Displacement convexity
of generalized relative entropies. II

Shin-ichi Ohta[*]and Asuka Takatsu[†]

April 9, 2013

Abstract

We introduce a class of generalized relative entropies (inspired by the Bregman divergence in information theory) on the Wasserstein space over a weighted Riemannian or Finsler manifold. We prove that the convexity of all the entropies in this class is equivalent to the combination of the nonnegative weighted Ricci curvature and the convexity of another weight function used in the definition of the generalized relative entropies. This convexity condition corresponds to Lott and Villani's version of the curvature-dimension condition. As applications, we obtain appropriate variants of the Talagrand, HWI and logarithmic Sobolev inequalities, as well as the concentration of measures. We also investigate the gradient flow of our generalized relative entropy.

Contents

1 Introduction

This is a continuation of our work [OT] on the displacement convexity of generalized entropies and its applications. We consider more general entropies than [OT] and generalize most results in appropriate ways. Some of our observation shall shed new light on [OT].

It has been known since the celebrated work of McCann [Mc1] that the convexity of an energy (entropy) functional along geodesics in the Wasserstein space plays a vital role in the study of the existence and the uniqueness of a ground state (a minimizer of the energy). Here the (quadratic) *Wasserstein space* over a complete separable metric space (X, d) is the space $\mathcal{P}^2(X)$ of Borel probability measures on X having finite second moments, endowed with the *Wasserstein distance function* W_2 derived from the Monge–Kantorovich mass transport problem (see Subsection 2.2). We say that a functional S on $\mathcal{P}^2(X)$ is *displacement K-convex* for $K \in \mathbb{R}$ (Hess $S \geq K$ for short) if any pair

$\mu_0, \mu_1 \in \mathcal{P}^2(X)$ can be joined by a minimal geodesic $(\mu_t)_{t \in [0,1]}$ in $(\mathcal{P}^2(X), W_2)$ such that

$$S(\mu_t) \leq (1-t)S(\mu_0) + tS(\mu_1) - \frac{K}{2}(1-t)tW_2(\mu_0, \mu_1)^2$$

holds for all $t \in [0,1]$. As usual, the displacement 0-convexity may be simply called the *displacement convexity*. The word 'displacement' is inserted for avoiding a confusion with the convexity along the linear interpolation $S((1-t)\mu_0 + t\mu_1) \leq (1-t)S(\mu_0) + tS(\mu_1)$. Since we deal with only the displacement convexity, we may sometimes omit 'displacement'.

As any geodesic in the Wasserstein space is written as the transport along geodesics in the underlying metric space, the displacement convexity of an energy functional can be derived from the convexity of its generating function. For instance, let S_u^Ψ be the *free energy functional* on $\mathcal{P}^2(\mathbb{R}^n)$ consisting of the internal energy and the potential energy as

$$S_u^\Psi(\mu) := \int_{\mathbb{R}^n} u\left(\frac{d\mu}{d\mathcal{L}^n}\right) d\mathcal{L}^n + \int_{\mathbb{R}^n} \Psi \, d\mu$$

for absolutely continuous probability measures μ on $\mathbb{R}^n$ with respect to the Lebesgue measure $\mathcal{L}^n$, where the energy density u is a function on $\mathbb{R}$ and the potential Ψ is a function on $\mathbb{R}^n$. Then S_u^Ψ is strictly displacement convex if u is convex (and satisfies certain additional conditions to be precise, see Definition 3.1) and Ψ is strictly convex, and the unique ground state $\nu := \sigma\mathcal{L}^n$ satisfies $u'(\sigma) = -\Psi + \lambda$ with a normalizing constant λ. We mention that the uniqueness is measured at the level of the energy functional, that is, we have $S_u^\Psi(\mu) - S_u^\Psi(\nu) \geq 0$ and equality holds if and only if $\mu = \nu$. Moreover, the displacement convexity of the free energy S_u^Ψ is a crucial tool also in the investigation of the asymptotic behavior of the solution to the associated evolution equation

$$\frac{\partial \rho}{\partial t} = \mathrm{div}\left(\rho \nabla[u'(\rho)] + \rho \nabla \Psi\right)$$

by regarding it as the gradient flow of S_u^Ψ in the Wasserstein space (see [JKO], [AGS1], [CMV1] and [CMV2] among others). In particular, the heat flow is regarded as the gradient flow of the *relative entropy* (with respect to the Lebesgue measure)

$$\mathrm{Ent}_{\mathcal{L}^n}(\mu) := \int_{\mathbb{R}^n} \frac{d\mu}{d\mathcal{L}^n} \ln\left(\frac{d\mu}{d\mathcal{L}^n}\right) d\mathcal{L}^n,$$

which is also called the *Kullback–Leibler divergence* in information theory.

On curved spaces such as Riemannian manifolds, the displacement convexity of energy functionals is related to the curvature of the underlying space, that is a crucial difference from the convexity along linear interpolations $(1-t)\mu_0 + t\mu_1$. On a Riemannian manifold equipped with the Riemannian volume measure vol_g, the relative entropy is similarly defined by

$$\mathrm{Ent}_{\mathrm{vol}_g}(\mu) := \int_M \frac{d\mu}{d\mathrm{vol}_g} \ln\left(\frac{d\mu}{d\mathrm{vol}_g}\right) d\mathrm{vol}_g.$$

It has been shown by von Renesse and Sturm [vRS] (inspired by [CMS1] and [OV]) that for any $K \in \mathbb{R}$ the following are mutually equivalent:

- The relative entropy $\mathrm{Ent}_{\mathrm{vol}_g}$ is displacement K-convex on $(\mathcal{P}^2(M), W_2)$.

- The Ricci curvature is bounded from below by K as $\mathrm{Ric}_g(\mathbf{v}, \mathbf{v}) \geq K\langle \mathbf{v}, \mathbf{v}\rangle$ for all $\mathbf{v} \in TM$.

- The heat flow is K-*contractive* in the sense that $W_2(\mu_t, \tilde{\mu}_t) \leq e^{-Kt} W_2(\mu_0, \tilde{\mu}_0)$ holds for all $t \geq 0$ and for any weak solutions $(\rho_t)_{t \geq 0}$, $(\tilde{\rho}_t)_{t \geq 0}$ to the heat equation $\partial \rho / \partial t = \Delta \rho$ such that $\mu_t := \rho_t \, \mathrm{vol}_g, \tilde{\mu}_t := \tilde{\rho}_t \, \mathrm{vol}_g \in \mathcal{P}^2(M)$.

See also [AGS1], [AGS3], [Oh1], [Sa] and [Vi2, Chapter 23] for the connection between the K-convexity of the functional and the K-contraction property of its gradient flow.

The displacement K-convexity of $\mathrm{Ent}_{\mathrm{vol}_g}$ ($\mathrm{Hess}\,\mathrm{Ent}_{\mathrm{vol}_g} \geq K$) is called the *curvature-dimension condition* $\mathsf{CD}(K, \infty)$ after Bakry and Émery's pioneering work [BE]. One remarkable point of $\mathsf{CD}(K, \infty)$ is that it can be formulated on general metric measure spaces without any differentiable (manifold) structure. Such metric measure spaces *with Ricci curvature bounded below* are independently investigated by Sturm [St2] and Lott and Villani [LV2], and known to enjoy several properties common to Riemannian manifolds of $\mathrm{Ric}_g \geq K$. For example, as was indicated by Otto and Villani [OV], $\mathsf{CD}(K, \infty)$ with $K > 0$ implies various functional inequalities such as the *Talagrand inequality*, the *HWI inequality*, the *logarithmic Sobolev inequality* and the *global Poincaré inequality* ([LV2, Section 6]).

The curvature-dimension condition $\mathsf{CD}(K, \infty)$ is generalized to $\mathsf{CD}(K, N)$ for each $K \in \mathbb{R}$ and $N \in (1, \infty]$. On an n-dimensional complete connected Riemannian manifold (M, g) of $n \geq 2$ equipped with a weighted measure $\omega = e^{-f} \mathrm{vol}_g$ with $f \in C^\infty(M)$, the condition $\mathsf{CD}(K, N)$ is known to be equivalent to the lower bound of the N-Ricci curvature $\mathrm{Ric}_N(\mathbf{v}, \mathbf{v}) \geq K\langle \mathbf{v}, \mathbf{v}\rangle$ ([St1], [St3], [LV1], [LV2], see Definition 2.1 for the definition of Ric_N). In particular, an unweighted Riemannian manifold (M, vol_g) satisfies $\mathsf{CD}(K, N)$ if and only if its Ricci curvature is bounded below by K and its dimension is bounded above by N. We remark that Sturm's and Lott and Villani's definitions of the curvature-dimension condition are slightly different, though they are equivalent on non-branching spaces such as Riemannian or Finsler manifolds. In both cases it is a certain convexity condition of a class of entropies, and Lott and Villani's class is larger than Sturm's one.

On non-branching metric measure spaces, the condition $\mathsf{CD}(0, N)$ for $N \in [n, \infty)$ is equivalent to the displacement convexity of the *Rényi entropy*

$$S_N(\mu) := -\int_M \left(\frac{d\mu}{d\omega}\right)^{(N-1)/N} d\omega.$$

For $K \neq 0$, however, $\mathsf{CD}(K, N)$ is not simply the displacement K-convexity of S_N. In fact, it was shown in [St1] (see also [OT, Theorem 4.1, Remark 4.3(2)] and [BS]) that, on a weighted Riemannian manifold (M, ω), $\mathrm{Hess}\,S_N \geq K$ can hold only for $K \leq 0$ and is equivalent to $\mathrm{Ric}_N \geq 0$ regardless of the value of $K \leq 0$. It was also observed in [St1, Theorem 1.7] for unweighted Riemannian manifolds that there are some functionals whose displacement K-convexity characterizes the combination of $\mathrm{Ric} \geq K$ and $\dim \leq N$, whereas it is unclear if there are any applications of these entropies.

In our previous work [OT], we introduced the *m-relative entropy* H_m for the parameter $m \in [(n-1)/n, 1) \cup (1, \infty)$ inspired by the Bregman divergence in information

theory/geometry (see [Am], [AN]) as well as the Tsallis entropy in statistical mechanics (see [Ts1], [Ts2]). We fix a reference measure $\nu = \exp_m(-\Psi)\omega$ on a weighted Riemannian manifold (M, ω) involving the *m-exponential function*

$$\exp_m(t) := \max\{1 + (m-1)t, 0\}^{1/(m-1)},$$

then the *m-relative entropy* of an absolutely continuous measure $\mu \in \mathcal{P}^2(M)$ with respect to ν is given by (up to an additive constant)

$$H_m(\mu) := \frac{1}{m(m-1)} \int_M \left\{ \left(\frac{d\mu}{d\omega}\right)^m - m\frac{d\mu}{d\omega}\left(\frac{d\nu}{d\omega}\right)^{m-1} \right\} d\omega.$$

This generalizes the relative and the Rényi entropies in the sense that $\lim_{m\to 1} H_m(\mu) = \mathrm{Ent}_\nu(\mu) - 1$ and that $H_m(\mu) = N\{m^{-1}S_N(\mu) + 1\}$ with $N = 1/(1-m)$ if $\Psi \equiv 0$ (i.e., $\nu = \omega$).

Then the displacement K-convexity of H_m is equivalent to the combination of $\mathrm{Ric}_N \geq 0$ (of (M, ω)) and $\mathrm{Hess}\,\Psi \geq K$ ([OT, Theorem 4.1]). We stress that N becomes *negative* for $m > 1$, then Ric_N is defined in the same form as the case of $N \in (n, \infty)$ (see Definition 2.1). Similarly to $\mathsf{CD}(K, \infty)$, we can derive from $\mathrm{Hess}\,H_m \geq K > 0$ the associated functional inequalities (see also [AGK], [CGH], [Ta1] for related works) and the concentration of measures (in terms of $\exp_m$). Furthermore, the gradient flow of H_m produces weak solutions to the fast diffusion equation ($m < 1$) or the porous medium equation ($m > 1$) with drift of the form

$$\frac{\partial \rho}{\partial t} = \mathrm{div}_\omega \left(\frac{1}{m}\nabla(\rho^m) + \rho\nabla\Psi \right),$$

where div_ω is the divergence of (M, ω) (see also [Ot], [Vi2, Theorem 23.19]). We remark that Sturm [St1] studied a more general class of entropies on unweighted Riemannian manifolds, where $\mathrm{Ric}_N = \mathrm{Ric}$ for all N. Compared to it, [OT] gave a detailed investigation of a concrete class of entropies, on more general weighted Riemannian manifolds (by choosing appropriate parameters N).

In this article, we introduce the more general class of entropies, called the *φ-relative entropies* H_φ, again inspired by information theory/geometry. Here $\varphi : (0, \infty) \longrightarrow (0, \infty)$ is a non-decreasing, positive, continuous function. Roughly speaking, our new class corresponds to Lott and Villani's class of entropies in their definition of the curvature-dimension condition, while the *m-relative entropies* in [OT] correspond to Sturm's class. The definition of H_φ (see Definition 5.3 for details) involves $\nu = \exp_\varphi(-\Psi)$ with the *φ-exponential function* $\exp_\varphi$ which is the inverse function of the *φ-logarithmic function* $\ln_\varphi(t) := \int_1^t \varphi(s)^{-1}\,ds$. We recover $\exp_m$ and H_m from $\varphi(s) = s^{2-m}$.

Our first main theorem (Theorem 5.7) asserts that $\mathrm{Hess}\,H_m \geq K$ is equivalent to $\mathrm{Hess}\,H_\varphi \geq K$ for all φ's in a certain class. This actually corresponds to the equivalence between Sturm's and Lott and Villani's curvature-dimension conditions on weighted Riemannian manifolds. This reveals that H_m is an extremal element among H_φ's in the appropriate class, see [Ta2] for a related work. Similarly to H_m, we can derive from $\mathrm{Hess}\,H_\varphi \geq K > 0$ the variants of the Talagrand, HWI, logarithmic Sobolev, and global

Poincaré inequalities (Theorem 6.3) as well as the concentration of measures in terms of $\exp_m$ for some $m = m(\varphi)$ (Theorem 7.9). Moreover, the gradient flow of H_φ in $(\mathcal{P}^2(M), W_2)$ produces weak solutions to the φ-heat equation (Theorems 8.7, 9.7)

$$\frac{\partial \rho}{\partial t} = \operatorname{div}_\omega \left(\frac{\rho \nabla \rho}{\varphi(\rho)} + \rho \nabla \Psi \right).$$

The article is organized as follows: We first review the basic notions of weighted Riemannian geometry, Wasserstein geometry and information geometry in Section 2. Then, after preparing necessary notions in Sections 3, 4, we define H_φ and study its displacement convexity in Section 5. Section 6 is devoted to the functional inequalities and Section 7 is concerned with the concentration of measures. The gradient flow of H_φ is studied in Sections 8, 9 in the compact and noncompact cases, respectively. We extend most results to Finsler manifolds in Section 10. Finally in Appendix, we compare our concentration of measures derived from the generalized Talagrand inequality with the Herbst-type argument deriving the concentration from the u_φ-entropy inequality, which is a generalization of the logarithmic Sobolev inequality different from ours.

2 Preliminaries

2.1 Weighted Riemannian manifolds

Throughout the article except Section 10, (M, g) will be an n-dimensional complete connected Riemannian manifold without boundary. As we are interested in the role of the curvature, we will always assume $n \geq 2$. Denote by d_g and vol_g the Riemannian distance function and the Riemannian volume measure of (M, g). We fix an arbitrary measure

$$\omega = e^{-f} \operatorname{vol}_g, \quad f \in C^\infty(M),$$

as our base measure. To control the behavior of ω, we modify the Ricci curvature Ric_g of (M, g) as follows.

Definition 2.1 (Weighted Ricci curvature) Given $N \in (-\infty, 0) \cup [n, \infty]$, we define the *N-Ricci curvature tensor* of (M, ω) by

$$\operatorname{Ric}_N := \begin{cases} \operatorname{Ric}_g + \operatorname{Hess}_g f & \text{if } N = \infty, \\ \operatorname{Ric}_g + \operatorname{Hess}_g f - \dfrac{Df \otimes Df}{N - n} & \text{if } N \in (-\infty, 0) \cup (n, \infty), \\ \operatorname{Ric}_g + \operatorname{Hess}_g f - \infty \cdot (Df \otimes Df) & \text{if } N = n, \end{cases}$$

where by convention $\infty \cdot 0 = 0$.

We set $\operatorname{Ric}_N(\mathbf{v}) := \operatorname{Ric}_N(\mathbf{v}, \mathbf{v})$ and will say that $\operatorname{Ric}_N \geq K$ holds for some $K \in \mathbb{R}$ if $\operatorname{Ric}_N(\mathbf{v}) \geq K \langle \mathbf{v}, \mathbf{v} \rangle$ for every $\mathbf{v} \in TM$.

Remark 2.2 The tensor Ric_N was usually considered only for $N \in [n, \infty]$, and then the monotonicity $\operatorname{Ric}_N(\mathbf{v}) \leq \operatorname{Ric}_{N'}(\mathbf{v})$ for $N < N'$ clearly holds. Note that $\operatorname{Ric}_\infty$ is the

famous *Bakry–Émery tensor* and Ric_N for $N \in (n, \infty)$ was introduced by Qian (see [BE], [Qi] and [Lo] as well). Extending the range of N to $(-\infty, 0) \cup [n, \infty]$ violates the above monotonicity in N, however, observe that Ric_N is non-decreasing in the parameter

$$ m := 1 - \frac{1}{N} \in \left[1 - \frac{1}{n}, \infty \right), \quad \text{where } m := 1 \text{ if } N = \infty. $$

This observation will be helpful for understanding the validity of Theorem 5.7 below.

Note that, if (M, ω) satisfies $\mathrm{Ric}_N \geq K$ for some $K \in \mathbb{R}$ and $N \in [n, \infty)$, then it behaves like a Riemannian manifold with dimension bounded above by N and Ricci curvature bounded below by K (see [Qi], [Lo], as well as [St2], [St3], [LV1], [LV2], [Vi2, Part III] related to the curvature-dimension condition). For example, the following area growth inequality of Bishop type (numerically extended to non-integer N's) holds. Denote by $\mathrm{area}_\omega[S(x_0, r)]$ the area of the sphere $S(x_0, r) := \{x \in M \mid d_g(x_0, x) = r\}$ with respect to ω.

Theorem 2.3 ([Qi], [St3, Theorem 2.3]) *If (M, ω) satisfies $\mathrm{Ric}_N \geq 0$ for some $N \in [n, \infty)$, then*

$$ \mathrm{area}_\omega[S(x_0, R)] \leq \mathrm{area}_\omega[S(x_0, r)] \cdot \left(\frac{R}{r} \right)^{N-1} $$

holds for any $0 < r < R$ and $x_0 \in M$.

For $N = \infty$, we have the following global estimate.

Theorem 2.4 ([Vi2, Theorem 18.12]) *Under the nonnegativity of Ric_∞ of (M, ω),*

$$ \int_M \exp\left(-\lambda d_g(x_0, x)^2 \right) d\omega(x) < \infty $$

holds for any $\lambda > 0$ and $x_0 \in M$.

Though Theorems 2.3, 2.4 are generalized to $\mathrm{Ric}_N \geq K$ for $K \neq 0$, we will need only the above special cases.

2.2 Wasserstein geometry

Let us recall some basic notions and facts in optimal transport theory and Wasserstein geometry. See [AGS1], [Vi1] and [Vi2] for details and more information.

Let (X, d) be a metric space. A rectifiable curve $\gamma : [0, 1] \longrightarrow X$ is called a *geodesic* if it is locally minimizing and has a constant speed. We say that γ is *minimal* if it is globally minimizing, namely $d(\gamma(s), \gamma(t)) = |s - t| d(\gamma(0), \gamma(1))$ holds for all $s, t \in [0, 1]$. A subset Y of X is said to be *totally convex* if, for any $x, y \in Y$, any minimal geodesic in X from x to y is contained in Y.

For a complete Riemannian manifold (M, g), let $\mathcal{P}(M)$ be the set of all Borel probability measures on M. Given $\mu \in \mathcal{P}(M)$ and a measurable map $\mathcal{T} : M \longrightarrow M$, the *push forward measure* $\mathcal{T}_\sharp \mu$ of μ through $\mathcal{T}$ is defined by $\mathcal{T}_\sharp \mu[B] := \mu[\mathcal{T}^{-1}(B)]$ for all Borel sets $B \subset M$. For each $p \in [1, \infty)$, denote by $\mathcal{P}^p(M) \subset \mathcal{P}(M)$ the subset consisting of

measures μ of finite p-th moments, that is, $\int_M d_g(x_0, x)^p \, d\mu(x) < \infty$ for some (and hence all) $x_0 \in M$.

For $\mu, \nu \in \mathcal{P}(M)$, a probability measure $\pi \in \mathcal{P}(M \times M)$ is called a *coupling* of μ and ν if its projections are μ and ν, namely $\pi[B \times M] = \mu[B]$ and $\pi[M \times B] = \nu[B]$ hold for any Borel set $B \subset M$. We define the *L^p-Wasserstein distance* between $\mu, \nu \in \mathcal{P}^p(M)$ by

$$W_p(\mu, \nu) := \inf \left\{ \left(\int_{M \times M} d_g(x, y)^p \, d\pi(x, y) \right)^{1/p} \,\middle|\, \pi\text{: a coupling of } \mu \text{ and } \nu \right\}.$$

A coupling π is said to be *optimal* if it attains the infimum above. The function W_p is indeed a distance function on $\mathcal{P}^p(M)$. The metric space $(\mathcal{P}^p(M), W_p)$ is complete, separable and called the *L^p-Wasserstein space* over M. The Wasserstein space inherits several properties of M. For instance, if M is compact, then $(\mathcal{P}^p(M), W_p)$ is also compact and the topology induced from W_p coincides with the weak topology. We will mainly consider the quadratic case $p = 2$, and then we omit 'L^2-' and simply call W_2 and $(\mathcal{P}^2(M), W_2)$ the Wasserstein distance function and the Wasserstein space.

In view of optimal transport theory, $W_2(\mu_0, \mu_1)^2$ is regarded as the least cost of transporting μ_0 to μ_1, where the cost of transporting a unit mass from x to y is $d_g(x, y)^2$. A minimal geodesic $(\mu_t)_{t \in [0,1]}$ with respect to W_2 is then also called the *optimal transport* from μ_0 to μ_1, and it can be described by using a family of minimal geodesics in the underlying space M. We denote by $\Gamma(M)$ the set of all minimal geodesics $\gamma : [0, 1] \longrightarrow M$ endowed with the uniform topology induced from the distance function $d_{\Gamma(M)}(\gamma, \eta) := \sup_{t \in [0,1]} d_g(\gamma(t), \eta(t))$. For $t \in [0, 1]$, the *evaluation map* $\mathrm{ev}_t : \Gamma(M) \longrightarrow M$ is defined by $\mathrm{ev}_t(\gamma) := \gamma(t)$, which is clearly 1-Lipschitz.

Proposition 2.5 ([LV2, Proposition 2.10], [Vi2, Corollary 7.22]) *Given any minimal geodesic $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}^2(M)$, there exists $\Pi \in \mathcal{P}(\Gamma(M))$ such that $(\mathrm{ev}_t)_\sharp \Pi = \mu_t$ for all $t \in [0, 1]$ and that $(\mathrm{ev}_0 \times \mathrm{ev}_1)_\sharp \Pi$ is an optimal coupling of μ_0 and μ_1.*

In particular, for any totally convex set X of (M, d_g), $\mathcal{P}^2(X)$ is also totally convex in $(\mathcal{P}^2(M), W_2)$.

If one of μ_0 and μ_1 is absolutely continuous with respect to vol_g, then a more precise description of a minimal geodesic $(\mu_t)_{t \in [0,1]}$ is obtained via the gradient vector field of a *locally semi-convex* function ϕ (i.e., every point $x \in M$ admits a neighborhood on which ϕ is K-convex in the weak sense for some $K \in \mathbb{R}$, see Definition 4.1). For a measure ν on M, we denote by $\mathcal{P}_{\mathrm{ac}}(M, \nu) \subset \mathcal{P}(M)$ the subset of absolutely continuous measures with respect to ν. We also set $\mathcal{P}_{\mathrm{ac}}^2(M, \nu) := \mathcal{P}^2(M) \cap \mathcal{P}_{\mathrm{ac}}(M, \nu)$.

Theorem 2.6 ([FG, Theorem 1]) *Given any $\mu_0 \in \mathcal{P}_{\mathrm{ac}}^2(M, \mathrm{vol}_g)$ and $\mu_1 \in \mathcal{P}^2(M)$, there exists a locally semi-convex function $\phi : \Omega \longrightarrow \mathbb{R}$ on an open set $\Omega \subset M$ with $\mu_0[\Omega] = 1$ such that the map $\mathcal{T}_t(x) := \exp_x(t \nabla \phi(x))$, $t \in [0, 1]$, provides a unique minimal geodesic from μ_0 to μ_1. Precisely, $(\mathcal{T}_0 \times \mathcal{T}_1)_\sharp \mu_0$ is a unique optimal coupling of μ_0 and μ_1, and $\mu_t := (\mathcal{T}_t)_\sharp \mu_0$ is a unique minimal geodesic from μ_0 to μ_1 with respect to W_2.*

If M is compact, then the above theorem is due to McCann's celebrated work [Mc2] and we can take as the potential function $-\phi$ a *c-concave function* for the cost $c(x, y) =$

8

$d_g(x, y)^2/2$. (We do not give the definition of the c-concave function, what we need is only the fact that c-concave functions are locally semi-convex.) A locally semi-convex function is locally Lipschitz and twice differentiable almost everywhere by the Alexandrov–Bangert theorem. Thus $\mathcal{T}_t$ is differentiable μ_0-a.e. and the following *Jacobian* (or *Monge–Amperè*) *equation* holds.

Theorem 2.7 ([Vi2, Theorems 8.7, 11.1]) *Under the same assumptions as Theorem* 2.6 *above, we have* $\mu_t \in \mathcal{P}^2_{\mathrm{ac}}(M, \mathrm{vol}_g)$ *for all* $t \in [0, 1)$. *Moreover, by putting*

$$\rho_t \omega := \mu_t = (\mathcal{T}_t)_\sharp \mu, \qquad \mathbf{J}^\omega_t(x) := e^{f(x) - f(\mathcal{T}_t(x))} \det\big(D\mathcal{T}_t(x)\big),$$

we have $\rho_t(\mathcal{T}_t(x))\mathbf{J}^\omega_t(x) = \rho_0(x)$ *and* $\mathbf{J}^\omega_t(x) > 0$ *for all* $t \in [0, 1)$ *at* μ_0-*a.e.* $x \in \Omega$. *In the case of* $\nu \in \mathcal{P}^2_{\mathrm{ac}}(M, \mathrm{vol}_g)$, *the above assertions hold also at* $t = 1$.

Note that $\mathbf{J}^\omega_t$ should be understood as the Jacobian with respect to ω, and its behavior is naturally controlled by the weighted Ricci curvature. This is a fundamental geometric intuition behind the curvature-dimension condition (see Section 5).

2.3   Information geometry

We briefly summarize some notions in information geometry associated with a non-decreasing, positive, continuous function $\varphi : (0, \infty) \longrightarrow (0, \infty)$. We refer to [Na1] and [Na2] for further discussion.

We define the *φ-logarithmic function* on $(0, \infty)$ by

$$\ln_\varphi(t) := \int_1^t \frac{1}{\varphi(s)} \, ds,$$

which is clearly strictly increasing. We will denote by l_φ and L_φ the infimum and the supremum of $\ln_\varphi$, that is,

$$l_\varphi := \inf_{t>0} \ln_\varphi(t) = \lim_{t \downarrow 0} \ln_\varphi(t) \in [-\infty, 0), \quad L_\varphi := \sup_{t>0} \ln_\varphi(t) = \lim_{t \uparrow \infty} \ln_\varphi(t) \in (0, \infty].$$

The inverse function of $\ln_\varphi$ is called the *φ-exponential function*. We extend it to the function on $\mathbb{R}$ as

$$\exp_\varphi(\tau) := \begin{cases} 0 & \text{if } \tau \leq l_\varphi, \\ \ln_\varphi^{-1}(\tau) & \text{if } \tau \in (l_\varphi, L_\varphi), \\ \infty & \text{if } \tau \geq L_\varphi. \end{cases}$$

We also introduce the strictly convex function

$$u_\varphi(r) := \int_0^r \ln_\varphi(t) \, dt, \qquad r \in [0, \infty),$$

provided that it is well-defined (i.e., $\ln_\varphi$ is integrable on $(0, 1)$).

Lemma 2.8 *The function* u_φ *is well-defined if*

$$\inf\left\{ \delta \in \mathbb{R} \;\middle|\; \frac{s^{1+\delta}}{\varphi(s)} \text{ is bounded on } (0, 1) \right\} < 1. \tag{2.1}$$

Proof. As φ is positive and non-decreasing, it suffices to see $u_\varphi(1) > -\infty$. We deduce from the hypothesis that $s/\varphi(s)$ is integrable on $(0,1)$. This shows the claim since

$$u_\varphi(1) = -\int_0^1 \int_t^1 \frac{1}{\varphi(s)} \, ds dt = -\int_0^1 \frac{s}{\varphi(s)} \, ds > -\infty.$$

$\square$

Entropy is a function measuring the uncertainty of an event, and the divergence in information theory is a quantity expressing the difference between a pair of probability measures. In this spirit, the φ-*entropy* for $\rho\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ is defined by

$$E_\varphi(\rho\omega) := -\int_M u_\varphi(\rho) \, d\omega$$

(provided that it is well-defined). Then we define the *Bregman divergence* between $\rho\omega, \sigma\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ by

$$D_\varphi(\rho\omega|\sigma\omega) := \int_M \left\{ u_\varphi(\rho) - u_\varphi(\sigma) - u_\varphi'(\sigma)(\rho - \sigma) \right\} \, d\omega. \tag{2.2}$$

The strict convexity of u_φ guarantees $D_\varphi(\rho\omega|\sigma\omega) > 0$ unless $\rho = \sigma$ ω-a.e.. Furthermore, the square root of the divergence D_φ satisfies a generalized Pythagorean theorem and hence it can be regarded as a kind of distance function, though it is not symmetric (i.e., $D_\varphi(\rho\omega|\sigma\omega) \neq D_\varphi(\sigma\omega|\rho\omega)$ in general). We shall explain some more details of these facts.

Let us consider an $\mathbb{R}^k$-valued random variable $X = (X_i)_{i=1}^k$ on M, and a convex open set $\Xi \subset \mathbb{R}^k$ on which a partition function $\Lambda_\varphi : \Xi \longrightarrow \mathbb{R}$ can be chosen so that the total mass of ρ_ξ is normalized to unity for any $\xi \in \Xi$, where ρ_ξ is defined by

$$\nu_\xi = \rho_\xi \omega := \exp_\varphi(\Lambda_\varphi(\xi) - \langle \xi, X \rangle)\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega).$$

Then ν_ξ is a conditional maximizer of E_φ (see [Na2, Theorem 7.2]). We set $\mathcal{M}_\varphi$ as the set of conditional maximizers for E_φ,

$$\mathcal{M}_\varphi := \{\nu_\xi = \exp_\varphi(\Lambda_\varphi(\xi) - \langle \xi, X \rangle)\omega \,|\, \xi \in \Xi \,\} \subset \mathcal{P}_{\mathrm{ac}}(M, \omega).$$

Through the coordinate $\xi \mapsto \nu_\xi$, we introduce the Riemannian metric g_φ of $\mathcal{M}_\varphi$ (associated with the Fisher metric when $\varphi(s) = s$) as

$$g_\varphi\left(\frac{\partial}{\partial\xi^i}, \frac{\partial}{\partial\xi^j}\right) := \int_\Omega \frac{\partial\rho_\xi}{\partial\xi^i} \frac{\partial(\ln_\varphi(\rho_\xi))}{\partial\xi^j} \, d\omega, \quad i, j = 1, \dots, k.$$

We consider another global coordinate $(\eta_\alpha)_{\alpha=1}^k$ of $\mathcal{M}_\varphi$ given by $\eta_\alpha(\mu) := \int_M X_\alpha \, d\mu$ provided that $\eta_\alpha(\mu) \in \mathbb{R}$ for all $\mu \in \mathcal{M}_\varphi$ and $\alpha = 1, \dots, k$. Then it holds

$$g_\varphi\left(\frac{\partial}{\partial\xi^i}, \frac{\partial}{\partial\eta_\alpha}\right) = \delta_\alpha^i.$$

This shows that geodesics generated by two coordinate systems ξ and η are orthogonal to each other. Then the square root of the Bregman divergence D_φ satisfies a generalized Pythagorean theorem for geodesic right triangles in the following sense.

10

Proposition 2.9 ([OW, Proposition 3]) *Fix $\mu \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ satisfying $\eta_\alpha(\mu) \in \mathbb{R}$ for any $\alpha = 1, \ldots, k$. We assume that, for any $\xi \in \Xi$, $\mathrm{supp}(\nu_\xi) = M$ holds and $D_\varphi(\mu | \cdot)$ and Λ_φ are differentiable on $\mathcal{M}_\varphi$. If some $\xi(\mu) \in \Xi$ satisfies $\inf_{\xi \in \Xi} D_\varphi(\mu | \nu_\xi) = D_\varphi(\mu | \nu_{\xi(\mu)})$, in other words, if $\nu_{\xi(\mu)}$ is the foot of the η-geodesic from μ meeting $\mathcal{M}_\varphi$ orthogonally, then we have*

$$\eta_\alpha(\mu) = \eta_\alpha(\nu_{\xi(\mu)}) = \int_M X_\alpha \, d\nu_{\xi(\mu)} \quad \text{for all } \alpha = 1, \ldots, k,$$

$$D_\varphi(\mu | \nu_\xi) = D_\varphi(\mu | \nu_{\xi(\mu)}) + D_\varphi(\nu_{\xi(\mu)} | \nu_\xi) \quad \text{for all } \xi \in \Xi.$$

We define three more quantities measuring the order of φ for later use:

$$\theta_\varphi := \sup \left\{ \frac{s}{\varphi(s)} \cdot \limsup_{t \downarrow 0} \frac{\varphi(s+t) - \varphi(s)}{t} \; \middle| \; s > 0 \right\} \in [0, \infty], \tag{2.3}$$

$$\delta_\varphi := \inf \left\{ \frac{s}{\varphi(s)} \cdot \limsup_{t \downarrow 0} \frac{\varphi(s+t) - \varphi(s)}{t} \; \middle| \; s > 0 \right\} \in [0, \infty), \tag{2.4}$$

$$N_\varphi := \begin{cases} (\theta_\varphi - 1)^{-1} & \text{if } \theta_\varphi \neq 1, \\ \infty & \text{if } \theta_\varphi = 1. \end{cases} \tag{2.5}$$

The following lemma will be useful.

Lemma 2.10 *The function $s^{\delta_\varphi}/\varphi(s)$ is non-increasing in $s \in (0, \infty)$. Moreover, if θ_φ is finite, then the function $s^{\theta_\varphi}/\varphi(s)$ is non-decreasing in $s \in (0, \infty)$.*

Proof. Assume $\theta_\varphi < \infty$ (which will not play any role in the discussion on δ_φ), and fix $s > 0$ and small $\varepsilon > 0$. By the definitions of θ_φ and δ_φ, there exists $r_\varepsilon(s) > 0$ such that

$$\delta_\varphi \leq \frac{s}{\varphi(s)} \cdot \sup_{t \in (0, r_\varepsilon(s)]} \frac{\varphi(s+t) - \varphi(s)}{t} \leq \theta_\varphi + \frac{\varepsilon}{2}.$$

Consider the functions

$$h_+(\tau) := \theta_\varphi + \frac{\varepsilon}{2} - \frac{(1 + \tau)^{\theta_\varphi + \varepsilon} - 1}{\tau}, \qquad h_-(\tau) := \delta_\varphi - \frac{(1 + \tau)^{\delta_\varphi - \varepsilon} - 1}{\tau}$$

for $\tau > 0$. Since h_+ and h_- are continuous and satisfy

$$\lim_{\tau \downarrow 0} h_+(\tau) = \theta_\varphi + \frac{\varepsilon}{2} - (\theta_\varphi + \varepsilon) < 0, \qquad \lim_{\tau \downarrow 0} h_-(\tau) = \delta_\varphi - (\delta_\varphi - \varepsilon) > 0,$$

there exists $\tau_\varepsilon > 0$ such that we have $h_+(\tau) < 0$ and $h_-(\tau) > 0$ for all $\tau \in (0, \tau_\varepsilon)$.

