# Displacement convexity of generalized entropies

Shin-ichi Ohta*and Asuka Takatsu†

May 7, 2010

**Abstract**

We investigate the $m$-relative entropy, which stems from the Bregman divergence, on weighted Riemannian and Finsler manifolds. We prove that the displacement convexity of the $m$-relative entropy is equivalent to the combination of a lower Ricci curvature bound and the convexity of the weight function. We use this to show appropriate variants of the Talagrand, HWI and the logarithmic Sobolev inequalities, as well as the concentration of measures. We also prove that the gradient flow of the $m$-relative entropy produces a solution to the porous medium equation.

## 1 Introduction

The *displacement convexity* was introduced in McCann's influential paper [Mc1] as the convexity along geodesics in the Wasserstein space. Recent astonishing development of optimal transport theory reveals that the displacement convexity of an entropy-type functional plays important roles in the theory of partial differential equations, probability theory and Riemannian geometry (see [Vi1], [Vi2] and the references therein). For instance, on a compact Riemannian manifold $(M, g)$ equipped with the Riemannian volume measure $\mathrm{vol}_g$, the gradient flow of the relative entropy $\mathrm{Ent}_{\mathrm{vol}_g}$ (see (3.2)) in the $L^2$-Wasserstein space $(\mathcal{P}(M), W_2)$ produces a weak solution to the heat equation ([Oh1], [GO], [Vi2, Chapter 23]). Then the $K$-convexity of $\mathrm{Ent}_{\mathrm{vol}_g}$ for some $K \in \mathbb{R}$ (denoted by $\mathrm{Hess}\,\mathrm{Ent}_{\mathrm{vol}_g} \ge K$ for short) implies the $K$-contraction property

$$W_2\big(p(t,x,\cdot)\,\mathrm{vol}_g, p(t,y,\cdot)\,\mathrm{vol}_g\big) \le e^{-tK} d(x,y)$$

of the heat kernel $p : (0,\infty) \times M \times M \longrightarrow (0,\infty)$ (and vice versa, [vRS]). The condition $\mathrm{Hess}\,\mathrm{Ent}_{\mathrm{vol}_g} \ge K$ is called the *curvature-dimension condition* $\mathsf{CD}(K,\infty)$ and known to be equivalent to the lower Ricci curvature bound $\mathrm{Ric} \ge K$ ([vRS]). There is a rich theory on general metric measure spaces satisfying $\mathsf{CD}(K,\infty)$ ([St1], [LV2], [Vi2, Part III]).

*Department of Mathematics, Kyoto University, Kyoto 606-8502, Japan (sohta@math.kyoto-u.ac.jp) & Max-Planck-Intitut für Mathematik, Vivatsgasse 7, 53111 Bonn, Germany; Supported in part by the Grant-in-Aid for Young Scientists (B) 20740036.

†Mathematical Institute, Tohoku University, Sendai 980-8578, Japan (sa6m21@math.tohoku.ac.jp); Supported in part by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

Especially, $\mathsf{CD}(K,\infty)$ with $K>0$ is an important condition which yields, among others, the logarithmic Sobolev inequality and the normal concentration of measures (a kind of large deviation principle).

The curvature-dimension condition is extended to $\mathsf{CD}(K,N)$ for general $K \in \mathbb{R}$ and $N \in (1,\infty]$, and then $\mathsf{CD}(K,N)$ is equivalent to the lower bound of the weighted Ricci curvature $\mathrm{Ric}_N \ge K$ of a weighted Riemannian manifold $(M,\omega)$, where $\omega$ is a conformal deformation of $\mathrm{vol}_g$ ([St2], [LV1]). However, $\mathsf{CD}(K,N)$ with $N<\infty$ is written as a simple convexity condition only when $K=0$ (and it causes some difficulties when $K \neq 0$, see [BS]). Precisely, $\mathsf{CD}(0,N)$ is defined as the convexity of the Rényi entropy $S_N$ ($=(m-1)E_m - 1$ with $m = 1-1/N$, see (3.1)), while $\mathsf{CD}(K,N)$ with $K \neq 0$ is a more subtle inequality involving the integrand of $S_N$. It will be observed in Remark 4.3(2) that $\mathrm{Hess}\, S_N \ge K$ is actually meaningless when $K \neq 0$.

In this article, we introduce and consider a different kind of entropy $H_m(\cdot|\nu)$ for $m \in [(n-1)/n, 1) \cup (1,\infty)$ —we call this the *m-relative entropy*— which is related to, but different from $S_N$. Here $\nu = \exp_m(-\Psi)\omega$ is a fixed conformal deformation of $\omega$, and $\exp_m$ is the $m$-exponential function (see Subsection 2.2). Our definition of $H_m(\cdot|\nu)$ stems from the Bregman divergence in information theory/geometry which is closely related to the Tsallis/Rényi entropy (see Subsection 3.1). Roughly speaking, $H_m(\mu|\nu)$ represents the difference between $\mu$ and $\nu$. Taking the limit as $m$ tends 1 recovers the usual relative entropy $\mathrm{Ent}_\nu$ (or the Kullback-Leibler divergence $H(\cdot|\nu)$). Our results will guarantee that $H_m(\cdot|\nu)$ is a natural and important object.

Our first main theorem asserts that $\mathrm{Hess}\, H_m(\cdot|\nu) \ge K$ in $(\mathcal{P}^2(M), W_2)$ is equivalent to the combination of $\mathrm{Ric}_N \ge 0$ with $N = 1/(1-m)$ and $\mathrm{Hess}\,\Psi \ge K$, where $\mathrm{Ric}_N$ is of $(M,\omega)$ (Theorem 4.1, we remark that $N$ can be negative, such $\mathrm{Ric}_N$ would be of independent interest). It is interesting to obtain such split curvature bound/convexity conditions from a single convexity condition of the entropy. Then, according to the technique similar to the curvature-dimension condition, we show that $\mathrm{Ric}_N \ge 0$ and $\mathrm{Hess}\,\Psi \ge K$ imply appropriate variants of the Talagrand, HWI, logarithmic Sobolev and the global Poincaré inequalities (Propositions 5.1, 5.4, Theorem 5.2), and also the concentration of measures (Theorem 6.1). Furthermore, the gradient flow of $H_m(\cdot|\nu)$ produces a weak solution to the porous medium equation of the form

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

where $\Delta^\omega$ and $\mathrm{div}_\omega$ are the Laplacian and the divergence associated with the measure $\omega$ (Theorem 7.6). Most results hold true also for Finsler manifolds thanks to the theory developed in [Oh2] and [OS] (see Section 8).

Former work on this kind of entropy has been concerned with only the unweighted Euclidean spaces $(\mathbb{R}^n, dx)$ where $\omega = dx$ is the Lebesgue measure (as far as the authors know). Among them, Otto [Ot] demonstrated that the gradient flow of the Tsallis entropy $E_m$ solves the porous medium equation, and it is shown in [AGK], [CGH] and [Ta2] that $\mathrm{Hess}\,\Psi \ge K > 0$ (with $\nu = \exp_m(-\Psi)dx$) implies various functional inequalities. Even in this unweighted Euclidean situation, however, the equivalence between $\mathrm{Hess}\,\Psi \ge K$ and $\mathrm{Hess}\, H_m(\cdot|\nu) \ge K$ and the concentration of measures are previously unknown.

The organization of the article is as follows. After preliminaries, we introduce the $m$-relative entropy $H_m(\cdot|\nu)$ in Section 3, and show that $\mathrm{Hess}\, H_m(\cdot|\nu) \ge K$ is equivalent to

$\operatorname{Hess}\Psi \ge K$ with $\mathrm{Ric}_N \ge 0$ in Section 4. Using this equivalence, we obtain several funcitonal inequalities in Section 5, and the concentration of measures in Section 6. Section 7 is devoted to the study of the gradient flow of $H_m(\cdot|\nu)$. Finally, we treat the Finsler case in Section 8.

# 2 Preliminaries

Throughout the article except Section 8, $(M, g)$ will be an $n$-dimensional Riemannian manifold with $n \ge 2$. We denote by $B(x, r)$ the open ball of center $x \in M$ and radius $r > 0$.

## 2.1 Weighted Ricci curvature

We fix a conformal change $\omega = e^{-\psi}\,\mathrm{vol}_g$, with $\psi \in C^\infty(M)$, of the Riemannian volume measure $\mathrm{vol}_g$ as our base measure. Given a unit vector $v \in T_xM$ and $N \in (-\infty, 0) \cup (n, \infty)$, we define the *weighted Ricci curvature* by

$$\mathrm{Ric}_N(v) := \mathrm{Ric}(v) + \operatorname{Hess}\psi(v, v) - \frac{\langle \nabla\psi, v\rangle^2}{N - n}. \tag{2.1}$$

We also set

$$\mathrm{Ric}_n(v) := \begin{cases} \mathrm{Ric}(v) + \operatorname{Hess}\psi(v,v) & \text{if } \langle \nabla\psi, v\rangle = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Observe that, if $\psi$ is constant, then $\mathrm{Ric}_N(v)$ coincides with $\mathrm{Ric}(v)$ for all $N$.

**Remark 2.1** We usually consider $\mathrm{Ric}_N$ only for $N \in [n, \infty]$ (where $\mathrm{Ric}_\infty(v) = \mathrm{Ric}(v) + \operatorname{Hess}\psi(v, v)$ is the *Bakry-Émery tensor*, see [BE], [Qi], [Lo]), and then it enjoys the monotonicity: $\mathrm{Ric}_N(v) \le \mathrm{Ric}_{N'}(v)$ for $N < N'$. Admitting $N < 0$ violates this monotonicity, but we abuse this notation for brevity. The reason why we consider this range of $N$ will be seen in (2.2).

As we mentioned in the introduction, $\mathrm{Ric}_N \ge K$ for some $K \in \mathbb{R}$ and $N \ge n$ is equivalent to Sturm's curvature-dimension condition $\mathsf{CD}(K, N)$. Spaces satisfying $\mathsf{CD}(K, N)$ behave like a space with dimension $\le N$ and Ricci curvature $\ge K$ (see [St2], [LV1], [Vi2, Part III]).

