathsf{X})\cap L^\infty(\mathsf{X}):|\mathrm{D}f|\le 1\text{ a.e.}\}\ ,$$

and the *local Sobolev-to-Lipschitz property* means the property described in Proposition 3.5: for any $f\in\mathbb{L}$, there exists a bounded 1-Lipschitz function $\tilde f$ such that $f=\tilde f$ m-a.e.

Now, Theorem 1.2 is proved in the same way as Proposition 4.1. We remark that the reflexivity of $W^{1,2}(\mathsf{X})$ is used in the proof of Proposition 3.1, and the local Sobolev-to-Lipschitz property is necessary to take a continuous version in the proof of Lemma 3.4. Concerning the lower estimate in the next section, our argument is applicable up to Subsection 5.3, whereas the use of a linearised heat semigroup in Lemma 5.7 prevents us from applying the same proof to the non-smooth case.

Remark 4.3. The nonsmooth calculus in this subsection is not yet generalised to the asymmetric setting. We refer to [KZ22] for a related study of the curvature-dimension condition in asymmetric metric measure spaces.

5. Lower estimate

This last section is devoted to the lower estimate of (1.5):

$$\liminf_{t\downarrow 0} t\log \mathsf{P}_t(A,B)\ge -\frac{\bar{\mathsf{d}}_{\mathsf{m}}(A,B)^2}{2}\tag{5.1}$$

for measurable sets $A,B\subset\mathsf{M}$ with $0<\mathsf{m}(A),\mathsf{m}(B)<\infty$. Let $(\mathsf{M},F,\mathsf{m})$ be a complete Finsler manifold with $\Lambda_F<\infty$ with a measure m. We will need the additional assumptions $\mathsf{C}_F,\mathsf{S}_F<\infty$ and $\mathsf{m}(\mathsf{M})<\infty$ only in a later step in Subsections 5.4, 5.5.

5.1. Outline of the proof.

Let us first remark that it is sufficient to show (5.1) in the case of $\mathsf{P}_t(A,B)<1$. This is seen by replacing m with $c\mathsf{m}$ for $c>0$ such that $c\mathsf{m}(A)<1$. Indeed, $(\mathsf{M},F,c\mathsf{m})$ has the same heat flow as $(\mathsf{M},F,\mathsf{m})$, thereby $\mathsf{P}^{c\mathsf{m}}_t(A,B)=c\mathsf{P}^{\mathsf{m}}_t(A,B)$, and clearly $\bar{\mathsf{d}}_{c\mathsf{m}}(A,B)=\bar{\mathsf{d}}_{\mathsf{m}}(A,B)$. Thus, we have $\mathsf{P}^{c\mathsf{m}}_t(A,B)\le c\mathsf{m}(A)<1$, and (5.1) for $c\mathsf{m}$ implies that for m.

Given $\varepsilon>0$, we set

$$D_\varepsilon:=\{x\in A:\bar{\mathsf{d}}_B(x)\le\bar{\mathsf{d}}_{\mathsf{m}}(A,B)+\varepsilon\}\ ,$$

and observe $\mathsf{m}(D_\varepsilon) > 0$ by Proposition 3.1. Then, recalling $\Phi_t = \Phi(-t\log \mathsf{T}_t\mathbf{1}_B)$ and noting $\log \mathsf{P}_t(A,B) < 0$ for small $t>0$, we have

$$\begin{aligned}
(5.2)\qquad \limsup_{t\downarrow 0} \Phi\big(-t\log \mathsf{P}_t(A,B)\big) &\le \limsup_{t\downarrow 0} \Phi\big(-t\log \mathsf{P}_t(D_\varepsilon,B)\big)\\
&= \limsup_{t\downarrow 0} \Phi\left(-t\log\left[\frac{1}{\mathsf{m}(D_\varepsilon)}\int_{D_\varepsilon} \mathsf{T}_t\mathbf{1}_B \,\mathrm{d}\mathsf{m}\right]\right)\\
&\le \limsup_{t\downarrow 0} \frac{1}{\mathsf{m}(D_\varepsilon)}\int_{D_\varepsilon} \Phi(-t\log \mathsf{T}_t\mathbf{1}_B)\,\mathrm{d}\mathsf{m}\\
&= \limsup_{t\downarrow 0} \frac{1}{\mathsf{m}(D_\varepsilon)}\int_{D_\varepsilon} \Phi_t\,\mathrm{d}\mathsf{m}\ ,
\end{aligned}$$

where the latter inequality is derived from Jensen's inequality since the function $s \mapsto \Phi(-t\log s)$ is convex on $[0,1]$ for sufficiently small $t>0$ (see [HR03, Lemma 2.1]). Now, suppose that the following inequality holds:

$$(5.3)\qquad \limsup_{t\downarrow 0}\int_{D_\varepsilon} \Phi_t\,\mathrm{d}\mathsf{m} \le \int_{D_\varepsilon} \Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right)\mathrm{d}\mathsf{m}\ .$$

Then, we can continue the estimation in (5.2) as, with the help of (2.9),

$$\le \frac{1}{\mathsf{m}(D_\varepsilon)}\int_{D_\varepsilon} \Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right)\mathrm{d}\mathsf{m} \le \frac{1}{\mathsf{m}(D_\varepsilon)}\int_{D_\varepsilon} \frac{\bar{\mathsf{d}}_B^2}{2}\,\mathrm{d}\mathsf{m} \le \frac{(\bar{\mathsf{d}}_{\mathsf{m}}(A,B)+\varepsilon)^2}{2}\ .$$

Since $\varepsilon>0$ was arbitrary and $\Phi(t) = \Phi^K(t) \to t$ as $K\to\infty$, we conclude (5.1).

The purpose of the rest of the section is to show (5.3), which completes the proof of (5.1). In fact, at the end of the section, we will prove that equality holds (see (5.11)).

5.2. **Uniform bounds.** For $\delta \in (0,1)$ and $t>0$, we put

$$u_t^\delta := -t\log\big((1-\delta)\mathsf{T}_t\mathbf{1}_B + \delta\big)\ ,\qquad e_t^\delta := -t\log\delta\ .$$

Note that $u_t^\delta \ge 0$ since $0 \le \mathsf{T}_t\mathbf{1}_B \le 1$. In addition, for ϕ, Φ and Ψ in Subsection 2.2, we will abbreviate as $\phi_t^\delta := \phi(u_t^\delta)$, $\Phi_t^\delta := \Phi(u_t^\delta)$ and $\Psi_t^\delta := \Psi(u_t^\delta)$.