Given any $t \in (0, \min\{r_\varepsilon(s), s\tau_\varepsilon\})$, we have

$$\frac{s^{\theta_\varphi + \varepsilon}}{\varphi(s)} - \frac{(s+t)^{\theta_\varphi + \varepsilon}}{\varphi(s+t)} = \frac{ts^{\theta_\varphi + \varepsilon - 1}}{\varphi(s+t)} \left\{ \frac{s}{\varphi(s)} \frac{\varphi(s+t) - \varphi(s)}{t} - \frac{(1 + ts^{-1})^{\theta_\varphi + \varepsilon} - 1}{ts^{-1}} \right\}$$

$$\leq \frac{ts^{\theta_\varphi + \varepsilon - 1}}{\varphi(s+t)} h_+(ts^{-1}) < 0.$$

11

As $s > 0$ was arbitrary, this shows that $s^{\theta_\varphi + \varepsilon} / \varphi(s)$ is strictly increasing in $s > 0$. We similarly obtain

$$\frac{s^{\delta_\varphi - \varepsilon}}{\varphi(s)} - \frac{(s+t)^{\delta_\varphi - \varepsilon}}{\varphi(s+t)} \geq \frac{ts^{\delta_\varphi - \varepsilon - 1}}{\varphi(s+t)} h_-(ts^{-1}) > 0,$$

so that $s^{\delta_\varphi - \varepsilon} / \varphi(s)$ is strictly decreasing. Letting $\varepsilon \downarrow 0$, we complete the proof. $\qquad \square$

Remark 2.11 The function φ will be sometimes normalized so as to satisfy $\varphi(1) = 1$. This costs no generality as we easily see the following relations for any $a > 0$:

$$\ln_{a\varphi}(t) = a^{-1} \ln_\varphi(t), \quad \exp_{a\varphi}(\tau) = \exp_\varphi(a\tau), \quad u_{a\varphi}(r) = a^{-1} u_\varphi(r),$$
$$l_{a\varphi}(\tau) = a^{-1} l_\varphi, \quad L_{a\varphi} = a^{-1} L_\varphi, \quad \theta_{a\varphi} = \theta_\varphi, \quad \delta_{a\varphi} = \delta_\varphi, \quad N_{a\varphi} = N_\varphi.$$

2.4 Information geometry continued: The case of $\varphi_m(s) = s^{2-m}$

In [OT], we considered the power function $\varphi_m(s) := s^{2-m}$ for $m \in (0, 2]$ and the corresponding *m-logarithmic* and *m-exponential functions*. (We have actually considered $m \in [(n-1)/n, \infty)$ in [OT], but φ_m is non-decreasing only when $m \leq 2$.) We summarize several facts in this especially important case. For brevity, we set

$$\ell_m := \ln_{\varphi_m}, \quad e_m := \exp_{\varphi_m}, \quad l_m := l_{\varphi_m}, \quad L_m := L_{\varphi_m}, \quad \theta_m := \theta_{\varphi_m}, \quad N_m := N_{\varphi_m}.$$

(A) In the case of $\varphi_1(s) = s$, ℓ_1 and e_1 coincide with the usual logarithmic and exponential functions, respectively. Thus we find $l_1 = -\infty$ and $L_1 = \infty$. We can easily observe $\theta_1 = 1$ and $N_1 = \infty$ as well. For $\rho\omega, \sigma\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$, we deduce from $u_{\varphi_1}(r) = r \ln r - r$ that

$$E_{\varphi_1}(\rho\omega) = -\int_M \rho \ln \rho \, d\omega + 1, \qquad D_{\varphi_1}(\rho\omega | \sigma\omega) = \int_M \rho \ln \frac{\rho}{\sigma} \, d\omega.$$

Namely E_{φ_1} is the *Boltzmann entropy* up to adding 1, and D_{φ_1} is the *Kullback–Leibler divergence*. By choosing $\sigma \equiv 1$ formally in the definition of $D_{\varphi_1}(\rho\omega | \sigma\omega)$, the *relative entropy* of $\mu \in \mathcal{P}^2(M)$ with respect to ω is defined by

$$\mathrm{Ent}_\omega(\mu) := \begin{cases} \lim_{\varepsilon \downarrow 0} \int_{\{\rho \geq \varepsilon\}} \rho \ln \rho \, d\omega & \text{if } \mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}^2(M, \omega), \\ \infty & \text{otherwise.} \end{cases} \tag{2.6}$$

In other words, the relative entropy is '$(-1)\times$ the Boltzmann entropy.'

(B) For $\varphi_m(s) = s^{2-m}$ with $m \in (0, 1) \cup (1, 2]$, ℓ_m and e_m are given by power functions as

$$\ell_m(t) = \frac{t^{m-1} - 1}{m - 1}, \qquad e_m(\tau) = [1 + (m-1)\tau]_+^{1/(m-1)},$$

where we set $[t]_+ := \max\{t, 0\}$ and by convention $0^a := \infty$ for $a < 0$. Observe

$$l_m = \begin{cases} -\infty & \text{if } m < 1, \\ -\dfrac{1}{m-1} & \text{if } m > 1, \end{cases} \qquad L_m = \begin{cases} \dfrac{1}{1-m} & \text{if } m < 1, \\ \infty & \text{if } m > 1, \end{cases} \tag{2.7}$$

$\theta_m = 2 - m$ and $N_m = (1 - m)^{-1}$. As $u_{\varphi_m}(r) = (r^m - mr)/\{m(m-1)\}$, the φ_m-entropy for $\rho\omega \in \mathcal{P}_{ac}(M, \omega)$ is given by

$$E_{\varphi_m}(\rho\omega) = -\int_M \frac{\rho^m}{m(m-1)} \, d\omega + \frac{1}{m-1}.$$

Up to additive and multiplicative constants, this coincides with the *Rényi(-Tsallis) entropy*

$$S_N(\rho\omega) := -\int_M \rho^{(N-1)/N} \, d\omega \tag{2.8}$$

with $N = N_m$, which is applied to complex (strongly correlated) systems. The Bregman divergence between $\rho\omega, \sigma\omega \in \mathcal{P}_{ac}(M, \omega)$ is given by

$$D_{\varphi_m}(\rho\omega|\sigma\omega) = \frac{1}{m(m-1)} \int_M \left[(\rho^m - \sigma^m) - m\sigma^{m-1}(\rho - \sigma) \right] d\omega.$$

This coincides with the *β-divergence*, whose strength is its robustness. For instance, we refer to [MTKE] for the roles and the differences of statistical divergences including the Bregman divergences. Note that as $m \to 1$ we have

$$\ell_m(t) \to \ell_1(t), \quad e_m(\tau) \to e_1(t), \quad E_{\varphi_m}(\rho\omega) \to E_{\varphi_1}(\rho\omega), \quad D_{\varphi_m}(\rho\omega|\sigma\omega) \to D_{\varphi_1}(\rho\omega|\sigma\omega).$$

The function $\varphi_m = s^{2-m}$ is an extremal element among those φ's satisfying $\theta_\varphi = 2 - m$ in several respects, as one can see in the next useful lemma for instance.

Lemma 2.12 *Assume $\theta_\varphi < 2$ and put $m = 2 - \theta_\varphi$. Then for any $t > 0$ and $r \in \mathbb{R}$ we have*

$$\frac{1}{\varphi(1)}\ell_m(t) \leq \ln_\varphi(t) \leq \frac{t^{\theta_\varphi}}{\varphi(t)}\ell_m(t), \tag{2.9}$$

$$\exp_\varphi(r) \leq e_m\big(\varphi(1)r\big). \tag{2.10}$$

Proof. It follows from Lemma 2.10 that, for any $t > 0$,

$$\frac{1}{\varphi(1)} \int_1^t s^{-\theta_\varphi} \, ds \leq \int_1^t \frac{1}{\varphi(s)} \, ds \leq \frac{t^{\theta_\varphi}}{\varphi(t)} \int_1^t s^{-\theta_\varphi} \, ds.$$

This is exactly (2.9) since $\theta_\varphi = 2 - m$.

As for (2.10), the assertion for $r \leq l_\varphi$ is trivial since $\exp_\varphi(r) = 0$ by definition. If $r \geq L_\varphi$, then we deduce from (2.9) that $\varphi(1)r \geq \varphi(1)L_\varphi \geq L_m$, which shows $e_m(\varphi(1)r) = \infty$. We therefore assume $l_\varphi < r < L_\varphi$ and set $t := \exp_\varphi(r) > 0$. Then we obtain again from (2.9) that

$$\exp_\varphi(r) = t = e_m\big(\ell_m(t)\big) \leq e_m\big(\varphi(1)\ln_\varphi(t)\big) = e_m\big(\varphi(1)r\big).$$

$\square$

Taking the limits as $t \downarrow 0$ or $t \uparrow \infty$ in (2.9), we obtain from (2.7) the following.

Lemma 2.13 *Suppose $\theta_\varphi < 2$.*

13

(i) *If $\theta_\varphi < 1$, then $l_\varphi > -\infty$ (equivalently, if $l_\varphi = -\infty$, then $\theta_\varphi \geq 1$).*

(ii) *If $\theta_\varphi \leq 1$, then $L_\varphi = \infty$ (equivalently, if $L_\varphi < \infty$, then $\theta_\varphi > 1$).*

We similarly find the corresponding estimates concerning δ_φ. Note that $\delta_{\varphi_m} = \theta_m = 2 - m$.

Lemma 2.14 *Assume $\delta_\varphi < 2$ and put $m = 2 - \delta_\varphi$. Then for any $t > 0$ and $r \in \mathbb{R}$ we have*

$$\frac{t^{\delta_\varphi}}{\varphi(t)}\ell_m(t) \leq \ln_\varphi(t) \leq \frac{1}{\varphi(1)}\ell_m(t), \qquad \exp_\varphi(r) \geq e_m\big(\varphi(1)r\big).$$

In particular,

(i) *If $\delta_\varphi > 1$, then $L_\varphi < \infty$ (equivalently, if $L_\varphi = \infty$, then $\delta_\varphi \leq 1$).*

(ii) *If $\delta_\varphi \geq 1$, then $l_\varphi = -\infty$ (equivalently, if $l_\varphi > -\infty$, then $\delta_\varphi < 1$).*

3 Displacement convexity classes $\mathcal{DC}_N$

In this section, we introduce the important classes of convex functions. These classes were first considered by McCann [Mc1] for $N \geq 1$ (see also [LV2, §5.1], [Vi1, Section 5.2], [Vi2, Chapter 16]), we adopt the same definition also for $N < 0$.

Definition 3.1 (Displacement convexity classes) For $N \in (-\infty, 0) \cup [1, \infty)$, we define $\mathcal{DC}_N$ as the set of all continuous convex functions $u : [0, \infty) \longrightarrow \mathbb{R}$ such that $u(0) = 0$ and that the function

$$\psi_N(r) := r^N u(r^{-N})$$

is convex on $(0, \infty)$. In a similar way, $\mathcal{DC}_\infty$ is defined as the set of all continuous convex functions $u : [0, \infty) \longrightarrow \mathbb{R}$ such that $u(0) = 0$ and that the function

$$\psi_\infty(r) := e^r u(e^{-r})$$

is convex on $\mathbb{R}$.

The following is well-known for $N \geq 1$, we give a proof for completeness.

Lemma 3.2 *If $u \in \mathcal{DC}_N$ for $N \in [1, \infty]$ (resp. $N \in (-\infty, 0)$), then the function ψ_N is non-increasing (resp. non-decreasing).*

Proof. For $N \in [1, \infty)$ and $0 < s < t$, the convexity of u and $u(0) = 0$ yield

$$\psi_N(t) = t^N u(t^{-N}) \leq t^N \left\{ \left(1 - \frac{s^N}{t^N}\right) u(0) + \frac{s^N}{t^N} u(s^{-N}) \right\} = \psi_N(s).$$

We similarly obtain for $N = \infty$ and $s, t \in \mathbb{R}$ with $s < t$ that

$$\psi_\infty(t) = e^t u(e^{-t}) \leq e^t \left\{ \left(1 - \frac{e^s}{e^t}\right) u(0) + \frac{e^s}{e^t} u(e^{-s}) \right\} = \psi_\infty(s).$$

14

Finally, for $N < 0$ and $0 < s < t$, it holds

$$\psi_N(s) = s^N u(s^{-N}) \leq s^N \left\{ \left(1 - \frac{s^{-N}}{t^{-N}}\right) u(0) + \frac{s^{-N}}{t^{-N}} u(t^{-N}) \right\} = \psi_N(t).$$

$\square$

It is also known that $\mathcal{DC}_{N'} \subset \mathcal{DC}_N$ for $1 \leq N < N' \leq \infty$. This monotonicity in N is violated by extending to $N < 0$, but the monotonicity in $m = (N-1)/N \in [0, \infty)$ holds instead. Compare this with the monotonicity of Ric_N in m (Remark 2.2).

Lemma 3.3 *For each $N, N' \in (-\infty, 0) \cup [1, \infty]$ with $m < m'$, we have $\mathcal{DC}_{N'} \subset \mathcal{DC}_N$, where we set $m = (N-1)/N$, $m' = (N'-1)/N'$ and $m = 1$ if $N = \infty$ (resp. $m' = 1$ if $N' = \infty$).*

Proof. We first consider the case of $0 \leq m < m' < 1$ (equivalently, $1 \leq N < N' < \infty$). For any $u \in \mathcal{DC}_{N'}$ and $r > 0$, we observe

$$\psi_N(r) = r^N u(r^{-N}) = (r^{N/N'})^{N'} u\big((r^{N/N'})^{-N'}\big) = \psi_{N'}(r^{N/N'}).$$

This is convex in r since the function $r \mapsto r^{N/N'}$ is concave and $\psi_{N'}$ is convex and non-increasing. Thus $u \in \mathcal{DC}_N$ and hence $\mathcal{DC}_{N'} \subset \mathcal{DC}_N$.

The other cases are similar. For $1 < m < m'$, we have $N < N' < 0$ so that $r \mapsto r^{N/N'}$ is convex and $\psi_{N'}$ is non-decreasing. When $m' = 1 > m$, $\psi_N(r) = \psi_\infty(N \log r)$ holds and note that $r \mapsto N \log r$ is concave and ψ_∞ is non-increasing. For $m = 1 < m'$, we find $\psi_\infty(r) = \psi_{N'}(e^{r/N'})$ and that $r \mapsto e^{r/N'}$ is convex and $\psi_{N'}$ is non-decreasing. $\square$

We shall write down a condition for $u_\varphi \in \mathcal{DC}_N$ on φ. As u_φ is continuous, convex and satisfies $u_\varphi(0) = 0$ by definition once it is well-defined, it is sufficient to check (2.1) and the convexity of ψ_N.

Proposition 3.4 *Assume that φ satisfies the condition (2.1). Then the function ψ_N for $N \in (-\infty, 0) \cup (1, \infty]$ is convex if and only if*

$$\int_0^t \frac{s}{\varphi(s)} \, ds \leq \frac{N}{N-1} \frac{t^2}{\varphi(t)} \tag{3.1}$$

holds for all $t > 0$, where $N/(N-1) = 1$ if $N = \infty$.

Proof. We first of all recall that u_φ is well-defined thanks to (2.1) (Lemma 2.8). For $N \in (-\infty, 0) \cup (1, \infty)$ and $r > 0$, we calculate

$$\psi_N'(r) = Nr^{N-1} u_\varphi(r^{-N}) - Nr^{-1} u_\varphi'(r^{-N}),$$
$$\psi_N''(r) = N(N-1)r^{N-2} u_\varphi(r^{-N}) + (N - N^2) r^{-2} u_\varphi'(r^{-N}) + N^2 r^{-N-2} u_\varphi''(r^{-N})$$
$$= N(N-1)r^{N-2} \left\{ u_\varphi(r^{-N}) - r^{-N} \ln_\varphi(r^{-N}) + \frac{N}{N-1} \frac{r^{-2N}}{\varphi(r^{-N})} \right\}.$$

15

Note that $N(N-1) > 0$. For any $t > 0$, we have

$$u_\varphi(t) - t\ln_\varphi(t) + \frac{N}{N-1}\frac{t^2}{\varphi(t)} = \int_0^t \{\ln_\varphi(\tau) - \ln_\varphi(t)\}\,d\tau + \frac{N}{N-1}\frac{t^2}{\varphi(t)}$$
$$= -\int_0^t \int_\tau^t \frac{1}{\varphi(s)}\,ds d\tau + \frac{N}{N-1}\frac{t^2}{\varphi(t)} = -\int_0^t \frac{s}{\varphi(s)}\,ds + \frac{N}{N-1}\frac{t^2}{\varphi(t)}.$$

Therefore $\psi_N'' \geq 0$ if and only if (3.1) holds. For $N = \infty$, we similarly obtain

$$\psi_\infty''(r) = e^r\left\{u_\varphi(e^{-r}) - e^{-r}\ln_\varphi(e^{-r}) + \frac{e^{-2r}}{\varphi(e^{-r})}\right\}$$

for $r \in \mathbb{R}$, and

$$u_\varphi(t) - t\ln_\varphi(t) + \frac{t^2}{\varphi(t)} = -\int_0^t \frac{s}{\varphi(s)}\,ds + \frac{t^2}{\varphi(t)}$$

for $t > 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Theorem 3.5 *If $\theta_\varphi < 2$, then the condition (2.1) holds and we have $u_\varphi \in \mathcal{DC}_{N_\varphi}$.*

Proof. We deduce from Lemma 2.10 that

$$0 \leq \frac{s^{\theta_\varphi}}{\varphi(s)} \leq \frac{1}{\varphi(1)}$$

for all $s \in (0,1)$, this implies (2.1) since $\theta_\varphi < 2$. Lemma 2.10 also yields that, for any $t > 0$,

$$\int_0^t \frac{s}{\varphi(s)}\,ds \leq \int_0^t \frac{t^{\theta_\varphi}}{\varphi(t)}s^{1-\theta_\varphi}\,ds = \frac{1}{2-\theta_\varphi}\frac{t^2}{\varphi(t)}.$$

This is nothing but (3.1) with $N = N_\varphi$ (recall the definition of N_φ in (2.5)), and hence $u_\varphi \in \mathcal{DC}_{N_\varphi}$ by Proposition 3.4. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Recall that $\varphi_m(s) = s^{2-m}$ with $m \in (0,2]$ satisfies $\theta_m = 2-m < 2$. Hence Theorem 3.5 shows $u_{\varphi_m} \in \mathcal{DC}_{N_m}$. We close the section with a partial converse of Theorem 3.5.

Proposition 3.6 *If the condition (2.1) holds, $\delta_\varphi < 2$ and if we have $u_\varphi \in \mathcal{DC}_N$ with some $N \in (-\infty, 0) \cup (1, \infty]$, then it holds $\delta_\varphi \leq (N+1)/N$ (where $(N+1)/N = 1$ for $N = \infty$).*

Proof. Lemma 2.10 with Proposition 3.4 yields that, for any $t > 0$,

$$\frac{N}{N-1}\frac{t^2}{\varphi(t)} \geq \int_0^t \frac{s}{\varphi(s)}\,ds \geq \int_0^t \frac{t^{\delta_\varphi}}{\varphi(t)}s^{1-\delta_\varphi}\,ds = \frac{1}{2-\delta_\varphi}\frac{t^2}{\varphi(t)},$$

which shows $\delta_\varphi \leq (N+1)/N$ as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

In particular, for $m \in (0,2]$, $u_{\varphi_m} \in \mathcal{DC}_N$ if and only if $m \geq (N-1)/N$.

16

4 Admissible spaces

This section is devoted to introducing the class of spaces admissible in our consideration. Recall our weighted Riemannian manifold (M, ω) and a function φ as in Subsection 2.3. From here on, we further fix the reference measure

$$\nu = \sigma\omega := \exp_\varphi(-\Psi)\omega,$$

where $\Psi \in C(M)$ such that

$$\Psi > -L_\varphi \text{ on } M, \qquad M_\varphi^\Psi := \Psi^{-1}\big((-L_\varphi, -l_\varphi)\big) \neq \emptyset. \tag{4.1}$$

Note that $\operatorname{supp}\nu = \overline{M_\varphi^\Psi} \neq \emptyset$. For later convenience, let us define the K-convexity of a function on a general metric space.

Definition 4.1 (K-convexity) Given $K \in \mathbb{R}$, we say that a function $\Psi : X \longrightarrow (-\infty, \infty]$ on a metric space (X, d) is K-*convex in the weak sense* (denoted by $\operatorname{Hess}\Psi \geq K$ by slight abuse of notation) if it is not identically $-\infty$ and, for any two points $x, y \in X$, there exists a minimal geodesic $\gamma : [0, 1] \longrightarrow X$ from x to y along which

$$\Psi\big(\gamma(t)\big) \leq (1-t)\Psi(x) + t\Psi(y) - \frac{K}{2}(1-t)td(x, y)^2 \tag{4.2}$$

holds for all $t \in [0, 1]$.

We remark that, on a Riemannian manifold M, (4.2) certainly holds for any minimal geodesic $\gamma : [0, 1] \longrightarrow M$ by approximation. Indeed, $\gamma|_{[\varepsilon, 1-\varepsilon]}$ is a unique minimal geodesic for any $\varepsilon > 0$ and Ψ is continuous. We are interested in the situation that $\operatorname{Ric}_{N_\varphi} \geq 0$ as well as $\operatorname{Hess}\Psi \geq K$ hold (see Theorem 5.7). Finer analysis is possible in the particular case of $K > 0$ (Sections 6, 7). We prove a lemma in such a case for later use. The open ball of center $x \in M$ and radius $r > 0$ will be denoted by $B(x, r)$.

Lemma 4.2 *Suppose that $\varphi(1) = 1$ (this costs no generality, see Remark 2.11), $\operatorname{Hess}\Psi \geq K$ for some $K > 0$, and take a minimizer $x_0 \in M$ of Ψ.*

(i) *If $l_\varphi > -\infty$, then the set M_φ^Ψ as in (4.1) is totally convex and $M_\varphi^\Psi \subset B(x_0, R)$ holds with $R = \sqrt{-2(l_\varphi + \Psi(x_0))/K}$. Moreover, $\operatorname{supp}\nu$ is also totally convex and compact.*

(ii) *If $l_\varphi = -\infty$, $N_\varphi \in [n, \infty)$ and if $\operatorname{Ric}_{N_\varphi} \geq 0$, then we have $M_\varphi^\Psi = M$, $\nu[M] < \infty$ and*

$$\int_M d_g(x_0, x)^p \sigma(x)^a \, d\omega(x) \leq C_1 \nu[M]^a + C_2 K^{-aN_\varphi} < \infty$$

for any $a \in (1/2, 1]$ and $p \in [0, \infty)$ satisfying $(2a-1)N_\varphi - p > 0$, where $C_1 = C_1(\omega)$ and $C_2 = C_2(a, p, \theta_\varphi, \omega)$. In particular, $\sigma \in L^a(M, \omega)$ for all $a \in (1/2, 1]$ and $\nu[M]^{-1} \cdot \nu \in \mathcal{P}_{\mathrm{ac}}^p(M, \omega)$ for all $p \in [0, N_\varphi)$.

(iii) *If $N_\varphi = \infty$ and $\operatorname{Ric}_{N_\varphi} \geq 0$, then $\sigma \in L^a(M, \omega)$ for any $a > 0$.*

Proof. We first remark that the assumption $\mathrm{Hess}\,\Psi \geq K > 0$ guarantees the unique existence of a point $x_0 \in M_\varphi^\Psi$ such that $\Psi(x_0) = \inf_M \Psi$. We deduce from the K-convexity (4.2) that

$$\Psi\big(\gamma(1)\big) \geq \Psi(x_0) + \frac{K}{2}d_g\big(x_0, \gamma(1)\big)^2 \tag{4.3}$$

holds for all minimal geodesics $\gamma : [0,1] \longrightarrow M$ with $\gamma(0) = x_0$.

(i) For any minimal geodesic $\gamma : [0,1] \longrightarrow M$ connecting two points $x, y \in M_\varphi^\Psi$, we have

$$\Psi\big(\gamma(t)\big) \leq (1-t)\Psi(x) + t\Psi(y) - \frac{K}{2}(1-t)td_g(x,y)^2 < -l_\varphi,$$

so that γ is contained in M_φ^Ψ. The total convexity of $\mathrm{supp}\,\nu$ can be seen similarly. Precisely, for any $x, y \in \Psi^{-1}((-L_\varphi, -l_\varphi])$, γ as above satisfies $\gamma((0,1)) \subset M_\varphi^\Psi$. This in fact implies that $\mathrm{supp}\,\nu = \overline{M_\varphi^\Psi} = \Psi^{-1}((-L_\varphi, -l_\varphi])$ and is totally convex. We moreover obtain $M_\varphi^\Psi \subset B(x_0, R)$ from (4.3), and thus $\mathrm{supp}\,\nu$ is compact.

(ii) The first assertion $M_\varphi^\Psi = M$ is obvious by definition (see (4.1)). Note that $N_\varphi \in [n, \infty)$ implies $\theta_\varphi \in (1, (n+1)/n]$ (see (2.5)). Set $m := 2 - \theta_\varphi < 1$ and take $a \in (1/2, 1]$ and $p \geq 0$ satisfying $(2a-1)N_\varphi - p > 0$. Then (2.10) and (4.3) imply

$$\int_M d_g(x_0, x)^p \sigma(x)^a \, d\omega(x)$$
$$\leq \int_{B(x_0,1)} \sigma^a \, d\omega + \int_1^\infty e_m\left(-\Psi(x_0) - \frac{K}{2}r^2\right)^a r^p \, \mathrm{area}_\omega[S(x_0, r)] \, dr.$$

We mention that $m = (N_\varphi - 1)/N_\varphi$ and, for $s < L_\varphi$ and $t < 0$, we have

$$e_m(s + t) = \{e_m(s)^{m-1} + (m-1)t\}^{-N_\varphi}.$$

Thanks to the hypothesis $\mathrm{Ric}_{N_\varphi} \geq 0$ with $N_\varphi \in [n, \infty)$, we can estimate the second term by Theorem 2.3 as

$$\int_1^\infty e_m\left(-\Psi(x_0) - \frac{K}{2}r^2\right)^a r^p \, \mathrm{area}_\omega[S(x_0, r)] \, dr$$
$$\leq \mathrm{area}_\omega[S(x_0, 1)] \int_1^\infty \left\{e_m\big(-\Psi(x_0)\big)^{m-1} + (1-m)\frac{K}{2}r^2\right\}^{-aN_\varphi} r^{p+N_\varphi-1} \, dr$$
$$= \mathrm{area}_\omega[S(x_0, 1)] \int_1^\infty \left\{e_m\big(-\Psi(x_0)\big)^{m-1}r^{-2} + (1-m)\frac{K}{2}\right\}^{-aN_\varphi} r^{-(2a-1)N_\varphi+p-1} \, dr$$
$$\leq \mathrm{area}_\omega[S(x_0, 1)] \left\{(1-m)\frac{K}{2}\right\}^{-aN_\varphi} \int_1^\infty r^{-(2a-1)N_\varphi+p-1} \, dr$$
$$= \frac{1}{(2a-1)N_\varphi - p} \mathrm{area}_\omega[S(x_0, 1)] \left\{(1-m)\frac{K}{2}\right\}^{-aN_\varphi}.$$

We used the condition $(2a-1)N_\varphi - p > 0$ in the last equality. Choosing $a = 1$ and $p = 0$, we in particular find $\nu[M] < \infty$. Then the Hölder inequality yields that

$$\int_{B(x_0,1)} \sigma^a \, d\omega \leq \left(\int_{B(x_0,1)} \sigma \, d\omega\right)^a \omega[B(x_0, 1)]^{1-a} \leq \nu[M]^a \omega[B(x_0, 1)]^{1-a}.$$

18

Thus choosing

$$C_1 := \max\{\omega[B(x_0, 1)], 1\} \ge \omega[B(x_0, 1)]^{1-a},$$

$$C_2 := \frac{1}{(2a-1)N_\varphi - p} \left(\frac{2}{1-m}\right)^{aN_\varphi} \text{area}_\omega[S(x_0, 1)]$$

gives the desired estimate.

(iii) Combining (4.3) with (2.10) provides, as $\theta_\varphi = 1$,

$$\int_M \sigma(x)^a \, d\omega(x) \le \int_M \exp_\varphi\left(-\Psi(x_0) - \frac{K}{2}d_g(x_0, x)^2\right)^a d\omega(x)$$

$$\le \exp\left(-a\Psi(x_0)\right) \int_M \exp\left(-\frac{aK}{2}d_g(x_0, x)^2\right) d\omega(x).$$

Hence the assertion follows from Theorem 2.4. □

Now we introduce the conditions for a quadruple $(M, \omega, \varphi, \Psi)$ to be admissible in our consideration.

Definition 4.3 (Admissibility) We say that a quadruple $(M, \omega, \varphi, \Psi)$ is *admissible* if all the following conditions hold:

(A-1) $\varphi(1) = 1$.

(A-2) $N_\varphi \in (-\infty, -1] \cup [n, \infty]$ and $N_\varphi \ne 2$ or, equivalently, $\theta_\varphi \in [0, (n+1)/n]$ and $\theta_\varphi < 3/2$.

(A-3) $\Psi > -L_{2-\theta_\varphi}$ on M and $M_\varphi^\Psi = \Psi^{-1}((-L_\varphi, -l_\varphi)) \ne \emptyset$.

(A-4) $h_\varphi(\sigma) \in L^1(M, \omega)$ and $\sigma \ln_\varphi(\sigma) \in L^1(M, \omega)$, where $h_\varphi := u_\varphi$ if $L_\varphi = \infty$ and $h_\varphi(r) := u_\varphi(r) - rL_\varphi$ if $L_\varphi < \infty$ (see (5.3) below).

We mention that $l_m \le l_\varphi$ and $L_m \le L_\varphi$ hold with $m = 2 - \theta_\varphi$ by (2.9), and hence $M_\varphi^\Psi \subset M_{\varphi_m}^\Psi$ by (A-3). The first condition $\varphi(1) = 1$ is merely the normalization (see Remark 2.11), and (A-4) is imposed for ν being adopted as a reference measure of the Bregman divergence (see (2.2)). The next lemma ensures that (A-4) automatically holds if $\text{Ric}_{N_\varphi} \ge 0$ and $\text{Hess}\,\Psi \ge K$ for some $K > 0$.

Lemma 4.4 *Suppose that $(M, \omega, \varphi, \Psi)$ satisfies (A-1), (A-2), $\Psi > -L_\varphi$ on M, $M_\varphi^\Psi \ne \emptyset$, $\text{Ric}_{N_\varphi} \ge 0$ and $\text{Hess}\,\Psi \ge K$ for some $K > 0$. Then (A-4) also holds.*

Proof. The case of $l_\varphi > -\infty$ is clear due to Lemma 4.2(i), so that we assume $l_\varphi = -\infty$ and then $\theta_\varphi \ge 1$ (Proposition 2.13(i)). Observe also that $u_\varphi(\sigma) \in L^1(M, \omega)$ implies $h_\varphi(\sigma) \in L^1(M, \omega)$ since $\nu[M] < \infty$ by Lemma 4.2(ii), (iii). Let x_0 be the minimizer of Ψ and set $R := \sqrt{\max\{1, -2\Psi(x_0)\}/K}$. Note that the K-convexity of Ψ (4.3) guarantees that, on $M \setminus B(x_0, R)$,

$$0 \le \sigma = \exp_\varphi(-\Psi) \le \exp_\varphi\left(-\Psi(x_0) - \frac{K}{2}R^2\right) \le \exp_\varphi(0) = 1.$$

19

We first consider the case of $\theta_\varphi > 1$. On $M \setminus B(x_0, R)$, (2.9) implies that

$$|\sigma \ln_\varphi(\sigma)| = -\sigma \ln_\varphi(\sigma) \leq -\sigma \ell_{2-\theta_\varphi}(\sigma) = N_\varphi(\sigma^{2-\theta_\varphi} - \sigma),$$

$$|u_\varphi(\sigma)| = -\int_0^\sigma \ln_\varphi(t) \, dt \leq \int_0^\sigma N_\varphi(t^{1-\theta_\varphi} - 1) \, dt = N_\varphi \left(\frac{\sigma^{2-\theta_\varphi}}{2 - \theta_\varphi} - \sigma \right).$$

Thus we have

$$\int_M |\sigma \ln_\varphi(\sigma)| \, d\omega \leq \int_{B(x_0,R)} |\sigma \ln_\varphi(\sigma)| \, d\omega + \int_{M \setminus B(x_0,R)} N_\varphi(\sigma^{2-\theta_\varphi} - \sigma) \, d\omega,$$

$$\int_M |u_\varphi(\sigma)| \, d\omega \leq \int_{B(x_0,R)} |u_\varphi(\sigma)| \, d\omega + \int_{M \setminus B(x_0,R)} N_\varphi \left(\frac{\sigma^{2-\theta_\varphi}}{2 - \theta_\varphi} - \sigma \right) \, d\omega.$$

As $2 - \theta_\varphi \in (1/2, 1)$ by $\theta_\varphi < 3/2$, Lemma 4.2(ii) ensures $u_\varphi(\sigma), \sigma \ln_\varphi(\sigma) \in L^1(M, \omega)$.

In the case of $\theta_\varphi = 1$, we similarly have on $M \setminus B(x_0, R)$

$$|\sigma \ln_\varphi(\sigma)| \leq -\sigma \ln \sigma \leq \sqrt{\sigma}, \qquad |u_\varphi(\sigma)| = -\int_0^\sigma \ln_\varphi(t) \, dt \leq \int_0^\sigma \frac{1}{\sqrt{t}} \, dt = 2\sqrt{\sigma}.$$

Then the claim follows from Lemma 4.2(iii). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

We close the section with an auxiliary lemma on how to normalize ν when $K > 0$.

Lemma 4.5 *Let $(M, \omega, \varphi, \Psi)$ be admissible, $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\, \Psi \geq K$ for some $K > 0$, and set $I = (l, L) := (l_\varphi + \inf_M \Psi, L_\varphi + \inf_M \Psi)$. We in addition assume that Ψ is differentiable at the minimizer of Ψ if $N_\varphi = n$. Then there exists some $\lambda \in I$ such that $\exp_\varphi(\lambda - \Psi)\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$.*

Proof. We first remark that $\inf_M \Psi > -\infty$, and that for any $\lambda \in I$

$$\Psi - \lambda > \inf_M \Psi - \left(L_\varphi + \inf_M \Psi \right) = -L_\varphi$$

as well as $\mathrm{Hess}(\Psi - \lambda) = \mathrm{Hess}\, \Psi \geq K$ hold. Thus

$$\Xi(\lambda) := \int_M \exp_\varphi(\lambda - \Psi) \, d\omega < \infty$$

by Lemma 4.2. Since Ξ is non-decreasing and continuous on I by Lebesgue's dominated convergence theorem (or the monotone convergence theorem), we are done if $\lim_{\lambda \downarrow l} \Xi(\lambda) < 1 < \lim_{\lambda \uparrow L} \Xi(\lambda)$ holds. We also deduce from the dominated convergence theorem that $\lim_{\lambda \downarrow l} \Xi(\lambda) = 0$. If $L_\varphi = \infty$, then we find $\lim_{\lambda \uparrow \infty} \Xi(\lambda) = \infty$ by the monotone convergence theorem.