## 2.2 Generalized exponential functions and Gaussian measures

We briefly recall the *m-calculus*, see [Ts2] for further discussion. We introduce a parameter $m$ such that

$$m \in [(n-1)/n, 1) \cup (1, \infty).$$

We sometimes eliminate the special case $m = 1/n = 1/2$ or restrict ourselves to $m \le 2$. We set

$$N = N(m) := 1/(1 - m) \in (-\infty, 0) \cup [n, \infty). \tag{2.2}$$

Define the *$m$-logarithmic function* by

$$\ln_m(t) := \frac{t^{m-1}-1}{m-1} \quad \text{for} \begin{cases} t > 0 & \text{if } m < 1, \\ t \geq 0 & \text{if } m > 1. \end{cases}$$

Note that $\ln_m$ is monotone increasing and that the image of $\ln_m$ is $(-\infty, 1/(1-m))$ if $m < 1$; $[-1/(m-1), \infty)$ if $m > 1$. We define the *$m$-exponential function* $\exp_m$ as the inverse of $\ln_m$, namely

$$\exp_m(t) := \{1 + (m-1)t\}^{1/(m-1)} \quad \text{for} \begin{cases} t \in (-\infty, 1/(1-m)) & \text{if } m < 1, \\ t \in [-1/(m-1), \infty) & \text{if } m > 1. \end{cases}$$

For the sake of simplicity, we set $\exp_m(t) := 0$ for $m > 1$, $t < -1/(m-1)$. For $t > 0$, we define

$$e_m(t) := t \ln_m(t) = \frac{t^m - t}{m-1}.$$

We also set $e_m(0) := 0$. Observe that

$$\lim_{m \to 1} \ln_m(t) = \ln(t), \qquad \lim_{m \to 1} \exp_m(t) := e^t, \qquad \lim_{m \to 1} e_m(t) = t \ln(t).$$

**Remark 2.2** (1) Taking $m < 1$ and $m > 1$ give rise to qualitatively different phenomena (see Lemma 2.4, Example 2.5 for instances). Nonetheless, most of our results cover both cases.

(2) In some notations, it is common to use the parameter $q = 2 - m$ instead of $m$ (e.g., $\exp_q$ and $q$-Gaussian measures). In the present paper, however, we shall use only $m$ for simplicity.

Using $\exp_m$ and $\omega = e^{-\psi} \operatorname{vol}_g$, we will fix

$$\nu = \sigma\omega := \exp_m(-\Psi)\omega$$

as our reference measure, where $\Psi \in C(M)$ such that $\Psi > -1/(1-m)$ if $m < 1$. For later convenience, we set

$$M_0 := \begin{cases} M & \text{for } m < 1, \\ \Psi^{-1}\big((-\infty, 1/(m-1))\big) & \text{for } m > 1, \end{cases}$$

and assume that $M_0$ is nonempty. Note that $\operatorname{supp}\nu = \overline{M_0}$ holds in both cases. We shall study how the convexity of $\Psi$ is related to the geometric/analytic structure of $(M, \nu)$. Given $K \in \mathbb{R}$, we say that $\Psi$ is *$K$-convex in the weak sense*, denoted by $\operatorname{Hess}\Psi \geq K$, if any two points $x, y \in M$ admits a minimal geodesic $\gamma : [0,1] \longrightarrow M$ between them along which

$$\Psi\big(\gamma(t)\big) \leq (1-t)\Psi(x) + t\Psi(y) - \frac{K}{2}(1-t)td(x,y)^2 \tag{2.3}$$

holds for all $t \in [0,1]$. Note that this is equivalent to saying that (2.3) holds along any minimal geodesic $\gamma$ between $x$ and $y$, for $\gamma|_{[\varepsilon, 1-\varepsilon]}$ is a unique minimal geodesic for all $\varepsilon > 0$ and $\Psi$ is continuous.

**Remark 2.3** Consider a different presentation $\nu = (c\sigma)(c^{-1}\omega) =: \tilde{\sigma}\tilde{\omega}$ of $\nu$ for some constant $c > 0$. Then the weighted Ricci curvature $\mathrm{Ric}_N$ is unchanged, while

$$\begin{aligned}\tilde{\sigma} &= c\exp_m(-\Psi) = \{c^{m-1} - (m-1)c^{m-1}\Psi\}^{1/(m-1)} \\ &= \left\{1 - (m-1)\left(c^{m-1}\Psi - \frac{c^{m-1}-1}{m-1}\right)\right\}^{1/(m-1)} =: \exp_m(-\widetilde{\Psi})\end{aligned}$$

and hence $\mathrm{Hess}\,\widetilde{\Psi} = c^{m-1}\,\mathrm{Hess}\,\Psi$.

Sections 5, 6 will be concerned with the case where $\mathrm{Hess}\,\Psi \geq K > 0$ as well as $\mathrm{Ric}_N \geq 0$. In such a situation, it turns out that $\nu$ has finite total mass. Here we give explicit estimates for later use (in Section 6).

**Lemma 2.4** *Assume that* $\mathrm{Hess}\,\Psi \geq K$ *holds for some* $K > 0$*, and take a unique minimizer* $x_0 \in M$ *of* $\Psi$*.*

(i) *If* $m < 1$ *and* $\mathrm{Ric}_N \geq 0$*, then* $\sigma \in L^c(M,\omega)$ *for all* $c \in (1/2, 1]$*. Moreover, we have*

$$\int_M \sigma^c \, d\omega \leq C_1(\omega)^{1-c}\nu(M)^c + C_2(m,c,\omega)K^{c/(m-1)}.$$

(ii) *If* $m < 1$ *and* $\mathrm{Ric}_N \geq 0$*, then* $\int_M d(x_0,x)^p \, d\nu < \infty$ *for all* $p \in [1, 1/(1-m))$*.*

(iii) *If* $m > 1$*, then* $M_0$ *and* $\mathrm{supp}\,\nu$ *are convex in the sense that any minimal geodesic connecting two points in* $M_0$ *or* $\mathrm{supp}\,\nu$ *is contained in* $M_0$ *or* $\mathrm{supp}\,\nu$*, respectively. In addition, we have*

$$\mathrm{supp}\,\nu \subset \overline{B\left(x_0, \left\{\frac{2}{K}\left(\frac{1}{m-1} - \Psi(x_0)\right)\right\}^{1/2}\right)}.$$

*Proof.* By our assumption $\mathrm{Hess}\,\Psi \geq K > 0$, we find a unique point $x_0 \in M_0$ such that $\Psi(x_0) = \inf_M \Psi$. Then we deduce from (2.3) that

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

holds for all minimal geodesics $\gamma$ with $\gamma(0) = x_0$. Thus we have

$$\sigma(x) = \exp_m\big(-\Psi(x)\big) \leq \exp_m\left(-\Psi(x_0) - \frac{K}{2}d(x_0,x)^2\right) \tag{2.4}$$

for all $x \in M_0$.

(i) Denote by $\mathrm{area}_\omega(S(x_0,r))$ the area of the sphere $S(x_0,r) := \{x \in M \mid d(x_0,x) = r\}$ with respect to $\omega$. Then (2.4) implies

$$\int_M \sigma^c \, d\omega \leq \int_{B(x_0,1)} \sigma^c \, d\omega + \int_1^\infty \exp_m\left(-\Psi(x_0) - \frac{K}{2}r^2\right)^c \mathrm{area}_\omega\big(S(x_0,r)\big)\, dr.$$

On the one hand, if follows from $\mathrm{Ric}_N \geq 0$ that, for $r \geq 1$,

$$\mathrm{area}_\omega\big(S(x_0,r)\big) \leq r^{N-1}\,\mathrm{area}_\omega\big(S(x_0,1)\big) = r^{m/(1-m)}\,\mathrm{area}_\omega\big(S(x_0,1)\big).$$

Therefore we obtain, putting $a := \exp_m(-\Psi(x_0))^{m-1} > 0$,

$$\begin{aligned}
&\int_1^\infty \exp_m\bigg(-\Psi(x_0) - \frac{K}{2}r^2\bigg)^c \mathrm{area}_\omega\big(S(x_0,r)\big)\,dr \\
&\leq \mathrm{area}_\omega\big(S(x_0,1)\big)\int_1^\infty \bigg\{a + (1-m)\frac{K}{2}r^2\bigg\}^{c/(m-1)} r^{m/(1-m)}\,dr \\
&= \mathrm{area}_\omega\big(S(x_0,1)\big)\int_1^\infty \bigg\{ar^{-2} + (1-m)\frac{K}{2}\bigg\}^{c/(m-1)} r^{(m-2c)/(1-m)}\,dr \\
&\leq \mathrm{area}_\omega\big(S(x_0,1)\big)\bigg\{(1-m)\frac{K}{2}\bigg\}^{c/(m-1)}\int_1^\infty r^{(m-2c)/(1-m)}\,dr \\
&= \mathrm{area}_\omega\big(S(x_0,1)\big)\frac{(1-m)^{c/(m-1)+1}}{2c-1}\bigg(\frac{K}{2}\bigg)^{c/(m-1)} \\
&=: C_2(m,c,\omega)K^{c/(m-1)} < \infty.
\end{aligned}$$

On the other hand, as $\nu(M) < \infty$ is already observed, the Hölder inequality yields

$$\begin{aligned}
\int_{B(x_0,1)} \sigma^c\,d\omega &\leq \bigg(\int_{B(x_0,1)} \sigma\,d\omega\bigg)^c \omega\big(B(x_0,1)\big)^{1-c} \leq \nu(M)^c\omega\big(B(x_0,1)\big)^{1-c} \\
&=: C_1(\omega)^{1-c}\nu(M)^c.
\end{aligned}$$

(ii) We similarly deduce that

$$\begin{aligned}
&\int_{M\setminus B(x_0,1)} d(x_0,x)^p\,d\nu(x) \\
&\leq \int_1^\infty r^p \exp_m\bigg(-\Psi(x_0) - \frac{K}{2}r^2\bigg)\mathrm{area}_\omega\big(S(x_0,r)\big)\,dr \\
&\leq \mathrm{area}_\omega\big(S(x_0,1)\big)\bigg\{(1-m)\frac{K}{2}\bigg\}^{1/(m-1)}\int_1^\infty r^{p+(m-2)/(1-m)}\,dr \\
&= \mathrm{area}_\omega\big(S(x_0,1)\big)\frac{(1-m)^{m/(m-1)}}{1-(1-m)p}\bigg(\frac{K}{2}\bigg)^{1/(m-1)} \\
&< \infty.
\end{aligned}$$

(iii) Recall that $\mathrm{supp}\,\nu = \overline{\Psi^{-1}((-\infty, 1/(m-1)))}$. Therefore $M_0$ and $\mathrm{supp}\,\nu$ are convex and (2.4) shows the desired estimate. $\Box$

**Example 2.5 ($m$-Gaussian measures)** One fundamental and important example is the *$m$-Gaussian measure* on $\mathbb{R}^n$ defined by

$$N_m(v,V) = \sigma dx := C_0(\det V)^{-1/2}\exp_m\bigg[-\frac{C_1}{2}\langle x-v, V^{-1}(x-v)\rangle\bigg]dx, \tag{2.5}$$

where $dx$ is the Lebesgue measure, $v \in \mathbb{R}^n$ is the mean, $V \in \mathrm{Sym}^+(n, \mathbb{R})$ is the covariance matrix, and $C_0, C_1$ are positive constants depending only on $n$ and $m$ (see [Ta2]). Then clearly $\mathrm{Hess}\,\Psi = C_0^{m-1}(\det V)^{(1-m)/2} \cdot C_1 V^{-1}$ and hence

$$\mathrm{Hess}\,\Psi \geq C_0^{m-1} C_1 (\det V)^{(1-m)/2} \Lambda^{-1}$$

holds by taking Remark 2.3 into account, where $\Lambda$ denotes the largest eigenvalue of $V$. Note that $N_m(v, V)$ has unbounded and bounded support for $m < 1$ and $m > 1$, respectively. The family of $m$-Gaussian measures will play interesting roles also in Sections 3, 5, 7.