We will denote by $(\cdot,\cdot)_{L^2}$ both the L^2-inner product and the paring between a one-form and a vector field with respect to m. We also introduce the following notation: For a nonnegative function ρ, define

$$\mathcal{E}_\rho(f) := \frac{1}{2}\int_{\mathsf{M}} F^*(\mathbf{d}f)^2\rho\,\mathrm{d}\mathsf{m}\ ,\qquad (f_1,f_2)_{L^2(\rho)} := \int_{\mathsf{M}} f_1f_2\rho\,\mathrm{d}\mathsf{m}\ .$$

Lemma 5.1. *For any $t>0$, we have $u_t^\delta - e_t^\delta \in H_0^1(\mathsf{M})$ and, for every bounded nonnegative function $\rho \in H_0^1(\mathsf{M}) \cap L^1(\mathsf{M})$,*

$$\partial_t(\rho,\Phi_t^\delta)_{L^2} = \frac{1}{t}(\rho,\Psi_t^\delta)_{L^2} + \frac{1}{2}\Big(\mathbf{d}\big(\phi'(u_t^\delta)^2\rho\big),\nabla(-u_t^\delta)\Big)_{L^2} - \frac{1}{t}\mathcal{E}_{\phi'(u_t^\delta)^2\rho}(-u_t^\delta)\ .$$

Proof. We put $f := (1-\delta)\mathbf{1}_B$ for brevity. Observe that $u_t^\delta - e_t^\delta = \vartheta(\mathsf{T}_tf)$, where $\vartheta(s) := -t\log((s+\delta)/\delta)$ is a Lipschitz function on $[0,\infty)$ with $\vartheta(0)=0$. Thus, since $\mathsf{T}_tf \in H_0^1(\mathsf{M})$, we have $u_t^\delta - e_t^\delta \in H_0^1(\mathsf{M})$. We deduce from the chain rule for $\mathbf{d}$ that

$$\mathbf{d}u_t^\delta = -\frac{t}{\mathsf{T}_tf+\delta}\mathbf{d}\mathsf{T}_tf\ ,\qquad \nabla(-u_t^\delta) = \frac{t}{\mathsf{T}_tf+\delta}\nabla\mathsf{T}_tf\ .$$

We remark that, to derive the latter equation, we needed $t/(\mathsf{T}_t f + \delta) > 0$ because of the irreversibility of F. Combining this with the Leibniz rule for $\mathbf{d}$, we have

$$\left(\mathbf{d}\left[\frac{\rho}{\mathsf{T}_t f + \delta}\right], \nabla \mathsf{T}_t f\right)_{L^2} = \left(\frac{\mathbf{d}\rho}{\mathsf{T}_t f + \delta}, \nabla \mathsf{T}_t f\right)_{L^2} - \left(\frac{\mathbf{d}\mathsf{T}_t f}{(\mathsf{T}_t f + \delta)^2}, \nabla \mathsf{T}_t f\right)_{L^2(\rho)}$$
$$= \frac{1}{t}\big(\mathbf{d}\rho, \nabla(-u_t^\delta)\big)_{L^2} + \frac{1}{t^2}\big(\mathbf{d}u_t^\delta, \nabla(-u_t^\delta)\big)_{L^2(\rho)} .$$

Hence, we obtain

$$\begin{aligned}(\rho, \partial_t u_t^\delta)_{L^2} &= \frac{1}{t}(\rho, u_t^\delta)_{L^2} - \frac{t}{2}\left(\rho, \frac{\Delta \mathsf{T}_t f}{\mathsf{T}_t f + \delta}\right)_{L^2} \\ &= \frac{1}{t}(\rho, u_t^\delta)_{L^2} + \frac{t}{2}\left(\mathbf{d}\left[\frac{\rho}{\mathsf{T}_t f + \delta}\right], \nabla \mathsf{T}_t f\right)_{L^2} \\ &= \frac{1}{t}(\rho, u_t^\delta)_{L^2} + \frac{1}{2}\big(\mathbf{d}\rho, \nabla(-u_t^\delta)\big)_{L^2} - \frac{1}{t}\mathcal{E}_\rho(-u_t^\delta) .\end{aligned}$$

Since $\partial_t(\rho, \Phi_t^\delta)_{L^2} = (\rho, \partial_t \Phi_t^\delta)_{L^2} = \big(\phi'(u_t^\delta)^2\rho, \partial_t u_t^\delta\big)_{L^2}$, replacing ρ with $\phi'(u_t^\delta)^2\rho$ in the above calculation completes the proof. ■

For a function $f : (0,\infty) \times \mathsf{M} \to \mathbb{R}$ (such as $(t,x) \mapsto \phi_t^\delta(x)$), we will denote its time average by

$$\bar{f}_t(x) := \frac{1}{t}\int_0^t f_s(x)\,\mathrm{d}s ,$$

where $f_s(x) := f(s,x)$. The next lemma is a standard fact of the ($H_0^1(\mathsf{M})$-valued) Bochner integral (cf. [HR03, Lemma 2.5]).

Lemma 5.2. *Let $f : (0,T] \times \mathsf{M} \to \mathbb{R}$ be a bounded jointly measurable function such that $f_t \in H_0^1(\mathsf{M})$ for all $t \in (0,T]$ and $\int_0^T \|f_t\|_{H^1}^2\,\mathrm{d}t < \infty$. Then, we have $\bar{f}_T \in H_0^1(\mathsf{M})$ and*

$$\mathcal{E}_\rho(\bar{f}_T) \le \frac{1}{T}\int_0^T \mathcal{E}_\rho(f_t)\,\mathrm{d}t$$

for any bounded nonnegative function $\rho \in L^1_{\mathrm{loc}}(\mathsf{M})$.

Proof. By hypothesis, $t \mapsto f_t \in H_0^1(\mathsf{M})$ is Bochner integrable and we have $\bar{f}_T \in H_0^1(\mathsf{M})$. Then, the claimed inequality is a consequence of the linearity of $\mathbf{d}$ and Jensen's inequality for the convex function $(F^*)^2$:

$$\begin{aligned}2\mathcal{E}_\rho(\bar{f}_T) &= \int_{\mathsf{M}} F^*\left(\frac{1}{T}\int_0^T \mathbf{d}f_t\,\mathrm{d}t\right)^2 \rho\,\mathrm{d}\mathsf{m} \le \frac{1}{T}\int_{\mathsf{M}}\int_0^T F^*(\mathbf{d}f_t)^2\rho\,\mathrm{d}t\,\mathrm{d}\mathsf{m} \\ &= \frac{2}{T}\int_0^T \mathcal{E}_\rho(f_t)\,\mathrm{d}t .\end{aligned}$$

■

The next proposition is the goal of this subsection.