The rest is to prove $\lim_{\lambda \uparrow L} \Xi(\lambda) > 1$ when $L_\varphi < \infty$. Note that $L_\varphi < \infty$ implies $\lim_{s \uparrow \infty} \varphi(s) = \infty$ by definition, and $\theta_\varphi > 1$ (i.e., $N_\varphi \in [n, \infty)$) by Proposition 2.13(ii). Let $x_0 \in M$ be the unique minimizer of Ψ and take $R_0 > 0$ such that $B(x_0, R_0) \subset M_\varphi^\Psi$ and that $B(x_0, R_0)$ contains no cut point of x_0. Then, for any $x \in S(x_0, r)$ with $0 < r < R \leq R_0$, the K-convexity of Ψ provides

$$\Psi(x) \leq \left(1 - \frac{r}{R} \right) \Psi(x_0) + \frac{r}{R} \sup_{S(x_0,R)} \Psi - \frac{K}{2} \left(1 - \frac{r}{R} \right) \frac{r}{R} R^2 = \Psi(x_0) + \frac{K}{2} r^2 + ar, \quad (4.4)$$

20

where we set

$$a = a(R) := \frac{1}{R} \left(\sup_{S(x_0, R)} \Psi - \Psi(x_0) - \frac{K}{2} R^2 \right) \geq 0.$$

Observe that $\lim_{R \downarrow 0} a(R) < \infty$ by the K-convexity of Ψ, and that $\lim_{R \downarrow 0} a(R) = 0$ holds if $N_\varphi = n$ since Ψ is assumed to be differentiable at x_0. In both cases ($N_\varphi > n$ or $N_\varphi = n$) we can choose $R \in (0, R_0]$ small enough to satisfy

$$\frac{K}{2} R^2 + aR < L_\varphi, \qquad 2a < \left(\frac{\text{area}_\omega[S(x_0, R)]}{N_\varphi R^{N_\varphi - 1}} \right)^{1/N_\varphi}. \tag{4.5}$$

Then take large $\lambda \in I$ such that

$$\lambda > \Psi(x_0) + \frac{K}{2} R^2 + aR.$$

Set

$$e_\varphi^\lambda(r) := \exp_\varphi \left(\lambda - \Psi(x_0) - \frac{K}{2} r^2 - ar \right)$$

and note that it is decreasing.

We deduce from Theorem 2.3 and (4.4) that

$$\begin{aligned}
\Xi(\lambda) &\geq \int_{B(x_0, R)} \exp_\varphi(\lambda - \Psi) \, d\omega \geq \int_0^R \text{area}_\omega[S(x_0, r)] e_\varphi^\lambda(r) \, dr \\
&\geq \frac{\text{area}_\omega[S(x_0, R)]}{R^{N_\varphi - 1}} \int_0^R r^{N_\varphi - 1} e_\varphi^\lambda(r) \, dr \\
&= \frac{\text{area}_\omega[S(x_0, R)]}{R^{N_\varphi - 1}} \left(\frac{R^{N_\varphi}}{N_\varphi} e_\varphi^\lambda(R) - \int_0^R \frac{r^{N_\varphi}}{N_\varphi} (e_\varphi^\lambda)'(r) \, dr \right).
\end{aligned}$$

The concavity of $\ln_\varphi$ yields

$$\ln_\varphi(t) \leq \ln_\varphi(e_\varphi^\lambda(0)) + \frac{t - e_\varphi^\lambda(0)}{\varphi(e_\varphi^\lambda(0))},$$

and hence it holds for $t := e_\varphi^\lambda(r)$ that

$$Kr = -a + \sqrt{a^2 + 2K \left\{ \ln_\varphi(e_\varphi^\lambda(0)) - \ln_\varphi(t) \right\}} \geq -a + \sqrt{a^2 + 2K \frac{e_\varphi^\lambda(0) - t}{\varphi(e_\varphi^\lambda(0))}}.$$

By the change of variables formula for $t = e_\varphi^\lambda(r)$, we have

$$\begin{aligned}
-\int_0^R r^{N_\varphi} (e_\varphi^\lambda)'(r) \, dr &= \frac{1}{K^{N_\varphi}} \int_{e_\varphi^\lambda(R)}^{e_\varphi^\lambda(0)} \left(-a + \sqrt{a^2 + 2K \left\{ \ln_\varphi(e_\varphi^\lambda(0)) - \ln_\varphi(t) \right\}} \right)^{N_\varphi} dt \\
&\geq \frac{1}{K^{N_\varphi}} \int_{e_\varphi^\lambda(R)}^{e_\varphi^\lambda(0)} \left(-a + \sqrt{a^2 + 2K \frac{e_\varphi^\lambda(0) - t}{\varphi(e_\varphi^\lambda(0))}} \right)^{N_\varphi} dt.
\end{aligned}$$

21

Combining the triangle inequality

$$
\left\{ \int_{e_\varphi^\lambda(R)}^{e_\varphi^\lambda(0)} \left(-a + \sqrt{a^2 + 2K\frac{e_\varphi^\lambda(0) - t}{\varphi(e_\varphi^\lambda(0))}} \right)^{N_\varphi} dt \right\}^{1/N_\varphi}
$$

$$
\geq \left\{ \int_{e_\varphi^\lambda(R)}^{e_\varphi^\lambda(0)} \left(\sqrt{a^2 + 2K\frac{e_\varphi^\lambda(0) - t}{\varphi(e_\varphi^\lambda(0))}} \right)^{N_\varphi} dt \right\}^{1/N_\varphi} - \{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)\}^{1/N_\varphi} a
$$

with Jensen's inequality for the convex function $s \mapsto (\sqrt{a^2 + s})^{N_\varphi}$ ($s \geq 0$), we deduce that

$$
\int_{e_\varphi^\lambda(R)}^{e_\varphi^\lambda(0)} \left(-a + \sqrt{a^2 + 2K\frac{e_\varphi^\lambda(0) - t}{\varphi(e_\varphi^\lambda(0))}} \right)^{N_\varphi} dt
$$

$$
\geq \{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)\} \left\{ \sqrt{a^2 + \frac{2K}{\varphi(e_\varphi^\lambda(0))} \int_{e_\varphi^\lambda(R)}^{e_\varphi^\lambda(0)} \frac{e_\varphi^\lambda(0) - t}{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)} dt} - a \right\}^{N_\varphi}
$$

$$
= \{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)\} \left(\sqrt{a^2 + K\frac{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)}{\varphi(e_\varphi^\lambda(0))}} - a \right)^{N_\varphi}.
$$

Hence we obtain, as $\varphi(s) \leq s^{\theta_\varphi}$ for $s \geq 1$ by Lemma 2.10,

$$
\lim_{\lambda\uparrow L} \Xi(\lambda) \geq \frac{\mathrm{area}_\omega[S(x_0,R)]}{N_\varphi R^{N_\varphi-1} K^{N_\varphi}} \lim_{\lambda\uparrow L} \left[\{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)\} \left(\sqrt{a^2 + K\frac{e_\varphi^\lambda(0) - e_\varphi^\lambda(R)}{\varphi(e_\varphi^\lambda(0))}} - a \right)^{N_\varphi} \right]
$$

$$
\geq \frac{\mathrm{area}_\omega[S(x_0,R)]}{N_\varphi R^{N_\varphi-1} K^{N_\varphi}} \lim_{s\uparrow\infty} \left[s \left(\sqrt{a^2 + \frac{Ks}{(s + e_\varphi^L(R))^{\theta_\varphi}}} - a \right)^{N_\varphi} \right]
$$

$$
= \frac{\mathrm{area}_\omega[S(x_0,R)]}{N_\varphi R^{N_\varphi-1} K^{N_\varphi}} \lim_{s\uparrow\infty} \left[s^{\theta_\varphi-1} (\sqrt{a^2 + Ks^{1-\theta_\varphi}} - a) \right]^{N_\varphi}
$$

$$
= \frac{\mathrm{area}_\omega[S(x_0,R)]}{N_\varphi R^{N_\varphi-1} K^{N_\varphi}} \lim_{t\downarrow 0} \left(\frac{\sqrt{a^2 + Kt} - a}{t} \right)^{N_\varphi} = \frac{\mathrm{area}_\omega[S(x_0,R)]}{N_\varphi R^{N_\varphi-1}} (2a)^{-N_\varphi}.
$$

This is bigger than 1 by the choice of R (recall (4.5)) and we complete the proof. $\square$

Remark 4.6 If $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ for some $K > 0$, then Lemma 4.2 yields $\nu[M] < \infty$. We will sometimes normalize $(M, \omega, \varphi, \Psi)$ so as to satisfy $\nu[M] = 1$ (Sections 6, 7). There are two ways of such a normalization:

- Put $a := \nu[M]^{-1}$ and consider $(M, \tilde{\omega}, \varphi, \Psi) := (M, a\omega, \varphi, \Psi)$, i.e., $\tilde{\omega} = e^{-f+\ln a} \mathrm{vol}_g$.

- Take λ as in Lemma 4.5 and consider $(M, \omega, \varphi, \widetilde{\Psi}) := (M, \omega, \varphi, \Psi - \lambda)$.

In both cases it is easily seen that the conditions $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ are preserved. These two normalizations are equivalent when $\varphi = \varphi_1$, where we indeed observe

$$
\exp(-\Psi)\tilde{\omega} = \exp(-\Psi - f + \ln a)\omega, \quad \exp(-\widetilde{\Psi})\omega = \exp(-\Psi - f + \lambda)\omega,
$$

and hence $\lambda = \ln a$.

5 φ-relative entropy and its displacement convexity

In this section, we introduce a generalization of the relative entropy, that we call the *φ-relative entropy*, associated with functions φ in an appropriate class. For the relative entropy on a (unweighted) Riemannian manifold, it is known by von Renesse and Sturm [vRS] that its K-convexity in the Wasserstein space $(\mathcal{P}^2(M), W_2)$ (in the sense of Definition 4.1) is equivalent to the lower Ricci curvature bound $\mathrm{Ric} \geq K$. Then it was shown by Lott and Villani [LV2] that $\mathrm{Ric} \geq K$ further implies a kind of convexity property of a class of entropies including the relative entropy. In this sense, the relative entropy is an extremal element in such a class of entropies. In the same spirit, our main theorem in the section (Theorem 5.7) asserts that the m-relative entropy induced from $\varphi = \varphi_m$ (studied in [OT], recall Subsection 2.4 as well) is an extremal one in an appropriate class of φ-relative entropies.

5.1 Curvature-dimension condition

We begin with a brief review of Lott, Sturm and Villani's curvature-dimension condition. To formulate it, we need to introduce the two functions often used in comparison theorems in Riemannian geometry.

For $K \in \mathbb{R}$, $N \in (1, \infty)$ and $0 < r \; (< \pi\sqrt{(N-1)/K}$ if $K > 0)$, we consider the function

$$
\mathbf{s}_{K,N}(r) := \begin{cases}
\sqrt{(N-1)/K}\,\sin(r\sqrt{K/(N-1)}) & \text{if } K > 0, \\
r & \text{if } K = 0, \\
\sqrt{-(N-1)/K}\,\sinh(r\sqrt{-K/(N-1)}) & \text{if } K < 0.
\end{cases}
$$

This is the solution to the differential equation

$$
\mathbf{s}''_{K,N} + \frac{K}{N-1}\mathbf{s}_{K,N} = 0
$$

with the initial conditions $\mathbf{s}_{K,N}(0) = 0$ and $\mathbf{s}'_{K,N}(0) = 1$. For $n \in \mathbb{N}$ with $n \geq 2$, $\mathbf{s}_{K,n}(r)^{n-1}$ is proportional to the area of the sphere of radius r in the n-dimensional space form of constant sectional curvature $K/(n-1)$. Using $\mathbf{s}_{K,N}$, we define

$$
\beta^t_{K,N}(r) := \left(\frac{\mathbf{s}_{K,N}(tr)}{t\mathbf{s}_{K,N}(r)}\right)^{N-1}, \qquad \beta^t_{K,\infty}(r) := e^{K(1-t^2)r^2/6}
$$

for K, N, r as above and $t \in (0, 1)$. Now, we are ready to state Sturm's *curvature-dimension condition* characterizing lower Ricci curvature bounds, developed in [vRS], [St1] and [St3] in gradually general situations (see also [CMS1], [CMS2] for related important work). Recall (2.8) and (2.6) for the definitions of the Rényi entropy S_N and the relative entropy Ent_ω.

Theorem 5.1 (Sturm's curvature-dimension condition) *Let (M, ω) be a weighted Riemannian manifold. We have $\mathrm{Ric}_N \geq K$ for some $K \in \mathbb{R}$ and $N \in [n, \infty)$ if and only*

if any pair of measures $\mu_0 = \rho_0\omega, \mu_1 = \rho_1\omega \in \mathcal{P}^2_{\mathrm{ac}}(M,\omega)$ satisfies

$$S_N(\mu_t) \le -(1-t)\int_{M \times M} \beta_{K,N}^{1-t}\big(d_g(x,y)\big)^{1/N}\rho_0(x)^{-1/N}\,d\pi(x,y)$$

$$-t\int_{M \times M}\beta_{K,N}^t\big(d_g(x,y)\big)^{1/N}\rho_1(y)^{-1/N}\,d\pi(x,y) \qquad (5.1)$$

for all $t \in (0,1)$, where $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2_{\mathrm{ac}}(M,\omega)$ is the unique minimal geodesic from μ_0 to μ_1, and π is the unique optimal coupling of μ_0 and μ_1.

Similarly, $\mathrm{Ric}_\infty \ge K$ *is equivalent to the K-convexity of* Ent_ω,

$$\mathrm{Ent}_\omega(\mu_t) \le (1-t)\,\mathrm{Ent}_\omega(\mu_0) + t\,\mathrm{Ent}_\omega(\mu_1) - \frac{K}{2}(1-t)tW_2(\mu_0,\mu_1)^2.$$

For $K = 0$, we find $\beta_{0,N}^t \equiv 1$ and (5.1) is nothing but the convexity of S_N,

$$S_N(\mu_t) \le (1-t)S_N(\mu_0) + tS_N(\mu_1).$$

For $K \ne 0$, however, (5.1) is not simply the K-convexity of S_N.

Lott and Villani's version of the curvature-dimension condition requires a similar convexity condition, but for all entropies induced from functions in $\mathcal{DC}_N$ ([LV1], [LV2], [Vi2, Part III]). For $U \in \mathcal{DC}_N$ and $\mu \in \mathcal{P}^2(M)$, we denote by $\mu = \rho\omega + \mu^s$ its Lebesgue decomposition into absolutely continuous and singular parts with respect to the base measure ω, and define

$$U_\omega(\mu) := \lim_{\varepsilon\downarrow 0}\int_{\{\rho\ge\varepsilon\}}U(\rho)\,d\omega + U'(\infty)\mu^s[M], \qquad U'(\infty) := \lim_{r\to\infty}\frac{U(r)}{r}.$$

We set $\infty \cdot 0 = 0$ by convention, and remark that $U'(\infty) = \lim_{r\to\infty}U'(r)$ holds due to the convexity of U.

Theorem 5.2 (Lott and Villani's version) *We have $\mathrm{Ric}_N \ge K$ for some $K \in \mathbb{R}$ and $N \in [n,\infty]$ if and only if, given any pair of measures $\mu_0, \mu_1 \in \mathcal{P}^2(M)$ decomposed as $\mu_i = \rho_i\omega + \mu_i^s$ $(i = 0,1)$, there is a minimal geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2(M)$ between them such that*

$$U_\omega(\mu_t) \le (1-t)\int_M\int_M \beta_{K,N}^{1-t}\big(d_g(x,y)\big)U\left(\frac{\rho_0(x)}{\beta_{K,N}^{1-t}(d_g(x,y))}\right)d\pi_x(y)d\omega(x)$$

$$+t\int_M\int_M \beta_{K,N}^t\big(d_g(x,y)\big)U\left(\frac{\rho_1(y)}{\beta_{K,N}^t(d_g(x,y))}\right)d\pi_y(x)d\omega(y)$$

$$+U'(\infty)\{(1-t)\mu_0^s[M] + t\mu_1^s[M]\} \qquad (5.2)$$

holds for all $U \in \mathcal{DC}_N$ and $t \in (0,1)$, where π_x $(\in \mathcal{P}(M)$ μ_0-a.e. $x)$ and π_y $(\in \mathcal{P}(M)$ μ_1-a.e. $y)$ denote the disintegrations of π by μ_0 and μ_1, i.e.,

$$d\pi(x,y) = d\pi_x(y)d\mu_0(x) = d\pi_y(x)d\mu_1(y).$$

24

Recall that $\mathcal{DC}_{N'} \subset \mathcal{DC}_N$ for $n \leq N < N'$ (Lemma 3.3). This agrees with the monotonicity $\mathrm{Ric}_N \leq \mathrm{Ric}_{N'}$ for $n \leq N \leq N'$. In the case where both μ_0 and μ_1 are absolutely continuous with respect to ω, we find

$$d\pi(x,y) = \rho_0(x)d\pi_x(y)d\omega(x) = \rho_1(y)d\pi_y(x)d\omega(y)$$

and (5.2) is rewritten in the more symmetric form

$$U_\omega(\mu_t) \leq (1-t)\int_{M\times M} \frac{\beta_{K,N}^{1-t}(d_g(x,y))}{\rho_0(x)} U\left(\frac{\rho_0(x)}{\beta_{K,N}^{1-t}(d_g(x,y))}\right) d\pi(x,y)$$
$$+ t\int_{M\times M} \frac{\beta_{K,N}^{t}(d_g(x,y))}{\rho_1(y)} U\left(\frac{\rho_1(y)}{\beta_{K,N}^{t}(d_g(x,y))}\right) d\pi(x,y).$$

Besides Riemannian manifolds, these two versions of the curvature-dimension condition are equivalent for metric measure spaces where geodesics do not branch, such as Finsler manifolds and Alexandrov spaces. In other words, Sturm's version implies Lott and Villani's one. Roughly speaking, this implication can be seen by localizing Sturm's (5.1) thanks to the non-branching property, and then integrating these local inequalities for each $U \in \mathcal{DC}_N$ yields (5.2). The same infinitesimal estimate (Claim 5.8) will appear in our discussion. Theorem 5.2 is extended to general Finsler manifolds by introducing the appropriate notion of the weighted Ricci curvature (see Section 10 and [Oh2]).

General (not necessarily differentiable) metric measure spaces satisfying the condition in Theorem 5.1 or 5.2 are known to behave like Riemannian manifolds of $\mathrm{Ric} \geq K$ and $\dim \leq N$ in geometric and analytic respects ([St2], [St3], [LV1], [LV2], [Vi2, Part III]). We shall generalize this technique to φ-relative entropies in the following sections.

5.2 φ-relative entropy H_φ

Let $(M, \omega, \varphi, \Psi)$ be an admissible space in the sense of Definition 4.3. We modify u_φ as, for $r \geq 0$,

$$h_\varphi(r) := \begin{cases} u_\varphi(r) & \text{if } L_\varphi = \infty, \\ u_\varphi(r) - rL_\varphi & \text{if } L_\varphi < \infty. \end{cases} \tag{5.3}$$

We also define

$$h'_\varphi(\infty) := \lim_{r\to\infty} h'_\varphi(r) = \begin{cases} \infty & \text{if } L_\varphi = \infty, \\ 0 & \text{if } L_\varphi < \infty. \end{cases}$$

Note that $u_\varphi \in \mathcal{DC}_{N_\varphi}$ (by Theorem 3.5 thanks to the admissibility) immediately implies $h_\varphi \in \mathcal{DC}_{N_\varphi}$. Moreover, if $L_\varphi < \infty$, then h_φ is non-increasing and hence nonpositive. We set

$$L^\varphi(M, \omega) := \{\rho : M \longrightarrow \mathbb{R} \mid \text{measurable}, h_\varphi(\rho) \in L^1(M, \omega)\},$$
$$\mathcal{P}^\Psi(M) := \{\mu \in \mathcal{P}(M) \mid \Psi \in L^1(M_\varphi^\Psi, \mu)\}$$

(we will use these notations only in Remark 5.4). Now the Bregman divergence (2.2) leads us to the following generalization of the relative entropy.

Definition 5.3 (φ-relative entropy) Given $\mu \in \mathcal{P}(M)$, letting $\mu = \rho\omega + \mu^s$ be its Lebesgue decomposition, we define the *φ-relative entropy* of μ by

$$H_\varphi(\mu) := \int_M \{h_\varphi(\rho) - h_\varphi'(\sigma)\rho\} \, d\omega - \int_M h_\varphi'(\sigma) \, d\mu^s + h_\varphi'(\infty)\mu^s[M]$$

$$= \int_M h_\varphi(\rho) \, d\omega - \int_M h_\varphi'(\sigma) \, d\mu + h_\varphi'(\infty)\mu^s[M] \tag{5.4}$$

if $h_\varphi(\rho) \in L^1(M, \omega)$ and $h_\varphi'(\sigma) \in L^1(M, \mu)$, otherwise we set $H_\varphi(\mu) := \infty$.

Let us summarize several remarks on Definition 5.3.

Remark 5.4 (1) In the second term of (5.4), to be precise, we set

$$h_\varphi'(\sigma) := \begin{cases} l_\varphi & \text{if } L_\varphi = \infty, \\ l_\varphi - L_\varphi & \text{if } L_\varphi < \infty \end{cases} \qquad \text{on } M \setminus M_\varphi^\Psi.$$

This causes no problem because $M = M_\varphi^\Psi$ if $l_\varphi = -\infty$. The additional condition $\mu[M_\varphi^\Psi] = 1$ (in other words, $\mu \in \mathcal{P}(M_\varphi^\Psi)$) will be imposed only when we compare the behavior of Ψ with that of H_φ (as in Theorems 5.7, 8.7 and so forth).

(2) We remark that the condition (A-4) in the admissibility guarantees that $\sigma \in L^\varphi(M, \omega)$ as well as $\Psi \in L^1(M_\varphi^\Psi, \nu)$. Thus we have $H_\varphi(\nu) \in \mathbb{R}$ (by extending the definition (5.4) verbatim).

(3) The validity of the definition of $h_\varphi'(\infty)$ (for the lower semi-continuity of H_φ, see Lemma 5.6) would be understood by the following observation: For small $\varepsilon > 0$, put $\mu_\varepsilon = \rho_\varepsilon\omega := \omega[B(x, \varepsilon)]^{-1}\chi_{B(x,\varepsilon)}\omega$, where $\chi_{B(x,\varepsilon)}$ stands for the characteristic function of $B(x, \varepsilon)$. Then we have

$$\int_{B(x,\varepsilon)} h_\varphi(\rho_\varepsilon) \, d\omega = \omega[B(x,\varepsilon)] \cdot h_\varphi\left(\frac{1}{\omega[B(x,\varepsilon)]}\right) \to h_\varphi'(\infty)$$

as ε tends to zero.

(4) Finally, as for the domain of H_φ, it is more consistent to set $H_\varphi(\mu) = \infty$ only if $h_\varphi(\rho) - \rho h_\varphi'(\sigma) \notin L^1(M, \omega)$. However, as we will sometimes treat the internal energy $\int_M h_\varphi(\rho) \, d\omega$ and the potential energy $\int_M h_\varphi'(\sigma) \, d\mu$ separately, we consider the smaller domain in Definition 5.3. This may cause a problem when considering the lower semi-continuity of H_φ, whereas we need it only for compact M (see Lemma 5.6 below) where $h_\varphi'(\sigma) \in L^1(M, \mu)$ is always true (so that $h_\varphi(\rho) \in L^1(M, \omega)$ if and only if $h_\varphi(\rho) - \rho h_\varphi'(\sigma) \in L^1(M, \omega)$).

Let us add a comment on the relation between $\rho \in L^\varphi(M, \omega)$ and $\mu \in \mathcal{P}^\Psi(M)$. Assume $L_\varphi < \infty$ and $\mu = \rho\omega + \mu^s \in \mathcal{P}^\Psi(M)$. The nonpositivity and the convexity of h_φ yield

$$\int_M |h_\varphi(\rho)| \, d\omega = -\int_M h_\varphi(\rho) \, d\omega \le -\int_M \{h_\varphi(\sigma) + h_\varphi'(\sigma)(\rho - \sigma)\} \, d\omega$$

$$= -\int_M \{h_\varphi(\sigma) - h_\varphi'(\sigma)\sigma\} \, d\omega - \int_M \rho h_\varphi'(\sigma) \, d\omega < \infty.$$

Hence $\rho \in L^\varphi(M, \omega)$ automatically holds. One can also see the converse implication ($\rho \in L^\varphi(M, \omega) \Rightarrow \mu \in \mathcal{P}^\Psi(M)$) for the special case $\varphi_m(s) = s^m$ with $m > 1$, where $L_m = \infty$ as in (2.7) ([OT, Remark 3.2(2)]).

It is easily observed that ν is a unique *ground state* of H_φ (provided that $\nu[M] = 1$).

Lemma 5.5 *Suppose $\nu[M] = 1$. For any $\mu = \rho\omega + \mu^s \in \mathcal{P}(M)$, we have $H_\varphi(\mu) \geq H_\varphi(\nu)$ and equality holds if and only if $\mu = \nu$.*

Proof. We assume $H_\varphi(\mu) < \infty$ without loss of generality. Observe that

$$
H_\varphi(\mu) - H_\varphi(\nu)
$$
$$
= \int_M \{h_\varphi(\rho) - h_\varphi(\sigma) - h'_\varphi(\sigma)(\rho - \sigma)\}\, d\omega - \int_M h'_\varphi(\sigma)\, d\mu^s + h'_\varphi(\infty)\mu^s[M]. \qquad (5.5)
$$

On the one hand, if $\mu^s[M] > 0$, then the singular part

$$
- \int_M h'_\varphi(\sigma)\, d\mu^s + h'_\varphi(\infty)\mu^s[M]
$$

in (5.5) is positive if $L_\varphi < \infty$ (recall that $h'_\varphi(\sigma) = l_\varphi - L_\varphi < 0$ on $M \setminus M_\varphi^\Psi$) or infinity if $L_\varphi = \infty$. On the other hand, the strict convexity of h_φ implies that the absolutely continuous part

$$
\int_M \{h_\varphi(\rho) - h_\varphi(\sigma) - h'_\varphi(\sigma)(\rho - \sigma)\}\, d\omega
$$

in (5.5) is nonnegative and equality holds if and only if $\rho = \sigma$ ω-a.e.. Thus $H_\varphi(\mu) \geq H_\varphi(\nu)$ and equality holds if and only if $\mu^s[M] = 0$ and $\rho = \sigma$ holds ω-a.e.. $\qquad \square$

Observe from (5.5) that it holds $D_\varphi(\mu|\nu) = H_\varphi(\mu) - H_\varphi(\nu)$ for any absolutely continuous measure μ with respect to ω. Thus the Bregman divergence $D_\varphi(\mu|\nu)$ measures the difference between the entropies at μ and the ground state ν. In [OT], we have studied the specific function $\varphi_m(s) = s^{2-m}$ and the associated *m-relative entropy* $H_m(\mu|\nu)$ for $m \in [(n-1)/n, 1) \cup (1, \infty)$. In the present context, $H_m(\mu|\nu)$ coincides with $H_{\varphi_m}(\mu) - H_{\varphi_m}(\nu)$.

The following lemma will be used in Section 8 (Claim 8.8) to construct a discrete approximation of the gradient flow of H_φ, where M is assumed to be compact.

Lemma 5.6 *Let M be compact. Then the φ-relative entropy H_φ is lower semi-continuous with respect to the weak topology, that is to say, if a sequence $\{\mu_i\}_{i \in \mathbb{N}} \subset \mathcal{P}(M)$ weakly converges to $\mu \in \mathcal{P}(M)$, then we have*

$$
H_\varphi(\mu) \leq \liminf_{i \to \infty} H_\varphi(\mu_i).
$$

Proof. We divide $H_\varphi(\mu)$ into two parts as

$$
H_\varphi^{(1)}(\mu) := \int_M h_\varphi(\rho)\, d\omega + h'_\varphi(\infty)\mu^s[M], \qquad H_\varphi^{(2)}(\mu) := - \int_M h'_\varphi(\sigma)\, d\mu,
$$

where $\mu = \rho\omega + \mu^s$. Note that $\|h'_\varphi(\sigma)\|_\infty < \infty$ thanks to the compactness of M (recall Remark 5.4(1)). Then $H_\varphi^{(2)}(\mu)$ is clearly continuous in μ and the lower semi-continuity of $H_\varphi^{(1)}(\mu)$ follows from [LV2, Theorem B.33]. $\qquad \square$

5.3 Displacement convexity of H_φ

In our previous work [OT], we showed that the displacement K-convexity of the m-relative entropy $H_m(\mu|\nu) = H_{\varphi_m}(\mu) - H_{\varphi_m}(\nu)$ with respect to $\mu \in \mathcal{P}^2(M)$ is equivalent to the combination of $\mathrm{Ric}_N \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$, where $N = 1/(1-m)$ ([OT, Theorem 4.1]). This characterization can be regarded as to correspond to Sturm's version of the curvature-dimension condition (5.1) ($\mathrm{Hess}\,H_m(\cdot|\nu) \geq K$ and (5.1) actually coincide if Ψ is constant and $K = 0$). In the reminder of the section, we shall consider the convexity of appropriate families of the φ-relative entropies corresponding to Lott and Villani's version of the curvature-dimension condition (5.2). Recall (2.3) for the definition of θ_φ.

Theorem 5.7 (Displacement convexity of families of H_φ) *Given $K \in \mathbb{R}$, $N \in \mathbb{R} \setminus (-1, n)$ and an admissible space $(M, \omega, \varphi_m, \Psi)$, the following three conditions are mutually equivalent, where $m = (N-1)/N$:*

(A) *We have $\mathrm{Ric}_N \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ on $M_{\varphi_m}^\Psi$.*

(B) *For any $\mu_0, \mu_1 \in \mathcal{P}^2(M)$ with $\mu_0[M_{\varphi_m}^\Psi] = \mu_1[M_{\varphi_m}^\Psi] = 1$ such that any pair of points $x_i \in \mathrm{supp}\,\mu_i \cap M_{\varphi_m}^\Psi$ $(i = 0, 1)$ are joined by some minimal geodesic contained in $M_{\varphi_m}^\Psi$, there exists a minimal geodesic $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}^2(M_{\varphi_m}^\Psi)$ along which*

$$H_{\varphi_m}(\mu_t) \leq (1-t)H_{\varphi_m}(\mu_0) + tH_{\varphi_m}(\mu_1) - \frac{K}{2}(1-t)tW_2(\mu_0, \mu_1)^2$$

holds for all $t \in [0, 1]$.

(C) *Take any φ with $\theta_\varphi \leq 2 - m$ such that $(M, \omega, \varphi, \Psi)$ is admissible. Then, for any $\mu_0, \mu_1 \in \mathcal{P}^2(M)$ with $\mu_0[M_\varphi^\Psi] = \mu_1[M_\varphi^\Psi] = 1$ such that any pair of points $x_i \in \mathrm{supp}\,\mu_i \cap M_\varphi^\Psi$ $(i = 0, 1)$ are joined by some minimal geodesic contained in M_φ^Ψ, there exists a minimal geodesic $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}^2(M_\varphi^\Psi)$ along which*

$$H_\varphi(\mu_t) \leq (1-t)H_\varphi(\mu_0) + tH_\varphi(\mu_1) - \frac{K}{2}(1-t)tW_2(\mu_0, \mu_1)^2 \qquad (5.6)$$

holds for all $t \in [0, 1]$.

Proof. The equivalence between (A) and (B) has been established in [OT, Theorem 4.1]. As (C) $\Rightarrow$ (B) is trivial (recall $\theta_{\varphi_m} = 2 - m$, see Subsection 2.4), it is enough to show (A) $\Rightarrow$ (C).

We can assume that both $H_\varphi(\mu_0)$ and $H_\varphi(\mu_1)$ are finite, otherwise the assertion (5.6) is obvious. We first consider the case where both μ_0 and μ_1 are absolutely continuous with respect to ω. By Theorem 2.6, there exists an almost everywhere twice differentiable function $\phi : \Omega \longrightarrow \mathbb{R}$ with $\mu_0[\Omega] = 1$ such that the map $\mathcal{T}_t(x) := \exp_x(t\nabla\phi(x))$ ($t \in [0, 1]$) gives the unique minimal geodesic $\mu_t := (\mathcal{T}_t)_\sharp \mu_0$ from μ_0 to μ_1. Given μ_0-a.e. x, $\mathcal{T}_1(x)$ is not a cut point of x due to [CMS1, Proposition 4.1], so that the geodesic $(\mathcal{T}_t(x))_{t \in [0,1]}$ is unique and contained in M_φ^Ψ. Put $\mu_t = \rho_t \omega$ and $\mathbf{J}_t^\omega(x) := e^{f(x)-f(\mathcal{T}_t(x))}\det(D\mathcal{T}_t(x))$.

By the change of variables formula with the Jacobian equation $(\rho_t \circ \mathcal{T}_t)\mathbf{J}_t^\omega = \rho_0$ μ_0-a.e. (Theorem 2.7), we deduce that

$$H_\varphi^{(1)}(\mu_t) := \int_M h_\varphi(\rho_t) \, d\omega = \int_M h_\varphi\big(\rho_t(\mathcal{T}_t)\big) \mathbf{J}_t^\omega \, d\omega$$

$$= \int_M h_\varphi\left(\frac{\rho_0}{\mathbf{J}_t^\omega}\right) \frac{\mathbf{J}_t^\omega}{\rho_0} \, d\mu_0 = \int_M \psi\left(\left(\frac{\mathbf{J}_t^\omega}{\rho_0}\right)^{1/N}\right) d\mu_0,$$

where we set $\psi(r) := r^N h_\varphi(r^{-N})$. As Theorem 3.5 together with the monotonicity of $\mathcal{DC}_N$ in m (Lemma 3.3) ensures $h_\varphi \in \mathcal{DC}_N$, the function $\psi(r)$ is non-increasing (resp. non-decreasing) and convex in r if $N \geq 1$ (resp. $N < 0$) due to Lemma 3.2.

Then the essential ingredient is the concavity of $N\mathbf{J}_t^\omega(x)^{1/N}$ as in the next claim. We give a sketch of the proof for completeness (see [St3] or [Oh2] for a detailed proof, where a more delicate estimate under $\mathrm{Ric}_N \geq K$ for general $K \in \mathbb{R}$ is discussed).

Claim 5.8 *Under* $\mathrm{Ric}_N \geq 0$, $N\mathbf{J}_t^\omega(x)^{1/N}$ *is concave in* t *for* μ_0-*a.e.* x.