## 2.3 Wasserstein geometry

We very briefly recall optimal transport theory and Wasserstein geometry. We refer to [Vi1], [Vi2] for basics as well as recent diverse development of them.

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 $\gamma$ is *minimal* if it is globally minimizing (i.e., $d(\gamma(s), \gamma(t)) = |s - t| d(\gamma(0), \gamma(1))$ for all $s, t \in [0, 1]$). If any two points in $X$ is connected by a minimal geodesic, then $(X, d)$ is called a *geodesic space*.

We denote by $\mathcal{P}(X)$ the set of all Borel probability measures on $X$, and by $\mathcal{P}^p(X) \subset \mathcal{P}(X)$ with $p \geq 1$ the subset consisting of measures $\mu$ of finite $p$-th moment, that is, $\int_X d(x, y)^p \, d\mu(y) < \infty$ for some (and hence all) $x \in X$. Given $\mu, \nu \in \mathcal{P}(X)$, a probability measure $\pi \in \mathcal{P}(X \times X)$ is called a *coupling* of $\mu$ and $\nu$ if its projections are $\mu, \nu$, namely $\pi(A \times X) = \mu(A)$ and $\pi(X \times A) = \nu(A)$ hold for any measurable set $A \subset X$. We define the *$L^p$-Wasserstein distance* between $\mu, \nu \in \mathcal{P}^p(X)$ by

$$W_p(\mu, \nu) := \inf_{\pi} \left( \int_{X \times X} d(x, y)^p \, d\pi(x, y) \right)^{1/p},$$

where $\pi$ runs over all couplings of $\mu$ and $\nu$. We call $\pi$ an *optimal coupling* if it attains the infimum above. We remark that $W_p(\mu, \nu)$ is finite since $\mu, \nu \in \mathcal{P}^p(X)$, and it is indeed a distance of $X$ if $X$ is complete and separable. Then the metric space $(\mathcal{P}^p(X), W_p)$ is called the *$L^p$-Wasserstein space* over $X$. If $X$ is compact, then $(\mathcal{P}^p(X), W_p)$ is also compact and the topology induced from $W_p$ coincides with the weak topology.

We will consider only the case of $p = 2$ that is suitable and important for applications in Riemannian geometry. A minimal geodesic between $\mu, \nu \in \mathcal{P}^2(X)$ amounts to an optimal way of transporting $\mu$ to $\nu$. Then it is natural to expect that such an optimal transport is performed along minimal geodesics in $X$, that is indeed the case as seen in the following proposition. We denote by $\Gamma(X)$ the set of all minimal geodesics $\gamma : [0, 1] \longrightarrow X$ endowed with the topology induced from the distance $d_{\Gamma(X)}(\gamma, \eta) := \sup_{t \in [0,1]} d(\gamma(t), \eta(t))$. For $t \in [0, 1]$, define the *evaluation map* $e_t : \Gamma(X) \longrightarrow X$ as $e_t(\gamma) := \gamma(t)$ and observe that it is 1-Lipschitz.

**Proposition 2.6** ([Vi2, Corollary 7.22]) *Let $(X, d)$ be a locally compact geodesic space. Then, for any $\mu, \nu \in \mathcal{P}^2(X)$ and any minimal geodesic $\alpha : [0, 1] \longrightarrow \mathcal{P}^2(X)$ between them, there exists $\Pi \in \mathcal{P}(\Gamma(X))$ such that $(e_0 \times e_1)_\sharp \Pi$ is an optimal coupling of $\mu$ and $\nu$ and that $(e_t)_\sharp \Pi = \alpha(t)$ holds for all $t \in [0, 1]$.*

We denoted by $(e_t)_\sharp \Pi$ the push-forward measure of $\Pi$ by $e_t$. In Riemannian manifolds, a more precise description of an optimal transport using a gradient vector field of some kind of convex function is known. We first recall McCann's original work on compact Riemannian manifolds. Denote by $\mathcal{P}_{\mathrm{ac}}(M, \mathrm{vol}_g) \subset \mathcal{P}(M)$ the set of absolutely continuous measures with respect to the volume measure $\mathrm{vol}_g$.

**Theorem 2.7** ([Mc2, Theorems 8, 9]) *Let $(M, g)$ be a compact Riemannian manifold. Then, for any $\mu \in \mathcal{P}_{\mathrm{ac}}(M, \mathrm{vol}_g)$ and $\nu \in \mathcal{P}(M)$, there exists a $(d^2/2)$-convex function $\varphi : M \longrightarrow \mathbb{R}$ such that the map $\mathcal{T}_t(x) := \exp_x(t\nabla\varphi(x))$, $t \in [0, 1]$, provides a unique minimal geodesic from $\mu$ to $\nu$. To be precise, $(\mathcal{T}_0 \times \mathcal{T}_1)_\sharp \mu$ is an optimal coupling of $\mu$ and $\nu$, and $\mu_t = (\mathcal{T}_t)_\sharp \mu$ is a minimal geodesic from $\mu_0 = \mu$ to $\mu_1 = \nu$.*

See [Vi2, Chapter 5] for the definition of the $(d^2/2)$-*convex function*, here we just remark that it is locally semi-convex in compact spaces. Such convexity is important as it implies the almost everywhere twice differentiability, and is generalized to noncompact spaces in [FG].

**Theorem 2.8** ([FG, Theorem 1]) *Let $(M, g)$ be a complete Riemannian manifold. Then, for any $\mu \in \mathcal{P}^2_{\mathrm{ac}}(M, \mathrm{vol}_g)$ and $\nu \in \mathcal{P}^2(M)$, there exists a locally semi-convex function $\varphi : M \longrightarrow \mathbb{R}$ such that the map $\mathcal{T}_t(x) := \exp_x(t\nabla\varphi(x))$, $t \in [0, 1]$, provides a unique minimal geodesic from $\mu$ to $\nu$.*

# 3 Generalized entropies

Before discussing the $m$-relative entropy, we briefly review the Boltzmann and the Tsallis entropies (see [Ts1], [Ts2]), and explain the motivation related to information geometry (see [Am], [AN]). The readers familiar with (or not interested in) these can skip to Subsection 3.2.

## 3.1 Background: Tsallis entropy and information geometry

Entropy is a functional playing prominent roles in thermodynamics, information theory (sometimes with the opposite sign) and many other fields. It describes how particles diffuse in thermodynamics, and measures the uncertainty of an event in information theory. The most fundamental entropy is the *Boltzmann(-Gibbs-Shannon) entropy* given by

$$E(\mu) = -\int_{\mathbb{R}^n} \rho \ln \rho \, dx$$

for $\mu = \rho dx \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, dx)$, where $dx$ is the Lebesgue measure.

Boltzmann entropy is thermodynamically *extensive* (and probabilistically *additive*) and suitable for the treatment of independent systems. Recently, there is a growing interest in strongly correlated systems and *nonextensive* entropies. Among them, we consider the *Tsallis entropy* defined by

$$E_m(\mu) := -\int_{\mathbb{R}^n} e_m(\rho) \, dx = -\int_{\mathbb{R}^n} \rho \ln_m \rho \, dx = -\int_{\mathbb{R}^n} \frac{\rho^m - \rho}{m - 1} \, dx \tag{3.1}$$

for $\mu = \rho dx \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, dx)$, where $m \in [(n-1)/n, 1) \cup (1, 2]$. Note that letting $m$ tend to 1 recovers the Boltzmann entropy $E(\mu)$. One can connect $E$ and $E_m$ via Gaussian measures as follows. On the one hand, given $v \in \mathbb{R}^n$ and $V \in \mathrm{Sym}^+(n, \mathbb{R})$, the (usual) Gaussian measure

$$N(v, V) = \frac{1}{(2\pi)^{n/2}} (\det V)^{-1/2} \exp\left[ -\frac{1}{2} \langle x - v, V^{-1}(x - v) \rangle \right] dx$$

maximizes $E$ among $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, dx)$ with mean $v$ and covariance matrix $V$. On the other hand, the $m$-Gaussian measure $N_m(v, V)$ defined in (2.5) similarly maximizes $E_{2-m}$ under the same constraint (for $m \neq 1/2, 2$).

In the following sections, we shall verify that a number of further geometric/analytic properties and applications of $E$ have counterparts for $E_m$. Precisely, since $E_m$ itself is not really interesting in our view (see Remark 4.3(2)), we modify $E_m$ in the manner of information geometry.