Proposition 5.3. *For sufficiently small $T_0 > 0$, the families $\{\bar{\phi}_t^\delta \chi_k\}_{0<t<T_0,\,0<\delta<1}$ and $\{\bar{\Phi}_t^\delta \chi_k\}_{0<t<T_0,\,0<\delta<1}$ are bounded in $H_0^1(\mathsf{M})$ for every $k \in \mathbb{N}$.*

Proof. Since ϕ and Φ are bounded, it is straightforward that both families are bounded in $L^2(\mathsf{M})$. Thus, we discuss only the bound for the energy. Put

$$U_t^\delta := 2\mathcal{E}(-\phi_t^\delta \chi_k) \ , \qquad a := \Lambda_F^2 \cdot \mathsf{m}(B_{k+1}) \ ,$$

for Λ_F in (2.3). Then, by (2.9) and the choice of χ_k as in Section 3, we have

$$\begin{aligned} U_t^\delta &\leq \left\| F^*\big(-\phi'(u_t^\delta)\chi_k \,\mathbf{d}u_t^\delta\big) + F^*(-\phi_t^\delta \,\mathbf{d}\chi_k) \right\|_{L^2}^2 \\ &\leq \left\| \phi'(u_t^\delta)\chi_k F^*(-\mathbf{d}u_t^\delta) + \phi_t^\delta F^*(-\mathbf{d}\chi_k) \right\|_{L^2}^2 \\ &\leq 4\mathcal{E}_{\phi'(u_t^\delta)^2\chi_k^2}(-u_t^\delta) + 2(KL)^2 \mathsf{m}(B_{k+1}) \ . \end{aligned} \tag{5.4}$$

Letting $\rho = \chi_k^2$ in Lemma 5.1, we find

$$\begin{aligned} V_t^\delta &:= 2\mathcal{E}_{\phi'(u_t^\delta)^2\chi_k^2}(-u_t^\delta) \\ &= -2t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + 2(\Psi_t^\delta, \chi_k^2)_{L^2} + t\Big(\mathbf{d}\big(\phi'(u_t^\delta)^2\chi_k^2\big), \nabla(-u_t^\delta)\Big)_{L^2} \ . \end{aligned}$$

Note that, in the RHS, $(\Psi_t^\delta, \chi_k^2)_{L^2} \leq KL\mathsf{m}(B_{k+1})$ by (2.9) and the choice of χ_k. For the third term, we deduce from the Leibniz rule for $\mathbf{d}$ and (2.9) that

$$\begin{aligned} &\Big(\mathbf{d}\big(\phi'(u_t^\delta)^2\chi_k^2\big), \nabla(-u_t^\delta)\Big)_{L^2} \\ &= 2\big(\mathbf{d}\chi_k, \nabla(-u_t^\delta)\big)_{L^2(\phi'(u_t^\delta)^2\chi_k)} + 2\big(\phi''(u_t^\delta)\,\mathbf{d}u_t^\delta, \nabla(-u_t^\delta)\big)_{L^2(\phi'(u_t^\delta)\chi_k^2)} \\ &\leq 2\Big(F^*(\mathbf{d}\chi_k), F\big(\nabla(-u_t^\delta)\big)\Big)_{L^2(\phi'(u_t^\delta)^2\chi_k)} + \frac{4C}{K}\mathcal{E}_{\phi'(u_t^\delta)^2\chi_k^2}(-u_t^\delta) \ . \end{aligned}$$

Since $\mathbf{d}(\phi_t^\delta\chi_k) = \phi'(u_t^\delta)\chi_k \,\mathbf{d}u_t^\delta + \phi_t^\delta \,\mathbf{d}\chi_k$ by the chain rule for $\mathbf{d}$, we also have

$$\phi'(u_t^\delta)\chi_k F\big(\nabla(-u_t^\delta)\big) \leq F\big(\nabla(-\phi_t^\delta\chi_k)\big) + \phi_t^\delta F(\nabla\chi_k) \ .$$

Combining these inequalities with $2\mathcal{E}(\chi_k) \leq 2\Lambda_F^2\mathcal{E}(-\chi_k) \leq a$ yields that

$$\begin{aligned} V_t^\delta &\leq -2t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + 2KLa + \frac{2Ct}{K}V_t^\delta \\ &\quad + 2t\Big(F^*(\mathbf{d}\chi_k), F\big(\nabla(-\phi_t^\delta\chi_k)\big)\Big)_{L^2(\phi'(u_t^\delta))} + 2t\big(F^*(\mathbf{d}\chi_k), F(\nabla\chi_k)\big)_{L^2(\phi_t^\delta\phi'(u_t^\delta))} \\ &\leq -2t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + 2KLa + \frac{2Ct}{K}V_t^\delta + 4t\sqrt{\mathcal{E}(\chi_k)\mathcal{E}(-\phi_t^\delta\chi_k)} + 4KLt\mathcal{E}(\chi_k) \\ &\leq -2t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + 2KLa + \frac{2Ct}{K}V_t^\delta + 2t\sqrt{aU_t^\delta} + 2KLat \ . \end{aligned}$$

Hence, we obtain

$$\left(1 - \frac{2Ct}{K}\right)V_t^\delta \leq -2t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + \frac{U_t^\delta}{8} + 8at^2 + 2KLa(t+1) \ . \tag{5.5}$$

Now, put $T_0 := K/(4C)$ and observe that $V_t^\delta/2 \leq (1 - (2Ct)/K)V_t^\delta$ for $t \in (0, T_0]$. Thus, for $t \in (0, T_0]$, we find from (5.4) and (5.5) that

$$\begin{aligned} U_t^\delta &\leq 2V_t^\delta + 2K^2L^2a \\ &\leq -8t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + \frac{U_t^\delta}{2} + 32aT_0^2 + 8KLa(T_0+1) + 2K^2L^2a \ . \end{aligned} \tag{5.6}$$

This implies

$$U_t^\delta \leq -16t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + c \ ,$$

where c is a constant independent of $t \in (0, T_0]$ and $\delta \in (0,1)$. Hence, for $0 < \varepsilon < t \le T_0$,

$$\int_\varepsilon^t 2\mathcal{E}(-\phi_s^\delta \chi_k)\,\mathrm{d}s = \int_\varepsilon^t U_s^\delta \,\mathrm{d}s \le -16 \int_\varepsilon^t s\partial_s(\Phi_s^\delta, \chi_k^2)_{L^2}\,\mathrm{d}s + c(t-\varepsilon)$$
$$= -16\Big[s(\Phi_s^\delta, \chi_k^2)_{L^2}\Big]_\varepsilon^t + 16\int_\varepsilon^t (\Phi_s^\delta, \chi_k^2)_{L^2}\,\mathrm{d}s + c(t-\varepsilon)\ .$$

Letting $\varepsilon \to 0$ yields, since $0 \le \Phi \le KL$,

$$\int_0^t 2\mathcal{E}(-\phi_s^\delta \chi_k)\,\mathrm{d}s \le 16KLat + ct\ .$$

Therefore, we conclude

$$2\mathcal{E}(-\bar\phi_t^\delta \chi_k) \le \frac{1}{t}\int_0^t 2\mathcal{E}(-\phi_s^\delta \chi_k)\,\mathrm{d}s \le 16KLa + c\ ,$$

where the first inequality follows from Lemma 5.2. This completes the proof of the boundedness of $\{\bar\phi_t^\delta \chi_k\}_{0<t<T_0,\,0<\delta<1}$ in $H_0^1(\mathsf{M})$.