Proof. Take an orthonormal basis $\{e_i\}_{i=1}^n$ of $T_x M$ and extend each e_i to the vector field $E_i(t) := D(\mathcal{T}_t)_x(e_i)$ for $t \in [0,1]$. Note that every E_i is a Jacobi field along the geodesic $\gamma(t) := \mathcal{T}_t(x)$, since $\mathcal{T}_t$ is a transport along geodesics. Let us consider the $n \times n$ matrix-valued functions $A(t) = (a_{ij}(t))$ and $B(t) = (b_{ij}(t))$ given by

$$a_{ij}(t) := \langle E_i(t), E_j(t) \rangle, \qquad D_{\dot{\gamma}} E_i(t) = \sum_{j=1}^n b_{ij}(t) E_j(t),$$

where $D_{\dot{\gamma}}$ is the covariant derivative along γ. Observe that $\mathbf{J}_t^\omega(x) = e^{f(x)-f(\gamma(t))}\sqrt{\det A(t)}$. We see by calculations $A' = 2BA$ and $A'' = -2\,\mathrm{Ric}_{\dot{\gamma}} + 2B^2 A$, where we set $\mathrm{Ric}_{\dot{\gamma}} := (\langle R(E_i, \dot{\gamma})\dot{\gamma}, E_j \rangle)_{i,j=1}^n$ and R stands for the Riemannian curvature tensor of (M, g). Combining these with the symmetry of B, we have

$$\frac{d^2}{dt^2}\left[(\det A)^{1/2n}\right] = \left\{\frac{(\mathrm{tr}\,B)^2}{n} - \mathrm{tr}(\mathrm{Ric}_{\dot{\gamma}}\,A^{-1}) - \mathrm{tr}(B^2)\right\}\frac{(\det A)^{1/2n}}{n} \leq -\frac{\mathrm{Ric}(\dot{\gamma})}{n}(\det A)^{1/2n}.$$

Put

$$v(t) := \mathbf{J}_t^\omega(x)^{1/N}, \quad v_1(t) := e^{\{f(x)-f(\gamma(t))\}/(N-n)}, \quad v_2(t) := \{\det A(t)\}^{1/2n}.$$

As $v = v_1^{(N-n)/N} v_2^{n/N}$, we obtain

$$Nv^{-1}v'' = (N-n)v_1^{-1}v_1'' + nv_2^{-1}v_2'' - \frac{(N-n)n}{N}(v_1^{-1}v_1' - v_2^{-1}v_2')^2$$

$$\leq -(f \circ \gamma)'' + \frac{\{(f \circ \gamma)'\}^2}{N-n} - \mathrm{Ric}(\dot{\gamma}) = -\,\mathrm{Ric}_N(\dot{\gamma}).$$

Note that the range of $N \in (-\infty, 0) \cup [n, \infty)$ is essential here for making $(N-n)/N$ nonnegative. Thus the assumption $\mathrm{Ric}_N \geq 0$ implies $Nv'' \leq 0$, so that $Nv = N\mathbf{J}_t^\omega(x)^{1/N}$ is concave in t. $\diamond$

Therefore we have, as $\mathbf{J}_0^\omega \equiv 1$,

$$\psi\big((\mathbf{J}_t^\omega/\rho_0)^{1/N}\big) \leq \psi\big((1-t)(1/\rho_0)^{1/N} + t(\mathbf{J}_1^\omega/\rho_0)^{1/N}\big)$$
$$\leq (1-t)\psi\big((1/\rho_0)^{1/N}\big) + t\psi\big((\mathbf{J}_1^\omega/\rho_0)^{1/N}\big) \tag{5.7}$$

μ_0-a.e.. This implies

$$H_\varphi^{(1)}(\mu_t) = \int_M \psi\left(\left(\frac{\mathbf{J}_t^\omega}{\rho_0}\right)^{1/N}\right) d\mu_0$$
$$\leq \int_M \left\{ (1-t)\psi\left(\left(\frac{1}{\rho_0}\right)^{1/N}\right) + t\psi\left(\left(\frac{\mathbf{J}_1^\omega}{\rho_0}\right)^{1/N}\right) \right\} d\mu_0$$
$$= (1-t)H_\varphi^{(1)}(\mu_0) + tH_\varphi^{(1)}(\mu_1). \tag{5.8}$$

On the other hand, it follows from $\operatorname{Hess}\Psi \geq K$ that

$$\int_M \Psi \, d\mu_t = \int_M \Psi(\mathcal{T}_t) \, d\mu_0 \leq \int_M \left\{ (1-t)\Psi(\mathcal{T}_0) + t\Psi(\mathcal{T}_1) - \frac{K}{2}(1-t)t d_g(\mathcal{T}_0, \mathcal{T}_1)^2 \right\} d\mu_0,$$

and hence

$$H_\varphi^{(2)}(\mu_t) := -\int_M h_\varphi'(\sigma) \, d\mu_t \leq (1-t)H_\varphi^{(2)}(\mu_0) + tH_\varphi^{(2)}(\mu_1) - \frac{K}{2}(1-t)tW_2(\mu_0, \mu_1)^2. \tag{5.9}$$

Combining (5.8) with (5.9), we obtain the desired inequality (5.6).

Let us next consider the case where μ_0 or μ_1 has nontrivial singular part. We may assume $L_\varphi < \infty$, otherwise (5.6) trivially holds by the definition of $h_\varphi'(\infty)$. Decompose μ_0 and μ_1 as $\mu_0 = \rho_0\omega + \mu_0^s$ and $\mu_1 = \rho_1\omega + \mu_1^s$, and take an optimal coupling π of μ_0 and μ_1. Let $p_1, p_2 : M \times M \longrightarrow M$ denote the projections to the first and second components. Now, π is decomposed into four parts

$$\pi = \pi_{aa} + \pi_{as} + \pi_{sa} + \pi_{ss}$$

such that $(p_1)_\sharp(\pi_{aa})$, $(p_1)_\sharp(\pi_{as})$, $(p_2)_\sharp(\pi_{aa})$ and $(p_2)_\sharp(\pi_{sa})$ are absolutely continuous and that $(p_1)_\sharp(\pi_{sa})$, $(p_1)_\sharp(\pi_{ss})$, $(p_2)_\sharp(\pi_{as})$ and $(p_2)_\sharp(\pi_{ss})$ are singular or null measures. We divide optimal transport between μ_0 and μ_1 into two parts, corresponding to $\pi - \pi_{ss}$ and π_{ss}. For $\hat{\mu}_0 := (p_1)_\sharp(\pi - \pi_{ss})$ and $\hat{\mu}_1 := (p_2)_\sharp(\pi - \pi_{ss})$, Theorem 2.6 guarantees the existence of a geodesic

$$\hat{\mu}_t \in (1 - \pi_{ss}[M \times M]) \cdot \mathcal{P}_{ac}^2(M, \omega), \qquad t \in (0, 1),$$

(i.e., $\hat{\mu}_t[M] = 1 - \pi_{ss}[M \times M]$) such that $\hat{\mu}_t[M_\varphi^\Psi] = \hat{\mu}_t[M]$. Setting $\hat{\mu}_t = \hat{\rho}_t\omega$, we observe

$$\int_M h_\varphi(\hat{\rho}_t) \, d\omega \leq (1-t)\int_M h_\varphi(\rho_0) \, d\omega + t\int_M h_\varphi(\rho_1) \, d\omega,$$
$$-\int_M h_\varphi'(\sigma) \, d\hat{\mu}_t \leq -(1-t)\int_M h_\varphi'(\sigma) \, d\hat{\mu}_0 - t\int_M h_\varphi'(\sigma) \, d\hat{\mu}_1$$
$$- \frac{K}{2}(1-t)t\int_{M \times M} d_g(x, y)^2 \, d(\pi - \pi_{ss})(x, y).$$

30

To be precise, in the first inequality, we used $h_\varphi \leq 0$ along the transports corresponding to π_{as} and π_{sa}. By Proposition 2.5, we find a minimal geodesic

$$\tilde{\mu}_t = \tilde{\rho}_t \omega + \tilde{\mu}_t^s \in \pi_{ss}[M \times M] \cdot \mathcal{P}^2(M)$$

from $\tilde{\mu}_0 := (p_1)_\sharp(\pi_{ss})$ to $\tilde{\mu}_1 := (p_2)_\sharp(\pi_{ss})$ realized through a family of geodesics in M_φ^Ψ. Then the condition Hess $\Psi \geq K$ implies

$$-\int_M h'_\varphi(\sigma)\, d\tilde{\mu}_t \leq -(1-t)\int_M h'_\varphi(\sigma)\, d\tilde{\mu}_0 - t\int_M h'_\varphi(\sigma)\, d\tilde{\mu}_1$$
$$-\frac{K}{2}(1-t)t\int_{M \times M} d_g(x,y)^2\, d\pi_{ss}(x,y).$$

We put $\mu_t := \hat{\mu}_t + \tilde{\mu}_t$ and conclude that

$$H_\varphi(\mu_t) = \int_M h_\varphi(\hat{\rho}_t + \tilde{\rho}_t)\, d\omega - \int_M h'_\varphi(\sigma)\, d\mu_t \leq \int_M h_\varphi(\hat{\rho}_t)\, d\omega - \int_M h'_\varphi(\sigma)\, d(\hat{\mu}_t + \tilde{\mu}_t)$$
$$\leq (1-t)H_\varphi(\mu_0) + tH_\varphi(\mu_1) - \frac{K}{2}(1-t)tW_2(\mu_0,\mu_1)^2,$$

where we used the fact that h_φ is non-increasing (since $L_\varphi < \infty$) in the second line. $\qquad \square$

Remark 5.9 Recall that $M_\varphi^\Psi = M_{\varphi_m}^\Psi = M$ if $l_\varphi = -\infty$ by definition (and admissibility). Hence the condition in (B) and (C) that $\text{supp}\,\mu_0$ and $\text{supp}\,\mu_1$ are connected in M_φ^Ψ is nontrivial only if $l_\varphi > -\infty$. Even when $l_\varphi > -\infty$, Lemma 4.2(i) guarantees that M_φ^Ψ is totally convex if Hess $\Psi \geq K > 0$.

In the limit case of $N = \infty$ $(m = 1)$, we can follow the proof of (A) $\Rightarrow$ (C) using $\psi(r) = e^r h_\varphi(e^{-r})$ as well as the concavity of $\log(\mathbf{J}_t^\omega(x))$. However, the implication (B) $\Rightarrow$ (A) is not true. This is because the two weights f and Ψ are synchronized as $\nu = e^{-f-\Psi}\text{vol}_g$ and we can control only the behavior of $f + \Psi$ (see the proof of (B) $\Rightarrow$ (A) sketched in the next subsection).

Instead, one can see from Theorem 5.2 that $\text{Ric}_\infty \geq K$ (of (M, ω)) implies the $\lambda(K, U)$-convexity of U_ω for all $U \in \mathcal{DC}_\infty$, where

$$\lambda(K, U) := \inf_{r>0} K\frac{rU'(r) - U(r)}{r} = \begin{cases} K\lim_{r\downarrow 0}\{rU'(r) - U(r)\}/r & \text{if } K > 0, \\ 0 & \text{if } K = 0, \\ K\lim_{r\to\infty}\{rU'(r) - U(r)\}/r & \text{if } K < 0 \end{cases}$$

([LV2, Theorem 7.3], [Vi2, Theorem 30.5]).

6 Functional inequalities

If $\text{Ric}_{N_\varphi} \geq 0$ and Hess $\Psi \geq K$ for some $K > 0$, then we can obtain variants of the Talagrand inequality, the HWI inequality, the logarithmic Sobolev inequality and the global Poincaré inequality. These are derived from fundamental properties of convex functions along the lines of [OV] and [LV2, Section 6] (see also [OT, Section 5] where

we studied the special case of the m-relative entropies). We will impose only the strictly weaker condition $\mathrm{Hess}\, H_\varphi \geq K > 0$ (for single φ) in the φ-Talagrand inequality for the use in the next section.

For $\mu = \rho\omega \in \mathcal{P}^2_{\mathrm{ac}}(M, \omega)$ with $\mu[M^\Psi_\varphi] = 1$, we define the φ-relative Fisher information with respect to $\nu = \sigma\omega$ by

$$I_\varphi(\mu) := \int_M |\nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)]|^2 \, d\mu = \int_M \left| \frac{\nabla\rho}{\varphi(\rho)} + \nabla\Psi \right|^2 d\mu \qquad (6.1)$$

provided that it is well-defined, otherwise we set $I_\varphi(\mu) := \infty$. This quantity describes the directional derivatives of H_φ as follows. (At this point the treatment in [OT] was somewhat too rough, the argument in the present paper is correct.)

Proposition 6.1 (Directional derivatives of H_φ) *Let $(M, \omega, \varphi, \Psi)$ be an admissible space with $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ on M^Ψ_φ for some $K \in \mathbb{R}$, and $\mu = \rho\omega \in \mathcal{P}^2_{\mathrm{ac}}(M, \omega)$ be such that $\mu[M^\Psi_\varphi] = 1$, $H_\varphi(\mu) < \infty$, $\rho h'_\varphi(\rho) - h_\varphi(\rho) \in H^1_{\mathrm{loc}}(M)$ and that $|\nabla\rho/\varphi(\rho)| + |\nabla\Psi| \in L^2(M, \mu)$. Take a minimal geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2(M)$ emanating from $\mu_0 = \mu$ generated by a locally semi-convex function $\phi : M \longrightarrow \mathbb{R}$ as $\mu_t = (\mathcal{T}_t)_\sharp \mu$ with $\mathcal{T}_t(x) = \exp_x(t\nabla\phi(x))$. If $\theta_\varphi < 1$, then we further suppose that $\mathrm{supp}\,\mu_0$ and $\mathrm{supp}\,\mu_1$ are compact. Then we have*

$$\liminf_{t\downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} \geq \int_M \left\langle \frac{\nabla\rho}{\varphi(\rho)} + \nabla\Psi, \nabla\phi \right\rangle d\mu. \qquad (6.2)$$

Moreover, equality holds in (6.2) (with $\lim_{t\downarrow 0}$ in place of $\liminf_{t\downarrow 0}$) if $\phi \in C^\infty_c(M)$.

Proof. We first deduce equality in (6.2) for $\phi \in C^\infty_c(M)$. Put $\mu_t = \rho_t \omega$ and $\mathbf{J}^\omega_t := e^{f - f(\mathcal{T}_t)}\det(D\mathcal{T}_t)$ as in the proof of Theorem 5.7. We follow the calculation in Theorem 5.7 and see

$$H_\varphi(\mu_t) - H_\varphi(\mu) = \int_M \left\{ h_\varphi\left(\frac{\rho}{\mathbf{J}^\omega_t} \right) \mathbf{J}^\omega_t - h_\varphi(\rho) \right\} d\omega - \int_M \{h'_\varphi(\sigma \circ \mathcal{T}_t) - h'_\varphi(\sigma)\} \, d\mu.$$

By the convexity (5.7) of $h_\varphi(\rho/\mathbf{J}^\omega_t)\mathbf{J}^\omega_t$ in t and $\mathrm{Hess}\,\Psi \geq K$, we can apply the monotone convergence theorem and obtain

$$\lim_{t\downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} = \int_M \lim_{t\downarrow 0} \frac{h_\varphi(\rho/\mathbf{J}^\omega_t)\mathbf{J}^\omega_t - h_\varphi(\rho)}{t} \, d\omega + \int_M \langle \nabla\Psi, \nabla\phi \rangle \, d\mu.$$

Note that, by using the weighted Laplacian Δ^ω introduced at the beginning of Subsection 8.3 below,

$$\lim_{t\downarrow 0} \frac{\mathbf{J}^\omega_t - 1}{t} = \lim_{t\downarrow 0} \frac{e^{f - f(\mathcal{T}_t)}\det(D\mathcal{T}_t) - 1}{t} = \mathrm{tr}(\mathrm{Hess}\,\phi) - \langle \nabla\phi, \nabla f \rangle$$
$$= \Delta\phi - \langle \nabla\phi, \nabla f \rangle = \Delta^\omega\phi.$$

Thus we have

$$\lim_{t\downarrow 0} \frac{h_\varphi(\rho/\mathbf{J}^\omega_t)\mathbf{J}^\omega_t - h_\varphi(\rho)}{t} = \{h_\varphi(\rho) - h'_\varphi(\rho)\rho\} \lim_{t\downarrow 0} \frac{\mathbf{J}^\omega_t - 1}{t} = \{h_\varphi(\rho) - h'_\varphi(\rho)\rho\}\Delta^\omega\phi.$$

32

Therefore we conclude, by the integration by parts for Δ^ω (since $\phi \in C_c^\infty(M)$),

$$\lim_{t \downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} = \int_M \langle \nabla[h_\varphi'(\rho)\rho - h_\varphi(\rho)] + \rho \nabla \Psi, \nabla \phi \rangle \, d\omega$$

$$= \int_M \left\langle \frac{\nabla \rho}{\varphi(\rho)} + \nabla \Psi, \nabla \phi \right\rangle d\mu.$$

In the case of $\phi \notin C_c^\infty(M)$, we need to take care about the last step of integration by parts. If $\theta_\varphi \geq 1$ (equivalently, $N_\varphi \in [n, \infty] \cap (2, \infty]$), then we can directly apply [Vi2, Theorem 23.14] to see (6.2). For $\theta_\varphi < 1$, the same proof (Step 3 in [Vi2, Theorem 23.14]) still works provided that $\operatorname{supp} \mu_0$ and $\operatorname{supp} \mu_1$ are compact. $\square$

Remark 6.2 Let us add some more remarks on the case of $\theta_\varphi < 1$. A large part of the proof of [Vi2, Theorem 23.14] also works in this case (even without the approximation procedure based on [Vi2, Proposition 17.7]). Proposition 2.13(i) ensures $l_\varphi > -\infty$ so that $u_\varphi(r) \geq l_\varphi r$ for all $r \geq 0$, and Lemma 2.10 shows

$$\frac{s}{\varphi(s)} \leq \left(\frac{s}{t} \right)^{1-\theta_\varphi} \frac{t}{\varphi(t)} \leq \frac{t}{\varphi(t)}$$

for all $0 < s < t$, which corresponds to [Vi2, (23.52)] with $A = \infty$ (hence (23.53) and (23.54) are unnecessary). Note that $p'(s)$ in [Vi2] is $s/\varphi(s)$ in our context. The only problem is that $s/\varphi(s)$ is never bounded for large s, as seen from the model case of $s/\varphi_m(s) = s^{m-1}$ with $m > 1$. The boundedness is used to guarantee $p'(\rho) \in L^2(M, \mu)$, so that we can assume $\rho/\varphi(\rho) \in L^2(M, \mu)$ instead of the compactness of $\operatorname{supp} \mu_0 \cup \operatorname{supp} \mu_1$ in Proposition 6.1 above.

We assume $\nu[M] = 1$ by scaling (recall Remark 4.6), and prove functional inequalities associated with H_φ.

Theorem 6.3 Let $(M, \omega, \varphi, \Psi)$ be admissible with $\nu \in \mathcal{P}_{\mathrm{ac}}^2(M, \omega)$. We set $H_\nu := H_\varphi(\nu)$ for brevity.

(i) (φ-Talagrand inequality) Suppose that $\operatorname{Hess} H_\varphi \geq K$ for some $K > 0$. For any $\mu \in \mathcal{P}^2(M)$, we have

$$W_2(\mu, \nu) \leq \sqrt{\frac{2}{K}(H_\varphi(\mu) - H_\nu)}. \tag{6.3}$$

(ii) (φ-HWI, φ-logarithmic Sobolev inequalities) Assume $\operatorname{Ric}_{N_\varphi} \geq 0$ and $\operatorname{Hess} \Psi \geq K$ on M_φ^Ψ for some $K > 0$. Given $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}^2(M)$ with $\mu[M_\varphi^\Psi] = 1$ such that $H_\varphi(\mu) < \infty$ and that ρ is locally Lipschitz, we have

$$H_\varphi(\mu) - H_\nu \leq \sqrt{I_\varphi(\mu)} \cdot W_2(\mu, \nu) - \frac{K}{2} W_2(\mu, \nu)^2, \tag{6.4}$$

$$H_\varphi(\mu) - H_\nu \leq \frac{1}{2K} I_\varphi(\mu). \tag{6.5}$$

(iii) (φ-global Poincaré inequality) *Let (M, g) be compact and φ be C^1, and assume* $\mathrm{Ric}_{N_\varphi} \geq 0$ *and* $\mathrm{Hess}\,\Psi \geq K$ *on* M_φ^Ψ *for some* $K > 0$. *Then for any Lipschitz function* $w : M_\varphi^\Psi \longrightarrow \mathbb{R}$ *such that* $\int_{M_\varphi^\Psi} w \, d\nu = 0$, *we have*

$$\int_{M_\varphi^\Psi} \frac{w^2 \sigma}{\varphi(\sigma)} \, d\nu \leq \frac{1}{K} \int_{M_\varphi^\Psi} \left| \nabla \left(\frac{w\sigma}{\varphi(\sigma)} \right) \right|^2 d\nu. \tag{6.6}$$

Proof. We first remark that M_φ^Ψ is totally convex if $\mathrm{Hess}\,\Psi \geq K > 0$ (see Lemma 4.2 and Remark 5.9 as well). Thus M_φ^Ψ is totally convex in (ii) and (iii).

(i) There is nothing to prove if $H_\varphi(\mu) = \infty$, so that we assume $H_\varphi(\mu) < \infty$. By the hypothesis $\mathrm{Hess}\,H_\varphi \geq K$, there is a minimal geodesic $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}^2(M)$ from $\mu_0 = \mu$ to $\mu_1 = \nu$ such that

$$0 \leq H_\varphi(\mu_t) - H_\nu \leq (1-t)H_\varphi(\mu) - (1-t)H_\nu - \frac{K}{2}(1-t)tW_2(\mu, \nu)^2 \tag{6.7}$$

for all $t \in [0, 1]$. Dividing both sides with $1 - t$ and letting t go to 1, we obtain the desired inequality (6.3).

(ii) As the case of $I_\varphi(\mu) = \infty$ is trivial, we assume $I_\varphi(\mu) < \infty$. For the minimal geodesic $(\mu_t)_{t \in [0,1]}$ from $\mu_0 = \mu$ to $\mu_1 = \nu$, Theorem 2.6 ensures that $\mu_t \in \mathcal{P}^2_{\mathrm{ac}}(M, \omega)$ for all $t \in [0, 1]$ and there is a locally semi-convex function ϕ such that $\mu_t = (\mathcal{T}_t)_\sharp \mu$ with $\mathcal{T}_t(x) = \exp_x(t\nabla\phi(x))$. Due to (6.7), we have

$$\frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} \leq -H_\varphi(\mu) + H_\nu - \frac{K}{2}(1-t)W_2(\mu, \nu)^2. \tag{6.8}$$

Moreover, Proposition 6.1 shows that

$$\liminf_{t \downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} \geq \int_M \langle \nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)], \nabla\phi \rangle \, d\mu.$$

We remark that M_φ^Ψ is bounded if $\theta_\varphi < 1$ (by Proposition 2.13(i) and Lemma 4.2(i)), so that Proposition 6.1 is certainly available. We obtain from the Cauchy–Schwarz inequality that

$$\liminf_{t \downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} \geq -\left(\int_M |\nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)]|^2 \, d\mu \right)^{1/2} \left(\int_M |\nabla\phi|^2 \, d\mu \right)^{1/2}$$

$$= -\sqrt{I_\varphi(\mu)} \cdot W_2(\mu, \nu),$$

where the last equality follows from

$$|\nabla\phi(x)| = d_g\big(x, \exp_x(\nabla\phi)\big) = d_g\big(\mathcal{T}_0(x), \mathcal{T}_1(x)\big) \qquad \mu\text{-a.e. } x.$$

Combining this with (6.8), we obtain (6.4). By completing the square, we deduce (6.5) from (6.4) as

$$\sqrt{I_\varphi(\mu)} \cdot W_2(\mu, \nu) - \frac{K}{2}W_2(\mu, \nu)^2 = -\frac{K}{2}\left(W_2(\mu, \nu) - \frac{1}{K}\sqrt{I_\varphi(\mu)} \right)^2 + \frac{I_\varphi(\mu)}{2K} \leq \frac{I_\varphi(\mu)}{2K}.$$

(iii) For small $\varepsilon > 0$, we put $\mu = \rho\omega := (1 + \varepsilon w)\sigma\omega$. We remark that $H_\varphi(\mu)$ and $I_\varphi(\mu)$ are finite as M is compact. It follows from (6.5) that

$$\int_{M_\varphi^\Psi} \{u_\varphi(\rho) - u_\varphi(\sigma) - u'_\varphi(\sigma)(\rho - \sigma)\}\, d\omega \leq \frac{1}{2K} \int_{M_\varphi^\Psi} |\nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)]|^2\, d\mu.$$

On the one hand, we have by expansion

$$u_\varphi(\rho) - u_\varphi(\sigma) - u'_\varphi(\sigma)(\rho - \sigma) = \frac{(\rho - \sigma)^2}{2} u''_\varphi(\sigma) + O\left((\rho - \sigma)^3\right) = \frac{\varepsilon^2 w^2 \sigma^2}{2\varphi(\sigma)} + O(\varepsilon^3),$$

where $O(\varepsilon^3)$ is uniform on M (for fixed w) thanks to the compactness of M. On the other hand, it holds

$$|\nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)]|^2 = \left|\nabla\left[(\rho - \sigma)\ln'_\varphi(\sigma) + O\left((\rho - \sigma)^2\right)\right]\right|^2 = \left|\nabla\left(\frac{\varepsilon w \sigma}{\varphi(\sigma)}\right) + O(\varepsilon^2)\right|^2$$

$$= \varepsilon^2 \left|\nabla\left(\frac{w\sigma}{\varphi(\sigma)}\right)\right|^2 + O(\varepsilon^3).$$

Thus we have, letting ε go to zero,

$$\int_{M_\varphi^\Psi} \frac{w^2 \sigma}{\varphi(\sigma)}\, d\nu = \int_{M_\varphi^\Psi} \frac{w^2 \sigma^2}{\varphi(\sigma)}\, d\omega \leq \frac{1}{K} \int_{M_\varphi^\Psi} \left|\nabla\left(\frac{w\sigma}{\varphi(\sigma)}\right)\right|^2\, d\nu.$$

$\qquad\qquad\square$

The φ-Talagrand inequality (6.3) is regarded as a comparison between the distance functions appearing in Wasserstein geometry and information geometry, since the square root of the Bregman divergence behaves like a distance function (see Subsection 2.3). Note that, in the φ-global Poincaré inequality (6.6), the usual global Poincaré inequality $\int_M w^2\, d\nu \leq K^{-1} \int_M |\nabla w|^2\, d\nu$ is indeed recovered when $\varphi(s) = \varphi_1(s) = s$. Other inequalities are also clearly reduced to the usual ones for $\varphi = \varphi_1$.

7 Concentration of measures

The aim of this section is to derive the concentration of measures from the φ-Talagrand inequality (6.3). Let us assume $\nu[M] = 1$ (see Remark 4.6) and define the *concentration function* by

$$\alpha(r) = \alpha_{(M,\nu)}(r) := \sup\left\{1 - \nu[B(A, r)] \,\middle|\, A \subset M: \text{ measurable, } \nu[A] \geq 1/2\right\}$$

for $r > 0$, where $B(A, r) := \{y \in M \mid \inf_{x \in A} d_g(x, y) < r\}$. The function α describes how the probability measure ν is concentrating on the neighborhood of an *arbitrary* set A of half the total measure in a quantitative way (in other words, a kind of large (or moderate) deviation principle). Equivalently, α measures how any 1-Lipschitz function is close to the constant function at its mean. We refer to the excellent book [Le] for an introduction to the concentration of measure phenomenon.

In the classical case of $\varphi_1(s) = s$, the Talagrand inequality (6.3) implies the *normal concentration* $\alpha(r) \leq C \exp(-cr^2)$ or equivalently $\alpha(r)^{-1} \geq C^{-1} \exp(cr^2)$ with constants $c, C > 0$ depending only on K (see [Le, Section 6.1]). For general φ, we will similarly derive from (6.3) the *m-normal concentration* involving the *m*-exponential function e_m (see Subsection 2.4). Precisely, we have $\alpha(r) \leq C e_m(-cr^2)$ with $m = m(\varphi) \leq 2 - \theta_\varphi$ if $\theta_\varphi > 1$, and $\alpha(r)^{-1} \geq C^{-1} e_m(cr^2)$ with $m = m(\varphi) \geq 2 - \theta_\varphi$ if $\theta_\varphi \leq 1$ (Theorem 7.9).

7.1 General estimate

For each measurable set $A \subset M$ with $0 < \nu[A] < \infty$, denote the normalized restriction of ν on A by

$$\nu_A := \frac{\chi_A}{\nu[A]} \nu \in \mathcal{P}_{\mathrm{ac}}(M, \omega).$$

To analyze its entropy $H_\varphi(\nu_A)$, we introduce the function

$$U(\xi, t) := u_\varphi\left(\frac{\xi}{t}\right) - \frac{\xi}{t} \ln_\varphi(\xi), \qquad (\xi, t) \in (0, \infty) \times (0, 1].$$

Note that $H_\varphi(\nu_A) = \int_A U(\sigma, \nu[A]) \, d\omega$. To be precise, we can set $U(\sigma, \nu[A]) := 0$ on $M \setminus M_\varphi^\Psi$ thanks to the following lemma.

Lemma 7.1 *If $\theta_\varphi < 2$, then we have $\lim_{\xi\downarrow 0} U(\xi, t) = 0$ for every $t \in (0, 1]$.*

Proof. Note that

$$U(\xi, t) = \int_0^{\xi/t} \{\ln_\varphi(s) - \ln_\varphi(\xi)\} \, ds = \int_0^{\xi/t} \int_\xi^s \frac{1}{\varphi(r)} \, dr ds$$

$$= -\int_0^\xi \int_s^\xi \frac{1}{\varphi(r)} \, dr ds + \int_\xi^{\xi/t} \int_\xi^s \frac{1}{\varphi(r)} \, dr ds = -\int_0^{\xi/t} \frac{r}{\varphi(r)} \, dr + \frac{\xi}{t} \int_\xi^{\xi/t} \frac{1}{\varphi(r)} \, dr. \tag{7.1}$$

Then (2.1) (deduced from Theorem 3.5) shows the claim. $\square$

Lemma 7.2 *Assume $\theta_\varphi < 2$. For any measurable set A with $\nu[A] \geq 1/2$, we have $H_\varphi(\nu_A) \leq 0$. In particular, it holds $H_\varphi(\nu) \leq 0$ if $\nu[M] = 1$.*

Proof. For any $\xi > 0$, $U(\xi, t)$ is non-increasing in $t \in (0, 1]$ since we have

$$\frac{\partial U}{\partial t}(\xi, t) = -\frac{\xi}{t^2} \ln_\varphi\left(\frac{\xi}{t}\right) + \frac{\xi}{t^2} \ln_\varphi(\xi) = \frac{\xi}{t^2} \int_{\xi/t}^\xi \frac{1}{\varphi(s)} \, ds \leq 0.$$

Similarly, $U(\xi, 1/2)$ is non-increasing in ξ due to

$$\frac{dU}{d\xi}\left(\xi, \frac{1}{2}\right) = 2 \ln_\varphi(2\xi) - 2 \ln_\varphi(\xi) - \frac{2\xi}{\varphi(\xi)} = 2 \int_\xi^{2\xi} \left(\frac{1}{\varphi(s)} - \frac{1}{\varphi(\xi)}\right) ds \leq 0.$$

Thus we deduce from Lemma 7.1 that, for any $\xi > 0$ and $t \geq 1/2$,

$$U(\xi, t) \leq U(\xi, 1/2) \leq \lim_{\xi\downarrow 0} U(\xi, 1/2) = 0,$$

which shows $H_\varphi(\nu_A) = \int_A U(\sigma, \nu[A]) \, d\omega \leq 0$. $\square$

36

Next we give an estimate on $H_\varphi(\nu_B)$ for $B \subset M$ not necessarily $\nu[B] \geq 1/2$. Recall (2.4) for the definition of δ_φ.

Lemma 7.3 *Assume $\theta_\varphi < \infty$ and $\|\sigma\|_\infty < \infty$. Given any measurable set $B \subset M$ with $0 < \nu[B] < \infty$ and any $\xi_0 \geq \max\{\nu[B], \|\sigma\|_\infty\}$, we have*

$$H_\varphi(\nu_B) \leq -\nu[B]^{\delta_\varphi - 2} \ln_\varphi(\nu[B]) \xi_0^{\theta_\varphi - \delta_\varphi} \int_B \sigma^{2 - \theta_\varphi} \, d\omega. \tag{7.2}$$

Proof. For $t, \xi \in (0, \xi_0]$, we deduce from (7.1) that

$$U(\xi, t) \leq \frac{\xi}{t} \int_\xi^{\xi/t} \frac{1}{\varphi(r)} \, dr = \frac{\xi^2}{t^2} \int_t^1 \frac{1}{\varphi(s\xi t^{-1})} \, ds,$$

where we changed the variables as $r = s\xi t^{-1}$. Note that Lemma 2.10 shows that for all $s > 0$

$$\frac{\xi^2}{\varphi(s\xi t^{-1})} \leq \frac{\xi^{2 - \theta_\varphi} \xi_0^{\theta_\varphi}}{\varphi(s\xi_0 t^{-1})} \leq \frac{\xi^{2 - \theta_\varphi} \xi_0^{\theta_\varphi}}{\varphi(s)} \left(\frac{\xi_0}{t}\right)^{-\delta_\varphi}.$$

Thus we find

$$U(\xi, t) \leq -t^{\delta_\varphi - 2} \ln_\varphi(t) \xi_0^{\theta_\varphi - \delta_\varphi} \xi^{2 - \theta_\varphi}.$$

This implies

$$H_\varphi(\nu_B) = \int_B U(\sigma, \nu[B]) \, d\omega \leq -\nu[B]^{\delta_\varphi - 2} \ln_\varphi(\nu[B]) \xi_0^{\theta_\varphi - \delta_\varphi} \int_B \sigma^{2 - \theta_\varphi} \, d\omega$$

as desired. $\square$

We remark that, if $\delta_\varphi \leq 2$, then we have at any $s \in (0, 1)$

$$\frac{d}{ds} \left[s^{\delta_\varphi - 2} \ln_\varphi(s)\right] = s^{\delta_\varphi - 3} \left\{(\delta_\varphi - 2) \ln_\varphi(s) + \frac{s}{\varphi(s)}\right\} > 0.$$

Therefore the right hand side of (7.2) is non-increasing in $\nu[B]$ provided that ν is a probability measure.

Now we show a general estimate of $\alpha(r)$ under the strict convexity of H_φ.

Proposition 7.4 *Assume that $(M, \omega, \varphi, \Psi)$ is admissible, $\nu \in \mathcal{P}_{ac}(M, \omega)$, $\mathrm{Hess}\, H_\varphi \geq K$ for some $K > 0$ and that $\|\sigma\|_\infty < \infty$. We set $H_\nu := H_\varphi(\nu)$ (≤ 0) as in Theorem 6.3. Then, for any $\xi_0 \geq \max\{1/2, \|\sigma\|_\infty\}$ and any $r > 0$ with $\alpha(r) > 0$, we have*

$$\alpha(r)^{\delta_\varphi - 2} \ln_\varphi\big(\alpha(r)\big) \xi_0^{\theta_\varphi - \delta_\varphi} \cdot \sup_A \int_B \sigma^{2 - \theta_\varphi} \, d\omega \leq -\left(\sqrt{\frac{K}{2}} r - \sqrt{-H_\nu}\right)^2 - H_\nu, \tag{7.3}$$

where $A \subset M$ runs over all measurable sets of $\nu[A] \geq 1/2$ and we set $B := M \setminus B(A, r)$.

37

Proof. Since $\alpha(r) \leq 1$ by definition, the left hand side of (7.3) is always nonpositive. Therefore the assertion is clear if $r^2 \leq -8H_\nu/K$. Suppose $r^2 > -8H_\nu/K$, take a measurable set $A \subset M$ with $\nu[A] \geq 1/2$ and put $B := M \setminus B(A, r)$. We also assume $\nu[B] > 0$ since we have $\alpha(r) = 0$ if $\nu[B] = 0$ for all such A.

Observe that $W_1(\nu_A, \nu_B) \geq r$ as $d_g(x, y) \geq r$ for all $x \in A$ and $y \in B$. Note also that $W_1 \leq W_2$ holds by the Cauchy–Schwarz inequality. Then the triangle inequality for W_1 and the φ-Talagrand inequality (6.3) yield

$$r \leq W_1(\nu_A, \nu_B) \leq W_1(\nu_A, \nu) + W_1(\nu, \nu_B) \leq \sqrt{\frac{2}{K}\left(H_\varphi(\nu_A) - H_\nu\right)} + \sqrt{\frac{2}{K}\left(H_\varphi(\nu_B) - H_\nu\right)}.$$

Applying Lemma 7.2 gives, as $r^2 > -8H_\nu/K$ ensures $\sqrt{K/2}\,r \geq \sqrt{-H_\nu}$,

$$H_\varphi(\nu_B) \geq \left(\sqrt{\frac{K}{2}}r - \sqrt{-H_\nu}\right)^2 + H_\nu.$$

Since A is arbitrary and $\nu[B] \leq 1/2 \leq \xi_0$, combining the above estimate with Lemma 7.3 yields

$$-\left(\sqrt{\frac{K}{2}}r - \sqrt{-H_\nu}\right)^2 - H_\nu \geq \sup_A \left\{\nu[B]^{\delta_\varphi - 2}\ln_\varphi(\nu[B])\xi_0^{\theta_\varphi - \delta_\varphi}\int_B \sigma^{2-\theta_\varphi}\,d\omega\right\}$$

$$\geq \alpha(r)^{\delta_\varphi - 2}\ln_\varphi(\alpha(r))\xi_0^{\theta_\varphi - \delta_\varphi} \cdot \sup_A \int_B \sigma^{2-\theta_\varphi}\,d\omega.$$

$\square$

7.2 Concentration of measures

We shall obtain the concentration of $\{(M, \omega, \varphi, \Psi_i)\}_{i\in\mathbb{N}}$, i.e., $\lim_{i\to\infty} \alpha_{(M,\nu_i)}(r) = 0$ for all $r > 0$ with $\nu_i := \exp_\varphi(-\Psi_i)\omega$, under an appropriate condition on the convexity of H_φ associated with Ψ_i. We first prove an auxiliary lemma.