We start from the family of Gaussian measures

$$\mathcal{N}(n) := \{ N(v, V) \,|\, v \in \mathbb{R}^n,\ V \in \mathrm{Sym}^+(n, \mathbb{R}) \}$$

as an $((n^2 + 3n)/2)$-dimensional manifold. In information geometry, we equip $\mathcal{N}(n)$ with the *Fisher information metric* $m_F$ which is different from the Wasserstein metric $W_2$. In fact, $(\mathcal{N}(1), m_F)$ has the negative constant sectional curvature ([Am]) and $(\mathcal{N}(1), W_2)$ is flat (cf. [Ta1, Theorem 2.2] and the references therein). The Fisher metric admits a pair of dually flat connections (*exponential* and *mixture connections*) and the *Kullback-Leibler divergence*

$$H(\mu|\nu) = \int_{\mathbb{R}^n} \frac{\rho}{\sigma} \ln\left( \frac{\rho}{\sigma} \right) d\nu$$

for $\nu = \sigma dx \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, dx)$ and $\mu = \rho dx \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, \nu)$. Note that $H(\mu|\nu)$ is nonnegative by Jensen's inequality. The square root of the divergence $H(\mu|\nu)$ can be regarded as a kind of distance between $\mu$ and $\nu$. It certainly satisfies a generalized Pythagorean theorem, though it does not satisfy symmetry nor the triangle inequality. The Kullback-Leibler divergence $H(\mu|\nu)$ coincides with the *relative entropy* $\mathrm{Ent}_\nu(\mu)$ of $\mu$ with respect to $\nu$. More precisely, $\mathrm{Ent}_\nu(\mu)$ is defined for $\mu \in \mathcal{P}(\mathbb{R}^n)$ and Borel measure $\nu$ on $\mathbb{R}^n$ by

$$\mathrm{Ent}_\nu(\mu) := \begin{cases} \int_{\mathbb{R}^n} \varsigma \ln \varsigma \, d\nu & \text{for } \mu = \varsigma\nu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, \nu), \\ \infty & \text{otherwise}, \end{cases} \tag{3.2}$$

and then $\mathrm{Ent}_\nu(\mu) \geq -\ln \nu(M)$.

The family of $m$-Gaussian measures

$$\mathcal{N}(n, m) := \{ N_m(v, V) \,|\, v \in \mathbb{R}^n,\ V \in \mathrm{Sym}^+(n, \mathbb{R}) \}$$

similarly admits dually flat connections and the corresponding *Bregman divergence* (*$\beta$-divergence*) is

$$H_m(\mu|\nu) = \frac{1}{m(m-1)} \int_{\mathbb{R}^n} \{ \rho^m - m\rho\sigma^{m-1} + (m-1)\sigma^m \} \, dx \tag{3.3}$$

for $\nu = \sigma dx \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, dx)$ and $\mu = \rho dx \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^n, \nu)$. We rewrite this by using $e_m$ as

$$H_m(\mu|\nu) = \frac{1}{m} \int_{\mathbb{R}^n} \{e_m(\rho) - e_m(\sigma) - e_m'(\sigma)(\rho - \sigma)\} \, dx$$

and recover the Kullback-Leibler divergence as the limit:

$$\lim_{m \to 1} H_m(\mu|\nu) = \int_{\mathbb{R}^n} \{\rho \ln \rho - \sigma \ln \sigma - (\ln \sigma + 1)(\rho - \sigma)\} \, dx = H(\mu|\nu).$$

It will turn out that the entropy induced from (3.3) is appropriate for our purpose (see Theorem 4.1). We remark that the division by $m$ in (3.3) is unessential, we prefer this form merely for aesthetic reasons of the presentation of Theorem 4.1 as well as functional inequalities in Section 5.

### 3.2 $m$-relative entropy

Recall our weighted Riemannian manifold $(M, \omega)$ and reference measure $\nu = \sigma\omega$. The Bregman divergence (3.3) leads us to the following generalization of the relative entropy.

**Definition 3.1 ($m$-relative entropy)** Assume $\sigma \in L^m(M, \omega)$. Given $\mu \in \mathcal{P}(M)$, let $\mu = \rho\omega + \mu^s$ be its Lebesgue decomposition into absolutely continuous and singular parts with respect to $\omega$. Then we define the *$m$-relative entropy* by

$$\begin{aligned} &H_m(\mu|\nu) \\ &:= \frac{1}{m} \int_M \{e_m(\rho) - e_m(\sigma) - e_m'(\sigma)(\rho - \sigma)\} \, d\omega - \frac{1}{m-1} \int_M \sigma^{m-1} \, d\mu^s + H_m(\infty)\mu^s(M) \\ &= \frac{1}{m(m-1)} \int_M \{\rho^m + (m-1)\sigma^m\} \, d\omega - \frac{1}{m-1} \int_M \sigma^{m-1} \, d\mu + H_m(\infty)\mu^s(M) \qquad (3.4) \end{aligned}$$

if $\rho \in L^m(M, \omega)$, where $H_m(\infty) := 0$ for $m < 1$, $H_m(\infty) := \infty$ for $m > 1$, and $\infty \cdot 0 = 0$ as convention. We set $H_m(\mu|\nu) := \infty$ for $\mu \in \mathcal{P}(M)$ with $\rho \notin L^m(M, \omega)$.

For $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ with $\rho \in L^m(M, \omega)$, $H_m(\mu|\nu)$ has the simplified form

$$H_m(\mu|\nu) = \frac{1}{m(m-1)} \int_M \{\rho^m - m\rho\sigma^{m-1} + (m-1)\sigma^m\} \, d\omega$$

as in (3.3). Note that this is indeed well-defined.

**Remark 3.2** (1-a) For $m < 1$, moreover, the Hölder inequality implies

$$\int_M \rho^m \, d\omega = \int_M (\rho\sigma^{m-1})^m \sigma^{m(1-m)} \, d\omega \le \left( \int_M \rho\sigma^{m-1} \, d\omega \right)^m \left( \int_M \sigma^m \, d\omega \right)^{1-m}.$$

Thus we have, for $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$,

$$\begin{aligned} &H_m(\mu|\nu) - \frac{1}{m} \int_M \sigma^m \, d\omega \\ &\ge \frac{1}{m(m-1)} \int_M \rho^m \, d\omega + \frac{1}{1-m} \left( \int_M \rho^m \, d\omega \right)^{1/m} \left( \int_M \sigma^m \, d\omega \right)^{(m-1)/m} \\ &= \frac{1}{m(1-m)} \int_M \rho^m \, d\omega \cdot \left\{ m \left( \int_M \rho^m \, d\omega \right)^{(1-m)/m} \left( \int_M \sigma^m \, d\omega \right)^{(m-1)/m} - 1 \right\}, \end{aligned}$$

and it is natural to define $H_m(\mu|\nu) = \infty$ for $\rho \notin L^m(M, \omega)$.

(1-b) For $m > 1$, the Hölder inequality

$$\int_M \rho\sigma^{m-1}\, d\omega \le \left(\int_M \rho^m\, d\omega\right)^{1/m} \left(\int_M \sigma^m\, d\omega\right)^{(m-1)/m}$$

similarly yields

$$\begin{aligned}&H_m(\mu|\nu) - \frac{1}{m}\int_M \sigma^m\, d\omega \\ &\ge \frac{1}{m(m-1)}\left(\int_M \rho^m\, d\omega\right)^{1/m}\left\{\left(\int_M \rho^m\, d\omega\right)^{(m-1)/m} - m\left(\int_M \sigma^m\, d\omega\right)^{(m-1)/m}\right\}.\end{aligned}$$

Hence it is again natural to set $H_m(\mu|\nu) = \infty$ for $\rho \notin L^m(M, \omega)$.

(2) The validity of the definition of $H_m(\infty)$ would be understood by the following observation (putting $\rho = \chi_{B(x,\varepsilon)}/\omega(B(x,\varepsilon))$ so that $\chi_{B(x,\varepsilon)}$ is the characteristic function of $B(x,\varepsilon)$):

$$\int_{B(x,\varepsilon)} \frac{1}{\omega(B(x,\varepsilon))^m}\, d\omega = \omega\big(B(x,\varepsilon)\big)^{1-m} \quad \to \quad \begin{cases} 0 & \text{if } m < 1, \\ \infty & \text{if } m > 1 \end{cases}$$

as $\varepsilon$ tends to zero.

We remark that the seemingly unessential term $m^{-1}\int_M \sigma^m\, d\omega$ was inserted in $H_m(\mu|\nu)$ for the sake of nonnegativity.

**Lemma 3.3** *We have $H_m(\mu|\nu) \ge 0$ for all $\mu \in \mathcal{P}(M)$, and the equality holds if and only if $\mu = \nu$.*

*Proof.* Note that, if $\mu^s(M) > 0$, then the singular part

$$-\frac{1}{m-1}\int_M \sigma^{m-1}\, d\mu^s + H_m(\infty)\mu^s(M)$$

in (3.4) is positive for $m < 1$ (since $\sigma > 0$ on $M$) and infinity for $m > 1$, respectively. Hence it is sufficient to consider the absolutely continuous part. As the function $e_m(t) = (t^m - t)/(m-1)$ is strictly convex on $(0, \infty)$, we have

$$e_m(\rho) - e_m(\sigma) - e_m'(\sigma)(\rho - \sigma) \ge 0$$

in (3.4) and equality holds if and only if $\rho = \sigma$. Therefore $H_m(\mu|\nu) \ge 0$ and equality holds if and only if $\mu^s(M) = 0$ and $\rho = \sigma$ a.e. $\square$

The following lemma will be used in Section 7 (Claim 7.7) where $M$ is assumed to be compact. This also guarantees the validity of $H_m(\infty)$.

**Lemma 3.4** *Let $(M, g)$ be compact. Then the entropy $H_m(\cdot|\nu)$ 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*

$$H_m(\mu|\nu) \le \liminf_{i\to\infty} H_m(\mu_i|\nu).$$

*Proof.* We divide $H_m(\mu|\nu) - m^{-1}\int_M \sigma^m \, d\omega$ into two parts:

$$h_1(\mu) := \frac{1}{m(m-1)} \int_M \rho^m \, d\omega + H_m(\infty)\mu^s(M), \quad h_2(\mu) := -\frac{1}{m-1}\int_M \sigma^{m-1} \, d\mu.$$

Then $h_2(\mu)$ is clearly continuous in $\mu$. In addition, the lower semi-continuity of $h_1(\mu)$ follows from [LV2, Theorem B.33] since the function $U_m(t) := t^m/m(m-1)$ is continuous, convex and satisfies $U_m(0) = 0$ as well as $\lim_{t\to\infty} U_m(t)/t = H_m(\infty)$. $\square$

# 4 Displacement convexity

In this section, we prove our first main theorem on a characterization of the displacement convexity of $H_m(\cdot|\nu)$ along the lines of [CMS], [vRS] and [St2]. This should be compared with the equivalence between $\mathrm{Ric}_N \ge K$ and $\mathsf{CD}(K,N)$ for $(M,\omega)$ ([St2, Theorem 1.7], [LV1, Theorem 4.22]). Such a characterization has motivated the general theory of metric measure spaces satisfying $\mathsf{CD}(K,N)$, as what is called a synthetic lower Ricci curvature bound ([St1], [St2], [LV1], [LV2], [Vi2, Part III]). Recall that $M_0 = M$ for $m < 1$, $M_0 = \Psi^{-1}((-\infty, 1/(m-1)))$ for $m > 1$, and that $\overline{M_0} = \operatorname{supp}\nu$ in both cases.