By a similar calculation to (5.4), we find

$$2\mathcal{E}(-\Phi_t^\delta \chi_k) \le \left\|\phi'(u_t^\delta)^2 \chi_k F^*(-\mathbf{d}u_t^\delta) + \Phi_t^\delta F^*(-\mathbf{d}\chi_k)\right\|_{L^2}^2 \le 2V_t^\delta + 2(KL)^2 a\ .$$

Thus, we deduce from (5.6) that, for $t \in (0, T_0]$,

$$2\mathcal{E}(-\Phi_t^\delta \chi_k) \le -8t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + \frac{U_t^\delta}{2} + \frac{c}{2} \le -16t\partial_t(\Phi_t^\delta, \chi_k^2)_{L^2} + c\ ,$$

which yields

$$\int_0^t 2\mathcal{E}(-\Phi_s^\delta \chi_k)\,\mathrm{d}s \le 16KLat + ct \tag{5.7}$$

and the boundedness of $\{\bar\Phi_t^\delta \chi_k\}_{0<t<T_0,\,0<\delta<1}$ in $H_0^1(\mathsf{M})$ by the same reasoning as above. ∎

5.3. **Limit functions.** Thanks to Proposition 5.3 and the reflexivity of $H_0^1(\mathsf{M}) = H^1(\mathsf{M})$, up to extracting a (non-relabelled) subsequence, we have

$$\bar\phi_t^\delta \chi_k \xrightarrow{\delta\to 0,\, t\to 0} \exists \bar\phi_k\ , \qquad \text{weakly in } H_0^1(\mathsf{M})\ .$$

Since $\bar\phi_k = \bar\phi_l$ on B_k for $k \le l$, there exists a bounded nonnegative function $\bar\phi_0 \in H^1_{\mathrm{loc}}(\mathsf{M})$ such that $\bar\phi_0 = \bar\phi_k$ on B_k for every $k \in \mathbb{N}$. By the boundedness of Φ and Ψ, we may take a further (non-relabelled) subsequence such that, by passing $\delta \to 0$ and then $t \to 0$,

$$\Phi_t^\delta \xrightarrow{\delta\to 0,\, t\to 0} \exists \Phi_0\ , \qquad \bar\Phi_t^\delta \xrightarrow{\delta\to 0,\, t\to 0} \exists \bar\Phi_0\ , \qquad \bar\Psi_t^\delta \xrightarrow{\delta\to 0,\, t\to 0} \exists \bar\Psi_0$$

for some nonnegative functions $\Phi_0, \bar\Phi_0, \bar\Psi_0 \in L^\infty(\mathsf{M})$, both in the weak $L^2(\nu)$ sense for any finite measure ν mutually absolutely continuous with m and in the weak-star $L^\infty(\mathsf{m})$ sense.

Then the goal of this subsection is to prove the next proposition.

Proposition 5.4. *We have*

$$\bar\Phi_0 = \Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right) \quad a.e.$$

Recall Proposition 3.1 for $\bar{\mathsf{d}}_B$. To this end, we first observe the following.

Lemma 5.5. *We have*

$$\bar{\Phi}_0 \le \bar{\phi}_0 \le \frac{\bar{\mathsf{d}}_B^2}{2} \quad a.e.$$

Proof. We will use the same symbols as in Proposition 5.3. The former inequality $\bar{\Phi}_0 \le \bar{\phi}_0$ follows from $\Phi \le \phi$. To show the latter inequality, take a bounded nonnegative function $\rho \in H^1_0(\mathsf{M})$ with $\operatorname{supp}\rho \subset B_k$. Then, it follows from $\chi_k \equiv 1$ on B_k and Lemma 5.1 that

$$\begin{aligned} 2\mathcal{E}_\rho(-\phi^\delta_t\chi_k) &= 2\mathcal{E}_{\phi'(u^\delta_t)^2\rho}(-u^\delta_t) \\ &= -2t\partial_t(\Phi^\delta_t,\rho)_{L^2} + t\Big(\mathbf{d}\big(\phi'(u^\delta_t)^2\rho\big), \nabla(-u^\delta_t)\Big)_{L^2} + 2(\Psi^\delta_t,\rho)_{L^2} \ . \end{aligned}$$

By the Leibniz rule for $\mathbf{d}$ and (2.9), the second term can be bounded as

$$\begin{aligned} &\Big(\mathbf{d}\big(\phi'(u^\delta_t)^2\rho\big), \nabla(-u^\delta_t)\Big)_{L^2} \\ &= \big(\phi'(u^\delta_t)^2\,\mathbf{d}\rho, \nabla(-u^\delta_t)\big)_{L^2} + 2\big(\phi''(u^\delta_t)\,\mathbf{d}u^\delta_t, \nabla(-u^\delta_t)\big)_{L^2(\phi'(u^\delta_t)\rho)} \\ &\le \big(\mathbf{d}\rho, \nabla(-\Phi^\delta_t)\big)_{L^2} + \frac{4C}{K}\mathcal{E}_{\phi'(u^\delta_t)^2\rho}(-u^\delta_t) \\ &= \big(\mathbf{d}\rho, \nabla(-\Phi^\delta_t\chi_k)\big)_{L^2} + \frac{4C}{K}\mathcal{E}_\rho(-\phi^\delta_t) \ . \end{aligned}$$

Combining this with the above calculation and observing $\mathcal{E}_\rho(-\phi^\delta_t) = \mathcal{E}_\rho(-\phi^\delta_t\chi_k)$, we find

$$2\Big(1 - \frac{2Ct}{K}\Big)\mathcal{E}_\rho(-\phi^\delta_t\chi_k) \le -2t\partial_t(\Phi^\delta_t,\rho)_{L^2} + t\big(\mathbf{d}\rho, \nabla(-\Phi^\delta_t\chi_k)\big)_{L^2} + 2(\Psi^\delta_t,\rho)_{L^2} \ .$$

This implies, with the help of Fubini's theorem,

(5.8)

$$\begin{aligned} &2\Big(1 - \frac{2CT}{K}\Big)\frac{1}{T}\int_0^T \mathcal{E}_\rho(-\phi^\delta_t\chi_k)\,\mathrm{d}t \\ &\le -\frac{2}{T}\Big[t(\Phi^\delta_t,\rho)_{L^2}\Big]_0^T + \frac{2}{T}\int_0^T (\Phi^\delta_t,\rho)_{L^2}\,\mathrm{d}t + \frac{1}{T}\int_0^T t\big(\mathbf{d}\rho, \nabla(-\Phi^\delta_t\chi_k)\big)_{L^2}\,\mathrm{d}t + 2(\bar{\Psi}^\delta_T,\rho)_{L^2} \\ &= -2(\Phi^\delta_T,\rho)_{L^2} + 2(\bar{\Phi}^\delta_T,\rho)_{L^2} + \frac{1}{T}\int_0^T t\big(\mathbf{d}\rho, \nabla(-\Phi^\delta_t\chi_k)\big)_{L^2}\,\mathrm{d}t + 2(\bar{\Psi}^\delta_T,\rho)_{L^2} \ . \end{aligned}$$