Lemma 7.5 *Assume that $(M, \omega, \varphi, \Psi)$ is admissible, $\nu \in \mathcal{P}_{ac}(M, \omega)$ and that $\|\sigma\|_\infty < \infty$. Set $H_\nu := H_\varphi(\nu)$ and take arbitrary $\xi_0 \geq \|\sigma\|_\infty$.*

(i) *If $\theta_\varphi \leq 1$, then we have*

$$\int_M \sigma^{2-\theta_\varphi}\,d\omega \leq \xi_0^{1-\theta_\varphi}, \qquad H_\nu \geq -\frac{\xi_0}{(2-\theta_\varphi)\varphi(\xi_0)}.$$

(ii) *If $\theta_\varphi \in (1, 3/2)$ and $\omega[M] < \infty$, then we have*

$$\int_M \sigma^{2-\theta_\varphi}\,d\omega \leq \omega[M]^{\theta_\varphi - 1}, \qquad H_\nu \geq -\frac{\xi_0^{\theta_\varphi}\omega[M]^{\theta_\varphi - 1}}{(2-\theta_\varphi)\varphi(\xi_0)}.$$

Proof. It follows from (7.1) and Lemma 2.10 that

$$H_\nu = \int_M U(\sigma, 1)\, d\omega \geq -\int_M \int_0^\sigma \frac{r}{\varphi(r)}\, dr d\omega \geq -\int_M \int_0^\sigma \frac{\xi_0^{\theta_\varphi} r^{1-\theta_\varphi}}{\varphi(\xi_0)}\, dr d\omega$$

$$= -\frac{\xi_0^{\theta_\varphi}}{\varphi(\xi_0)} \int_M \frac{\sigma^{2-\theta_\varphi}}{2 - \theta_\varphi}\, d\omega.$$

(i) The assertion immediately follows from

$$\int_M \sigma^{2-\theta_\varphi}\, d\omega \leq \|\sigma\|_\infty^{1-\theta_\varphi} \int_M \sigma\, d\omega \leq \xi_0^{1-\theta_\varphi}.$$

(ii) The Hölder inequality yields that

$$\int_M \sigma^{2-\theta_\varphi}\, d\omega \leq \left(\int_M \sigma\, d\omega\right)^{2-\theta_\varphi} \omega[M]^{\theta_\varphi-1} = \omega[M]^{\theta_\varphi-1},$$

which shows the claim. $\square$

Theorem 7.6 *Let $\{(M, \omega, \varphi, \Psi_i)\}_{i\in\mathbb{N}}$ be a sequence of admissible spaces such that*

(a) $\omega[M] < \infty$ if $\theta_\varphi > 1$,

(b) $\nu_i = \sigma_i \omega := \exp_\varphi(-\Psi_i)\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ for all i,

(c) $\xi_i := \max\{1, \|\sigma_i\|_\infty\} < \infty$ for all i,

(d) $\mathrm{Hess}\, H_\varphi^i \geq K_i$ for some $K_i > 0$, where H_φ^i is the φ-relative entropy for $(M, \omega, \varphi, \Psi_i)$,

(e) $\lim_{i\to\infty} K_i \xi_i^{\delta_\varphi-1} = \infty$ if $\theta_\varphi \leq 1$, and $\lim_{i\to\infty} K_i \xi_i^{\delta_\varphi-\theta_\varphi} = \infty$ if $\theta_\varphi > 1$.

Then the concentration function $\alpha_i(r) := \alpha_{(M,\nu_i)}(r)$ satisfies $\lim_{i\to\infty} \alpha_i(r) = 0$ for all $r > 0$.

Proof. Fix $r > 0$ and put $\alpha_i := \alpha_i(r)$ and $H_{\nu_i} := H_\varphi^i(\nu_i)$ for brevity. It follows from Proposition 7.4 that

$$\alpha_i^{\delta_\varphi-2} \ln_\varphi(\alpha_i) \xi_i^{\theta_\varphi-\delta_\varphi} \leq -\left\{\left(\sqrt{\frac{K_i}{2}} r - \sqrt{-H_{\nu_i}}\right)^2 + H_{\nu_i}\right\} \left(\int_M \sigma_i^{2-\theta_\varphi}\, d\omega\right)^{-1}.$$

We have $\ln_\varphi(\alpha_i) \geq \ell_{2-\theta_\varphi}(\alpha_i)$ by (2.9), so that

$$\alpha_i^{\delta_\varphi-2} \ell_{2-\theta_\varphi}(\alpha_i) \leq -\xi_i^{\delta_\varphi-\theta_\varphi}\left(\frac{K_i}{2} r^2 - \sqrt{2K_i}\sqrt{-H_{\nu_i}} r\right)\left(\int_M \sigma_i^{2-\theta_\varphi}\, d\omega\right)^{-1}. \qquad (7.4)$$

Now, for $\theta_\varphi \leq 1$, Lemma 7.5(i) and Lemma 2.10 yield

$$\int_M \sigma_i^{2-\theta_\varphi}\, d\omega \leq \xi_i^{1-\theta_\varphi}, \qquad H_{\nu_i} \geq -\frac{\xi_i}{(2-\theta_\varphi)\varphi(\xi_i)} \geq -\frac{\xi_i^{1-\delta_\varphi}}{2-\theta_\varphi}$$

39

since $\xi_i \geq 1$. Hence the right hand side of (7.4) is bounded from above by (for large i)

$$-\xi_i^{\delta_\varphi - 1}\left(\frac{K_i}{2}r^2 - \sqrt{2K_i}\sqrt{\frac{\xi_i^{1-\delta_\varphi}}{2-\theta_\varphi}}r\right) = -\frac{K_i}{\xi_i^{1-\delta_\varphi}}\left(\frac{r^2}{2} - \sqrt{\frac{2}{2-\theta_\varphi}\frac{\xi_i^{1-\delta_\varphi}}{K_i}}r\right) \to -\infty$$

as i goes to infinity due to the condition (e). Therefore we obtain

$$\lim_{i\to\infty} \alpha_i^{-2}\ell_{2-\theta_\varphi}(\alpha_i) \leq \lim_{i\to\infty} \alpha_i^{\delta_\varphi - 2}\ell_{2-\theta_\varphi}(\alpha_i) = -\infty,$$

and hence $\lim_{i\to\infty} \alpha_i = 0$.

For $\theta_\varphi > 1$, we similarly deduce from Lemma 7.5(ii) that

$$\int_M \sigma_i^{2-\theta_\varphi}\, d\omega \leq \omega[M]^{\theta_\varphi - 1}, \qquad H_{\nu_i} \geq -\frac{\xi_i^{\theta_\varphi}\omega[M]^{\theta_\varphi-1}}{(2-\theta_\varphi)\varphi(\xi_i)} \geq -\frac{\xi_i^{\theta_\varphi-\delta_\varphi}\omega[M]^{\theta_\varphi-1}}{2-\theta_\varphi}.$$

Hence the right hand side of (7.4) is bounded from above by

$$-\frac{\xi_i^{\delta_\varphi-\theta_\varphi}}{\omega[M]^{\theta_\varphi-1}}\left(\frac{K_i}{2}r^2 - \sqrt{2K_i}\sqrt{\frac{\xi_i^{\theta_\varphi-\delta_\varphi}\omega[M]^{\theta_\varphi-1}}{2-\theta_\varphi}}r\right)$$

$$= -\frac{K_i\xi_i^{\delta_\varphi-\theta_\varphi}}{\omega[M]^{\theta_\varphi-1}}\left(\frac{r^2}{2} - \sqrt{\frac{2\omega[M]^{\theta_\varphi-1}}{2-\theta_\varphi}\frac{\xi_i^{\theta_\varphi-\delta_\varphi}}{K_i}}r\right) \to -\infty \quad (i \to \infty).$$

Thus we have $\lim_{i\to\infty} \alpha_i = 0$. $\qquad \square$

Remark 7.7 (1) For φ_m with $m < 1$, we have $\theta_{\varphi_m} = \delta_{\varphi_m}$ and hence the condition (e) is reduced to $\lim_{i\to\infty} K_i = \infty$, whereas the condition $\|\sigma_i\|_\infty < \infty$ was implicitly used in our discussion. See [OT, Section 6] for a more precise estimate associated with φ_m without assuming $\|\sigma_i\|_\infty < \infty$.

(2) We stress that only $\mathrm{Hess}\, H_\varphi^i \geq K_i$ for single φ is assumed in Theorem 7.6, rather than $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\, \Psi_i \geq K_i$. If $\mathrm{Hess}\, \Psi_i \geq K_i$ and $l_\varphi > -\infty$, then Lemma 4.2(i) gives a stronger estimate on the diameter of $M_\varphi^{\Psi_i}$ as

$$\mathrm{diam}\, M_\varphi^{\Psi_i} \leq 2\sqrt{\frac{2}{K_i}}\left\{\ln_\varphi(\|\sigma_i\|_\infty) - l_\varphi\right\} \leq 2\sqrt{\frac{2}{K_i}}\left\{\ln_\varphi(\xi_i) - l_\varphi\right\}.$$

Indeed, we observe from (2.9) and Lemma 2.10 that

$$\frac{\ln_\varphi(\xi_i)}{K_i} \leq \frac{\xi_i^{\theta_\varphi}\ell_{2-\theta_\varphi}(\xi_i)}{K_i\varphi(\xi_i)} \leq \frac{\xi_i^{\theta_\varphi-\delta_\varphi}\ell_{2-\theta_\varphi}(\xi_i)}{K_i} = \frac{\xi_i^{\theta_\varphi-\delta_\varphi}(\xi_i^{1-\theta_\varphi} - 1)}{K_i(1-\theta_\varphi)},$$

provided that $\theta_\varphi \neq 1$. If $\theta_\varphi < 1$, then the leading term (as $\xi_i \to \infty$) is

$$\frac{1}{1-\theta_\varphi}\frac{\xi_i^{1-\delta_\varphi}}{K_i} \to 0 \quad (i \to \infty)$$

under the condition (e) in Theorem 7.6. Similarly, for $\theta_\varphi > 1$ the leading term is

$$\frac{1}{\theta_\varphi - 1}\frac{\xi_i^{\theta_\varphi-\delta_\varphi}}{K_i} \to 0 \quad (i \to \infty).$$

Therefore, in both cases, $\lim_{i\to\infty} \mathrm{diam}\, M_\varphi^{\Psi_i} = 0$ holds and it is obviously stronger than $\lim_{i\to\infty} \alpha_{(M,\nu_i)}(r) = 0$.

7.3 $m(\varphi)$-normal concentration

In order to derive the m-normal concentration for some $m = m(\varphi)$ from the general estimate (7.3), we prove a computational lemma on e_m (see also [OT, Lemma 6.4]). Recall from Subsection 2.4 that $e_m(\tau) = \exp_{\varphi_m}(\tau) = [1 + (m-1)\tau]_+^{1/(m-1)}$.

Lemma 7.8 (i) *Given $0 < m \le m' < 1$ with $m + m' > 1$, set*

$$\beta = \beta(m, m') = 1 + \frac{1 - m'}{1 - m} \in (1, 2].$$

Then we have $\beta m' > 1$ and, for any $a, r > 0$,

$$e_m\left(-\left(ar - \frac{1}{\sqrt{m'}}\right)^2 + \frac{1}{m'}\right) \le e_m(\beta)e_m\left(-\left(1 - \frac{1}{\beta m'}\right)\frac{a^2 r^2}{m + m' - 1}\right).$$

(ii) *For any $m \in [1, 2)$ and $a, r > 0$, we have*

$$e_m\left((ar - 1)^2 - 1\right) \ge e_m\left(-\frac{2}{m}\right)e_m\left(\frac{a^2}{2}r^2\right).$$

Proof. (i) The assumptions $m' < 1$ and $m + m' > 1$ yields $(m + m')(1 - m') > (1 - m')$, and hence

$$\beta m' = \frac{m'\{2 - (m + m')\}}{1 - m} = \frac{(m' + m)(1 - m') + m' - m}{1 - m} > \frac{1 - m' + m' - m}{1 - m} = 1.$$

From the direct calculation

$$-\left(ar - \frac{1}{\sqrt{m'}}\right)^2 + \frac{1}{m'} = -a^2 r^2 + \frac{2ar}{\sqrt{m'}} \le -a^2 r^2 + \frac{a^2 r^2}{\beta m'} + \beta$$

and the monotonicity of e_m, we deduce that

$$e_m\left(-\left(ar - \frac{1}{\sqrt{m'}}\right)^2 + \frac{1}{m'}\right) \le e_m\left(-\left(1 - \frac{1}{\beta m'}\right)a^2 r^2 + \beta\right)$$

$$= \left[1 + (m-1)\left\{-\left(1 - \frac{1}{\beta m'}\right)a^2 r^2 + \beta\right\}\right]^{1/(m-1)}$$

$$= \{1 + (m-1)\beta\}^{1/(m-1)}\left\{1 - (m-1)\left(1 - \frac{1}{\beta m'}\right)\frac{a^2 r^2}{m + m' - 1}\right\}^{1/(m-1)}.$$

(ii) The assertion for $m = 1$ (with $e_1(\tau) = e^\tau$) is easily checked. For $m \in (1, 2)$, we deduce from

$$(ar - 1)^2 - 1 = a^2 r^2 - 2ar \ge a^2 r^2 - \frac{m}{2}\left\{a^2 r^2 + \left(\frac{2}{m}\right)^2\right\} = \left(1 - \frac{m}{2}\right)a^2 r^2 - \frac{2}{m}$$

41

that

$$e_m\big((ar-1)^2-1\big) \geq e_m\left(\left(1-\frac{m}{2}\right)a^2r^2 - \frac{2}{m}\right)$$

$$= \left[1 + (m-1)\left\{\left(1-\frac{m}{2}\right)a^2r^2 - \frac{2}{m}\right\}\right]^{1/(m-1)}$$

$$= \left\{1 - (m-1)\frac{2}{m}\right\}^{1/(m-1)} \left\{1 + \frac{m-1}{2}ma^2r^2\right\}^{1/(m-1)}$$

$$= e_m\left(-\frac{2}{m}\right)e_m\left(\frac{ma^2}{2}r^2\right) \geq e_m\left(-\frac{2}{m}\right)e_m\left(\frac{a^2}{2}r^2\right).$$

$\square$

Theorem 7.9 ($m(\varphi)$-normal concentration) *Assume that $(M, \omega, \varphi, \Psi)$ is admissible, $\nu \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$, $\mathrm{Hess}\, H_\varphi \geq K$ for some $K > 0$ and that $\|\sigma\|_\infty < \infty$. Fix arbitrary $\xi_0 \geq \max\{1, \|\sigma\|_\infty\}$.*

(i) *If $\theta_\varphi < 1$ and $\delta_\varphi > 0$, then we have for any $r > 0$*

$$\alpha(r)^{-1} \geq \left\{\frac{\delta_\varphi(1-\theta_\varphi)}{(1-\delta_\varphi)(2-\delta_\varphi)}\right\}^{1/(1-\delta_\varphi)} \cdot e_{2-\delta_\varphi}\left(\frac{K}{4}\xi_0^{\delta_\varphi-1}r^2\right).$$

(ii) *If $\theta_\varphi \in (1, 3/2)$, $\delta_\varphi > 3(\theta_\varphi - 1)$ and if $\omega[M] < \infty$, then we have for any $r > 0$*

$$\alpha(r) \leq \left(\frac{(\theta_\varphi-1)(3-3\theta_\varphi+\delta_\varphi)}{2\theta_\varphi-\delta_\varphi-1}\right)\right\}^{1/(1-2\theta_\varphi+\delta_\varphi)}$$

$$\times e_{2(1-\theta_\varphi)+\delta_\varphi}\left(-\frac{K}{2}\frac{\theta_\varphi-1}{(2-\theta_\varphi)(3\theta_\varphi-\delta_\varphi-2)}\xi_0^{\delta_\varphi-\theta_\varphi}\omega[M]^{1-\theta_\varphi}r^2\right).$$

(iii) *If $\theta_\varphi = 1$ and $\delta_\varphi > 1/2$, then we have for any $r > 0$*

$$\alpha(r)^{-1} \geq e_{3-2\delta_\varphi}\left(\frac{-2}{3-2\delta_\varphi}\right) \cdot e_{3-2\delta_\varphi}\left(\frac{K}{4}\xi_0^{\delta_\varphi-1}r^2\right).$$

Proof. We abbreviate $\alpha(r)$ as α in this proof, and assume $\alpha > 0$ without loss of generality. Let $A \subset M$ be a measurable set of $\nu[A] \geq 1/2$ and put $B := M \setminus B(A, r)$.
   (i) We first observe

$$\int_B \sigma^{2-\theta_\varphi}\,d\omega \leq \|\sigma\|_\infty^{1-\theta_\varphi} \int_B \sigma\,d\omega \leq \alpha\xi_0^{1-\theta_\varphi}.$$

Then (7.3) yields

$$\alpha^{\delta_\varphi-1}\ln_\varphi(\alpha) \leq \xi_0^{\delta_\varphi-1}\left\{-\left(\sqrt{\frac{K}{2}}r - \sqrt{-H_\nu}\right)^2 - H_\nu\right\}, \tag{7.5}$$

42

where $H_\nu := H_\varphi(\nu)$. On the one hand, it follows from (2.9) that

$$\alpha^{\delta_\varphi - 1} \ln_\varphi(\alpha) \geq \alpha^{\delta_\varphi - 1} \ell_{2-\theta_\varphi}(\alpha) = \frac{\alpha^{\delta_\varphi - \theta_\varphi} - \alpha^{\delta_\varphi - 1}}{1 - \theta_\varphi}.$$

Since $\alpha^{\delta_\varphi - \theta_\varphi} \geq 1 \geq (1 - \theta_\varphi)/(1 - \delta_\varphi)$, we obtain

$$\alpha^{\delta_\varphi - 1} \ln_\varphi(\alpha) \geq \frac{1 - (C\alpha^{-1})^{1-\delta_\varphi}}{1 - \delta_\varphi} = -\ell_{2-\delta_\varphi}(C\alpha^{-1}), \quad C := \left(\frac{1 - \delta_\varphi}{1 - \theta_\varphi} \right)^{1/(1-\delta_\varphi)} \geq 1.$$

On the other hand, Lemmas 7.5(i) and 2.10 give

$$-\xi_0^{\delta_\varphi - 1} H_\nu \leq \frac{\xi_0^{\delta_\varphi}}{(2 - \theta_\varphi)\varphi(\xi_0)} \leq \frac{1}{2 - \theta_\varphi} \leq 1. \tag{7.6}$$

Hence we have

$$\ell_{2-\delta_\varphi}(C\alpha^{-1}) \geq \left(\sqrt{\frac{K\xi_0^{\delta_\varphi - 1}}{2}} r - 1 \right)^2 - 1.$$

We apply Lemma 7.8(ii) and obtain

$$\alpha^{-1} \geq C^{-1} e_{2-\delta_\varphi} \left(\left(\sqrt{\frac{K}{2}} \xi_0^{(\delta_\varphi - 1)/2} r - 1 \right)^2 - 1 \right)$$

$$\geq C^{-1} e_{2-\delta_\varphi} \left(\frac{2}{\delta_\varphi - 2} \right) e_{2-\delta_\varphi} \left(\frac{K}{4} \xi_0^{\delta_\varphi - 1} r^2 \right)$$

$$= \left\{ \frac{\delta_\varphi(1 - \theta_\varphi)}{(1 - \delta_\varphi)(2 - \delta_\varphi)} \right\}^{1/(1-\delta_\varphi)} \cdot e_{2-\delta_\varphi} \left(\frac{K}{4} \xi_0^{\delta_\varphi - 1} r^2 \right).$$

(ii) We deduce from the Hölder inequality that

$$\int_B \sigma^{2-\theta_\varphi} \, d\omega \leq \left(\int_B \sigma \, d\omega \right)^{2-\theta_\varphi} \omega[B]^{\theta_\varphi - 1} \leq \alpha^{2-\theta_\varphi} \omega[M]^{\theta_\varphi - 1}.$$

Then (7.3) gives

$$\alpha^{\delta_\varphi - \theta_\varphi} \ln_\varphi(\alpha) \leq \xi_0^{\delta_\varphi - \theta_\varphi} \omega[M]^{1-\theta_\varphi} \left\{ - \left(\sqrt{\frac{K}{2}} r - \sqrt{-H_\nu} \right)^2 - H_\nu \right\}.$$

Set $m := 2(1 - \theta_\varphi) + \delta_\varphi$ and $m' := 2 - \theta_\varphi$, and observe $0 < m \leq m' < 1$ as well as $m + m' > 1$. Similarly to (i), (2.9) yields

$$\alpha^{\delta_\varphi - \theta_\varphi} \ln_\varphi(\alpha) \geq \alpha^{\delta_\varphi - \theta_\varphi} \ell_{2-\theta_\varphi}(\alpha) = \frac{\alpha^{\delta_\varphi - 2\theta_\varphi + 1} - \alpha^{\delta_\varphi - \theta_\varphi}}{1 - \theta_\varphi} = \frac{\alpha^{m-1} - \alpha^{m-m'}}{m' - 1}.$$

As $\alpha^{m-m'} \geq 1 \geq (1 - m')/(1 - m)$, we find

$$\alpha^{\delta_\varphi - \theta_\varphi} \ln_\varphi(\alpha) \geq \frac{(c\alpha)^{m-1} - 1}{m - 1} = \ell_m(c\alpha), \quad c := \left(\frac{1 - m}{1 - m'} \right)^{1/(m-1)} \leq 1.$$

43

Lemmas 7.5(ii) and 2.10 imply

$$-\xi_0^{\delta_\varphi-\theta_\varphi}\omega[M]^{1-\theta_\varphi}H_\nu \le \frac{\xi_0^{\delta_\varphi}}{(2-\theta_\varphi)\varphi(\xi_0)} \le \frac{1}{2-\theta_\varphi} = \frac{1}{m'},$$

and hence

$$\ell_m(c\alpha) \le -\left(\sqrt{\frac{K\xi_0^{\delta_\varphi-\theta_\varphi}\omega[M]^{1-\theta_\varphi}}{2}}\,r - \frac{1}{\sqrt{m'}}\right)^2 + \frac{1}{m'}.$$

Then we apply Lemma 7.8(i) to have, with $\beta = (2-m-m')/(1-m)$,

$$\alpha \le c^{-1}e_m(\beta)e_m\left(-\left(1-\frac{1}{\beta m'}\right)\frac{K\xi_0^{\delta_\varphi-\theta_\varphi}\omega[M]^{1-\theta_\varphi}}{2(m+m'-1)}r^2\right).$$

(iii) It immediately follows from (7.5) and (7.6) that

$$\alpha^{\delta_\varphi-1}\ln_\varphi(\alpha) \le \xi_0^{\delta_\varphi-1}\left\{-\left(\sqrt{\frac{K}{2}}\,r - \sqrt{-H_\nu}\right)^2 - H_\nu\right\} \le -\left(\sqrt{\frac{K\xi_0^{\delta_\varphi-1}}{2}}\,r - 1\right)^2 + 1.$$

Note that (2.9) provides $\alpha^{\delta_\varphi-1}\ln_\varphi(\alpha) \ge \alpha^{\delta_\varphi-1}\ln(\alpha)$. If $\delta_\varphi = 1$, then it holds $\alpha^{\delta_\varphi-1}\ln(\alpha) = -\ln(\alpha^{-1})$. Otherwise, the numerical estimate

$$\ln(t) \ge \frac{t^s - t^{-s}}{2s} \qquad \text{for } t \in (0,1],\ s > 0,$$

shows $\alpha^{\delta_\varphi-1}\ln(\alpha) \ge -\ell_{3-2\delta_\varphi}(\alpha^{-1})$ (let $s = 1-\delta_\varphi$ and $t = \alpha$). Therefore we have, thanks to Lemma 7.8(ii) with $m = 3 - 2\delta_\varphi < 2$,

$$\alpha^{-1} \ge e_{3-2\delta_\varphi}\left(\left(\sqrt{\frac{K\xi_0^{\delta_\varphi-1}}{2}}\,r - 1\right)^2 - 1\right) \ge e_{3-2\delta_\varphi}\left(\frac{-2}{3-2\delta_\varphi}\right) \cdot e_{3-2\delta_\varphi}\left(\frac{K}{4}\xi_0^{\delta_\varphi-1}r^2\right).$$

$\square$

Note that $3(\theta_\varphi - 1) < \theta_\varphi$ in (ii) by $\theta_\varphi < 3/2$, so that the condition $\delta_\varphi > 3(\theta_\varphi - 1)$ is not vacuous.

Remark 7.10 Letting $\delta_\varphi = \theta_\varphi$ and then $\theta_\varphi \to 1$, all of the estimates (i)–(iii) in Theorem 7.9 tend to the normal concentration $\alpha(r) \le e^2 \exp(-Kr^2/4)$.

8 Gradient flow of H_φ: Compact case

In this and the next sections, we show that the gradient flow of the φ-relative entropy produces weak solutions to the nonlinear evolution equation

$$\frac{\partial \rho}{\partial t} = \mathrm{div}_\omega\left(\frac{\rho\nabla\rho}{\varphi(\rho)} + \rho\nabla\Psi\right)$$

44

on the weighted Riemannian manifold (M, ω). See the beginning of Subsection 8.3 for more explanation and background. This kind of interpretation of evolution equations has turned out extremely useful after the pioneering work due to Jordan et al. [JKO]. There are several ways of interpreting this coincidence. In this section, we adapt the rather 'metric geometric' approach developed in [Oh1] inspired by [PP] and [Ly] (see also [Pe]). This formulation of gradient flows requires a strong structure theorem (Theorem 8.1) of the Wasserstein space, which is known only for compact spaces. The noncompact situation will be treated in the next section in a different strategy along [AGS1] and [Er].

Before beginning the review of the structure of Wasserstein spaces, let us recall basic notions of calculus on our weighted Riemannian manifold (M, ω) with $\omega = e^{-f} \operatorname{vol}_g$. For a differentiable vector field V on M, the *weighted divergence* is defined as

$$\operatorname{div}_\omega V := \operatorname{div} V - \langle V, \nabla f \rangle,$$

where $\operatorname{div} V$ denotes the usual divergence of V for the unweighted space $(M, \operatorname{vol}_g)$. Note that $\operatorname{div}_\omega V = e^f \operatorname{div}(e^{-f} V)$ and, for any $w \in C_c^1(M)$, the integration by parts holds:

$$\int_M \langle \nabla w, V \rangle \, d\omega = \int_M \langle \nabla w, e^{-f} V \rangle \, d\operatorname{vol}_g = -\int_M w \operatorname{div}(e^{-f} V) \, d\operatorname{vol}_g = -\int_M w \operatorname{div}_\omega V \, d\omega.$$

Through this formula, the weighted divergence is defined in the weak sense also for measurable vector fields. For $\rho \in H_{\mathrm{loc}}^1(M)$, the *weighted Laplacian* is defined in the weak form by

$$\Delta^\omega \rho := \operatorname{div}_\omega(\nabla \rho) = \Delta \rho - \langle \nabla \rho, \nabla f \rangle.$$

8.1 Geometric structure of $(\mathcal{P}(M), W_2)$

Let M be compact throughout the section, so that $\mathcal{P}(M) = \mathcal{P}^2(M)$. It is known that $(\mathcal{P}(M), W_2)$ is an Alexandrov space of nonnegative curvature if and only if (M, g) has the nonnegative sectional curvature ([St2, Proposition 2.10], [LV2, Theorem A.8]). Alexandrov spaces are metric spaces whose sectional curvature is bounded from below by a constant in the sense of the triangle comparison property, and such spaces are known to possess nice infinitesimal structures (we refer to [BBI] for the basic theory). We remark that it is in most cases impossible to bound the curvature of $\mathcal{P}(M)$ from above (cf. [AGS1, Example 7.3.3]). In the case where (M, g) is not nonnegatively curved, although $(\mathcal{P}(M), W_2)$ does not admit any lower curvature bound in the sense of Alexandrov ([St2, Proposition 2.10]), we can consider the 'angle' between geodesics (see also [Oh1, Theorem 3.6]).

Theorem 8.1 ([Gi2, Theorem 3.4, Remark 3.5]) *For any $\mu \in \mathcal{P}(M)$ and unit speed geodesics $\alpha, \beta : [0, \delta) \longrightarrow \mathcal{P}(M)$ with $\alpha(0) = \beta(0) = \mu$, the joint limit*

$$\lim_{s,t \downarrow 0} \frac{s^2 + t^2 - W_2(\alpha(s), \beta(t))^2}{2st} \in [-1, 1]$$

exists.

Theorem 8.1 in particular guarantees that the scaling limit

$$\lim_{\varepsilon \downarrow 0} \frac{(s\varepsilon)^2 + (t\varepsilon)^2 - W_2(\alpha(s\varepsilon), \beta(t\varepsilon))^2}{2st\varepsilon^2}$$

exists, and is independent of the choices of the parameters $s, t > 0$. This means that an angle between α and β makes sense, so that $(\mathcal{P}(M), W_2)$ looks like a Riemannian space (rather than a Finsler space). This observation makes it possible to investigate the infinitesimal structure of $(\mathcal{P}(M), W_2)$ in the manner of the theory of Alexandrov spaces. For $\mu \in \mathcal{P}(M)$, denote by $\Sigma'_\mu[\mathcal{P}(M)]$ the set of all nontrivial unit speed minimal geodesics emanating from μ. Given $\alpha, \beta \in \Sigma'_\mu[\mathcal{P}(M)]$, Theorem 8.1 verifies that the *angle*

$$\angle_\mu(\alpha, \beta) := \arccos\left(\lim_{s,t \downarrow 0} \frac{s^2 + t^2 - W_2(\alpha(s), \beta(t))^2}{2st}\right) \in [0, \pi]$$

is well-defined. We define the *space of directions* $(\Sigma_\mu[\mathcal{P}(M)], \angle_\mu)$ as the completion of $(\Sigma'_\mu[\mathcal{P}(M)]/\sim, \angle_\mu)$, where $\alpha \sim \beta$ holds if $\angle_\mu(\alpha, \beta) = 0$. The angle $\angle_\mu$ provides a natural distance structure of $\Sigma_\mu[\mathcal{P}(M)]$. The *tangent cone* $(C_\mu[\mathcal{P}(M)], \sigma_\mu)$ is defined as the Euclidean cone over $(\Sigma_\mu[\mathcal{P}(M)], \angle_\mu)$, i.e.,

$$C_\mu[\mathcal{P}(M)] := \big(\Sigma_\mu[\mathcal{P}(M)] \times [0, \infty)\big) / \big(\Sigma_\mu[\mathcal{P}(M)] \times \{0\}\big),$$

$$\sigma_\mu\big((\alpha, s), (\beta, t)\big) := \sqrt{s^2 + t^2 - 2st \cos \angle_\mu(\alpha, \beta)}.$$

By means of this infinitesimal structure, we introduce a class of 'differentiable curves'.

Definition 8.2 (Right differentiability) We say that a curve $\xi : [0, l] \longrightarrow \mathcal{P}(M)$ is *right differentiable* at $t \in [0, l]$ if there is $\mathbf{v} \in C_{\xi(t)}[\mathcal{P}(M)]$ such that, for any sequences $\{\varepsilon_i\}_{i \in \mathbb{N}}$ of positive numbers tending to zero and $\{\alpha_i\}_{i \in \mathbb{N}}$ of unit speed minimal geodesics from $\xi(t)$ to $\xi(t+\varepsilon_i)$, the sequence $\{(\alpha_i, W_2(\xi(t), \xi(t+\varepsilon_i))/\varepsilon_i)\}_{i \in \mathbb{N}} \subset C_{\xi(t)}[\mathcal{P}(M)]$ converges to $\mathbf{v}$. Such $\mathbf{v}$ is clearly unique if it exists, and then we write $\dot{\xi}(t) = \mathbf{v}$.

8.2 Gradient flows in $(\mathcal{P}(M), W_2)$

Consider a lower semi-continuous function $H : \mathcal{P}(M) \longrightarrow (-\infty, +\infty]$ which is K-convex in the weak sense for some $K \in \mathbb{R}$. We in addition suppose that H is not identically $+\infty$, and define $\mathcal{P}^*_H(M) := \{\mu \in \mathcal{P}(M) \,|\, H(\mu) < \infty\}$.

Given $\mu \in \mathcal{P}^*_H(M)$ and $\alpha \in \Sigma_\mu[\mathcal{P}(M)]$, we set

$$D_\mu H(\alpha) := \liminf_{\Sigma'_\mu[\mathcal{P}(M)] \ni \beta \to \alpha} \lim_{t \downarrow 0} \frac{H(\beta(t)) - H(\mu)}{t},$$

where the convergence $\beta \to \alpha$ is with respect to $\angle_\mu$. Define the *absolute gradient* (also called the *local slope*) of H at $\mu \in \mathcal{P}^*_H(M)$ by

$$|\nabla_- H|(\mu) := \max\left\{0, \limsup_{\tilde{\mu} \to \mu} \frac{H(\mu) - H(\tilde{\mu})}{W_2(\mu, \tilde{\mu})}\right\},$$

where $\tilde{\mu} \to \mu$ is with respect to W_2. Note that $-D_\mu H(\alpha) \leq |\nabla_- H|(\mu)$ for any $\alpha \in \Sigma_\mu[\mathcal{P}(M)]$. The K-convexity of H guarantees the unique existence of the direction along which H decreases the most.

Lemma 8.3 ([Oh1, Lemma 4.2]) *For each $\mu \in \mathcal{P}_H^*(M)$ with $0 < |\nabla\!\!\!\!\!\!-\, H|(\mu) < \infty$, there exists a unique direction $\alpha \in \Sigma_\mu[\mathcal{P}(M)]$ satisfying $D_\mu H(\alpha) = -|\nabla\!\!\!\!\!\!-\, H|(\mu)$.*

Using α in the above lemma, we define the *negative gradient vector* of H at μ by

$$\nabla\!\!\!\!\!\!-\, H(\mu) := \big(\alpha, |\nabla\!\!\!\!\!\!-\, H|(\mu)\big) \in C_\mu[\mathcal{P}(M)].$$

If $|\nabla\!\!\!\!\!\!-\, H|(\mu) = 0$, then we simply define $\nabla\!\!\!\!\!\!-\, H(\mu)$ as the origin of $C_\mu[\mathcal{P}(M)]$. A trajectory of the gradient flow of H (which will be called a gradient curve) should be understood as a curve ξ solving $\dot{\xi}(t) = \nabla\!\!\!\!\!\!-\, H(\xi(t))$. Precisely, we adopt the following definition.

Definition 8.4 (Gradient curves) A continuous curve $\xi : [0, l) \longrightarrow \mathcal{P}_H^*(M)$ which is locally Lipschitz on $(0, l)$ is called a *gradient curve* of H if $|\nabla\!\!\!\!\!\!-\, H|(\xi(t)) < \infty$ for all $t \in (0, l)$ and if it is right differentiable with $\dot{\xi}(t) = \nabla\!\!\!\!\!\!-\, H(\xi(t))$ at all $t \in (0, l)$. We say that a gradient curve ξ is *complete* if it is defined on entire $[0, \infty)$.

By virtue of the K-convexity of H as well as the compactness of M, there starts a unique gradient curve from an arbitrary initial point $\mu \in \mathcal{P}_H^*(M)$ enjoying the K-*contraction property* as follows.