**Theorem 4.1** *Let $(M,\omega,\nu)$ and $m \in [(n-1)/n, 1) \cup (1,\infty)$ with $\sigma \in L^m(M,\omega)$ be given. Then, for $K \in \mathbb{R}$, the following three conditions are mutually equivalent:*

(A) *We have $\mathrm{Ric}_N \ge 0$ on $M_0$ with $N = 1/(1-m)$ as well as $\operatorname{Hess}\Psi \ge K$ on $M_0$ in the weak sense.*

(B) *For any $\mu_0, \mu_1 \in \mathcal{P}^2_{\mathrm{ac}}(M,\omega)$ such that $\operatorname{supp}\mu_0, \operatorname{supp}\mu_1 \subset M_0$ and that any two points $x_0 \in \operatorname{supp}\mu_0$, $x_1 \in \operatorname{supp}\mu_1$ are joined by some geodesic contained in $M_0$, there is a minimal geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2_{\mathrm{ac}}(M_0,\omega)$ along which we have*

$$H_m(\mu_t|\nu) \le (1-t)H_m(\mu_0|\nu) + tH_m(\mu_1|\nu) - \frac{K}{2}(1-t)tW_2(\mu_0,\mu_1)^2 \tag{4.1}$$

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

(C) *For any $\mu_0, \mu_1 \in \mathcal{P}^2(M)$ such that $\operatorname{supp}\mu_0, \operatorname{supp}\mu_1 \subset M_0$ and that any two points $x_0 \in \operatorname{supp}\mu_0$, $x_1 \in \operatorname{supp}\mu_1$ are joined by some geodesic contained in $M_0$, there is a minimal geodesic $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2(M_0)$ along which we have (4.1) for all $t \in [0,1]$.*

*Proof.* Note that (C) $\Rightarrow$ (B) is clear. Thus we will show (A) $\Rightarrow$ (B), (A) $\Rightarrow$ (C) and finally (B) $\Rightarrow$ (A). The part (A) $\Rightarrow$ (C) is somewhat technical and may be skipped at the first reading.

(A) $\Rightarrow$ (B): Since the assertion (4.1) is clear if $H_m(\mu_0|\nu) = \infty$ or $H_m(\mu_1|\nu) = \infty$, we assume that both $H_m(\mu_0|\nu)$ and $H_m(\mu_1|\nu)$ are finite. Theorem 2.8 ensures that there is an almost everywhere twice differentiable function $\varphi : M \longrightarrow \mathbb{R}$ such that the map $\mathcal{T}_t(x) := \exp_x(t\nabla\varphi(x))$ gives the unique minimal geodesic $\mu_t := (\mathcal{T}_t)_\sharp\mu_0$ from $\mu_0$ to $\mu_1$. Due to [CMS, Proposition 4.1], $\mathcal{T}_1(x)$ is not a cut point of $x$ for $\mu_0$-a.e. $x$, and hence the minimal geodesic $(\mathcal{T}_t(x))_{t\in[0,1]}$ is unique and contained in $M_0$. Note also that, putting

$\mu_t = \rho_t\omega$ and $\mathbf{J}_t^\omega(x) := e^{\psi(x)-\psi(\mathcal{T}_t(x))}\det(D\mathcal{T}_t(x))$, the Jacobian (Monge-Amperè) equation $(\rho_t \circ \mathcal{T}_t)\mathbf{J}_t^\omega = \rho_0$ holds $\mu_0$-a.e. ([CMS, Theorem 4.2]) Recall that

$$H_m(\mu_t|\nu) = \frac{1}{m(m-1)}\int_M (\rho_t^{m-1} - m\sigma^{m-1})\rho_t\, d\omega + \frac{1}{m}\int_M \sigma^m\, d\omega.$$

By the change of variables formula, we deduce that

$$\begin{aligned}\int_M (\rho_t^{m-1} - m\sigma^{m-1})\rho_t\, d\omega &= \int_M \{\rho_t(\mathcal{T}_t)^{m-1} - m\sigma(\mathcal{T}_t)^{m-1}\}\rho_t(\mathcal{T}_t)\mathbf{J}_t^\omega\, d\omega \\ &= \int_M \left\{\left(\frac{\mathbf{J}_t^\omega}{\rho_0}\right)^{1-m} - m\sigma(\mathcal{T}_t)^{m-1}\right\} d\mu_0.\end{aligned}$$

**Claim 4.2** *For $\mu_0$-a.e. $x \in M$, the function $\mathbf{J}_t^\omega(x)^{1-m}/(m-1) = -N\mathbf{J}_t^\omega(x)^{1/N}$ is convex in $t$.*

*Proof.* For $m < 1$ (and hence $N \ge n$), this is proved in [St2, Theorem 1.7] (see also [Oh2, Section 8.2]). We can apply the same calculation to $m > 1$ (and $N < 0$). Indeed, with the notations in [Oh2, Section 8.2], we observe that $\mathrm{Ric}_N \ge 0$ implies $(N-1)h_3''h_3^{-1} \le 0$. Thus $h_3$ is convex and $e^\beta$ is concave, therefore

$$\{e^{-\psi(x)}\mathbf{J}_t^\omega(x)\}^{1/N} = h(t) = (e^{\beta(t)})^{1/N}h_3(t)^{(N-1)/N}$$

is convex in $t$ (just calculate $h''$ or use the Hölder inequality). $\diamondsuit$

In order to estimate the term $\sigma(\mathcal{T}_t)^{m-1}/(1-m)$, we observe

$$\begin{aligned}\frac{\sigma(\mathcal{T}_t)^{m-1}}{1-m} &= \frac{1}{1-m} + \Psi(\mathcal{T}_t) \\ &\le \frac{1}{1-m} + (1-t)\Psi(\mathcal{T}_0) + t\Psi(\mathcal{T}_1) - \frac{K}{2}(1-t)td(\mathcal{T}_0,\mathcal{T}_1)^2 \\ &= (1-t)\frac{\sigma(\mathcal{T}_0)^{m-1}}{1-m} + t\frac{\sigma(\mathcal{T}_1)^{m-1}}{1-m} - \frac{K}{2}(1-t)td(\mathcal{T}_0,\mathcal{T}_1)^2.\end{aligned}$$

Combining this with Claim 4.2 and integrating with $\mu_0$ yield the desired inequality (4.1).

(A) $\Rightarrow$ (C): Suppose that $\mu_0$ or $\mu_1$ has nontrivial singular part. There is nothing to prove for $m > 1$. For $m < 1$, we decompose as $\mu_0 = \rho_0\omega + \mu_0^s$ and $\mu_1 = \rho_1\omega + \mu_1^s$, and take an optimal coupling $\pi$ of $\mu_0$ and $\mu_1$. Now, $\pi$ 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. Here $p_1, p_2 : M \times M \longrightarrow M$ denote projections to the first and second elements.

We divide optimal transport between $\mu_0$ and $\mu_1$ into two parts, corresponding to $\pi - \pi_{ss}$ and $\pi_{ss}$. As for $\hat{\mu}_0 := (p_1)_\sharp(\pi - \pi_{ss})$ and $\hat{\mu}_1 := (p_2)_\sharp(\pi - \pi_{ss})$, Theorem 2.8 is again applicable and gives a minimal geodesic $\hat{\mu}_t = \hat{\rho}_t\omega \in (1 - \pi_{ss}(M \times M)) \cdot \mathcal{P}^2_{\mathrm{ac}}(M_0, \omega)$ (i.e.,

$\hat{\mu}_t(M) = 1 - \pi_{ss}(M \times M)$) satisfying

$$\int_M \hat{\rho}_t^m \, d\omega \geq (1-t) \int_M \rho_0^m \, d\omega + t \int_M \rho_1^m \, d\omega,$$

$$\begin{aligned}\int_M \sigma^{m-1} \, d\hat{\mu}_t &\leq (1-t) \int_M \sigma^{m-1} \, d\hat{\mu}_0 + t \int_M \sigma^{m-1} \, d\hat{\mu}_1 \\ &\quad - \frac{(1-m)K}{2}(1-t)t \int_{M\times M} d(x,y)^2 \, d(\pi - \pi_{ss})(x,y).\end{aligned}$$

We then choose an arbitrary minimal geodesic $\tilde{\mu}_t = \tilde{\rho}_t \omega + \tilde{\mu}_t^s \in \pi_{ss}(M \times M) \cdot \mathcal{P}^2(M_0)$ from $\tilde{\mu}_0 := (p_1)_\sharp(\pi_{ss})$ to $\tilde{\mu}_1 := (p_2)_\sharp(\pi_{ss})$. Thanks to Proposition 2.6, $\tilde{\mu}_t$ is also realized through a family of geodesics in $M_0$, and hence $\operatorname{Hess} \Psi \geq K$ implies

$$\begin{aligned}\int_M \sigma^{m-1} \, d\tilde{\mu}_t &\leq (1-t) \int_M \sigma^{m-1} \, d\tilde{\mu}_0 + t \int_M \sigma^{m-1} \, d\tilde{\mu}_1 \\ &\quad - \frac{(1-m)K}{2}(1-t)t \int_{M\times M} d(x,y)^2 \, d\pi_{ss}(x,y).\end{aligned}$$

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

$$\begin{aligned}H_m(\mu_t|\nu) &= \frac{1}{m(m-1)} \int_M \{(\hat{\rho}_t + \tilde{\rho}_t)^m + (m-1)\sigma^m\} \, d\omega + \frac{1}{1-m} \int_M \sigma^{m-1} \, d\mu_t \\ &\leq \frac{1}{m(m-1)} \int_M \{\hat{\rho}_t^m + (m-1)\sigma^m\} \, d\omega + \frac{1}{1-m} \int_M \sigma^{m-1} \, d(\hat{\mu}_t + \tilde{\mu}_t) \\ &\leq (1-t)H_m(\mu_0|\nu) + tH_m(\mu_1|\nu) - \frac{K}{2}(1-t)tW_2(\mu_0,\mu_1)^2.\end{aligned}$$