Using the integration by parts, we can estimate the third term of the RHS as

$$\begin{aligned} &\frac{1}{T}\int_0^T t\big(\mathbf{d}\rho, \nabla(-\Phi^\delta_t\chi_k)\big)_{L^2}\,\mathrm{d}t \\ &= \frac{1}{T}\left[t\int_0^t \big(\mathbf{d}\rho, \nabla(-\Phi^\delta_s\chi_k)\big)_{L^2}\,\mathrm{d}s\right]_{t=0}^{t=T} - \frac{1}{T}\int_0^T\int_0^t \big(\mathbf{d}\rho, \nabla(-\Phi^\delta_s\chi_k)\big)_{L^2}\,\mathrm{d}s\,\mathrm{d}t \\ &\le \int_0^T 2\sqrt{\mathcal{E}(\rho)\mathcal{E}(-\Phi^\delta_s\chi_k)}\,\mathrm{d}s + \frac{1}{T}\int_0^T\int_0^t 2\sqrt{\mathcal{E}(-\rho)\mathcal{E}(-\Phi^\delta_s\chi_k)}\,\mathrm{d}s\,\mathrm{d}t \\ &\xrightarrow{\delta\to 0,\, T\to 0} 0 \ , \end{aligned}$$

where the convergence in the last line follows from the boundedness (5.7).

By the lower semi-continuity of $\mathcal{E}_\rho$, taking the limit of (5.8) as $\delta \to 0$, $T \to 0$ along the subsequence taken in the beginning of Subsection 5.3 yields

$$2\mathcal{E}_\rho(-\bar{\phi}_0) \le \big(-2\Phi_0 + 2\bar{\Phi}_0 + 2\bar{\Psi}_0, \rho\big)_{L^2} \le \big(4\bar{\phi}_0, \rho\big)_{L^2} \ ,$$

where we used $\rho \geq 0$ and $0 \leq \Psi \leq \Phi \leq \phi$ in (2.9) in the latter inequality. This implies that, for any $\varepsilon > 0$,

$$2\mathcal{E}_\rho\Big(-\sqrt{\bar\phi_0+\varepsilon}\Big) = \frac{1}{2}\mathcal{E}_{\rho/(\bar\phi_0+\varepsilon)}(-\bar\phi_0) \leq \left(\bar\phi_0, \frac{\rho}{\bar\phi_0+\varepsilon}\right)_{L^2} \leq \|\rho\|_{L^1} .$$

Letting $\varepsilon \to 0$ and noting that ρ was arbitrary, we find that $f_0 := \sqrt{\bar\phi_0} \in H^1_{\mathrm{loc}}(\mathsf{M})$ with $F^*(-\mathbf{d}f_0) \leq 1$ a.e. Moreover, a similar proof to [AH05, Lemma 4.4] shows that $\bar\phi_0 = 0$ a.e. on B. Thus, we have $f_0 \in \mathbb{L}_{B,\sqrt{KL}}$, and then it follows from Proposition 3.1 that

$$\bar\phi_0 \leq \bar{\mathsf{d}}_B^2 \quad \text{a.e.}$$

Finally, one can improve $\bar\phi_0 \leq \bar{\mathsf{d}}_B^2$ to $\bar\phi_0 \leq \bar{\mathsf{d}}_B^2/2$ by the same argument as in [AH05, Lemma 4.5] (see also [HR03, Lemma 2.12]) thanks to Proposition 4.1. ■

We can also show a partial converse inequality as follows.

Lemma 5.6. *We have*

$$\bar\Phi_0 \geq \Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right) \quad a.e.$$

Proof. We can follow the same lines as in [HR03, Lemma 2.13] by using Corollary 4.2. ■

Now, we are ready to prove Proposition 5.4.

Proof of Proposition 5.4. Thanks to Lemmas 5.5, 5.6, we can apply the argument in [HR03, Lemma 2.14] by replacing m with an equivalent finite measure ν. ■

5.4. **Tauberian argument.** By a Tauberian argument, we shall get rid of the time average from Proposition 5.4. Here we need additional assumptions to make use of linearised heat semigroups. Note that, for the linearised gradient operator as in (2.8), we have

$$\mathbf{d}h(\nabla f) = \sum_{i,j=1}^n g^*_{ij}(\mathbf{d}f)\frac{\partial f}{\partial x^j}\frac{\partial h}{\partial x^i} = \mathbf{d}f(\nabla^{\nabla f}h) . \tag{5.9}$$

Recall also that we set $u_t = \mathsf{T}_t\mathbf{1}_B$ and $\Phi_t = \Phi(-t\log\mathsf{T}_t\mathbf{1}_B)$ in Section 4.

Lemma 5.7. *Assume* $\mathsf{C}_F, \mathsf{S}_F < \infty$ *and* $\mathsf{m}(\mathsf{M}) < \infty$. *Then, for every* $\tau > 0$ *and measurable set* $D \subset \mathsf{M}$ *with* $\mathsf{m}(D) > 0$, *we have*

$$\lim_{t\downarrow 0}(\Phi_t, h^{\tau,\sigma}_{\tau-t})_{L^2} = (\bar\Phi_0, h^{\tau,\sigma}_\tau)_{L^2} ,$$

where $(h^{\tau,\sigma}_t)_{t\in[0,\tau]}$ *is the solution to the linearised heat equation*

$$\partial_t h^{\tau,\sigma}_t = \frac{1}{2}\Delta^{\nabla u_{\tau-t}}h^{\tau,\sigma}_t , \qquad h^{\tau,\sigma}_0 = \mathsf{T}_\sigma\mathbf{1}_D . \tag{5.10}$$

Remark 5.8. In (5.10), to be precise, we choose a measurable one-parameter family $(V_t)_{t\geq 0}$ of nowhere vanishing vector fields with $V_t(x) = \nabla u_t(x)$ when $\mathbf{d}u_t(x) \neq 0$, and replace $\nabla u_{\tau-t}$ with $V_{\tau-t}$. The unique existence of a solution $(h^{\tau,\sigma}_t)_{t\in[0,\tau]}$ is guaranteed in the same manner as that for (2.7) (see [Oht21, Proposition 13.20]) by virtue of the hypothesis $\mathsf{C}_F, \mathsf{S}_F < \infty$. We remark that $t \mapsto h^{\tau,\sigma}_t$ is L^2-continuous on $[0,\tau]$, and $\int_\mathsf{M} h^{\tau,\sigma}_t \,\mathrm{d}\mathsf{m} = \mathsf{m}(D)$ holds for all t (since constant functions belong to $L^2(\mathsf{M})$).

We also remark that choosing $\mathsf{T}_\sigma\mathbf{1}_D$ as the initial point is unessential; we may take any nonnegative function in $H^1_0(\mathsf{M})$ converging to $\mathbf{1}_D$ (see the very last step in Subsection 5.5).

Proof. We follow the argument in [HR03, Lemma 2.15] (see also [AH05, Lemma 4.6]), however, a modification is needed because of the asymmetry (2.6).