Theorem 8.5 ([Oh1, Theorem 5.11, Corollary 6.3], [GO, Theorem 4.2]) *Let M be compact and $H : \mathcal{P}(M) \longrightarrow (-\infty, +\infty]$ be a K-convex function for some $K \in \mathbb{R}$.*

(i) *From any $\mu \in \mathcal{P}_H^*(M)$, there exists a unique complete gradient curve $\xi : [0, \infty) \longrightarrow \mathcal{P}_H^*(M)$ of H with $\xi(0) = \mu$.*

(ii) *(K-contraction property) Given any two gradient curves $\xi, \zeta : [0, \infty) \longrightarrow \mathcal{P}_H^*(M)$ of H, we have*

$$W_2\big(\xi(t), \zeta(t)\big) \leq e^{-Kt} W_2\big(\xi(0), \zeta(0)\big) \tag{8.1}$$

for all $t \in [0, \infty)$.

The uniqueness in (i) is indeed a consequence of the K-contraction property (8.1). Thus the *gradient flow* $G : [0, \infty) \times \mathcal{P}_H^*(M) \longrightarrow \mathcal{P}_H^*(M)$ of H, given as $G(t, \mu) = \xi(t)$ for ξ in Theorem 8.5(i), is uniquely determined and continuously extended to the closure $G : [0, \infty) \times \overline{\mathcal{P}_H^*(M)} \longrightarrow \overline{\mathcal{P}_H^*(M)}$.

8.3 H_φ and the φ-heat equation

It is an established fact that the gradient flow of the relative entropy (or the *free energy*) with respect to ω,

$$\mathrm{Ent}_\omega(\rho\omega) = \int_M \rho \ln \rho \, d\omega = \int_M (\rho e^{-f}) \ln(\rho e^{-f}) \, d\mathrm{vol}_g + \int_M f \, d\mu,$$

produces solutions to the associated *heat equation* (or the *Fokker–Planck equation*)

$$\frac{\partial \rho}{\partial t} = \Delta^\omega \rho = e^f \big\{ \Delta(\rho e^{-f}) + \mathrm{div}\big((\rho e^{-f})\nabla f\big) \big\}.$$

See [JKO, Theorem 5.1], [Vi1, Subsection 8.4.2] for the Euclidean case, [Oh1, Theorem 6.6], [GO, Theorem 4.6], [Vi2, Corollary 23.23] for the Riemannian case, [OS1, Section 7] for the Finsler case, and [FSS], [Ju], [GKO], [Ma], [AGS2] for further related work on various kinds of spaces.

We shall see that a similar argumentation gives weak solutions to the equation

$$\frac{\partial \rho}{\partial t} = \operatorname{div}_\omega \left(\frac{\rho \nabla \rho}{\varphi(\rho)} + \rho \nabla \Psi \right) \tag{8.2}$$

as the gradient flow of the φ-relative entropy H_φ. We will call (8.2) the φ-heat equation. In the special case of $\varphi_m(s) = s^{2-m}$, (8.2) is called the *fast diffusion equation* (for $m < 1$) or the *porous medium equation* (for $m > 1$). Then this identification was demonstrated by Otto [Ot] on $(\mathbb{R}^n, \mathcal{L}^n)$, and by [Vi2, Theorem 23.19] as well as [OT] on weighted Riemannian manifolds (by the different means). We can follow the strategy of [OT] for general φ, up to some technical difficulties.

We first observe $|\nabla_- H_\varphi|(\mu) = \sqrt{I_\varphi(\mu)}$ as Proposition 6.1 suggests.

Proposition 8.6 *Let $(M, \omega, \varphi, \Psi)$ be a compact admissible space such that $\operatorname{Ric}_{N_\varphi} \geq 0$ and $\operatorname{Hess} \Psi \geq K$ for some $K \in \mathbb{R}$. Take $\mu = \rho \omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ with $\mu[M_\varphi^\Psi] = 1$, $H_\varphi(\mu) < \infty$, $\rho h'_\varphi(\rho) - h_\varphi(\rho) \in H^1(M)$ and with $|\nabla \rho / \varphi(\rho)| \in L^2(M, \mu)$. Then we have $|\nabla_- H_\varphi|(\mu) = \sqrt{I_\varphi(\mu)}$, and the negative gradient vector $\nabla_- H_\varphi(\mu)$ is given by $-\nabla \rho / \varphi(\rho) - \nabla \Psi$.*

Proof. Given any $\mu_1 \in \mathcal{P}(M)$ with $H_\varphi(\mu_1) < \infty$, let $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}(M)$ be a minimal geodesic from $\mu_0 = \mu$ to μ_1 along which H_φ is K-convex (Theorem 5.7). Letting $\mu_t = (\mathcal{T}_t)_\sharp \mu$ with $\mathcal{T}_t(x) = \exp_x(t \nabla \phi(x))$, we deduce from the K-convexity of H_φ that

$$\lim_{t \downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} \leq H_\varphi(\mu_1) - H_\varphi(\mu) - \frac{K}{2} W_2(\mu, \mu_1)^2.$$

Combining this with Proposition 6.1, we have

$$\frac{H_\varphi(\mu) - H_\varphi(\mu_1)}{W_2(\mu, \mu_1)} \leq -\frac{1}{W_2(\mu, \mu_1)} \int_M \left\langle \frac{\nabla \rho}{\varphi(\rho)} + \nabla \Psi, \nabla \phi \right\rangle d\mu - \frac{K}{2} W_2(\mu, \mu_1)$$

$$\leq \sqrt{I_\varphi(\mu)} - \frac{K}{2} W_2(\mu, \mu_1).$$

Thus we obtain $|\nabla_- H_\varphi|(\mu) \leq \sqrt{I_\varphi(\mu)}$, and equality follows also from Proposition 6.1 by choosing $\{\phi_i\}_{i \in \mathbb{N}} \subset C^\infty(M)$ which approximates $-\ln_\varphi(\rho) + \ln_\varphi(\sigma)$ in $H^1(M, \mu)$. Then, moreover, $\nabla_- H_\varphi(\mu)$ is achieved by $-\nabla \rho / \varphi(\rho) - \nabla \Psi$ (to be precise, $((\mu_t)_{t \in [0,1]}, W_2(\mu, \mu_1))$ associated with ϕ_i converges to $\nabla_- H_\varphi(\mu)$ in $C_\mu[\mathcal{P}(M)]$). $\qquad \square$

Now we are ready to show the main theorem of the section.

Theorem 8.7 (Gradient flow of H_φ) *Let $(M, \omega, \varphi, \Psi)$ be a compact admissible space such that $\operatorname{Ric}_{N_\varphi} \geq 0$ and $\operatorname{Hess} \Psi \geq K$ on M_φ^Ψ for some $K \in \mathbb{R}$. We in addition assume that $\theta_\varphi \in (0, (n+1)/n)$, $\lim_{s \to \infty} s^{\theta_\varphi} / \varphi(s) < \infty$ and that Ψ is Lipschitz. If a curve $(\mu_t)_{t \in [0,\infty)} \subset \mathcal{P}_{\mathrm{ac}}(M, \omega)$ with $\mu_t[M_\varphi^\Psi] \equiv 1$ is a gradient curve of H_φ, then its density*

function ρ_t is a weak solution to the φ-heat equation (8.2). To be precise, ρ_t is weakly differentiable as well as $|\nabla\rho_t/\varphi(\rho_t)| \in L^2(M, \mu_t)$ a.e. t, and we have

$$\int_M w_{t_1} \, d\mu_{t_1} - \int_M w_{t_0} \, d\mu_{t_0} = \int_{t_0}^{t_1} \int_M \left\{ \frac{\partial w_t}{\partial t} - \left\langle \frac{\nabla \rho_t}{\varphi(\rho_t)} + \nabla\Psi, \nabla w_t \right\rangle \right\} d\mu_t \, dt \qquad (8.3)$$

for all $0 \le t_0 < t_1 < \infty$ and $w \in C^\infty(\mathbb{R} \times M)$, where $\mu_t = \rho_t \omega$ and $w_t = w(t, \cdot)$.

Proof. First of all, the weak differentiability of ρ_t and $|\nabla\rho_t/\varphi(\rho_t)| \in L^2(M, \mu_t)$ follow from (I) $\Rightarrow$ (II) of Proposition 9.6 below. Fix $t \in (0, \infty)$ and, given small $\delta > 0$, choose $\mu^\delta \in \mathcal{P}(M)$ as a minimizer of the function

$$\mu \longmapsto H_\varphi(\mu) + \frac{W_2(\mu, \mu_t)^2}{2\delta}. \qquad (8.4)$$

We postpone the proof of the following technical claim until the end of the section. The condition $\theta_\varphi < (n+1)/n$ will come into play in (i), while $\theta_\varphi > 0$ and $\lim_{s\to\infty} s^{\theta_\varphi}/\varphi(s) < \infty$ will be used in (iii).

Claim 8.8 (i) *Such a minimizer μ^δ of (8.4) indeed exists and is absolutely continuous with respect to ω.*

(ii) *We have*

$$\lim_{\delta\downarrow 0} \frac{W_2(\mu^\delta, \mu_t)^2}{\delta} = 0, \qquad \lim_{\delta\downarrow 0} H_\varphi(\mu^\delta) = H_\varphi(\mu_t).$$

In particular, μ^δ converges to μ_t weakly.

(iii) *Moreover, by putting $\mu^\delta = \rho^\delta \omega$, $h_\varphi(\rho^\delta) - h'_\varphi(\rho^\delta)\rho^\delta$ converges to $h_\varphi(\rho_t) - h'_\varphi(\rho_t)\rho_t$ in $L^1(M, \omega)$ as $\delta \downarrow 0$.*

Take a semi-convex function $\phi : M \longrightarrow \mathbb{R}$ such that $\mathcal{T}(x) := \exp_x(\nabla\phi(x))$ gives the optimal transport from μ^δ to μ_t (recall Theorem 2.6). We also consider the transport $\mu_\varepsilon^\delta := (\mathcal{F}_\varepsilon)_\sharp \mu^\delta$ in another direction for small $\varepsilon > 0$, where $\mathcal{F}_\varepsilon(x) := \exp_x(\varepsilon\nabla w_t(x))$. It immediately follows from the choice of μ^δ that

$$H_\varphi(\mu_\varepsilon^\delta) + \frac{W_2(\mu_\varepsilon^\delta, \mu_t)^2}{2\delta} \ge H_\varphi(\mu^\delta) + \frac{W_2(\mu^\delta, \mu_t)^2}{2\delta}. \qquad (8.5)$$

We first estimate the difference of the Wasserstein distances. Observe that, as $(\mathcal{F}_\varepsilon \times \mathcal{T})_\sharp \mu^\delta$ is a (not necessarily optimal) coupling of μ_ε^δ and μ_t,

$$\limsup_{\varepsilon\downarrow 0} \frac{W_2(\mu_\varepsilon^\delta, \mu_t)^2 - W_2(\mu^\delta, \mu_t)^2}{\varepsilon}$$

$$\le \limsup_{\varepsilon\downarrow 0} \frac{1}{\varepsilon} \int_M \left\{ d_g\big(\mathcal{F}_\varepsilon(x), \mathcal{T}(x)\big)^2 - d_g\big(x, \mathcal{T}(x)\big)^2 \right\} d\mu^\delta(x) = -\int_M 2\langle \nabla w_t, \nabla\phi \rangle \, d\mu^\delta.$$

We used the first variation formula for the Riemannian distance function d_g in the last line (cf., e.g., [Chv, Theorem II.4.1]). Thanks to the compactness of M, there is a constant $C > 0$ (depending only on (M, g) and w) such that

$$w_t\big(\mathcal{T}(x)\big) \le w_t(x) + \langle \nabla w_t(x), \nabla\phi(x) \rangle + C d_g\big(x, \mathcal{T}(x)\big)^2$$

49

for a.e. $x \in M$. Thus we obtain, by virtue of Claim 8.8(ii),

$$\liminf_{\delta\downarrow 0} \frac{1}{2\delta} \limsup_{\varepsilon\downarrow 0} \frac{W_2(\mu_\varepsilon^\delta, \mu_t)^2 - W_2(\mu^\delta, \mu_t)^2}{\varepsilon} \leq -\limsup_{\delta\downarrow 0} \frac{1}{\delta} \int_M \langle \nabla w_t, \nabla \phi \rangle \, d\mu^\delta$$

$$\leq \liminf_{\delta\downarrow 0} \frac{1}{\delta} \left[\int_M \{w_t - w_t(\mathcal{T})\} \, d\mu^\delta + CW_2(\mu^\delta, \mu_t)^2 \right] = \liminf_{\delta\downarrow 0} \frac{1}{\delta} \left\{ \int_M w_t \, d\mu^\delta - \int_M w_t \, d\mu_t \right\}.$$

Next we calculate the difference of the entropies in (8.5). We put $\mu^\delta = \rho^\delta \omega$, $\mu_\varepsilon^\delta = \rho_\varepsilon^\delta \omega$ and $\mathbf{J}_\varepsilon^\omega := e^{f - f(\mathcal{F}_\varepsilon)} \det(D\mathcal{F}_\varepsilon)$. Then we obtain from Proposition 6.1 that, as $w_t \in C^\infty(M)$,

$$\lim_{\varepsilon\downarrow 0} \frac{H_\varphi(\mu^\delta) - H_\varphi(\mu_\varepsilon^\delta)}{\varepsilon} = \int_M \left[\{h_\varphi'(\rho^\delta)\rho^\delta - h_\varphi(\rho^\delta)\}\Delta^\omega w_t + \langle \rho^\delta \nabla[\ln_\varphi(\sigma)], \nabla w_t \rangle \right] d\omega$$

(we need the conditions $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ only here for applying Proposition 6.1). Hence we deduce that, together with Claim 8.8(ii), (iii),

$$\lim_{\delta\downarrow 0} \lim_{\varepsilon\downarrow 0} \frac{H_\varphi(\mu^\delta) - H_\varphi(\mu_\varepsilon^\delta)}{\varepsilon} = \int_M \left[\{h_\varphi'(\rho_t)\rho_t - h_\varphi(\rho_t)\}\Delta^\omega w_t - \langle \rho_t \nabla\Psi, \nabla w_t \rangle \right] d\omega$$

$$= -\int_M \langle \nabla[h_\varphi'(\rho_t)\rho_t - h_\varphi(\rho_t)] + \rho_t \nabla\Psi, \nabla w_t \rangle \, d\omega = -\int_M \left\langle \frac{\nabla\rho_t}{\varphi(\rho_t)} + \nabla\Psi, \nabla w_t \right\rangle d\mu_t.$$

These together imply

$$\liminf_{\delta\downarrow 0} \frac{1}{\delta} \left\{ \int_M w_t \, d\mu^\delta - \int_M w_t \, d\mu_t \right\} \geq -\int_M \left\langle \frac{\nabla\rho_t}{\varphi(\rho_t)} + \nabla\Psi, \nabla w_t \right\rangle d\mu_t.$$

Moreover, equality holds since we can change w into $-w$. Recall from [GO, (5)] (see also [Oh1, Lemma 6.4]) that

$$\lim_{\delta\downarrow 0} \frac{1}{\delta} \left\{ \int_M \eta \, d\mu_{t+\delta} - \int_M \eta \, d\mu^\delta \right\} = 0$$

holds for all $\eta \in C^\infty(M)$. Therefore we conclude

$$\lim_{\delta\downarrow 0} \frac{1}{\delta} \left\{ \int_M w_{t+\delta} \, d\mu_{t+\delta} - \int_M w_t \, d\mu_t \right\}$$

$$= \lim_{\delta\downarrow 0} \frac{1}{\delta} \left\{ \int_M (w_{t+\delta} - w_t) \, d\mu_{t+\delta} + \int_M w_t \, d\mu_{t+\delta} - \int_M w_t \, d\mu_t \right\}$$

$$= \int_M \left\{ \frac{\partial w_t}{\partial t} - \left\langle \frac{\nabla\rho_t}{\varphi(\rho_t)} + \nabla\Psi, \nabla w_t \right\rangle \right\} d\mu_t.$$

This shows (8.3) by integration in t. $\qquad\square$

Remark 8.9 In Theorem 8.7, assuming that μ_t is absolutely continuous is in fact redundant. If $L_\varphi = \infty$, then $H_\varphi(\mu_t) < \infty$ guarantees $\mu_t \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ by definition. As for $L_\varphi < \infty$, if μ_t with $t > 0$ has a nontrivial singular part μ^s, then we can modify μ_t as in

the proof of Claim 8.8(i) below (with $\mu^\delta = \mu_t$ and $\pi = \mathrm{diag}_\sharp \mu_t$ where $\mathrm{diag}(x) := (x, x)$) and obtain $\hat{\mu}_r \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ for small $r > 0$ such that

$$W_2(\hat{\mu}_r, \mu_t)^2 \le \mu^s[M]r^2, \qquad \lim_{r\downarrow 0} \frac{H_\varphi(\hat{\mu}_r) - H_\varphi(\mu_t)}{r} = -\infty.$$

This yields $|\nabla_- H_\varphi|(\mu_t) = \infty$ and contradicts the definition of gradient curves (compare this discussion with [AGS1, Theorem 10.4.8]).

Combining Theorems 5.7, 8.5, 8.7, we obtain the following.

Corollary 8.10 *Let $(M, \omega, \varphi, \Psi)$ be an admissible space as in Theorem 8.7, and further suppose that M_φ^Ψ is totally convex. Then the weak solution $(\mu_t)_{t \in [0,\infty)} \subset \mathcal{P}_{\mathrm{ac}}(M_\varphi^\Psi, \omega)$ to the φ-heat equation constructed in Theorem 8.7 satisfies the K-contraction property* (8.1).

8.4 Proof of Claim 8.8

(i) The existence follows from the compactness of $\mathcal{P}(M)$ and the lower semi-continuity of H_φ (Lemma 5.6). The absolute continuity is obvious if $L_\varphi = \infty$.

Assume $L_\varphi < \infty$, so that $\theta_\varphi \in (1, (n+1)/n)$ and $N_\varphi = (\theta_\varphi - 1)^{-1} \in (n, \infty)$ (Proposition 2.13(ii)). We decompose μ^δ into absolutely continuous and singular parts $\mu^\delta = \rho\omega + \mu^s$ and suppose $\mu^s[M] > 0$. For small $r > 0$, we modify μ^δ into $\hat{\mu}_r = \hat{\rho}_r\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ as

$$\hat{\rho}_r(x) := \rho(x) + \int_M \frac{\chi_{B(y,r)}(x)}{\omega[B(y,r)]} \, d\mu^s(y).$$

We shall show that $\hat{\mu}_r$ gives a better choice than μ^δ in our approximation scheme (8.4), which is a contradiction and hence $\mu^s[M] = 0$. We first observe

$$\int_M h'_\varphi(\sigma) \, d\hat{\mu}_r \ge \int_M h'_\varphi(\sigma) \, d\mu^\delta - \int_M \left| h'_\varphi(\sigma(y)) - \frac{1}{\omega[B(y,r)]} \int_{B(y,r)} h'_\varphi(\sigma) \, d\omega \right| d\mu^s(y)$$

$$\ge \int_M h'_\varphi(\sigma) \, d\mu^\delta - \left\{ \sup_M |\nabla(h'_\varphi \circ \sigma)| \cdot r \right\} \mu^s[M]. \tag{8.6}$$

Note that, on M_φ^Ψ, $h'_\varphi(\sigma) = -\Psi - L_\varphi$ is Lipschitz since Ψ is Lipschitz. Given an optimal coupling $\pi = \pi_1 + \pi_2$ of μ^δ and μ_t such that $(p_1)_\sharp \pi_1 = \rho\omega$ and $(p_1)_\sharp \pi_2 = \mu^s$,

$$d\hat{\pi}_r(x, z) := d\pi_1(x, z) + \int_{y \in M} \frac{\chi_{B(y,r)}(x)}{\omega[B(y,r)]} \, d\omega(x) \, d\pi_2(y, z)$$

is a coupling of $\hat{\mu}_r$ and μ_t. Hence we find

$$W_2(\hat{\mu}_r, \mu_t)^2 \le \int_{M \times M} d_g(x, z)^2 \, d\pi_1(x, z) + \int_{M \times M} \{d_g(y, z) + r\}^2 \, d\pi_2(y, z)$$

$$\le \int_{M \times M} d_g(x, z)^2 \, d\pi(x, z) + \{2 \operatorname{diam} M + r\} r \pi_2[M \times M]$$

$$\le W_2(\mu^\delta, \mu_t)^2 + \{3 \operatorname{diam} M \cdot r\} \mu^s[M]. \tag{8.7}$$

Next, observe that

$$\int_M h_\varphi(\hat{\rho}_r)\, d\omega = \int_M h_\varphi\left(\int_M \left\{\frac{\rho(x)}{\mu^s[M]} + \frac{\chi_{B(y,r)}(x)}{\omega[B(y,r)]}\right\} d\mu^s(y)\right) d\omega(x).$$

As h_φ is convex, Jensen's inequality shows

$$h_\varphi\left(\int_M \left\{\frac{\rho(x)}{\mu^s[M]} + \frac{\chi_{B(y,r)}(x)}{\omega[B(y,r)]}\right\} d\mu^s(y)\right)$$
$$\leq \frac{1}{\mu^s[M]} \int_M h_\varphi\left(\rho(x) + \frac{\chi_{B(y,r)}(x)}{\omega[B(y,r)]}\mu^s[M]\right) d\mu^s(y).$$

Since h_φ is non-increasing, we deduce from the Fubini theorem that

$$\int_M h_\varphi(\hat{\rho}_r)\, d\omega \leq \frac{1}{\mu^s[M]} \int_M \left\{\int_{M\backslash B(y,r)} h_\varphi(\rho)\, d\omega + \int_{B(y,r)} h_\varphi\left(\frac{\mu^s[M]}{\omega[B(y,r)]}\right) d\omega\right\} d\mu^s(y)$$
$$\leq \int_M h_\varphi(\rho)\, d\omega - \frac{1}{\mu^s[M]} \int_M \left(\int_{B(y,r)} h_\varphi(\rho)\, d\omega\right) d\mu^s(y)$$
$$+ \sup_{y\in M}\left\{\omega[B(y,r)]\cdot h_\varphi\left(\frac{\mu^s[M]}{\omega[B(y,r)]}\right)\right\}.$$

By virtue of the compactness of M, there are constants $0 < C_1 \leq C_2$ such that

$$C_1 r^n \leq \omega[B(y,r)] \leq C_2 r^n$$

for all $y \in M$ and small $r > 0$. Hence we have, as h_φ is non-increasing and nonpositive,

$$\sup_{y\in M}\left\{\omega[B(y,r)]\cdot h_\varphi\left(\frac{\mu^s[M]}{\omega[B(y,r)]}\right)\right\} \leq C_1 r^n h_\varphi\left(\frac{\mu^s[M]}{C_2 r^n}\right).$$

We find, by the monotonicity of $\ln_\varphi$, Lemma 2.10 and $N_\varphi = (\theta_\varphi - 1)^{-1}$,

$$\limsup_{r\downarrow 0} r^{N_\varphi-1} h_\varphi(r^{-N_\varphi}) = \limsup_{r\downarrow 0}\left\{r^{N_\varphi-1}\int_0^{r^{-N_\varphi}} \ln_\varphi(s)\, ds - r^{-1}L_\varphi\right\}$$
$$\leq \limsup_{r\downarrow 0}\left\{r^{-1}\ln_\varphi(r^{-N_\varphi}) - r^{-1}L_\varphi\right\} = -\liminf_{r\downarrow 0}\int_{r^{-N_\varphi}}^\infty \frac{1}{r\varphi(s)}\, ds$$
$$\leq -\lim_{r\downarrow 0}\int_{r^{-N_\varphi}}^\infty \frac{s^{-\theta_\varphi}}{r}\, ds = \lim_{r\downarrow 0}\frac{r^{N_\varphi(\theta_\varphi-1)}}{(1-\theta_\varphi)r} = \frac{1}{1-\theta_\varphi} < 0.$$

Hence we obtain, since $n < N_\varphi < \infty$,

$$r^{n-1}h_\varphi(r^{-n}) = r^{(n-N_\varphi)/N_\varphi}\cdot\left(r^{n/N_\varphi}\right)^{N_\varphi-1}h_\varphi\left(\left(r^{n/N_\varphi}\right)^{-N_\varphi}\right) \to -\infty$$

as $r \downarrow 0$ (here we need the hypothesis $\theta_\varphi < (n+1)/n$). Finally, for all $y \in \operatorname{supp}\mu^s$, the convexity of h_φ yields

$$\int_{B(y,r)} h_\varphi(\rho)\, d\omega \geq \int_{B(y,r)} \{h_\varphi(\sigma) + h'_\varphi(\sigma)(\rho-\sigma)\}\, d\omega$$
$$= \int_{B(y,r)} \{h_\varphi(\sigma) - h'_\varphi(\sigma)\sigma\}\, d\omega + \int_{B(y,r)} h'_\varphi(\sigma)\, d\mu.$$

52

We therefore obtain

$$\frac{1}{r}\left\{ \int_M h_\varphi(\hat{\rho}_r)\, d\omega - \int_M h_\varphi(\rho)\, d\omega \right\}$$

$$\leq -\frac{1}{r} \inf_{y \in M}\left[\int_{B(y,r)} \{h_\varphi(\sigma) - h'_\varphi(\sigma)\sigma\}\, d\omega + \int_{B(y,r)} h'_\varphi(\sigma)\, d\mu \right] + C_1 r^{n-1} h_\varphi\left(\frac{\mu^s[M]}{C_2 r^n} \right)$$

$$\to -\infty$$

as $r \downarrow 0$. Combining this with (8.6) and (8.7), we conclude that

$$\lim_{r \downarrow 0} \frac{1}{r}\left\{ H_\varphi(\hat{\mu}_r) + \frac{W_2(\hat{\mu}_r, \mu_t)^2}{2\delta} - H_\varphi(\mu^\delta) - \frac{W_2(\mu^\delta, \mu_t)^2}{2\delta} \right\} = -\infty.$$

This contradicts the choice of μ^δ as a minimizer of (8.4), so that it holds $\mu^s[M] = 0$.

(ii) By the choice of μ^δ, we have

$$H_\varphi(\mu^\delta) + \frac{W_2(\mu^\delta, \mu_t)^2}{2\delta} \leq H_\varphi(\mu_t).$$

Together with $H_\varphi(\mu^\delta) \geq H_\varphi(\nu)$ (Lemma 5.5), we immediately observe

$$\lim_{\delta \downarrow 0} W_2(\mu^\delta, \mu_t)^2 \leq \lim_{\delta \downarrow 0} 2\delta\{H_\varphi(\mu_t) - H_\varphi(\nu)\} = 0.$$

Thus μ^δ converges to μ_t weakly, and hence

$$\limsup_{\delta \downarrow 0} \frac{W_2(\mu^\delta, \mu_t)^2}{2\delta} \leq H_\varphi(\mu_t) - \liminf_{\delta \downarrow 0} H_\varphi(\mu^\delta) \leq 0$$

by the lower semi-continuity of H_φ (Lemma 5.6). These further yield

$$H_\varphi(\mu_t) \leq \liminf_{\delta \downarrow 0} H_\varphi(\mu^\delta) \leq \limsup_{\delta \downarrow 0} H_\varphi(\mu^\delta) \leq H_\varphi(\mu_t).$$

(iii) This is a consequence of the following lemma.

Lemma 8.11 *Assume that $\theta_\varphi \in (0, 2)$ and*

$$C_\varphi := \lim_{s \uparrow \infty} \frac{s^{\theta_\varphi}}{\varphi(s)} < \infty.$$

If a sequence $\{\mu_i\}_{i \in \mathbb{N}} \subset \mathcal{P}_{\mathrm{ac}}(M, \omega)$ converges to $\mu \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ weakly and satisfies $\lim_{i \to \infty} H_\varphi(\mu_i) = H_\varphi(\mu) < \infty$, then, by setting $\mu_i = \rho_i \omega$ and $\mu = \rho \omega$, the function $h_\varphi(\rho_i) - \rho_i h'_\varphi(\rho_i)$ converges to $h_\varphi(\rho) - \rho h'_\varphi(\rho)$ in $L^1(M, \omega)$.

Proof. We first show the following claim by using $\theta_\varphi < 2$.

Claim 8.12 *For any $C > 0$, it holds*

$$\lim_{i \to \infty} \| \min\{\rho, C\} - \min\{\rho_i, C\} \|_{L^2(M,\omega)} = 0.$$

Proof. Assume the contrary, that is, there are some constants $C, \varepsilon > 0$ such that, taking a subsequence of $\{\rho_i\}_{i \in \mathbb{N}}$ if necessary, we have

$$\| \min\{\rho, C\} - \min\{\rho_i, C\} \|_{L^2(M,\omega)} \geq \varepsilon \tag{8.8}$$

for all i. Now, since $h_\varphi''(s) = \varphi(s)^{-1}$ is positive and non-increasing, we find

$$h_\varphi\left(\frac{\rho + \rho_i}{2}\right) \leq \frac{h_\varphi(\rho) + h_\varphi(\rho_i)}{2} - \frac{|\rho - \rho_i|^2}{8 \max\{\varphi(\rho), \varphi(\rho_i)\}}.$$

We shall further deduce from $\theta_\varphi < 2$ that

$$\frac{|\rho - \rho_i|^2}{\max\{\varphi(\rho), \varphi(\rho_i)\}} \geq \frac{|\min\{\rho, C\} - \min\{\rho_i, C\}|^2}{\varphi(C)}. \tag{8.9}$$

This is clear if $\max\{\rho, \rho_i\} \leq C$ or $\min\{\rho, \rho_i\} \geq C$. Otherwise, (8.9) is reduced to

$$\frac{(\tau - \varepsilon)^2}{\varphi(\tau)} \geq \frac{(C - \varepsilon)^2}{\varphi(C)}, \qquad \varepsilon \leq C \leq \tau,$$

and to the monotonicity of the function $s \mapsto (s - \varepsilon)^2/\varphi(s)$ for $s > \varepsilon$. This monotonicity is easily seen by Lemma 2.10, since $\theta_\varphi < 2$ and

$$\frac{(s - \varepsilon)^2}{\varphi(s)} = \frac{s^{\theta_\varphi}}{\varphi(s)} \cdot s^{2 - \theta_\varphi} \cdot \left(\frac{s - \varepsilon}{s}\right)^2.$$

Thus we obtain from the hypothesis (8.8) that

$$\int_M h_\varphi\left(\frac{\rho + \rho_i}{2}\right) d\omega \leq \int_M \frac{h_\varphi(\rho) + h_\varphi(\rho_i)}{2} d\omega - \frac{1}{8\varphi(C)} \| \min\{\rho, C\} - \min\{\rho_i, C\} \|_{L^2(M,\omega)}^2$$

$$\leq \frac{1}{2} \int_M h_\varphi(\rho) \, d\omega + \frac{1}{2} \int_M h_\varphi(\rho_i) \, d\omega - \frac{1}{8\varphi(C)} \varepsilon^2.$$

However, as $\lim_{i \to \infty} H_\varphi(\mu_i) = H_\varphi(\mu)$ by assumption, this means that $\bar{\mu}_i := \{(\rho + \rho_i)/2\}\omega$ satisfies

$$\limsup_{i \to \infty} H_\varphi(\bar{\mu}_i) \leq H_\varphi(\mu) - \frac{1}{8\varphi(C)} \varepsilon^2.$$

This contradicts the lower semi-continuity of H_φ (Lemma 5.6) and we complete the proof of Claim 8.12. $\diamondsuit$

Observe that

$$h_\varphi(r) - r h_\varphi'(r) = \int_0^r \{\ln(s) - \ln(r)\} \, ds = -\int_0^r \int_s^r \frac{1}{\varphi(t)} \, dt \, ds = -\int_0^r \frac{t}{\varphi(t)} \, dt.$$

Combining this with Lemma 2.10, we have for any $r, s > 0$

$$|h_\varphi(r) - r h_\varphi'(r) - h_\varphi(s) - s h_\varphi'(s)| = \left| \int_s^r \frac{t}{\varphi(t)} \, dt \right| \leq C_\varphi \left| \int_s^r t^{1 - \theta_\varphi} \, dt \right| = \frac{C_\varphi}{2 - \theta_\varphi} |r^m - s^m|,$$

where we set $m = 2 - \theta_\varphi > 0$. Thus we deduce that

$$\int_M \left| h_\varphi(\rho_i) - \rho_i h_\varphi'(\rho_i) - h_\varphi(\rho) - \rho h_\varphi'(\rho) \right| d\omega \leq \frac{C_\varphi}{2 - \theta_\varphi} \int_M |\rho_i^m - \rho^m| \, d\omega. \tag{8.10}$$

We are done if the right hand side tends to zero as $i \to \infty$.

Claim 8.13 *For $m = 2 - \theta_\varphi \in (0, 2)$, we have*

$$\rho, \rho_i \in L^m(M, \omega), \qquad \lim_{i \to \infty} \|\rho_i - \rho\|_{L^m(M,\omega)} = 0.$$

Proof. The first assertion is clear when $m \leq 1$. For $m > 1$, it is a consequence of $h_\varphi(\rho), h_\varphi(\rho_i) \in L^1(M, \omega)$ (guaranteed by $H_\varphi(\mu), H_\varphi(\mu_i) < \infty$). Indeed, by Lemma 2.10 and (2.9), we have on $\{x \in M \mid \rho(x) \geq C\}$ for any $C > 0$

$$u_\varphi(\rho) - u_\varphi(C) = \int_C^\rho \ln_\varphi(s) \, ds \geq \int_C^\rho \ell_m(s) \, ds = \frac{\rho^m - C^m - m(\rho - C)}{m(m-1)},$$

which implies $\max\{\rho, C\} \in L^m(M, \omega)$ since $m - 1 > 0$, $u_\varphi(\rho) - u_\varphi(C) \geq 0$ and $u_\varphi(\rho) \in L^1(M, \omega)$. Thus we obtain $\rho \in L^m(M, \omega)$ and $\rho_i \in L^m(M, \omega)$ similarly. We remark that, as $\lim_{i \to \infty} H_\varphi(\mu_i) = H_\varphi(\mu)$ by assumption, we have $\lim_{i \to \infty} \int_M u_\varphi(\rho_i) \, d\omega = \int_M u_\varphi(\rho) \, d\omega$ so that $\int_M \rho_i^m \, d\omega$ is uniformly bounded in i.