(B) $\Rightarrow$ (A): We first consider the case of $m < 1$. Fix a unit vector $v \in T_xM$ with $x \in M_0$ and put $\gamma(t) := \exp_x(tv)$, $B_\pm := B(\gamma(\pm\delta), (1 \mp a\delta)\varepsilon)$ with a constant $a \in \mathbb{R}$ chosen later. Set

$$\mu_0 = \rho_0\omega := \frac{\chi_{B_-}}{\omega(B_-)}\omega, \quad \mu_1 = \rho_1\omega := \frac{\chi_{B_+}}{\omega(B_+)}\omega \tag{4.2}$$

for $0 < \varepsilon \ll \delta \ll 1$, where $\chi_A$ stands for the characteristic function of a set $A$. Let $(\mu_t)_{t\in[0,1]}$ be the unique optimal transport from $\mu_0$ to $\mu_1$. Recall that

$$H_m(\mu_t|\nu) - \frac{1}{m}\int_M \sigma^m \, d\omega = \frac{1}{m(m-1)} \int_M \{\rho_0^{m-1}(\mathbf{J}_t^\omega)^{1-m} - m\sigma(\mathcal{T}_t)^{m-1}\} \, d\mu_0. \tag{4.3}$$

Note that

$$\rho_0^{m-1} = \{c_n e^{-\psi(\gamma(-\delta))}(1+a\delta)^n\varepsilon^n + O(\varepsilon^{n+1})\}^{1-m}\chi_{B_-},$$

where $c_n$ denotes the volume of the unit ball in $\mathbb{R}^n$. Hence, since $1 - m > 0$, the leading term of (4.3) (as $\varepsilon \to 0$) is

$$\frac{1}{1-m}\int_M \sigma(\mathcal{T}_t)^{m-1} \, d\mu_0.$$

Thus we obtain from (4.1) that, by letting $\varepsilon$ go to zero,

$$\sigma\big(\gamma(0)\big)^{m-1} \leq \frac{\sigma(\gamma(-\delta))^{m-1} + \sigma(\gamma(\delta))^{m-1}}{2} - (1-m)\frac{K}{2}\delta^2.$$

This means that

$$\operatorname{Hess}\Psi = \frac{1}{1-m}\operatorname{Hess}(\sigma^{m-1}) \geq K$$

in the weak sense.

In order to show $\operatorname{Ric}_N(v) \geq 0$, we choose a point $y$ with $d(x,y) \gg \delta$ and modify $\mu_0$ and $\mu_1$ into

$$\tilde{\mu}_i := (1-\varepsilon^{n+1})\frac{\chi_{B(y,\delta)}}{\omega(B(y,\delta))}\omega + \varepsilon^{n+1}\mu_i \tag{4.4}$$

for $i = 0, 1$. Then $W_2(\tilde{\mu}_0, \tilde{\mu}_1) = \varepsilon^{(n+1)/2} \cdot W_2(\mu_0, \mu_1)$ and

$$\tilde{\mu}_t := (1-\varepsilon^{n+1})\frac{\chi_{B(y,\delta)}}{\omega(B(y,\delta))}\omega + \varepsilon^{n+1}\mu_t$$

is the unique optimal transport from $\tilde{\mu}_0$ to $\tilde{\mu}_1$, so that (4.3) is modified into

$$\begin{aligned}
&H_m(\tilde{\mu}_t|\nu) - \frac{1}{m}\int_M \sigma^m \, d\omega \\
&= \frac{\varepsilon^{n+1}}{m(m-1)}\int_M \{(\varepsilon^{n+1}\rho_0)^{m-1}(\mathbf{J}_t^\omega)^{1-m} - m\sigma(\mathcal{T}_t)^{m-1}\} \, d\mu_0 \\
&\quad + \frac{1}{m(m-1)}\frac{1-\varepsilon^{n+1}}{\omega(B(y,\delta))}\int_{B(y,\delta)} \left\{\left(\frac{1-\varepsilon^{n+1}}{\omega(B(y,\delta))}\right)^{m-1} - m\sigma^{m-1}\right\} d\omega.
\end{aligned}$$

We rewrite this as

$$\begin{aligned}
&H_m(\tilde{\mu}_t|\nu) - \frac{1}{m}\int_M \sigma^m \, d\omega \\
&\quad - \frac{1-\varepsilon^{n+1}}{m(m-1)}\left\{\left(\frac{1-\varepsilon^{n+1}}{\omega(B(y,\delta))}\right)^{m-1} - \frac{m}{\omega(B(y,\delta))}\int_{B(y,\delta)} \sigma^{m-1} \, d\omega\right\} \\
&= \frac{\varepsilon^{n+1}}{m(m-1)}\int_M \{(\varepsilon^{n+1}\rho_0)^{m-1}(\mathbf{J}_t^\omega)^{1-m} - m\sigma(\mathcal{T}_t)^{m-1}\} \, d\mu_0.
\end{aligned} \tag{4.5}$$

Since $(\varepsilon^{n+1}\rho_0)^{m-1} = \{c_n e^{-\psi(\gamma(-\delta))}(1+a\delta)^n\varepsilon^{-1} + O(1)\}^{1-m}$, the leading term of (4.5) is

$$\frac{\varepsilon^{m(n+1)}}{m(m-1)}\int_M \rho_0^{m-1}(\mathbf{J}_t^\omega)^{1-m} \, d\mu_0.$$

Therefore (4.1) and the Jacobian equation yield that

$$\begin{aligned}
\mathbf{J}_{1/2}^\omega\big(\gamma(-\delta)\big)^{1-m} &\geq \frac{1}{2}\left\{\mathbf{J}_0^\omega\big(\gamma(-\delta)\big)^{1-m} + \mathbf{J}_1^\omega\big(\gamma(-\delta)\big)^{1-m}\right\} \\
&= \frac{1}{2}\left\{1 + \left(\frac{1-a\delta}{1+a\delta}\right)^{n/N} e^{\{\psi(\gamma(-\delta))-\psi(\gamma(\delta))\}/N}\right\}.
\end{aligned}$$

As

$$\mathbf{J}_{1/2}^\omega\big(\gamma(-\delta)\big) = \left(\frac{1}{1+a\delta}\right)^n e^{\psi(\gamma(-\delta))-\psi(x)}\left\{1 + \frac{1}{2}\operatorname{Ric}(v)\delta^2 + O(\delta^3)\right\},$$

this together with the Taylor expansion shows

$$
\begin{aligned}
&1+\frac{1}{2N}\operatorname{Ric}(v)\delta^2 \\
&\ge \frac{1}{2}\Big\{(1+a\delta)^{n/N}e^{\{\psi(x)-\psi(\gamma(-\delta))\}/N}+(1-a\delta)^{n/N}e^{\{\psi(x)-\psi(\gamma(\delta))\}/N}\Big\}+O(\delta^3) \\
&=1+\frac{\delta^2}{2}\bigg[\frac{n}{N}\Big(\frac{n}{N}-1\Big)a^2-\bigg\{\frac{(\psi\circ\gamma)''(0)}{N}-\frac{(\psi\circ\gamma)'(0)^2}{N^2}\bigg\}+\frac{2na}{N}\frac{(\psi\circ\gamma)'(0)}{N}\bigg]+O(\delta^3) \\
&=1+\frac{\delta^2}{2N}\bigg\{-(\psi\circ\gamma)''(0)+\frac{n(n-N)}{N}a^2+\frac{2n(\psi\circ\gamma)'(0)}{N}a+\frac{(\psi\circ\gamma)'(0)^2}{N}\bigg\}+O(\delta^3).
\end{aligned}
$$

Therefore we obtain

$$
\operatorname{Ric}(v)+(\psi\circ\gamma)''(0)-\frac{n(n-N)}{N}a^2-\frac{2n(\psi\circ\gamma)'(0)}{N}a-\frac{(\psi\circ\gamma)'(0)^2}{N}\ge 0. \tag{4.6}
$$

If $N>n$, then choosing the minimizer $a=(\psi\circ\gamma)'(0)/(N-n)$ gives the desired curvature bound $\operatorname{Ric}_N(v)\ge 0$. If $N=n$, then we consider $a$ going to $\infty$ or $-\infty$ and find $(\psi\circ\gamma)'(0)=0$ and $\operatorname{Ric}_n(v)\ge 0$.

In the case of $m>1$, we use the same transport (4.2) and then the leading term of (4.3) changes into

$$
\frac{1}{m(m-1)}\int_M \rho_0^{m-1}(\mathbf{J}_t^{\omega})^{1-m}\,d\mu_0.
$$

Thus calculations as above yield (4.6) with $N<0$. We choose the minimizer $a=(\psi\circ\gamma)'(0)/(N-n)$ and find $\operatorname{Ric}_N(v)\ge 0$. Similarly, for the transport (4.4), the leading term of (4.5) changes into

$$
\frac{\varepsilon^{n+1}}{1-m}\int_M \sigma(\mathcal{T}_t)^{m-1}\,d\mu_0,
$$

and then (4.1) yields $\operatorname{Hess}\Psi=\operatorname{Hess}(\sigma^{m-1}/(1-m))\ge K$. $\square$

**Remark 4.3** (1) If we admit $m\in(0,(n-1)/n)$ and generalize $\operatorname{Ric}_N$ in (2.1) to $N\in(1,n)$, then Claim 4.2 is false. Moreover, as the coefficient of $a^2$ in (4.6) is negative, (4.1) is never satisfied (let $a\to\infty$).

(2) Note that the special case $\nu=\omega$ (i.e., $\Psi\equiv 0$) in Theorem 4.1 makes sense only for $K=0$. Then the assertion of Theorem 4.1 corresponds to the equivalence between $\operatorname{Ric}_N\ge 0$ and the convexity of the Rényi entropy $S_N$, i.e., the curvature-dimension condition $\mathsf{CD}(0,N)$ of $(M,\omega)$.