Since $\tau, \sigma > 0$ are fixed, we will denote $h_t^{\tau,\sigma}$ by h_t for brevity. For $t \in (0,\tau)$, put $H(t) := (\Phi_t, h_{\tau-t})_{L^2}$. For applying the Tauberian-type theorem in [Ram01, Lemma 3.11] which implies the claim $\lim_{t\downarrow 0} H(t) = (\bar{\Phi}_0, h_\tau)_{L^2}$, it suffices to see the following:

(a) $T^{-1}\int_0^T H(t)\,\mathrm{d}t \to (\bar{\Phi}_0, h_\tau)_{L^2}$ as $T \to 0$;
(b) There exist $M, t_0 > 0$ such that $H(t) - H(s) \le M(t-s)/s$ for all $0 < s < t \le t_0$.

The condition (a) follows from

$$\begin{aligned}\left|\frac{1}{T}\int_0^T H(t)\,\mathrm{d}t - (h_\tau, \bar{\Phi}_0)_{L^2}\right| &\le \frac{1}{T}\int_0^T \big|H(t) - (h_\tau, \Phi_t)_{L^2}\big|\,\mathrm{d}t + \big|(h_\tau, \bar{\Phi}_T - \bar{\Phi}_0)_{L^2}\big| \\ &\le \frac{KL}{T}\int_0^T \|h_{\tau-t} - h_\tau\|_{L^1}\,\mathrm{d}t + \big|(h_\tau, \bar{\Phi}_T - \bar{\Phi}_0)_{L^2}\big| \\ &\xrightarrow{T\to 0} 0\ .\end{aligned}$$

To see (b), we observe from Lemma 5.1 and (5.10) that

$$\begin{aligned}\Big[(h_{\tau-r}, \Phi_r^\delta)_{L^2}\Big]_{r=s}^{r=t} &= \int_s^t (h_{\tau-r}, \partial_r\Phi_r^\delta)_{L^2}\,\mathrm{d}r + \int_s^t (\partial_r h_{\tau-r}, \Phi_r^\delta)_{L^2}\,\mathrm{d}r \\ &= \int_s^t \frac{1}{r}(h_{\tau-r}, \Psi_r^\delta)_{L^2}\,\mathrm{d}r + \int_s^t \frac{1}{2}\Big(\mathbf{d}\big(\phi'(u_r^\delta)^2 h_{\tau-r}\big), \nabla(-u_r^\delta)\Big)_{L^2}\,\mathrm{d}r \\ &\quad - \int_s^t \frac{1}{r}\mathcal{E}_{\phi'(u_r^\delta)^2 h_{\tau-r}}(-u_r^\delta)\,\mathrm{d}r - \int_s^t \frac{1}{2}(\Delta^{\nabla u_r} h_{\tau-r}, \Phi_r^\delta)_{L^2}\,\mathrm{d}r \\ &=: I_1 + I_2 - I_3 - I_4\ .\end{aligned}$$

Note first that I_1 can be estimated as

$$I_1 \le \int_s^t \frac{1}{r} KL\,\mathsf{m}(D)\,\mathrm{d}r \le KL\,\mathsf{m}(D)\frac{t-s}{s}\ .$$

Next, since

$$\mathbf{d}\Phi_r^\delta = \phi'(u_r^\delta)^2\,\mathbf{d}u_r^\delta = -\phi'(u_r^\delta)^2\frac{t(1-\delta)}{(1-\delta)u_r + \delta}\cdot \mathbf{d}u_r$$

implies $g_{ij}^*(\mathbf{d}u_r) = g_{ij}^*(-\mathbf{d}\Phi_r^\delta)$, we deduce from (5.9) that

$$\begin{aligned}2I_4 &= \int_s^t \big(\mathbf{d}(-\Phi_r^\delta), \nabla^{\nabla u_r} h_{\tau-r}\big)_{L^2}\,\mathrm{d}r = \int_s^t \big(\mathbf{d}h_{\tau-r}, \nabla(-\Phi_r^\delta)\big)_{L^2}\,\mathrm{d}r \\ &= \int_s^t \big(\mathbf{d}h_{\tau-r}, \nabla(-u_r^\delta)\big)_{L^2(\phi'(u_r^\delta)^2)}\,\mathrm{d}r\end{aligned}$$

(we remark that the integrands in the above calculation vanish on the set $\{\mathbf{d}u_r = 0\}$). Then, using the Leibniz rule for $\mathbf{d}$ and (2.9), we find

$$\begin{aligned}I_2 &= \int_s^t \frac{1}{2}\big(\mathbf{d}h_{\tau-r}, \nabla(-u_r^\delta)\big)_{L^2(\phi'(u_r^\delta)^2)}\,\mathrm{d}r + \int_s^t \big(\phi''(u_r^\delta)\,\mathbf{d}u_r^\delta, \nabla(-u_r^\delta)\big)_{L^2(\phi'(u_r^\delta)h_{\tau-r})}\,\mathrm{d}r \\ &\le I_4 + \frac{2C}{K}\int_s^t \mathcal{E}_{\phi'(u_r^\delta)^2 h_{\tau-r}}(-u_r^\delta)\,\mathrm{d}r\ .\end{aligned}$$

Setting $t \le t_0 := K/(2C)$, we have $I_2 \le I_4 + I_3$. Therefore, we obtain

$$\Big[(h_{\tau-r}, \Phi_r^\delta)_{L^2}\Big]_{r=s}^{r=t} \le I_1 \le KL\,\mathsf{m}(D)\frac{t-s}{s} \quad \text{for all } 0 < s < t \le t_0\ .$$

Letting $\delta \to 0$ completes the proof. ■

5.5. **Proof of** (5.3)**.** We are now in a position to prove (5.3). We shall in fact show that

$$\lim_{t\downarrow 0} \int_D \Phi_t \,\mathsf{dm} = \int_D \Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right) \mathsf{dm} \tag{5.11}$$

for every measurable set $D \subset \mathsf{M}$ with $\mathsf{m}(D) > 0$, namely $\Phi(\bar{\mathsf{d}}_B^2/2)$ is the weak L^2-limit of Φ_t. For $\sigma > 0$ and $\tau > t > 0$, we decompose as

$$\int_D \Phi_t \,\mathsf{dm} = (\Phi_t, h^{\tau,\sigma}_{\tau-t})_{L^2} + (\Phi_t, \mathbf{1}_D - h^{\tau,\sigma}_{\tau-t})_{L^2} \ .$$

Taking the limit as $t \to 0$, we find from Lemma 5.7, Proposition 5.4 and $0 \le \Phi \le KL$ that

$$\left| \lim_{t\downarrow 0} \int_D \Phi_t \,\mathsf{dm} - \left(\Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2}\right), h^{\tau,\sigma}_\tau\right)_{L^2} \right| \le KL \|\mathbf{1}_D - h^{\tau,\sigma}_\tau\|_{L^1} \ . \tag{5.12}$$