As for the second estimate, thanks to Claim 8.12 and $m < 2$, it suffices to show that $\rho_i - \min\{\rho_i, C\}$ converges to $\rho - \min\{\rho, C\}$ in $L^m(M, \omega)$ for some (arbitrarily fixed) $C > 0$. Note first that

$$|(\rho_i - \min\{\rho_i, C\}) - (\rho - \min\{\rho, C\})| = |\max\{\rho_i, C\} - \max\{\rho, C\}|.$$

We put $\rho_i^C := \max\{\rho_i, C\}$ and $\rho^C := \max\{\rho, C\}$ for brevity. By the same argumentation as Claim 8.12, $\lim_{i \to \infty} H_\varphi(\mu_i) = H_\varphi(\mu)$ yields

$$\lim_{i \to \infty} \int_M \frac{|\rho_i - \rho|^2}{\max\{\varphi(\rho_i), \varphi(\rho)\}} \, d\omega = 0.$$

Since φ is positive and non-decreasing, it holds

$$\int_M \frac{|\rho_i - \rho|^2}{\max\{\varphi(\rho_i), \varphi(\rho)\}} \, d\omega \geq \int_M \frac{|\rho_i^C - \rho^C|^2}{\varphi(\rho_i^C) + \varphi(\rho^C)} \, d\omega.$$

It follows from the Hölder inequality that

$$\|\rho_i^C - \rho^C\|_{L^m(M,\omega)}^m \leq \left(\int_M \frac{|\rho_i^C - \rho^C|^2}{\varphi(\rho_i^C) + \varphi(\rho^C)} \, d\omega \right)^{m/2} \left(\int_M \{\varphi(\rho_i^C) + \varphi(\rho^C)\}^{m/\theta_\varphi} \, d\omega \right)^{\theta_\varphi/2}.$$

Observe that

$$\{\varphi(\rho_i^C) + \varphi(\rho^C)\}^{m/\theta_\varphi} \leq \begin{cases} \varphi(\rho_i^C)^{m/\theta_\varphi} + \varphi(\rho^C)^{m/\theta_\varphi} & \text{for } m \leq 1, \\ 2^{m/\theta_\varphi - 1}\{\varphi(\rho_i^C)^{m/\theta_\varphi} + \varphi(\rho^C)^{m/\theta_\varphi}\} & \text{for } m > 1. \end{cases}$$

We deduce from Lemma 2.10 that

$$\varphi(\rho_i^C)^{m/\theta_\varphi} + \varphi(\rho^C)^{m/\theta_\varphi} \leq \frac{\varphi(C)^{m/\theta_\varphi}}{C^m} \{(\rho_i^C)^m + (\rho^C)^m\}.$$

Since $\int_M (\rho_i^C)^m \, d\omega$ is uniformly bounded in i, we find

$$\limsup_{i \to \infty} \int_M \{\varphi(\rho_i^C) + \varphi(\rho^C)\}^{m/\theta_\varphi} \, d\omega < \infty,$$

and hence $\lim_{i \to \infty} \|\rho_i^C - \rho^C\|_{L^m(M,\omega)} = 0$. $\diamond$

Now we obtain, for $m \leq 1$,

$$\int_M |\rho_i^m - \rho^m| \, d\omega \leq \int_M |\rho_i - \rho|^m \, d\omega \to 0 \quad (i \to \infty)$$

with the help of Claim 8.13. Similarly, it holds for $m > 1$ that

$$\int_M |\rho_i^m - \rho^m| \, d\omega \leq m \int_M |\rho_i - \rho| \max\{\rho_i, \rho\}^{m-1} \, d\omega$$

$$\leq m \left(\int_M |\rho_i - \rho|^m \, d\omega \right)^{1/m} \left(\int_M (\rho_i + \rho)^m \, d\omega \right)^{(m-1)/m} \to 0 \quad (i \to \infty).$$

$\square$

We remark that, in Lemma 8.11 and hence in Theorem 8.7, the assumptions $\theta_\varphi \in (0, 2)$ and $C_\varphi < \infty$ can be replaced with

$$\delta_\varphi \in (0, 2), \qquad D_\varphi := \lim_{s \downarrow 0} \frac{s^{\delta_\varphi}}{\varphi(s)} < \infty, \qquad d_\varphi := \lim_{s \uparrow \infty} \frac{s^{\delta_\varphi}}{\varphi(s)} > 0.$$

Indeed, then we have

$$\frac{s^{\delta_\varphi}}{D_\varphi} \leq \varphi(s) \leq \frac{s^{\delta_\varphi}}{d_\varphi}$$

for all $s > 0$, and (8.10) becomes

$$\int_M \left| h_\varphi(\rho_i) - \rho_i h_\varphi'(\rho_i) - h_\varphi(\rho) - \rho h_\varphi'(\rho) \right| \, d\omega \leq \frac{D_\varphi}{2 - \delta_\varphi} \int_M |\rho_i^m - \rho^m| \, d\omega$$

for $m := 2 - \delta_\varphi$. With this $m \in (0, 2)$, Claim 8.13 follows from Proposition 2.14 and $\varphi(s) \leq s^{\delta_\varphi}/d_\varphi$ (Claim 8.12 is unnecessary in this case since we can treat ρ and ρ_i themselves instead of ρ^C and ρ_i^C).

Note that $C_\varphi = D_\varphi = d_\varphi = 1 < \infty$ for $\varphi_m(s) = s^{2-m}$. For

$$\varphi(s) := \begin{cases} \sqrt{s} & \text{for } 0 < s < 1, \\ s & \text{for } s \geq 1, \end{cases}$$

we have $\theta_\varphi = 1$, $\delta_\varphi = 1/2$, $C_\varphi = D_\varphi = 1$ and $d_\varphi = 0$. An example of φ with $C_\varphi = \infty$ is

$$\varphi(s) := \begin{cases} \sqrt{s} & \text{for } 0 < s < 1, \\ s & \text{for } 1 \leq s \leq 2, \\ \sqrt{2s} & \text{for } s > 2, \end{cases}$$

for which $\theta_\varphi = 1$, $\delta_\varphi = 1/2$, $D_\varphi = 1$ and $d_\varphi = 1/\sqrt{2}$.

9 Gradient flow of H_φ: Noncompact case

We continue the study of gradient flows in the Wasserstein space $(\mathcal{P}^2(M), W_2)$. For noncompact M, we can not follow the intrinsic argument in Subsection 8.1 since Theorem 8.1 is unavailable. We can nevertheless introduce a Riemannian structure of $\mathcal{P}^2(M)$ using the underlying Riemannian structure of M. Then gradient flows in $\mathcal{P}^2(M)$ are also formulated with the help of the underlying Riemannian/differentiable structure of M. In order to see that the analogue of Theorem 8.7 holds true, we follow the argumentation in [AGS1], [Er] and [Vi2, Chapter 23]. We refer to [AGS1] for the further deep theory of gradient flows.

9.1 Riemannian structure of $(\mathcal{P}^2(M), W_2)$

Recall that minimal geodesics in $\mathcal{P}^2(M)$ emanating from absolutely continuous measures are described by the gradient vector fields of appropriate functions (Theorem 2.6). This leads the following definitions due to Otto [Ot] of the tangent spaces and the Riemannian structure.

Definition 9.1 (Otto's Riemannian structure) We set

$$\hat{T}\mathcal{P} := \{\Phi = \nabla\phi \,|\, \phi \in C_c^\infty(M)\}$$

and define the *tangent space* $(T_\mu\mathcal{P}^2, \langle\cdot, \cdot\rangle_\mu)$ of $\mathcal{P}^2(M)$ at $\mu \in \mathcal{P}^2(M)$ as the completion of $\hat{T}\mathcal{P}$ with respect to the norm $\|\cdot\|_\mu$ induced from the inner product

$$\langle\Phi_1, \Phi_2\rangle_\mu := \int_M \langle\Phi_1, \Phi_2\rangle \, d\mu, \qquad \Phi_1, \Phi_2 \in \hat{T}\mathcal{P}.$$

Note that $\langle\cdot, \cdot\rangle$ is extended to the whole space $T_\mu\mathcal{P}^2$ as the limit, and $(T_\mu\mathcal{P}^2, \langle\cdot, \cdot\rangle_\mu)$ is a Hilbert space. We next introduce the class of 'differentiable curves' in a purely metric way (cf. [AGS1, Section 1.1]).

Definition 9.2 (Absolutely continuous curves) For $p \in [1, \infty]$, a curve $(\mu_t)_{t \in I} \subset \mathcal{P}^2(M)$ on an open interval $I \subset \mathbb{R}$ is said to be *p-absolutely continuous* if there is some $\eta \in L^p_{\text{loc}}(I)$ such that

$$W_2(\mu_s, \mu_t) \leq \int_s^t \eta(r) \, dr \tag{9.1}$$

holds for all $s, t \in I$ with $s < t$.

Note that p-absolutely continuous curves are continuous. We will consider only 2-absolutely continuous curves, so that we simply call them absolutely continuous curves. For any absolutely continuous curve $(\mu_t)_{t \in I} \subset \mathcal{P}^2(M)$, the *metric derivative*

$$|\dot\mu_t| := \lim_{s \to t} \frac{W_2(\mu_s, \mu_t)}{|t - s|}$$

exists for a.e. $t \in I$, and $\eta(t) = |\dot\mu_t|$ is a minimal function satisfying (9.1) (cf. [AGS1, Theorem 1.1.2]). We can associate a one-parameter family of vector fields on M with an absolutely continuous curve in $\mathcal{P}^2(M)$ via the continuity equation on M.

Proposition 9.3 ([AGS1, Theorem 8.3], [Er, Proposition 2.5]) *Given an absolutely continuous curve $(\mu_t)_{t \in I} \subset \mathcal{P}^2(M)$, there exists a Borel vector field $\Phi : I \times M \longrightarrow TM$ (with $\Phi_t(x) := \Phi(t,x) \in T_x M$) satisfying $\Phi_t \in T_{\mu_t} \mathcal{P}^2$ for a.e. $t \in I$ as well as the continuity equation*

$$\frac{\partial \mu_t}{\partial t} + \mathrm{div}(\Phi_t \mu_t) = 0$$

in the weak sense that

$$\int_I \int_M \left\{ \frac{\partial w_t}{\partial t} + \langle \Phi_t, \nabla w_t \rangle \right\} d\mu_t dt = 0 \tag{9.2}$$

holds for all $w \in C_c^\infty(I \times M)$. Such a vector field Φ (satisfying $\Phi_t \in T_{\mu_t} \mathcal{P}^2$ and (9.2)) is uniquely determined up to a difference on a null measure set with respect to $d\mu_t dt$, and we have $\|\Phi_t\|_{\mu_t} = |\dot{\mu}_t|$ for a.e. $t \in I$.

Conversely, if a curve $(\mu_t)_{t \in I} \subset \mathcal{P}^2(M)$ admits a Borel vector field $\Phi : I \times M \longrightarrow TM$ satisfying (9.2) and $\int_{t_0}^{t_1} \|\Phi_t\|_{\mu_t}^2 dt < \infty$ for all $t_0, t_1 \in I$ with $t_0 < t_1$, then $(\mu_t)_{t \in I}$ is absolutely continuous and $|\dot{\mu}_t| \leq \|\Phi_t\|_{\mu_t}$ at a.e. $t \in I$.

Definition 9.4 (Tangent vector fields) We say that the vector field Φ as in Proposition 9.3 is the *tangent vector field* of the absolutely continuous curve $(\mu_t)_{t \in I}$, and write $\dot{\mu}_t = \Phi_t$ (for a.e. $t \in I$).

It is guaranteed by the following *Benamou–Brenier formula* ([BB]) that Otto's Riemannian structure is compatible with the W_2-structure,

$$W_2(\mu_0, \mu_1) = \inf_{(\mu_t)_{t \in [0,1]}} \left(\int_0^1 \|\dot{\mu}_t\|_{\mu_t}^2 \, dt \right)^{1/2}$$

for any $\mu_0, \mu_1 \in \mathcal{P}^2(M)$, where the infimum is taken over all absolutely continuous curves $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}^2(M)$ from μ_0 to μ_1.

9.2 Gradient flow of H_φ

Using the Riemannian structure of $\mathcal{P}^2(M)$ in the previous subsection, we can formulate gradient curves (trajectories of gradient flow) in a way different from the previous section. We first define gradient vectors.

Definition 9.5 (Gradient vectors) Given a functional $H : \mathcal{P}^2(M) \longrightarrow (-\infty, \infty]$ and $\mu \in \mathcal{P}_{\mathrm{ac}}^2(M)$ with $H(\mu) < \infty$, we say that H is *differentiable* at μ if there is $\Phi \in T_\mu \mathcal{P}^2$ such that

$$\limsup_{t \downarrow 0} \frac{H(\mu_t) - H(\mu)}{t} \leq \int_M \langle \Phi, \nabla \phi \rangle \, d\mu$$

along all minimal geodesics $(\mu_t)_{t \in [0,1]} \subset \mathcal{P}^2(M)$ with $\mu_0 = \mu$, where $\mu_t = (\mathcal{T}_t)_\sharp \mu$ with $\mathcal{T}_t(x) = \exp_x(t \nabla \phi(x))$, and if equality holds for $\phi \in C_c^\infty(M)$ (with $\lim_{t \downarrow 0}$ in place of $\limsup_{t \downarrow 0}$). Such Φ is unique if it exists, so that we will write $\nabla_W H(\mu) = \Phi$.

Note that $|\nabla_-(-H)|(\mu) \le \|\nabla_W H(\mu)\|_\mu$ holds by the Cauchy–Schwarz inequality. A gradient curve of the φ-relative entropy H_φ should be understood as a solution to $\dot\mu_t = \nabla_W[-H_\varphi](\mu_t)$. Compare the next proposition with Proposition 8.6.

Proposition 9.6 *Let $(M, \omega, \varphi, \Psi)$ be admissible, assume $\mathrm{Ric}_{N_\varphi} \ge 0$ and $\mathrm{Hess}\,\Psi \ge K$ on M_φ^Ψ for some $K \in \mathbb{R}$ ($K > 0$ if M is noncompact and $\theta_\varphi < 1$). Fix $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}^2(M, \omega)$ with $\mu[M_\varphi^\Psi] = 1$, $H_\varphi(\mu) < \infty$ and with $|\nabla\Psi| \in L^2(M, \mu)$. Then the following are equivalent:*

(I) $|\nabla_- H_\varphi|(\mu) < \infty$,

(II) $\rho \in H_{\mathrm{loc}}^1(M)$ *and*

$$\frac{\nabla\rho}{\varphi(\rho)} + \nabla\Psi = -\Phi$$

holds μ-a.e. for some $\Phi \in T_\mu\mathcal{P}^2$.

Moreover, then we have $\Phi = \nabla_W[-H_\varphi](\mu)$ and $\|\Phi\|_\mu = |\nabla_- H_\varphi|(\mu)$.

Proof. (I) $\Rightarrow$ (II): Note that, by the calculation (before the integration by parts) in the proof of Proposition 6.1,

$$\left| \int_M \left[\{h_\varphi'(\rho)\rho - h_\varphi(\rho)\} \mathrm{div}_\omega V - \langle \rho\nabla\Psi, V\rangle \right] d\omega \right|$$
$$= \lim_{t \downarrow 0} \left\{ \frac{H_\varphi(\mu) - H_\varphi(\mu_t)}{W_2(\mu, \mu_t)} \frac{W_2(\mu, \mu_t)}{t} \right\} \le |\nabla_- H_\varphi|(\mu)\|V\|_\mu$$

for all C^∞-vector fields V of compact support, where we put $\mu_t = (\mathcal{T}_t)_\sharp\mu$ with $\mathcal{T}_t(x) = \exp_x(tV(x))$. Hence the hypothesis (I) together with $\Psi \in H_{\mathrm{loc}}^1(M)$ ensures that the function $h_\varphi'(\rho)\rho - h_\varphi(\rho)$ is weakly differentiable. Since the function $s \longmapsto h_\varphi'(s)s - h_\varphi(s)$ is differentiable and increasing in $s > 0$, this implies $\rho \in H_{\mathrm{loc}}^1(M)$, and we observe

$$\nabla[h_\varphi'(\rho)\rho - h_\varphi(\rho)] = \frac{\rho}{\varphi(\rho)}\nabla\rho.$$

Moreover, the above estimate shows that the function

$$\hat{T}\mathcal{P} \ni \nabla\phi \longmapsto \int_M \langle \nabla[h_\varphi'(\rho)\rho - h_\varphi(\rho)] + \rho\nabla\Psi, \nabla\phi\rangle\, d\omega = \int_M \left\langle \frac{\nabla\rho}{\varphi(\rho)} + \nabla\Psi, \nabla\phi\right\rangle d\mu$$

is extended to a bounded linear operator on the closure $T_\mu\mathcal{P}^2$. Therefore the Riesz representation theorem shows that there exists $\Phi \in T_\mu\mathcal{P}^2$ with

$$\|\Phi\|_\mu \le |\nabla_- H_\varphi|(\mu), \qquad \int_M \left\langle \frac{\nabla\rho}{\varphi(\rho)} + \nabla\Psi, \Xi\right\rangle d\mu = \int_M \langle -\Phi, \Xi\rangle\, d\mu \tag{9.3}$$

for all $\Xi \in T_\mu\mathcal{P}^2$. Thus we have $\nabla\rho/\varphi(\rho) + \nabla\Psi = -\Phi$ μ-a.e..

(II) $\Rightarrow$ (I): We remark that the condition $K > 0$ for $\theta_\varphi < 1$ makes Proposition 6.1 applicable. Thus we obtain

$$\limsup_{t\downarrow 0} \frac{H_\varphi(\mu) - H_\varphi(\mu_t)}{t} \leq \int_M \langle \Phi, \nabla\phi \rangle \, d\mu \tag{9.4}$$

along every minimal geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2(M)$ with $\mu_0 = \mu$, where $\mu_t = (\mathcal{T}_t)_\sharp \mu$ and $\mathcal{T}_t(x) = \exp_x(t\nabla\phi(x))$, and equality holds if $\phi \in C_c^\infty(M)$. Hence $|\nabla_- H_\varphi|(\mu) < \infty$ follows from the hypothesis $\Phi \in T_\mu \mathcal{P}^2$, and we find $\Phi = \nabla_W[-H_\varphi](\mu)$ in the sense of Definition 9.5. We have $\|\Phi\|_\mu \leq |\nabla_- H_\varphi|(\mu)$ by (9.3), and $|\nabla_- H_\varphi|(\mu) \leq \|\Phi\|_\mu$ by (9.4), so that $\|\Phi\|_\mu = |\nabla_- H_\varphi|(\mu)$ holds. $\qquad\square$

Now, we are ready to show the main result of the section. We remark that the roles of the conditions $\mathrm{Ric}_{N_\varphi} \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ are implicit at this stage, whereas they were necessary for applying Proposition 6.1.

Theorem 9.7 (Gradient flow of H_φ) *Suppose that $(M, \omega, \varphi, \Psi)$ is admissible and satisfies $\mathrm{Ric}_{N_\varphi} \geq 0$ as well as $\mathrm{Hess}\,\Psi \geq K$ on M_φ^Ψ for some $K \in \mathbb{R}$ ($K > 0$ if M is noncompact and $\theta_\varphi < 1$). Let $(\mu_t)_{t\in[0,\infty)} \subset \mathcal{P}_{\mathrm{ac}}^2(M, \omega)$ be a continuous curve such that $\mu_t[M_\varphi^\Psi] = 1$, $H_\varphi(\mu_t) < \infty$ and $|\nabla\Psi| \in L^2(M, \mu_t)$ for all $t > 0$. Then $(\mu_t)_{t\in(0,\infty)}$ is an absolutely continuous curve satisfying*

$$\dot\mu_t = \nabla_W[-H_\varphi](\mu_t) \in T_{\mu_t}\mathcal{P}^2$$

at a.e. $t \in (0, \infty)$ if and only if $(\rho_t)_{t\in[0,\infty)}$ is a weak solution to the φ-heat equation (8.2) with $\int_{t_0}^{t_1} |\nabla\rho_t/\varphi(\rho_t)|^2 \, d\mu_t dt < \infty$ for all $0 < t_0 < t_1 < \infty$, where $\mu_t = \rho_t\omega$.

Proof. Suppose $\dot\mu_t = \nabla_W[-H_\varphi](\mu_t)$ a.e. t. Since $|\nabla_- H_\varphi|(\mu_t) \leq \|\nabla_W[-H_\varphi](\mu_t)\|_{\mu_t} < \infty$ by definition, Proposition 9.6 yields

$$\dot\mu_t = -\left(\frac{\nabla\rho_t}{\varphi(\rho_t)} + \nabla\Psi\right) \in T_{\mu_t}\mathcal{P}^2 \qquad \text{a.e. } t.$$

Then it follows from the continuity equation (9.2) that

$$\int_0^\infty \int_M \frac{\partial w_t}{\partial t} \, d\mu_t dt = \int_0^\infty \int_M \left\langle \frac{\nabla\rho_t}{\varphi(\rho_t)} + \nabla\Psi, \nabla w_t \right\rangle d\mu_t dt$$

for all $w \in C_c^\infty((0, \infty) \times M)$. Therefore ρ_t weakly solves (8.2).

Conversely, if ρ_t is a weak solution to (8.2) with $\int_{t_0}^{t_1} |\nabla\rho_t/\varphi(\rho_t)|^2 \, d\mu_t dt < \infty$, then the same calculation implies that

$$\Phi_t = -\left(\frac{\nabla\rho_t}{\varphi(\rho_t)} + \nabla\Psi\right)$$

satisfies the continuity equation (9.2), and hence $(\mu_t)_{t\in(0,\infty)}$ is absolutely continuous by Proposition 9.3. As Proposition 6.1 guarantees $|\nabla_- H_\varphi|(\mu_t) \leq \|\Phi_t\|_{\mu_t} < \infty$ a.e. t (by (9.4)), Proposition 9.6 shows $\Phi_t = \nabla_W[-H_\varphi](\mu_t) \in T_{\mu_t}\mathcal{P}^2$ and then the uniqueness of a solution to the continuity equation (Proposition 9.3) yields $\dot\mu_t = \Phi_t = \nabla_W[-H_\varphi](\mu_t)$ a.e. t. $\qquad\square$

9.3 Remarks on construction and contraction

We can construct the gradient flow of H_φ along the line of [Er, Section 5], provided that M is compact. Precisely, we need the compactness for applying Lemma 5.6, Claim 8.8(i) and Lemma 8.11. As for the contractivity (see (8.1)), the usual technique starts from the first variation formula for the distance $W_2(\mu_t^1, \mu_t^2)$ between two gradient curves (see, e.g., [Er, Proposition 4.4]). To follow this line, however, we need (at least) the C^1-regularity of the density functions ρ_t^i. The authors do not know if such a regularity can be expected for our (nonlinear, scale-variant) φ-heat equation (8.2).

Another recipe (for construction as well as contraction) would be to apply the general theory of Savaré [Sa]. Under $\mathrm{Hess}\, H_\varphi \geq K$ and the additional *semi-concavity condition* of the squared distance function (which is always true for compact Riemannian manifolds), one can construct a unique gradient flow of H_φ enjoying the K-contractivity (8.1). However, we should take care about the point that his (metric) definition of gradient flows is different from the one discussed in Theorem 9.7. Thus, in particular, the existence of a gradient flow in our sense does not follow from Savaré's result.

We also mention another interesting contribution due to Gigli [Gi1], he showed the unique existence of the gradient flow of the relative entropy in a quite general situation without relying on the contractivity. As mentioned at the end of [Gi1], his technique uses some special properties of the generating function $u_{\varphi_1}(s) = s \log s - s$ and is not applicable to all φ's in our consideration (e.g., φ_m for $m < 1$ is excluded).

10 Finsler case

Most results in this article are extended to Finsler manifolds according to the theory of Ricci curvature developed in [Oh2], [OS1] (see also a survey [Oh3]). A Finsler manifold is a differentiable manifold equipped with a (Minkowski) norm on each tangent space. Restricting these norms to those coming from inner products, we have the family of Riemannian manifolds as a subclass. We refer to [BCS] and [Sh] for the basics of Finsler geometry.

10.1 Finsler manifolds

Let M be a connected, n-dimensional C^∞-manifold without boundary. Given a local coordinate $(x^i)_{i=1}^n$ on an open set $U \subset M$, we will always use the coordinate $(x^i, \mathbf{v}^j)_{i,j=1}^n$ of TU such that

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

Definition 10.1 (Finsler structures) We say that a nonnegative function $F : TM \longrightarrow [0, \infty)$ is a C^∞-*Finsler structure* of M if the following three conditions hold:

(1) (Regularity) F is C^∞ on $TM \setminus 0$, where $0 \subset TM$ stands for the zero section.

(2) (*Positive* 1-*homogeneity*) It holds $F(c\mathbf{v}) = cF(\mathbf{v})$ for all $\mathbf{v} \in TM$ and $c > 0$.

(3) (*Strong convexity*) The $n \times n$ symmetric matrix

$$\big(g_{ij}(\mathbf{v})\big)_{i,j=1}^n := \left(\frac{1}{2} \frac{\partial^2 (F^2)}{\partial \mathbf{v}^i \partial \mathbf{v}^j}(\mathbf{v}) \right)_{i,j=1}^n \tag{10.1}$$

is positive-definite for all $\mathbf{v} \in T_x M \setminus 0$.

We call such a pair (M, F) a C^∞-*Finsler manifold.*

In other words, F provides a C^∞-Minkowski norm (see Example 10.2(a) below) on each tangent space $T_x M$ which varies smoothly also in the horizontal direction. For $x, y \in M$, we define the distance from x to y in a natural way by $d_F(x, y) := \inf_\gamma \int_0^1 F\big(\dot{\gamma}(t)\big) \, dt$, where the infimum is taken over all C^1-curves $\gamma : [0, 1] \longrightarrow M$ such that $\gamma(0) = x$ and $\gamma(1) = y$. Note that d_F is not necessarily symmetric, namely $d_F(y, x) \neq d_F(x, y)$ can happen, since F is only positively homogeneous. A C^∞-curve γ on M is called a *geodesic* if it is locally distance minimizing and has a constant speed (i.e., $F(\dot{\gamma})$ is constant). We remark that $t \longmapsto \gamma(1 - t)$ may not be a geodesic. Given $\mathbf{v} \in T_x M$, if there is a geodesic $\gamma : [0, 1] \longrightarrow M$ with $\dot{\gamma}(0) = \mathbf{v}$, then we define the *exponential map* by $\exp_x(\mathbf{v}) := \gamma(1)$. We say that (M, F) is *forward complete* if the exponential map is defined on whole TM. Then the Hopf–Rinow theorem ensures that any pair of points is connected by a minimal geodesic (cf. [BCS, Theorem 6.6.1]).

We define the K-convexity of a function $\Psi : M \longrightarrow \mathbb{R}$ in the weak sense similarly to the case of symmetric distances (Definition 4.1), i.e., for any $x, y \in M$ there is a minimal geodesic $\gamma : [0, 1] \longrightarrow M$ from x to y such that

$$\Psi\big(\gamma(t)\big) \leq (1 - t)\Psi(x) + t\Psi(y) - \frac{K}{2}(1 - t)t d_F(x, y)^2$$

for all $t \in [0, 1]$.

For each $\mathbf{v} \in T_x M \setminus 0$, the positive-definite matrix $(g_{ij}(\mathbf{v}))_{i,j=1}^n$ in (10.1) induces the Riemannian structure $g_\mathbf{v}$ of $T_x M$ via

$$g_\mathbf{v}\left(\sum_{i=1}^n a_i \frac{\partial}{\partial x^i}\Big|_x, \sum_{j=1}^n b_j \frac{\partial}{\partial x^j}\Big|_x \right) := \sum_{i,j=1}^n g_{ij}(\mathbf{v})a_i b_j. \tag{10.2}$$

This is regarded as the best Riemannian approximation of $F|_{T_x M}$ in the direction $\mathbf{v}$. In fact, the unit sphere of $g_\mathbf{v}$ is tangent to that of $F|_{T_x M}$ at $\mathbf{v}/F(\mathbf{v})$ up to the second order. In particular, we have $g_\mathbf{v}(\mathbf{v}, \mathbf{v}) = F(\mathbf{v})^2$.

Let us denote by $\mathcal{L}^* : T^* M \longrightarrow TM$ the *Legendre transform*. Precisely, $\mathcal{L}^*$ is sending $\alpha \in T_x^* M$ to the unique element $\mathbf{v} \in T_x M$ such that $\alpha(\mathbf{v}) = F^*(\alpha)^2$ and $F(\mathbf{v}) = F^*(\alpha)$, where F^* stands for the dual norm of F. Note that $\mathcal{L}^*|_{T_x^* M}$ is a linear operator only when $F|_{T_x M}$ comes from an inner product. For a differentiable function $\rho : M \longrightarrow \mathbb{R}$, the *gradient vector* of ρ at x is defined as the Legendre transform of the derivative of ρ,

$$\nabla \rho(x) := \mathcal{L}^*\big(D\rho(x)\big) \in T_x M.$$

If $D\rho(x) = 0$, then clearly $\nabla \rho(x) = 0$. If $D\rho(x) \neq 0$, then we can write in coordinates

$$\nabla \rho = \sum_{i,j=1}^n g^{ij}(\nabla \rho) \frac{\partial \rho}{\partial x^j} \frac{\partial}{\partial x^i},$$

where (g^{ij}) stands for the inverse matrix of (g_{ij}). We must be careful when $D\rho(x) = 0$, because $g_{ij}(\nabla\rho(x))$ is not defined as well as the Legendre transform $\mathcal{L}^*$ being only continuous at the zero section. We also remark that the gradient ∇ is a nonlinear operator (i.e., $\nabla(\rho_1 + \rho_2)(x) \neq \nabla\rho_1(x) + \nabla\rho_2(x)$ and $\nabla(-\rho)(x) \neq -\nabla\rho(x)$ in general), since the Legendre transform is nonlinear unless F happens to be Riemannian.

We mention some of basic examples of non-Riemannian Finsler manifolds.

Example 10.2 (a) (Minkowski spaces) A *Minkowski norm* $|\cdot|$ on $\mathbb{R}^n$ is a nonnegative function on $\mathbb{R}^n$ satisfying the conditions in Definition 10.1. Note that the unit ball of $|\cdot|$ is a strictly convex (but not necessarily symmetric to the origin) domain containing the origin in its interior. A Minkowski norm induces a Finsler structure in a natural way through the identification between $T_x\mathbb{R}^n$ and $\mathbb{R}^n$. Then $(\mathbb{R}^n, |\cdot|)$ has the flat flag curvature (the *flag curvature* is a generalization of the sectional curvature).

(b) (Randers spaces) A *Randers space* (M, F) is a special kind of Finsler manifold given by $F(\mathbf{v}) = \sqrt{g(\mathbf{v}, \mathbf{v})} + \beta(\mathbf{v})$ for some Riemannian metric g and a one-form β, where we suppose $|\beta(\mathbf{v})|^2 < g(\mathbf{v}, \mathbf{v})$ unless $\mathbf{v} = 0$, for F being positive on $TM \setminus 0$. Randers spaces are important in applications and reasonable for concrete calculations. Sometimes β is regarded as the effect of wind blowing on the Riemannian manifold (M, g).

(c) (Hilbert geometry) Let $D \subset \mathbb{R}^n$ be a bounded open set with smooth boundary such that its closure $\overline{D}$ is strictly convex. Then the associated *Hilbert distance function* is defined by

$$d_H(x_1, x_2) := \log\left(\frac{|x_1 - x_2'| \cdot |x_2 - x_1'|}{|x_1 - x_1'| \cdot |x_2 - x_2'|}\right)$$

for distinct $x_1, x_2 \in D$, where $|\cdot|$ is the standard Euclidean norm and x_1', x_2' are intersections of ∂D and the line passing through x_1, x_2 such that x_i' is on the side of x_i. Hilbert geometry is known to be realized by a Finsler structure with constant negative flag curvature, and gives the Klein model of hyperbolic space if D is an ellipsoid.

(d) (Teichmüller space) Teichmüller metric on Teichmüller space is arguably one of the most famous Finsler structures in differential geometry. It is known to be complete, while, e.g., the Weil–Petersson metric is incomplete and Riemannian.

10.2 Weighted Ricci curvature and nonlinear Laplacian

Different from the Riemannian situation, one can not choose a unique canonical measure on a Finsler manifold. There are several constructive measures, such as the *Busemann–Hausdorff measure* and the *Holmes–Thompson measure*, which are canonical in their own ways (see, e.g., [AT]). Thus we will fix an arbitrary positive C^∞-measure ω on M as our base measure, like the theory of weighted Riemannian manifolds.

The *Ricci curvature* (as the trace of the flag curvature) on a Finsler manifold is defined by using the Chern connection (there are other connections but the flag and Ricci curvatures are in fact independent of the choice of connection). Instead of giving a precise definition in coordinates, here we explain a useful interpretation due to Shen [Sh, §6.2]. Given a unit vector $\mathbf{v} \in T_xM \cap F^{-1}(1)$, we extend it to a C^∞-vector field V on a neighborhood of x in such a way that every integral curve of V is geodesic, and consider the Riemannian structure g_V induced from (10.2). Then the Ricci curvature $\mathrm{Ric}(\mathbf{v})$ of $\mathbf{v}$ with

respect to F coincides with the Ricci curvature of $\mathbf{v}$ with respect to g_V (in particular, it is independent of the choice of V).

Inspired by the above interpretation of the Ricci curvature as well as the theory of weighted Riemannian manifolds, the *weighted Ricci curvature* for (M, F, ω) was introduced in [Oh2] as follows.

Definition 10.3 (Weighted Ricci curvature) Given a unit vector $\mathbf{v} \in T_x M$, let $\gamma : (-\varepsilon, \varepsilon) \longrightarrow M$ be the geodesic such that $\dot{\gamma}(0) = \mathbf{v}$. We decompose ω as $\omega = e^{-f} \operatorname{vol}_{\dot{\gamma}}$ along γ, where $\operatorname{vol}_{\dot{\gamma}}$ is the volume form of $g_{\dot{\gamma}}$. Define

(1) $\operatorname{Ric}_n(\mathbf{v}) := \begin{cases} \operatorname{Ric}(\mathbf{v}) + (f \circ \gamma)''(0) & \text{if } (f \circ \gamma)'(0) = 0, \\ -\infty & \text{otherwise,} \end{cases}$

(2) $\operatorname{Ric}_N(\mathbf{v}) := \operatorname{Ric}(\mathbf{v}) + (f \circ \gamma)''(0) - \dfrac{(f \circ \gamma)'(0)^2}{N - n}$ for $N \in (-\infty, 0) \cup (n, \infty)$,

(3) $\operatorname{Ric}_\infty(\mathbf{v}) := \operatorname{Ric}(\mathbf{v}) + (f \circ \gamma)''(0)$.

For $c \geq 0$, we set $\operatorname{Ric}_N(c\mathbf{v}) := c^2 \operatorname{Ric}_N(\mathbf{v})$.

It is established in [Oh2, Theorem 1.2] that, for $K \in \mathbb{R}$ and $N \in [n, \infty]$, the bound $\operatorname{Ric}_N(\mathbf{v}) \geq KF(\mathbf{v})^2$ is equivalent to the curvature-dimension condition $\mathsf{CD}(K, N)$ (note that (M, d_F) is non-branching and thus Sturm's and Lott–Villani's conditions are equivalent). This extends the corresponding result on weighted Riemannian manifolds (Theorems 5.1, 5.2). There are further applications of Ric_N beyond the curvature-dimension condition, e.g., a Bochner-type formula and gradient estimates ([OS3]).

Remark 10.4 For a Riemannian manifold $(M, g, \operatorname{vol}_g)$ endowed with the Riemannian volume measure, clearly we have $f \equiv 0$ and hence $\operatorname{Ric}_N = \operatorname{Ric}$ for all N. It is also known that, for Finsler manifolds of *Berwald type*, the Busemann–Hausdorff measure satisfies $(f \circ \gamma)' \equiv 0$ (in other words, Shen's **S**-*curvature* vanishes, see [Sh, §7.3]). In general, however, there may not exist any measure ω of vanishing **S**-curvature, see [Oh4] for such an example. This means that, on a general Finsler manifold, there is no measure as good as the Riemannian volume measure. This is a reason why we began with an arbitrary measure ω.

Define the *divergence* of a differentiable vector field V on M with respect to the base measure ω by

$$\operatorname{div}_\omega V := \sum_{i=1}^n \left(\frac{\partial V_i}{\partial x^i} + V_i \frac{\partial \eta}{\partial x^i} \right),$$

where we decompose ω in coordinates as $d\omega = e^\eta \, dx^1 dx^2 \cdots dx^n$. Similarly to the Riemannian case, this can be rewritten (and extended to weakly differentiable vector fields) in the weak form as

$$\int_M w \operatorname{div}_\omega V \, d\omega = -\int_M Dw(V) \, d\omega$$

for all $w \in C_c^\infty(M)$. Then we define the corresponding *Laplacian* of $\rho \in H_{\mathrm{loc}}^1(M)$ by $\Delta^\omega \rho := \mathrm{div}_\omega(\nabla\rho)$ in the distributional sense that

$$\int_M w\Delta^\omega\rho\,d\omega := -\int_M Dw(\nabla\rho)\,d\omega$$

for $w \in C_c^\infty(M)$. We remark that $H_{\mathrm{loc}}^1(M)$ is defined solely in terms of the differentiable structure of M. It is established in [OS1] and [OS3] that this nonlinear Laplacian works quite well with the weighted Ricci curvature.

For later convenience, we introduce the following notations.

Definition 10.5 (Reverse Finsler structure) Define the *reverse Finsler structure* $\overleftarrow{F}$ of F by $\overleftarrow{F}(\mathbf{v}) := F(-\mathbf{v})$. We will put arrows $\leftarrow$ on those quantities associated with $\overleftarrow{F}$, for example, $\overleftarrow{d}_F(x,y) = d_F(y,x)$, $\overleftarrow{\nabla}\rho = -\nabla(-\rho)$ and $\overleftarrow{\mathrm{Ric}}_N(\mathbf{v}) = \mathrm{Ric}_N(-\mathbf{v})$.

10.3 Displacement convexity of H_φ and applications

From now on, we consider only compact Finsler manifolds for simplicity. We remark that all compact Finsler manifolds are forward complete.