(3) In the case of $m=1$, two weights $\psi$ and $\Psi$ are synchronized as $\nu=e^{-\psi-\Psi}\operatorname{vol}_g$, and $\operatorname{Hess}\operatorname{Ent}_\nu\ge K$ (i.e., $\mathsf{CD}(K,\infty)$ for $(M,\nu)$) is equivalent to the single condition $\operatorname{Ric}+\operatorname{Hess}(\psi+\Psi)\ge K$. For $m\ne 1$, however, $\psi$ and $\Psi$ keep separate and they measure different phases of $(M,\omega,\nu)$, as indicated in Theorem 4.1.

# 5 Functional inequalities

Since Otto and Villani's celebrated work [OV], the displacement convexity of entropy-type functionals has played a significant role in the study of functional inequalities (and the concentration of measures). In this section, we follow the argument in [LV2, Section 6] (with $m = 1$) that the direct application of the displacement convexity of the entropy implies various functional inequalities. Our proofs use only fundamental properties of convex functions. In more analytic context, related results for $m \neq 1$ in the Euclidean spaces $(M, \omega) = (\mathbb{R}^n, dx)$ can be found in [AGK], [CGH] and [Ta2]. See especially [AGK, Section 4] and [CGH, Section 3] for various generalizations of the Talagrand (transport) inequality, logarithmic Sobolev (entropy-information) inequality, HWI inequality and the Poincaré inequality. The relation among these inequalities are also discussed there.

Throughout the section, we suppose that $m > 1/2$, $\mathrm{Ric}_N \geq 0$ and that $\mathrm{Hess}\,\Psi \geq K$ holds in the weak sense for some $K > 0$. Recall from Lemma 2.4(i), (iii) that $\nu(M) < \infty$ automatically follows from these hypotheses, so that the normalization gives

$$\bar{\nu} = \bar{\sigma}\omega = \exp_m(-\overline{\Psi})\omega := \nu(M)^{-1}\nu \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$$

with $\mathrm{Hess}\,\overline{\Psi} \geq \nu(M)^{1-m}K$ according to Remark 2.3. Moreover, Lemma 2.4(i), (ii) ensure that $\bar{\sigma} \in L^m(M, \omega)$ as well as $\bar{\nu} \in \mathcal{P}^2_{\mathrm{ac}}(M, \omega)$, for $m > 1/2$. Keeping this in mind, we will take $\nu$ with $\nu(M) = 1$ for simplicity.

**Proposition 5.1 (Talagrand inequality)** *Assume that* $m > 1/2$*,* $\nu(M) = 1$*,* $\mathrm{Ric}_N \geq 0$ *and that* $\mathrm{Hess}\,\Psi \geq K$ *holds for some* $K > 0$*. Then, for any* $\mu \in \mathcal{P}^2(M_0)$*, we have*

$$W_2(\mu, \nu) \leq \sqrt{\frac{2}{K} H_m(\mu|\nu)}.$$

*Proof.* Recall from Lemma 2.4(iii) that $M_0$ is convex. Let $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}^2(M_0)$ be an optimal transport from $\mu_0 = \mu$ to $\mu_1 = \nu$. It follows from (4.1) and $H_m(\nu|\nu) = 0$ that

$$H_m(\mu_t|\nu) \leq (1-t)H_m(\mu|\nu) - \frac{K}{2}(1-t)tW_2(\mu, \nu)^2. \tag{5.1}$$

Since $H_m(\mu_t|\nu) \geq 0$ (Lemma 3.3), we obtain $H_m(\mu|\nu) \geq (K/2)W_2(\mu, \nu)^2$ by dividing (5.1) with $1 - t$ and letting $t$ go to 1. $\square$

For $\mu = \rho\omega \in \mathcal{P}^2_{\mathrm{ac}}(M, \omega)$ such that $\rho$ is Lipschitz, we define the *$m$-relative Fisher information* by

$$I_m(\mu|\nu) := \frac{1}{m^2}\int_M \big|\nabla[e'_m(\rho) - e'_m(\sigma)]\big|^2 \rho\, d\omega = \frac{1}{(m-1)^2}\int_M \big|\nabla(\rho^{m-1} - \sigma^{m-1})\big|^2\, d\mu. \tag{5.2}$$

It will be demonstrated in Proposition 7.9 that $\sqrt{I_m(\mu|\nu)}$ is the absolute gradient of $H_m(\cdot|\nu)$ at $\mu$. Then it is natural to expect that the convexity of $H_m(\cdot|\nu)$ yields the following inequality.

**Theorem 5.2 (HWI and Logarithmic Sobolev inequalities)** *We assume that* $m > 1/2$*,* $\nu(M) = 1$*,* $\mathrm{Ric}_N \geq 0$ *and that* $\mathrm{Hess}\,\Psi \geq K$ *holds for some* $K > 0$*. Then, for any* $\mu = \rho\omega \in \mathcal{P}^2_{\mathrm{ac}}(M_0, \omega)$ *such that* $\rho$ *is Lipschitz, we have*

$$H_m(\mu|\nu) \leq \sqrt{I_m(\mu|\nu)} \cdot W_2(\mu, \nu) - \frac{K}{2}W_2(\mu, \nu)^2, \tag{5.3}$$

$$H_m(\mu|\nu) \leq \frac{1}{2K} I_m(\mu|\nu). \tag{5.4}$$

*Proof.* Let $\mu_t = \rho_t\omega \in \mathcal{P}^2_{\mathrm{ac}}(M_0,\omega)$, $t \in [0,1]$, be the optimal transport from $\mu_0 = \mu$ to $\mu_1 = \nu$ given by $\mu_t = \rho_t\omega = (\mathcal{T}_t)_\sharp\mu$ with $\mathcal{T}_t(x) = \exp_x(t\nabla\varphi(x))$, and put $H(t) := H_m(\mu_t|\nu)$. Then it follows from (5.1) that

$$H(0) \le \frac{H(0)-H(t)}{t} - \frac{K}{2}(1-t)W_2(\mu,\nu)^2. \tag{5.5}$$

We shall estimate the term

$$H(0)-H(t) = \frac{1}{m(m-1)}\int_M \{(\rho^m - \rho_t^m) - m(\rho-\rho_t)\sigma^{m-1}\}\,d\omega.$$

Since the function $f(s) := s^m/(m-1)$ is convex, we have

$$\frac{\rho^m - \rho_t^m}{m-1} \le f'(\rho)(\rho-\rho_t) = \frac{m}{m-1}\rho^{m-1}(\rho-\rho_t),$$

and hence

$$H(0)-H(t) \le \frac{1}{m-1}\int_M (\rho^{m-1}-\sigma^{m-1})(\rho-\rho_t)\,d\omega.$$

By the change of variables formula along with the Jacobian equation $\rho_t(\mathcal{T}_t)\mathrm{det}(D\mathcal{T}_t) = \rho$, we observe

$$\begin{aligned}\int_M (\rho^{m-1}-\sigma^{m-1})\rho_t\,d\omega &= \int_M \{\rho(\mathcal{T}_t)^{m-1} - \sigma(\mathcal{T}_t)^{m-1}\}\rho_t(\mathcal{T}_t)\mathrm{det}(D\mathcal{T}_t)\,d\omega \\ &= \int_M \{\rho(\mathcal{T}_t)^{m-1} - \sigma(\mathcal{T}_t)^{m-1}\}\,d\mu.\end{aligned}$$

This yields

$$H(0)-H(t) \le \frac{1}{m-1}\int_M \big\{(\rho^{m-1}-\sigma^{m-1}) - \big(\rho(\mathcal{T}_t)^{m-1} - \sigma(\mathcal{T}_t)^{m-1}\big)\big\}\,d\mu.$$

Thus we obtain

$$\begin{aligned}&\limsup_{t\to 0}\frac{H(0)-H(t)}{t} \le \frac{1}{|m-1|}\int_M |\nabla(\rho^{m-1}-\sigma^{m-1})|\cdot d(\mathcal{T}_0,\mathcal{T}_1)\,d\mu \\ &\le \frac{1}{|m-1|}\bigg(\int_M |\nabla(\rho^{m-1}-\sigma^{m-1})|^2\,d\mu\bigg)^{1/2}\bigg(\int_M d(\mathcal{T}_0,\mathcal{T}_1)^2\,d\mu\bigg)^{1/2} \\ &= \sqrt{I_m(\mu|\nu)}\cdot W_2(\mu,\nu).\end{aligned}$$

Combining this with (5.5), we conclude that

$$H_m(\mu|\nu) \le \sqrt{I_m(\mu|\nu)}\cdot W_2(\mu,\nu) - \frac{K}{2}W_2(\mu,\nu)^2 \le \frac{1}{2K}I_m(\mu|\nu).$$

□

**Remark 5.3** It is established in [Ta2] that, in the Euclidean space $(M,\omega) = (\mathbb{R}^n, dx)$, equality of both (5.3) and (5.4) is characterized by using $m$-Gaussian measures.

We finally show a kind of Poincaré inequality. Observe that putting $m = 1$ recovers the usual global Poincaré inequality $\int_M f^2 \, d\nu \le K^{-1} \int_M |\nabla f|^2 \, d\nu$.