We next consider the limit as $\tau \to 0$. Put $w_t := \mathsf{T}_t \mathbf{1}_D$ and observe that, for $t \in (0,\tau)$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\|w_{\sigma+t} - h^{\tau,\sigma}_t\|^2_{L^2}\right] \\
&= \int_{\mathsf{M}} (w_{\sigma+t} - h^{\tau,\sigma}_t)\big(\Delta w_{\sigma+t} - \Delta^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t\big) \,\mathsf{dm} \\
&= -\int_{\mathsf{M}} \mathbf{d}(w_{\sigma+t} - h^{\tau,\sigma}_t)\big(\nabla w_{\sigma+t} - \nabla^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t\big) \,\mathsf{dm} \\
&= -2\mathcal{E}(w_{\sigma+t}) - 2\mathcal{E}^{\nabla u_{\tau-t}}(h^{\tau,\sigma}_t) + \int_{\mathsf{M}} \big\{\mathbf{d}h^{\tau,\sigma}_t(\nabla w_{\sigma+t}) + \mathbf{d}w_{\sigma+t}\big(\nabla^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t\big)\big\} \,\mathsf{dm} \\
&\le -2\mathcal{E}(w_{\sigma+t}) - 2\mathcal{E}^{\nabla u_{\tau-t}}(h^{\tau,\sigma}_t) + \int_{\mathsf{M}} F^*(\mathbf{d}w_{\sigma+t})\big\{F^*(\mathbf{d}h^{\tau,\sigma}_t) + F\big(\nabla^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t\big)\big\} \,\mathsf{dm} \ ,
\end{aligned}$$

where we defined (recall (2.2) for $g^*_{\mathbf{d}u}$)

$$\mathcal{E}^{\nabla u}(h) := \frac{1}{2}\int_{\mathsf{M}} g^*_{\mathbf{d}u}(\mathbf{d}h, \mathbf{d}h) \,\mathsf{dm} = \frac{1}{2}\int_{\mathsf{M}} \mathbf{d}h(\nabla^{\nabla u} h) \,\mathsf{dm} \ .$$

Now, it follows from (2.5) that

$$F^*(\mathbf{d}h^{\tau,\sigma}_t)^2 \le \mathsf{S}_F \cdot g^*_{\mathbf{d}u_{\tau-t}}(\mathbf{d}h^{\tau,\sigma}_t, \mathbf{d}h^{\tau,\sigma}_t) \ ,$$

and similarly, by (2.4),

$$F\big(\nabla^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t\big)^2 \le \mathsf{C}_F \cdot g_{\nabla u_{\tau-t}}\big(\nabla^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t, \nabla^{\nabla u_{\tau-t}} h^{\tau,\sigma}_t\big) = \mathsf{C}_F \cdot g^*_{\mathbf{d}u_{\tau-t}}(\mathbf{d}h^{\tau,\sigma}_t, \mathbf{d}h^{\tau,\sigma}_t) \ .$$

Therefore,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\|w_{\sigma+t} - h^{\tau,\sigma}_t\|^2_{L^2}\right] \\
&\le -2\mathcal{E}(w_{\sigma+t}) - 2\mathcal{E}^{\nabla u_{\tau-t}}(h^{\tau,\sigma}_t) \\
&\quad + \int_{\mathsf{M}} \big(\sqrt{\mathsf{S}_F} + \sqrt{\mathsf{C}_F}\big) F^*(\mathbf{d}w_{\sigma+t}) \cdot \sqrt{g^*_{\mathbf{d}u_{\tau-t}}(\mathbf{d}h^{\tau,\sigma}_t, \mathbf{d}h^{\tau,\sigma}_t)} \,\mathsf{dm} \\
&\le -2\mathcal{E}(w_{\sigma+t}) - 2\mathcal{E}^{\nabla u_{\tau-t}}(h^{\tau,\sigma}_t) \\
&\quad + \int_{\mathsf{M}} \left\{\frac{1}{4}\big(\sqrt{\mathsf{S}_F} + \sqrt{\mathsf{C}_F}\big)^2 F^*(\mathbf{d}w_{\sigma+t})^2 + g^*_{\mathbf{d}u_{\tau-t}}(\mathbf{d}h^{\tau,\sigma}_t, \mathbf{d}h^{\tau,\sigma}_t)\right\} \mathsf{dm} \\
&= \left(\frac{(\sqrt{\mathsf{S}_F} + \sqrt{\mathsf{C}_F})^2}{2} - 2\right)\mathcal{E}(w_{\sigma+t}) \ .
\end{aligned}$$

Since $h_0^{\tau,\sigma} = w_\sigma$, integrating the above inequality in $t \in (0, \tau)$ yields

$$\|w_{\sigma+\tau} - h_\tau^{\tau,\sigma}\|_{L^2}^2 \leq \left(\frac{(\sqrt{\mathsf{S}_F} + \sqrt{\mathsf{C}_F})^2}{2} - 2 \right) \int_0^\tau \mathcal{E}(w_{\sigma+t}) \, \mathrm{d}t \ .$$

Plugging

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\|w_{\sigma+t}\|_{L^2}^2 \right] = \int_{\mathsf{M}} w_{\sigma+t} \Delta w_{\sigma+t} \, \mathrm{d}\mathsf{m} = -2\mathcal{E}(w_{\sigma+t})$$

into the RHS, we obtain

$$\|w_{\sigma+\tau} - h_\tau^{\tau,\sigma}\|_{L^2}^2 \leq \left(\frac{(\sqrt{\mathsf{S}_F} + \sqrt{\mathsf{C}_F})^2}{4} - 1 \right) \left(\|w_\sigma\|_{L^2}^2 - \|w_{\sigma+\tau}\|_{L^2}^2 \right) \ .$$

Hence, $h_\tau^{\tau,\sigma} \to w_\sigma$ in L^2 as $\tau \to 0$, and then (5.12) shows

$$\left| \lim_{t \downarrow 0} \int_D \Phi_t \, \mathrm{d}\mathsf{m} - \left(\Phi\left(\frac{\bar{\mathsf{d}}_B^2}{2} \right), w_\sigma \right)_{L^2} \right| \leq KL \|\mathbf{1}_D - w_\sigma\|_{L^1} \ .$$

Finally, as $\sigma \to 0$, we have $\|\mathbf{1}_D - w_\sigma\|_{L^1} \to 0$ since $w_\sigma \to \mathbf{1}_D$ in L^2 and $\mathsf{m}(\mathsf{M}) < \infty$. This completes the proof of (5.11) and hence the lower estimate (5.1).

5.6. **Further problems.** We conclude with some further problems.

First of all, in Theorem 1.1, we used the assumption $\mathsf{m}(\mathsf{M}) < \infty$ only for deducing the L^1-convergence from the L^2-convergence. We expect that a finer analysis of (non-linear and linearised) heat semigroups could remove it. Such an analysis will be helpful also for the further study of geometric analysis on noncompact Finsler manifolds, where some results are known only under seemingly artificial assumptions; we refer to [Oht22] for gradient estimates and an isoperimetric inequality, and to [Mai22] for a rigidity problem of the spectral gap.