Let us consider an admissible space $(M, \omega, \varphi, \Psi)$ in the sense of Definition 4.3 similarly to the Riemannian case. Then the analogue of Theorem 5.7 is demonstrated along the same line as the Riemannian case (see [Oh2] for details).

We can show the functional inequalities in Theorem 6.3 also in the same way by using the directional derivative of H_φ (see (6.2)) modified into

$$\liminf_{t\downarrow 0} \frac{H_\varphi(\mu_t) - H_\varphi(\mu)}{t} \geq \int_M \left(\frac{D\rho}{\varphi(\rho)} + D\Psi\right)(\nabla\phi)\,d\mu.$$

Precisely, the φ-relative Fisher information of $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ is defined by

$$I_\varphi(\mu) := \int_M F(\nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)])^2\,d\mu = \int_M F^*\left(\frac{D\rho}{\varphi(\rho)} + D\Psi\right)^2\,d\mu,$$

and the φ-global Poincaré inequality means

$$\int_{M_\varphi^\Psi} \frac{w^2\sigma}{\varphi(\sigma)}\,d\nu \leq \frac{1}{K}\int_{M_\varphi^\Psi} F\left(\nabla\left(\frac{w\sigma}{\varphi(\sigma)}\right)\right)^2\,d\nu.$$

We also remark that $W_2(\mu, \nu)$ in (i) and (ii) of Theorem 6.3 can be replaced with $W_2(\nu, \mu)$ since the curvature bound $\mathrm{Ric}_N \geq K$ for F is equivalent to that for its reverse $\overleftarrow{F}$. The above φ-Talagrand inequality shows the concentration of measures as in Section 7, where the open ball $B(A, r)$ in the definition of the concentration function $\alpha(r)$ is replaced with

$$B^+(A, r) := \left\{y \in M \,\Big|\, \inf_{x\in A} d_F(x, y) < r\right\} \text{ or } B^-(A, r) := \left\{y \in M \,\Big|\, \inf_{x\in A} d_F(y, x) < r\right\}.$$

10.4 Gradient flow of H_φ

As for the gradient flow of H_φ, due to the lack of the analogue of Theorem 8.1, the argument in Section 8 is unavailable. Nonetheless, one can apply the discussion in Section 9 using a (formal) Finsler structure of the Wasserstein space, and obtain a result corresponding to Theorem 9.7. We remark that, however, the K-contraction property (8.1) essentially depends on the Riemannian structure and can not be expected in the Finsler setting (see [OS2] for details).

Let (M, F) be compact again. We introduce a Finsler structure of $(\mathcal{P}(M), W_2)$ similarly to Section 9. Given $\mu \in \mathcal{P}(M)$, define the *tangent space* $(T_\mu \mathcal{P}, F_\mu)$ at μ by

$$F_\mu(\nabla\phi) := \left(\int_M F(\nabla\phi)^2 \, d\mu \right)^{1/2} \text{ for } \phi \in C^\infty(M), \quad T_\mu \mathcal{P} := \overline{\{\nabla\phi \,|\, \phi \in C^\infty(M)\}},$$

where the closure was taken with respect to the (Minkowski) norm F_μ. Then we can follow the line of Section 9 up to some computational differences. We denote by $\mathcal{L} := (\mathcal{L}^*)^{-1} : TM \longrightarrow T^*M$ the Legendre transform in the reverse direction.

Definition 10.6 (Gradient vectors) Given a functional $H : \mathcal{P}(M) \longrightarrow (-\infty, \infty]$ and $\mu \in \mathcal{P}(M)$ with $H(\mu) < \infty$, we say that H is *differentiable* at μ if there is $\Phi \in T_\mu \mathcal{P}$ such that

$$\limsup_{t\downarrow 0} \frac{H(\mu_t) - H(\mu)}{t} \leq \int_M \mathcal{L}(\Phi)(\nabla\phi) \, d\mu$$

along all minimal geodesics $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}(M)$ with $\mu_0 = \mu$, where $\mu_t = (\mathcal{T}_t)_\sharp \mu$ and $\mathcal{T}_t(x) := \exp_x(t\nabla\phi(x))$, and if equality holds for $\phi \in C^\infty(M)$ (with $\lim_{t\downarrow 0}$ in place of $\limsup_{t\downarrow 0}$). Such Φ is unique if it exists, and then we write $\nabla_W H(\mu) = \Phi$.

Proposition 10.7 *Let* $(M, \omega, \varphi, \Psi)$ *be compact, admissible and satisfy* $\mathrm{Ric}_{N_\varphi} \geq 0$ *and* $\mathrm{Hess}\,\Psi \geq K$ *on* M_φ^Ψ *for some* $K \in \mathbb{R}$, *and fix* $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ *with* $\mu[M_\varphi^\Psi] = 1$ *and* $H_\varphi(\mu) < \infty$. *Then the following are equivalent:*

(I) $|\nabla_- H_\varphi|(\mu) < \infty$,

(II) $\rho \in H^1(M)$ *and*
$$\Phi = \mathcal{L}^*\left(-\frac{D\rho}{\varphi(\rho)} - D\Psi \right) \qquad \mu\text{-a.e.}$$
for some $\Phi \in T_\mu \mathcal{P}$.

Moreover, then we have $\Phi = \nabla_W[-H_\varphi](\mu)$ *and* $F_\mu(\Phi) = |\nabla_- H_\varphi|(\mu)$.

Note that

$$\Phi = \mathcal{L}^*\left(-\frac{D\rho}{\varphi(\rho)} - D\Psi \right) = \mathcal{L}^*\big(D[-\ln_\varphi(\rho) - \Psi] \big) = \nabla[-\ln_\varphi(\rho) - \Psi].$$

Theorem 10.8 (Gradient flow of H_φ) *Let us suppose that* $(M, \omega, \varphi, \Psi)$ *is compact, admissible and satisfies* $\mathrm{Ric}_{N_\varphi} \geq 0$ *as well as* $\mathrm{Hess}\,\Psi \geq K$ *on* M_φ^Ψ *for some* $K \in \mathbb{R}$, *and*

let $(\mu_t)_{t \in [0,\infty)} \subset \mathcal{P}_{\mathrm{ac}}(M, \omega)$ be a continuous curve such that $\mu_t[M_\varphi^\Psi] = 1$ and $H_\varphi(\mu_t) < \infty$ for all $t > 0$. Then $(\mu_t)_{t \in (0,\infty)}$ is an absolutely continuous curve satisfying

$$\dot{\mu}_t = \nabla_W[-H_\varphi](\mu_t) \in T_{\mu_t}\mathcal{P}$$

at a.e. $t \in (0, \infty)$ if and only if $(\rho_t)_{t \in [0,\infty)}$ is a weak solution to the reverse φ-heat equation of the form

$$\frac{\partial \rho}{\partial t} = -\operatorname{div}_\omega\left(\rho\nabla[-\ln_\varphi(\rho) - \Psi]\right) \tag{10.3}$$

with $\int_{t_0}^{t_1} F(\nabla[-\ln_\varphi(\rho_t)])^2 \, d\mu_t dt < \infty$ for all $0 < t_0 < t_1 < \infty$, where $\mu_t = \rho_t\omega$.

Proof. If $\dot{\mu}_t = \nabla_W[-H_\varphi](\mu_t)$ a.e. t, then Proposition 10.7 yields $\dot{\mu}_t = \nabla[-\ln_\varphi(\rho_t) - \Psi] \in T_{\mu_t}\mathcal{P}$ a.e. t. Thus it follows from the continuity equation (9.2) that

$$\int_0^\infty \int_M \frac{\partial w_t}{\partial t} \, d\mu_t dt = -\int_0^\infty \int_M Dw_t\left(\nabla[-\ln_\varphi(\rho) - \Psi]\right) d\mu_t dt$$

for all $w \in C_c^\infty((0, \infty) \times M)$, and hence ρ_t weakly solves (10.3). Conversely, if ρ_t is a weak solution to (10.3), then the same calculation implies that $\Phi_t = \nabla[-\ln_\varphi(\rho_t) - \Psi]$ satisfies the continuity equation (9.2), and $(\mu_t)_{t \in (0,\infty)}$ is absolutely continuous. Therefore Proposition 10.7 shows $\dot{\mu}_t = \Phi_t = \nabla_W[-H_\varphi](\mu_t)$ a.e. t. $\qquad\square$

We meant by the *reverse φ-heat equation* the equation with respect to the reverse Finsler structure $\overleftarrow{F}(\mathbf{v}) = F(-\mathbf{v})$. Since the gradient vector for $\overleftarrow{F}$ is written as $\overleftarrow{\nabla}\rho = -\nabla(-\rho)$, (10.3) is indeed rewritten as

$$\frac{\partial \rho}{\partial t} = \operatorname{div}_\omega\left(\rho\overleftarrow{\nabla}[\ln_\varphi(\rho) + \Psi]\right).$$

A   Appendix: Measure concentration via u_φ-entropy inequality

Let us go back to the Riemannian situation. In Section 6, we introduced the φ-logarithmic Sobolev inequality (6.5) by generalizing the relative entropy to the φ-relative entropy associated with the Bregman divergence. Precisely, the classical logarithmic Sobolev inequality (corresponding to $\varphi_1(s) = s$) of the form

$$\operatorname{Ent}_\nu(\mu) - \operatorname{Ent}_\nu(\nu) \leq \frac{2}{K}\int_M \left|\nabla\left(\sqrt{\frac{\rho}{\sigma}}\right)\right|^2 d\nu = \frac{1}{2K}\int_M \left|\frac{\nabla\rho}{\rho} - \frac{\nabla\sigma}{\sigma}\right|^2 d\mu$$

is generalized to

$$H_\varphi(\mu) - H_\varphi(\nu) \leq \frac{1}{2K}\int_M \left|\nabla[\ln_\varphi(\rho) - \ln_\varphi(\sigma)]\right|^2 d\mu,$$

where $\mu = \rho\omega$, $\nu = \sigma\omega$ and K is a positive constant.

The logarithmic Sobolev inequality has the alternative form

$$\int_M w \ln(w) \, d\nu - \left(\int_M w \, d\nu \right) \ln \left(\int_M w \, d\nu \right) \le \frac{1}{2K} \int_M \frac{|\nabla w|^2}{w} \, d\nu$$

for nonnegative measurable functions $w : M \longrightarrow [0, \infty)$. Then the inequality

$$\int_M u_\varphi(w) \, d\nu - u_\varphi \left(\int_M w \, d\nu \right) \le \frac{1}{2K} \int_M u_\varphi''(w) |\nabla w|^2 \, d\nu = \frac{1}{2K} \int_M \frac{|\nabla w|^2}{\varphi(w)} \, d\nu$$

obtained by replacing the function $r \longmapsto r \ln r$ (generating the relative entropy) with u_φ is called the u_φ-*entropy inequality*, which provides a generalization of the logarithmic Sobolev inequality different from our φ-logarithmic Sobolev inequality. The function φ is usually imposed to be concave, that is equivalent to the convexity of the function

$$(s, t) \longmapsto d_\varphi(s + t, t) := u_\varphi(s + t) - u_\varphi(t) - \ln_\varphi(t)s.$$

Note that d_φ coincides with the density function of the Bregman divergence D_φ. We refer to [Chf1] and [Chf2] for details, where instead of u_φ it is treated C^2-strictly convex functions Φ such that $1/\Phi''$ is concave.

We demonstrated in Section 7 that the φ-Talagrand inequality leads the $m(\varphi)$-normal concentration of measures. In the classical case of $\varphi_1(s) = s$, it is known that the normal concentration also follows from the logarithmic Sobolev inequality by the Herbst argument (see, e.g., [Le, Chapter 5]). In the same spirit, we can deduce from the u_φ-entropy inequality the corresponding φ-normal concentration of measures. We first recall a kind of Chebyshev's inequality for later use.

Lemma A.1 (Chebyshev's inequality) *Let w be a measurable function on a measure space (X, μ). Then for any nonnegative, non-decreasing, measurable function v on $\mathbb{R}$,*

$$\mu \left[\{ x \in X \mid w(x) \ge t \} \right] \le \frac{1}{v(t)} \int_X v(w) \, d\mu$$

holds for any $t > 0$ with $v(t) > 0$.

We next show an auxiliary lemma. We will normalize φ as $\varphi(1) = 1$ for simplicity, recall that such a normalization does not change the value of θ_φ (Remark 2.11).

Lemma A.2 *Let $\varphi : (0, \infty) \longrightarrow (0, \infty)$ be a positive concave function with $\varphi(1) = 1$. Then we have $\theta_\varphi \le 1$ and*

$$u_\varphi(s) + a_\varphi s \ge a_\varphi \varphi(s) \ln_\varphi(s)$$

for any $s > 0$, where we set $a_\varphi := -u_\varphi(1) > 0$.

Proof. It follows from the concavity of φ that

$$\frac{\varphi(s + t) - \varphi(s)}{t} \le \frac{\varphi(s) - \varphi(\varepsilon)}{s - \varepsilon} < \frac{\varphi(s)}{s - \varepsilon}$$

68

for any $0 < \varepsilon < s < s + t$. Letting $\varepsilon \downarrow 0$ and then $t \downarrow 0$, we find

$$\frac{s}{\varphi(s)} \cdot \limsup_{t \downarrow 0} \frac{\varphi(s + t) - \varphi(s)}{t} \leq 1.$$

Since $s > 0$ is arbitrary, we obtain $\theta_\varphi \leq 1$.

Set $A(s) := u_\varphi(s) + a_\varphi s - a_\varphi \varphi(s) \ln_\varphi(s)$ and observe $A(1) = 0$ by the choice of a_φ. Proposition 2.14 implies

$$0 \geq \lim_{s \downarrow 0} \varphi(s) \ln_\varphi(s) \geq \lim_{s \downarrow 0} s^{\delta_\varphi} \ell_{2 - \delta_\varphi}(s) = 0,$$

so that $\lim_{s \downarrow 0} \varphi(s) \ln_\varphi(s) = 0$ and we can put $A(0) := 0$. Since the concavity of φ ensures that the right derivative

$$\varphi'_+(s) := \lim_{\varepsilon \downarrow 0} \frac{\varphi(s + \varepsilon) - \varphi(s)}{\varepsilon} \in \left[0, \frac{\varphi(s)}{s} \theta_\varphi\right]$$

is well-defined and non-increasing on $(0, \infty)$, a direct computation yields

$$A'_+(s) := \lim_{\varepsilon \downarrow 0} \frac{A(s + \varepsilon) - A(s)}{\varepsilon} = \ln_\varphi(s) \left\{1 - a_\varphi \varphi'_+(s)\right\}.$$

Note that (2.9) shows

$$1 - a_\varphi \varphi'_+(1) \geq 1 - a_\varphi \theta_\varphi = 1 + \theta_\varphi \int_0^1 \ln_\varphi(t) \, dt \geq 1 + \theta_\varphi \int_0^1 \ell_{2 - \theta_\varphi}(t) \, dt = 1 - \frac{\theta_\varphi}{2 - \theta_\varphi} \geq 0.$$

For $s \geq 1$, we deduce from $\ln_\varphi(s) \geq \ln_\varphi(1) = 0$ and $\varphi'_+(s) \leq \varphi'_+(1)$ that $A'_+(s) \geq 0$. Hence we have $A(s) \geq A(1) = 0$. On $(0, 1)$, since $\ln_\varphi < 0$, $A(0) = A(1) = 0$ and φ'_+ is non-increasing, A is identically zero or there is some $s_0 \in (0, 1)$ such that $A'_+ \geq 0$ on $(0, s_0)$ and that $A'_+ \leq 0$ on $(s_0, 1)$. Therefore we conclude that $A \geq 0$ on $(0, 1)$. □

Remark A.3 The condition $\theta_\varphi \leq 1$ does not imply the concavity of φ. For instance, let

$$\varphi(s) := \begin{cases} \sqrt{s} & \text{for } 0 < s < 1, \\ s & \text{for } s \geq 1. \end{cases}$$

Then we have $\theta_\varphi = 1$, whereas φ is clearly not concave.

Now we prove that the u_φ-entropy inequality implies the φ-normal concentration for φ as in Lemma A.2.

Theorem A.4 (φ-normal concentration from u_φ-entropy inequality) *Take a positive concave function $\varphi : (0, \infty) \longrightarrow (0, \infty)$ such that $\varphi(1) = 1$. For a Riemannian manifold (M, g) and $\nu \in \mathcal{P}(M)$, assume that there is a positive constant K such that the u_φ-entropy inequality*

$$\int_M u_\varphi(w) \, d\nu - u_\varphi\left(\int_M w \, d\nu\right) \leq \frac{1}{2K} \int_M u''_\varphi(w) |\nabla w|^2 \, d\nu \tag{A.1}$$

holds for every nonnegative measurable function $w \in L^1(M, \nu)$ satisfying $u''_\varphi(w)|\nabla w|^2 \in L^1(M, \nu)$. Then for any $r > 0$ we have

$$\alpha(r)^{-1} \geq \exp_\varphi\left(-\frac{u_\varphi(1)K}{8}r^2\right),$$

where α stands for the concentration function of (M, ν).

Proof. Fix arbitrary $A \subset M$ with $\nu[A] \geq 1/2$ and $r > 0$. Putting $B := M \setminus B(A, r)$, we also assume $\nu[B] > 0$ since we have $\alpha(r) = 0$ if $\nu[B] = 0$ for all such A. Set $F_r(x) := \min\{d_g(x, A), r\}$ for $x \in M$, and observe that F_r is 1-Lipschitz. Note also that the function

$$G_r(x) := F_r(x) - \int_M F_r \, d\nu$$

satisfies $G_r(x) \geq r/2$ for any $x \in B$ since $\int_M F_r \, d\nu \leq r \cdot \nu[M \setminus A] \leq r/2$. Applying Chebyshev's inequality (Lemma A.1) to the nonnegative, non-decreasing function

$$v_s(t) := \exp_\varphi\left(st - \frac{s^2}{2a_\varphi K}\right)$$

with $s > 0$ and $a_\varphi := -u_\varphi(1) > 0$, we have

$$\nu[B] \leq \nu\left[\left\{x \in M \ \Big| \ G_r(x) \geq \frac{r}{2}\right\}\right] \leq \frac{1}{v_s(r/2)}\int_M v_s(G_r)\,d\nu. \tag{A.2}$$

We shall show that $I(s) := \int_M v_s(G_r)\,d\nu \geq \int_B v_s(r/2)\,d\nu > 0$ is bounded above by 1.

Set

$$w_s(x) := v_s\big(G_r(x)\big) = \exp_\varphi\left(sG_r(x) - \frac{s^2}{2a_\varphi K}\right),$$

$$X_s := \left\{x \in M \ \Big| \ sG_r(x) - \frac{s^2}{2a_\varphi K} > l_\varphi\right\}.$$

For $s \in (0, a_\varphi Kr)$ and any $x \in B$, we have

$$sG_r(x) - \frac{s^2}{2a_\varphi K} \geq \frac{rs}{2} - \frac{s^2}{2a_\varphi K} = -\frac{1}{2a_\varphi K}\left(s - \frac{a_\varphi Kr}{2}\right)^2 + \frac{a_\varphi Kr^2}{8} \geq 0 > l_\varphi,$$

proving $B \subset X_s$. Let us introduce the strictly convex function $\Phi_\varphi(t) := u_\varphi(t) + a_\varphi t$ on $[0, \infty)$, and observe that $\Phi_\varphi \leq 0$ on $[0, 1]$ and $\Phi_\varphi > 0$ on $(1, \infty)$. Then the inequality (A.1) applied to $w = w_s$ can be rewritten as

$$\int_{X_s} \Phi_\varphi(w_s)\,d\nu - \frac{1}{2K}\int_{X_s} \Phi''_\varphi(w_s)|\nabla w_s|^2\,d\nu \leq \Phi_\varphi\left(\int_{X_s} w_s\,d\nu\right).$$

Note that w_s is bounded since G_r is bounded by definition, and hence $w_s \in L^1(M, \nu)$. Moreover, $u''_\varphi(w_s)|\nabla w_s|^2 \in L^1(M, \nu)$ is seen by

$$\int_{X_s} \Phi''_\varphi(w_s)|\nabla w_s|^2\,d\nu = \int_{X_s} s^2\varphi(w_s)|\nabla G_r|^2\,d\nu < s^2\int_{X_s} \varphi(w_s)\,d\nu$$

70

for $s \in (0, a_\varphi Kr)$, where we used the fact that $|\nabla G_r| \leq 1$ on whole M and $|\nabla G_r| \equiv 0$ on B. It follows from Lemma A.2 that

$$\int_{X_s} \left(\Phi_\varphi(w_s) - \frac{s^2}{2K}\varphi(w_s) \right) d\nu \geq \int_{X_s} \left(a_\varphi \varphi(w_s) \ln_\varphi(w_s) - \frac{s^2}{2K}\varphi(w_s) \right) d\nu$$
$$= \int_{X_s} \varphi(w_s) \left(sa_\varphi G_r - \frac{s^2}{K} \right) d\nu = sa_\varphi \frac{d}{ds} \left(\int_{X_s} w_s \, d\nu \right).$$

These together imply, as $\int_{X_s} w_s \, d\nu = \int_M w_s \, d\nu = I(s)$,

$$sa_\varphi I'(s) < \Phi_\varphi\big(I(s)\big) \qquad \text{for } s \in (0, a_\varphi Kr). \tag{A.3}$$

For $s_0 \in (0, a_\varphi Kr)$ chosen later, set

$$P(s) := \exp\left(\frac{1}{a_\varphi} \int_{s_0}^{s} \frac{\Phi_\varphi'(I(t))}{t} \, dt \right), \qquad Q(s) := \frac{\Phi_\varphi(I(s))}{P(s)}$$

for $s \in (0, s_0]$, and observe

$$Q'(s) = \frac{\Phi_\varphi'(I(s))}{P(s)} \left\{ I'(s) - \frac{\Phi_\varphi(I(s))}{sa_\varphi} \right\}.$$

Then we deduce from (A.3) that $Q'(s) = 0$ if and only if $\Phi_\varphi'(I(s)) = 0$.

Assume that $\sup_{s \in (0, a_\varphi Kr)} I(s) > I(0) = 1$ and choose $s_0 \in (0, a_\varphi Kr)$ such that $I(s_0) > 1$ and

$$c := \sup_{s \in (0, s_0]} \Phi_\varphi'\big(I(s)\big) \in (a_\varphi, 2a_\varphi)$$

(note that $\Phi_\varphi'(I(0)) = a_\varphi$). Then we have

$$P(s) \geq \exp\left(\frac{c}{a_\varphi} \int_{s_0}^{s} \frac{1}{t} \, dt \right) = \left(\frac{s}{s_0} \right)^{c/a_\varphi}, \qquad s \in (0, s_0]. \tag{A.4}$$

Moreover, since the convexity of Φ_φ and $\Phi_\varphi(0) = 0$ imply $I(s)\Phi_\varphi'(I(s)) \geq \Phi_\varphi(I(s))$, we find $\Phi_\varphi'(I(s_0)) > 0$ and hence $Q'(s_0) < 0$. Note that there does not exist $s \in (0, s_0)$ such that $Q' \leq 0$ on (s, s_0) as well as $Q'(s) = 0$, since then $I(s)\Phi_\varphi'(I(s)) = 0 \geq \Phi_\varphi(I(s))$ and $Q(s_0) \leq Q(s) \leq 0$ which contradicts $I(s_0) > 1$. Thus $Q' < 0$ on $(0, s_0)$, and by (A.4)

$$Q(s_0) \leq \limsup_{s \downarrow 0} Q(s) \leq s_0^{c/a_\varphi} \limsup_{s \downarrow 0} \frac{\Phi_\varphi(I(s))}{s^{c/a_\varphi}}.$$

Now, since

$$I'(0) = \int_M G_r \, d\nu = 0, \qquad \Phi_\varphi\big(I(s)\big) = \Phi_\varphi(1) + s\Phi_\varphi'(1)I'(0) + O(s^2) = O(s^2)$$

and $c < 2a_\varphi$, it holds $\lim_{s \downarrow 0} s^{-c/a_\varphi} \Phi_\varphi(I(s)) = 0$. This means $Q(s_0) \leq 0$ and hence $I(s_0) \leq 1$, which is a contradiction. We therefore obtain $I(s) \leq I(0) = 1$ for any $s \in (0, a_\varphi Kr)$ as desired.

Hence we deduce from (A.2) that $\nu[B] \leq v_s(r/2)^{-1}$ for any $s \in (0, a_\varphi K r)$. Choosing $s = a_\varphi K r/2$ and taking the supremum in A, we conclude that

$$\alpha(r) \leq \frac{1}{\exp_\varphi(a_\varphi K r^2/8)}.$$

$\square$

Remark A.5 Bolley and Gentil [BG] showed that if a probability measure on $\mathbb{R}^n$ satisfies $\mathsf{CD}(K, \infty)$ with $K > 0$, then it satisfies the u_φ-entropy inequality (A.1) with the same constant K. We remark that the condition $\mathsf{CD}(K, \infty)$ leads the normal concentration which is stronger than the φ-normal concentration for $\theta_\varphi < 1$ (since $\exp_\varphi(r)^{-1} \geq e_{2-\theta_\varphi}(r)^{-1} \geq e^{-r}$ by (2.10)), whereas there exists a probability measure which satisfies (A.1) and does not satisfy $\mathsf{CD}(K, \infty)$. See [LO, Theorem 2] for details, where they proved that the probability measure on $\mathbb{R}^n$ of the form

$$d\mu_a(x) := \left(\frac{a}{2\Gamma(1/a)} \right)^n \exp(-|x|^a) \, d\mathcal{L}^n(x)$$

with $a \in [1, 2)$ satisfies the u_{φ_m}-entropy inequality for $m \in (1, 2]$, while the concentration function $\alpha(r)$ of μ_a is dominated by $\exp(-r^2/3)$ (resp. $\exp(-r^a/3)$) for $r < 1$ (resp. $r > 1$).

References

[AGK]   M. Agueh, N. Ghoussoub and X. Kang, Geometric inequalities via a general comparison principle for interacting gases, Geom. Funct. Anal. **14** (2004), 215–244.

[AT]   J. C. Álvarez-Paiva and A. C. Thompson, Volumes in normed and Finsler spaces, A sampler of Riemann-Finsler geometry, 1–48, Math. Sci. Res. Inst. Publ., **50**, Cambridge Univ. Press, Cambridge, 2004.

[Am]   S. Amari, Differential-geometrical methods in statistics, Lecture Notes in Statistics, **28**. Springer-Verlag, New York, 1985.

[AN]   S. Amari and H. Nagaoka, Methods of information geometry. Translated from the 1993 Japanese original by Daishi Harada. American Mathematical Society, Providence, RI; Oxford University Press, Oxford, 2000.

[AGS1] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Birkhäuser Verlag, Basel, 2005.

[AGS2] L. Ambrosio, N. Gigli and G. Savaré, Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below, Preprint (2011). Available at arXiv:1106.2090

[AGS3] L. Ambrosio, N. Gigli and G. Savaré, Metric measure spaces with Riemannian Ricci curvature bounded from below, Preprint (2011). Available at arXiv:1109.0222

[BS]   K. Bacher and K.-T. Sturm, Localization and tensorization properties of the curvature-dimension condition for metric measure spaces, J. Funct. Anal. **259** (2010), 28–56.

[BE]   D. Bakry and M. Émery, Diffusions hypercontractives (French), Séminaire de probabilités, XIX, 1983/84, 177–206, Lecture Notes in Math. **1123**, Springer, Berlin, 1985.

[BCS]   D. Bao, S.-S. Chern and Z. Shen, An introduction to Riemann-Finsler geometry, Springer-Verlag, New York, 2000.

[BB]   J.-D. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. **84** (2000), 375–393.

[BG]   F. Bolley and I. Gentil, Phi-entropy inequalities for diffusion semigroups, J. Math. Pures Appl. (9) **93** (2010), no. 5, 449–473.

[BBI]   D. Burago, Yu. Burago and S. Ivanov, A course in metric geometry, American Mathematical Society, Providence, RI, 2001.

[CMV1]   J. A. Carrillo, R. J. McCann and C. Villani, Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates, Rev. Mat. Iberoamericana **19** (2003), no. 3, 971–1018.

[CMV2]   J. A. Carrillo, R. J. McCann and C. Villani, Contractions in the 2-Wasserstein length space and thermalization of granular media, Arch. Ration. Mech. Anal. **179** (2006), no. 2, 217–263.

[Chf1]   D. Chafaï, Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities, J. Math. Kyoto Univ. **44** (2004), no. 2, 325–363.

[Chf2]   D. Chafaï, Binomial-Poisson entropic inequalities and the $M/M/\infty$ queue, ESAIM Probab. Stat. **10** (2006), 317–339 (electronic).

[Chv]   I. Chavel, Riemannian geometry. A modern introduction. Second edition, Cambridge University Press, Cambridge, 2006.

[CGH]   D. Cordero-Erausquin, W. Gangbo and C. Houdré, Inequalities for generalized entropy and optimal transportation, Recent advances in the theory and applications of mass transport, 73–94, Contemp. Math. **353**, Amer. Math. Soc., Providence, RI, 2004.

[CMS1]   D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, A Riemannian interpolation inequality à la Borell, Brascamp and Lieb, Invent. Math. **146** (2001), 219–257.

[CMS2]   D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, Prékopa-Leindler type inequalities on Riemannian manifolds, Jacobi fields, and optimal transport, Ann. Fac. Sci. Toulouse Math. (6) **15** (2006), 613–635.

[Er]   M. Erbar, The heat equation on manifolds as a gradient flow in the Wasserstein space, Ann. Inst. Henri Poincaré Probab. Stat. **46** (2010), 1–23.

[FSS]   S. Fang, J. Shao and K.-T. Sturm, Wasserstein space over the Wiener space, Probab. Theory Related Fields **146** (2010), 535–565.

[FG]   A. Figalli and N. Gigli, Local semiconvexity of Kantorovich potentials on non-compact manifolds, ESAIM Control Optim. Calc. Var. **17** (2011), 648–653.

[Gi1]   N. Gigli, On the heat flow on metric measure spaces: existence, uniqueness and stability, Calc. Var. Partial Differential Equations **39** (2010), 101–120.

[Gi2]   N. Gigli, On the inverse implication of Brenier-McCann theorems and the structure of $(\mathscr{P}_2(M), W_2)$, Methods Appl. Anal. **18** (2011), 127–158.

[GKO]   N. Gigli, K. Kuwada and S. Ohta, Heat flow on Alexandrov spaces, to appear in Comm. Pure Appl. Math. Available at arXiv:1008.1319

[GO]   N. Gigli and S. Ohta, First variation formula in Wasserstein spaces over compact Alexandrov spaces, to appear in Canad. Math. Bull.

[Ju]   N. Juillet, Diffusion by optimal transport in Heisenberg groups, Preprint (2009). Available at http://www-irma.u-strasbg.fr/~ juillet/

[JKO]   R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. **29** (1998), 1–17.

[LO]   R. Latała and K. Oleszkiewicz, Between Sobolev and Poincaré, Geometric aspects of functional analysis, 147–168, Lecture Notes in Math., **1745**, Springer, Berlin, 2000.

[Le]   M. Ledoux, The concentration of measure phenomenon, American Mathematical Society, Providence, RI, 2001.

[Lo]   J. Lott, Some geometric properties of the Bakry-Émery-Ricci tensor, Comment. Math. Helv. **78** (2003), 865–883.

[LV1]   J. Lott and C. Villani, Weak curvature conditions and functional inequalities, J. Funct. Anal. **245** (2007), 311–333.

[LV2]   J. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. **169** (2009), 903–991.

[Ly]   A. Lytchak, Open map theorem for metric spaces, St. Petersburg Math. J. **17** (2006), 477–491.

[Ma]   J. Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. **261** (2011), 2250–2292.

[Mc1]   R. J. McCann, A convexity principle for interacting gases, Adv. Math. **128** (1997), 153–179.

[Mc2]   R. J. McCann, Polar factorization of maps on Riemannian manifolds, Geom. Funct. Anal. **11** (2001), 589–608.

[MTKE]   N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, Information geometry of U-boost and Bregman divergence, Neural Computation, **16** (2004), no. 7, 1437–1481.

[Na1]   J. Naudts, Continuity of a class of entropies and relative entropies, Rev. Math. Phys. **16** (2004), no. 6, 809–822.

[Na2]   J. Naudts, Estimators, escort probabilities, and φ-exponential families in statistical physics, JIPAM J. Inequal. Pure Appl. Math. **5** (2004), no. 4, Article 102, 15pp.

[OW]   A. Ohara and T. Wada, Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations, J. Phys. A **43** (2010), 035002, 18pp.

[Oh1]   S. Ohta, Gradient flows on Wasserstein spaces over compact Alexandrov spaces, Amer. J. Math. **131** (2009), 475–516.

[Oh2]   S. Ohta, Finsler interpolation inequalities, Calc. Var. Partial Differential Equations **36** (2009), 211–249.

[Oh3]   S. Ohta, Optimal transport and Ricci curvature in Finsler geometry, Adv. Stud. Pure Math. **57** (2010), 323–342.

[Oh4]   S. Ohta, Vanishing S-curvature of Randers spaces, Differential Geom. Appl. **29** (2011), 174–178.

[OS1]   S. Ohta and K.-T. Sturm, Heat flow on Finsler manifolds, Comm. Pure Appl. Math. **62** (2009), 1386–1433.

[OS2]   S. Ohta and K.-T. Sturm, Non-contraction of heat flow on Minkowski spaces, Arch. Ration. Mech. Anal. **204** (2012), 917–944.

[OS3]   S. Ohta and K.-T. Sturm, Bochner-Weitzenböck formula and Li-Yau estimates on Finsler manifolds, Preprint (2011). Available at arXiv:1104.5276

[OT]   S. Ohta and A. Takatsu, Displacement convexity of generalized relative entropies, Adv. Math. **228** (2011), 1742–1787.

[Ot]   F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations **26** (2001), 101–174.

[OV]   F. Otto and C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. **173** (2000), 361–400.

[PP]   G. Perel'man and A. Petrunin, Quasigeodesics and gradient curves in Alexandrov spaces, Unpublished preprint. Available at http://www.math.psu.edu/petrunin/

[Pe]   A. Petrunin, Semiconcave functions in Alexandrov's geometry, Surveys in differential geometry. Vol. XI, 137–201, Surv. Differ. Geom. **11**, Int. Press, Somerville, MA, 2007.

[Qi]   Z. Qian, Estimates for weighted volumes and applications, Quart. J. Math. Oxford Ser. (2) **48** (1997), 235–242.

[vRS]   M.-K. von Renesse and K.-T. Sturm, Transport inequalities, gradient estimates, entropy and Ricci curvature, Comm. Pure Appl. Math. **58** (2005), 923–940.

[Sa]   G. Savaré, Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds, C. R. Math. Acad. Sci. Paris **345** (2007), 151–154.

[Sh]   Z. Shen, Lectures on Finsler geometry, World Scientific Publishing Co., Singapore, 2001.

[St1]   K.-T. Sturm, Convex functionals of probability measures and nonlinear diffusions on manifolds, J. Math. Pures Appl. **84** (2005), 149–168.

[St2]   K.-T. Sturm, On the geometry of metric measure spaces. I, Acta Math. **196** (2006), 65–131.

[St3]   K.-T. Sturm, On the geometry of metric measure spaces. II, Acta Math. **196** (2006), 133–177.

[Ta1]   A. Takatsu, Wasserstein geometry of porous medium equation, Ann. Inst. H. Poincaré Anal. Non Linéaire **29** (2012), 217–232.

[Ta2]   A. Takatsu, Behaviors of φ-exponential distributions in Wasserstein geometry and an evolution equation, Preprint (2011). Available at arXiv:1109.6776

[Ts1]   C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, J. Statist. Phys. **52** (1988), 479–487.

[Ts2]   C. Tsallis, Introduction to nonextensive statistical mechanics, Springer, 2009.

[Vi1]   C. Villani, Topics in optimal transportation, American Mathematical Society, Providence, RI, 2003.

[Vi2]   C. Villani, Optimal transport, old and new, Springer-Verlag, Berlin, 2009.