In the non-smooth setting in Theorem 1.2, the lower bound estimate is an intriguing open problem. The main issue is whether we can avoid using a linearised heat semigroup in Lemma 5.7.

References

[ACD15] Ambrosio, L., Colombo, M., and Di Marino, S. Sobolev spaces in metric measure spaces: reflexivity and lower semicontinuity of slope. *Adv. Stud. Pure Math.*, 67:1–58, 2015.

[AGS08] Ambrosio, L., Gigli, N., and Savaré, G. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics - ETH Zürich. Birkhäuser, 2nd edition, 2008.

[AGS14] Ambrosio, L., Gigli, N., and Savaré, G. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 395:289–391, 2014.

[AH05] Ariyoshi, T. and Hino, M. Small-time asymptotic estimates in local Dirichlet spaces. *Electron. J. Probab.*, 10:1236–1259, 2005.

[AK01] Aida, S. and Kawabi, H. Short time asymptotics of a certain infinite dimensional diffusion process. In Decreusefond, L., Øksendal, B. K., and Üstünel, A. S., editors, *Stochastic Analysis and Related Topics VII – Proceedings of the Seventh*

Silivri Workshop, volume 48 of *Progress in Probability*, pages 77–124. Birkhäuser, 2001.

[AZ02] Aida, S. and Zhang, T. S. On the small time asymptotics of diffusion processes on path groups. *Potential Anal.*, 16:67–78, 2002.

[BN18] Bailleul, I. and Norris, J. R. Diffusion in small time in incomplete sub-Riemannian manifolds. *arXiv:1810.06328*, 2018.

[CKS87] Carlen, E. A., Kusuoka, S., and Stroock, D. W. Upper bounds for symmetric Markov transition functions. *Ann. I. H. Poincaré B*, 23(S2):245–287, 1987.

[DS21] Dello Schiavo, L. and Suzuki, K. On the Rademacher and Sobolev-to-Lipschitz properties for strongly local Dirichlet spaces. *J. Func. Anal.*, 281(11):Online first, 2021.

[DS22a] Dello Schiavo, L. and Suzuki, K. Configuration spaces over singular spaces II – Curvature. *arXiv:2205.01379*, 2022.

[DS22b] Dello Schiavo, L. and Suzuki, K. Sobolev-to-Lipschitz property on QCD-spaces and applications. *Math. Ann.*, 384:1815–1832, 2022.

[EG15] Evans, C. L. and Gariepy, F. R. *Measure theory and fine properties of functions.* Textbook in Mathematics, CRC Press, 2015.

[Fan94] Fang, S. On the Ornstein–Uhlenbeck process. *Stoch. Stoch. Rep.*, 46:141–159, 1994.

[FZ99] Fang, S. and Zhang, T. S. On the small time behavior of Ornstein–Uhlenbeck processes with unbounded linear drifts. *Probab. Theory Relat. Fields*, 114:487–504, 1999.

[Gig18] Gigli, N. Nonsmooth differential geometry – An approach tailored for spaces with Ricci curvature bounded from below. *Mem. Am. Math. Soc.*, 251(1196), 2018.

[GS01] Ge, Y. and Shen, Z. Eigenvalues and eigenfunctions of metric measure manifolds. *Proc. London Math. Soc. (3)*, 82:725–746, 2001.

[HM18] Hino, M. and Matsuura, K. An integrated version of Varadhan's asymptotics for lower-order perturbations of strong local Dirichlet forms. *Potent. Anal.*, 48:257–300, 2018.

[HR03] Hino, M. and Ramírez, J. A. Small-time Gaussian behavior of symmetric diffusion semigroups. *Ann. Probab.*, 31(3):1254–1295, 2003.

[KZ22] Kristály, A. and Zhao, W. On the geometry of irreversible metric-measure spaces: Convergence, stability and analytic aspects. *J. Math. Pures Appl. (9)*, 158:216–292, 2022.

[Léa87a] R. Léandre. Majoration en temps petit de la densité d'une diffusion dégénérée. *Probab. Theory Relat. Fields*, 74(2):289–294, 1987.

[Léa87b] R. Léandre. Minoration en temps petit de la densité d'une diffusion dégénérée. *J. Funct. Anal.*, 74(2):399–414, 1987.

[Mai22] Mai, C. H. A rigidity result of spectral gap on Finsler manifolds and its application. *arXiv:2204.02298*, 2022.

[Nor97] Norris, J. R. Heat kernel asymptotics and the distance function in Lipschitz Riemannian manifolds. *Acta Math.*, 179:79–103, 1997.

[Oht09] Ohta, S. Uniform convexity and smoothness, and their applications in Finsler geometry. *Math. Ann.*, 343:669–699, 2009.

[Oht17a] Ohta, S. Nonlinear geometric analysis on Finsler manifolds. *Eur. J. Math*, 3:916–952, 2017.

[Oht17b] Ohta, S. Some functional inequalities on non-reversible Finsler manifolds. *Proc. Indian Acad. Sci. Math. Sci.*, 127:833–855, 2017.

[Oht21] Ohta, S. *Comparison Finsler geometry.* Springer Monographs in Mathematics, 2021.

[Oht22] Ohta, S. A semigroup approach to Finsler geometry: Bakry–Ledoux's isoperimetric inequality. *Comm. Anal. Geom.*, 30:2347–2387, 2022.

[OS09] Ohta, S. and Sturm, K.-T. Heat flow on Finsler manifolds. *Comm. Pure Appl. Math.*, 62:1386–1433, 2009.

[OS14] Ohta, S. and Sturm, K.-T. Bochner–Weitzenböck formula and Li–Yau estimates on Finsler manifolds. *Adv. Math.*, 252:429–448, 2014.

[OZ22] Ohta, S. and Zhao, W. Gradient flows in asymmetric metric spaces and applications. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (to appear)*, arXiv:2206.07591, 2022.

[Ram01] Ramírez, J. Short time asymptotics in Dirichlet spaces. *Comm. Pure Appl. Math.*, 54(3):259–293, 2001.

[RS09] Renesse, M.-K. von and Sturm, K.-T. Entropic measure and Wasserstein diffusion. *Ann. Probab.*, 37(3):1114–1191, 2009.

[Var67] Varadhan, S. R. S. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Comm. Pure Appl. Math.*, 20(2):431–455, 1967.

[Zha01] Zhang, T. S. On the small time large deviations of diffusion processes on configuration spaces. *Stoch. Proc. Appl.*, 91(2):239–254, 2001.

Department of Mathematics, Osaka University, Osaka 560-0043, Japan, and RIKEN Center for Advanced Intelligence Project (AIP), 1-4-1 Nihonbashi, Tokyo 103-0027, Japan

Email address: `s.ohta@math.sci.osaka-u.ac.jp`

Department of Mathematical Science, Durham University, DH13LE, Durham, United Kingdom

Email address: `kohei.suzuki@durham.ac.uk`