**Proposition 5.4 (Global Poincaré inequality)** *Assume that* $m > 1/2$, $\nu(M) = 1$, $\mathrm{Ric}_N \ge 0$ *and that* $\mathrm{Hess}\,\Psi \ge K$ *holds for some* $K > 0$. *Then, for any Lipschitz function* $f : M_0 \longrightarrow \mathbb{R}$ *such that* $\int_{M_0} f \, d\nu = 0$, *we have*

$$\int_M f^2 \sigma^{m-1} \, d\nu \le \frac{1}{K} \int_M |\nabla (f \sigma^{m-1})|^2 \, d\nu.$$

*Proof.* We apply (5.4) to $\mu = \rho\omega := (1 + \varepsilon f)\sigma\omega$ for small $\varepsilon \ge 0$ and obtain

$$\frac{1}{m(m-1)} \int_M \{\rho^m - m\rho\sigma^{m-1} + (m-1)\sigma^m\} \, d\omega \le \frac{1}{2K} \frac{1}{(m-1)^2} \int_M |\nabla(\rho^{m-1} - \sigma^{m-1})|^2 \, d\mu.$$

On the one hand,

$$\begin{aligned} \rho^m - m\rho\sigma^{m-1} + (m-1)\sigma^m &= (1+\varepsilon f)^m \sigma^m - m(1+\varepsilon f)\sigma^m + (m-1)\sigma^m \\ &= \sigma^m \{(1+\varepsilon f)^m - 1 - m(\varepsilon f)\} \\ &= m(m-1)\sigma^m \frac{f^2}{2} \varepsilon^2 + O(\varepsilon^3). \end{aligned}$$

On the other hand,

$$\begin{aligned} |\nabla(\rho^{m-1} - \sigma^{m-1})|^2 &= \left| \nabla \left[ \left( (1+\varepsilon f)^{m-1} - 1 \right) \sigma^{m-1} \right] \right|^2 \\ &= |\nabla[(m-1) f \varepsilon \sigma^{m-1}] + O(\varepsilon^2)|^2 \\ &= (m-1)^2 \varepsilon^2 |\nabla (f \sigma^{m-1})|^2 + O(\varepsilon^3). \end{aligned}$$

Thus we have, letting $\varepsilon$ go to zero,

$$\int_M f^2 \sigma^m \, d\omega \le \frac{1}{K} \int_M |\nabla (f\sigma^{m-1})|^2 \, d\nu.$$

$\square$

# 6 Concentration of measures

This section is devoted to an application of Proposition 5.1 to the concentration of measures. Let us assume $\nu(M) = 1$ and define the *concentration function* by

$$\alpha_{(M,\nu)}(r) := \sup \left\{ 1 - \nu\big(B(A,r)\big) \mid A \subset M, \ \nu(A) \ge 1/2 \right\}$$

for $r > 0$, where $A$ is any measurable set and

$$B(A, r) := \{ y \in M \mid \inf_{x \in A} d(x,y) < r \}.$$

The function $\alpha_{(M,\nu)}$ describes how the probability measure $\nu$ concentrates on the neighborhood of an arbitrary set of half the total measure in a quantitative way (in other words, a kind of large deviation principle). An especially interesting situation is that a sequence $\{(M_i, \nu_i)\}_{i\in\mathbb{N}}$ satisfies $\lim_{i\to\infty} \alpha_{(M_i,\nu_i)}(r) = 0$ for all $r > 0$, that means that $(M_i, \nu_i)$ is getting more and more concentrated. We refer to [Le] for the basic theory and applications of the concentration of measure phenomenon.

In the classical case of $m = 1$, it is well-known that the concentration of measures has rich connections with functional inequalities appearing in Section 5. For instance, the $L^1$-transport inequality $W_1(\mu, \nu) \le \sqrt{(2/K) \operatorname{Ent}_\nu(\mu)}$ implies the *normal concentration* $\alpha(r) \le Ce^{-cr^2}$ with constants $c, C > 0$ depending only on $K$ ([Le, Section 6.1]). In the same spirit, we show that an application of Proposition 5.1 gives new examples of concentrating spaces.

We set $G_c = G_c(\nu) := \int_M \sigma^c \, d\omega$ for $c > 1/2$. Recall from Lemma 2.4(i) that, if $\operatorname{Hess} \Psi \ge K > 0$,

$$G_c(\nu) \le C_1(\omega)^{1-c} \nu(M)^c + C_2(m, c, \omega) K^{c/(m-1)} < \infty \tag{6.1}$$

holds for $m < 1$ and $c \in (1/2, 1]$.

**Theorem 6.1** (**$m < 1$ case**) *Let $(M, \omega)$ satisfy $\operatorname{Ric}_N \ge 0$ and $m \in [(n-1)/n, 1) \cap (1/2, 1)$.*

(i) *Assume that $\nu(M) = 1$ and $\operatorname{Hess} \Psi \ge K$ holds for some $K > 0$. Then we have*

$$\alpha_{(M,\nu)}(r)^{\theta-m} \ln_m \big(2\alpha_{(M,\nu)}(r)\big) \le -G^{\theta-1}_{(m-\theta)/(1-\theta)} \left\{ \left( \sqrt{\frac{mK}{2}} r - \sqrt{G_m} \right)^2 - G_m \right\} \tag{6.2}$$

*for all $r > 0$ and $\theta \in [0, 2m-1)$.*

(ii) *Take a sequence $\nu_i = \exp_m(-\Psi_i)\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$, $i \in \mathbb{N}$, such that $\operatorname{Hess} \Psi_i \ge K_i$ and $\lim_{i\to\infty} K_i = \infty$. Then we have $\lim_{i\to\infty} \alpha_{(M,\nu_i)}(r) = 0$ for any $r > 0$.*

*Proof.* (i) Note that $\nu \in \mathcal{P}^2_{\mathrm{ac}}(M, \omega)$ by Lemma 2.4(ii). We also remark that (6.2) clearly holds for $r \le 2\sqrt{2G_m/mK}$. Indeed, then the right-hand side is nonnegative, while $\alpha_{(M,\nu)}(r) \le 1/2$ implies $\ln_m(2\alpha_{(M,\nu)}(r)) \le 0$.

Suppose $r > 2\sqrt{2G_m/mK}$, take a measurable set $A \subset M$ with $\nu(A) \ge 1/2$ and put $B := M \setminus B(A, r)$, $a := \nu(A)$, $b := \nu(B)$,

$$\mu_A := \frac{\chi_A}{a} \nu, \qquad \mu_B := \frac{\chi_B}{b} \nu.$$

We assumed $b > 0$ since there is nothing to prove if $b = 0$ for all such $A$. The triangle inequality of $W_1$ and Proposition 5.1 together imply (since $W_1 \le W_2$ by the Schwarz inequality)

$$r \le W_1(\mu_A, \mu_B) \le W_1(\mu_A, \nu) + W_1(\nu, \mu_B) \le \sqrt{\frac{2}{K} H_m(\mu_A|\nu)} + \sqrt{\frac{2}{K} H_m(\mu_B|\nu)}.$$

Note that

$$H_m(\mu_A|\nu) = \frac{1}{m(1-m)} \int_A \frac{ma^{m-1} - 1}{a^m} \sigma^m \, d\omega + \frac{1}{m} G_m$$

and $ma^{m-1}-1<0$ since $a \ge 1/2 > m^{1/(1-m)}$. Thus we obtain

$$\sqrt{\frac{mK}{2}}r \le \sqrt{G_m} + \sqrt{G_m + b^{-m}\frac{mb^{m-1}-1}{1-m}\int_B \sigma^m \, d\omega}.$$

We observe from $r > 2\sqrt{2G_m/mK}$ that $\sqrt{mK/2}r > 2\sqrt{G_m}$ which yields $0 < mb^{m-1}-1 < (2b)^{m-1}-1$. Hence we have

$$\left(\sqrt{\frac{mK}{2}}r - \sqrt{G_m}\right)^2 - G_m \le -b^{-m}\ln_m(2b)\int_B \sigma^m \, d\omega. \tag{6.3}$$

It follows from the Hölder inequality that

$$\int_B \sigma^m \, d\omega = \int_B \sigma^{\theta+(m-\theta)} \, d\omega \le \left(\int_B \sigma \, d\omega\right)^\theta \left(\int_B \sigma^{(m-\theta)/(1-\theta)} \, d\omega\right)^{1-\theta} \le b^\theta G^{1-\theta}_{(m-\theta)/(1-\theta)},$$

where the assumption $\theta < 2m-1$ ensures $(m-\theta)/(1-\theta) > 1/2$. Therefore we obtain the desired inequality (6.2) by choosing $A_i \subset M$ such that $\lim_{i\to\infty}\nu(M\setminus B(A_i,r)) = \alpha_{(M,\nu)}(r)$.

(ii) Thanks to (6.1), we know that

$$\limsup_{i\to\infty} G_c(\nu_i) \le C_1(\omega)^{1-c} < \infty$$

for all $c \in (1/2,1]$. Therefore we deduce from (i) that, setting $\alpha_i := \alpha_{(M,\nu_i)}(r)$,

$$\lim_{i\to\infty} \alpha_i^{\theta-m}\ln_m(2\alpha_i) = -\lim_{i\to\infty}\frac{\alpha_i^{\theta-1}}{2^{1-m}}\frac{1-(2\alpha_i)^{1-m}}{1-m} = -\infty$$

which shows $\lim_{i\to\infty}\alpha_{(M,\nu_i)}(r) = 0$. $\square$

**Remark 6.2** (1) Taking the proof of Lemma 2.4(i) into account, we can generalize Theorem 6.1(ii) as follows. Suppose that a sequence $\{(M_i,\omega_i,\nu_i)\}_{i\in\mathbb{N}}$ satisfies

(a) $\mathrm{Ric}_N \ge 0$ for all $(M_i,\omega_i)$,

(b) $\nu_i = \exp_m(-\Psi_i)\omega_i \in \mathcal{P}_{\mathrm{ac}}(M_i,\omega_i)$ so that $\mathrm{Hess}\,\Psi_i \ge K_i$ and $\lim_{i\to\infty} K_i = \infty$,

(c) $\sup_{i\in\mathbb{N}}\sup_{x\in M_i}\omega_i(B(x,R)) < \infty$ and $\sup_{i\in\mathbb{N}}\sup_{x\in M_i}\mathrm{area}_{\omega_i}(S(x,R)) < \infty$ for some $R>0$.

Then we have $\lim_{i\to\infty}\alpha_{(M_i,\nu_i)}(r) = 0$ for all $r>0$.

(2) Taking the limit of (6.2) as $m\to 1$ and then $\theta \to 1$, we obtain

$$\ln\big(2\alpha(r)\big) \le -\left(\sqrt{\frac{K}{2}}r - 1\right)^2 + 1.$$

Here $\lim_{c\to 1} G_c = G_1 = 1$ follows from the dominated convergence theorem since $\sigma^c \le \max\{\sigma,\sigma^{c_0}\} \in L^1(M,\omega)$ for $1/2 < c_0 \le c < 1$. Therefore we recover the normal concentration

$$\alpha(r) \le \frac{1}{2}\exp\left[-\left(\sqrt{\frac{K}{2}}r - 1\right)^2 + 1\right] \le \frac{1}{2}e^{-Kr^2/4+2}$$

which is well-known to hold for $(M,\omega)$ with $\mathrm{Ric}_\infty \ge K > 0$.

Theorem 6.1(ii) is applicable to the fundamental example of $m$-Gaussian measures (see Example 2.5).

**Example 6.3** Let $\{N(v_i, V_i)\}_{i\in\mathbb{N}} \subset \mathcal{P}^2_{\mathrm{ac}}(\mathbb{R}^n, dx)$ be a sequence of $m$-Gaussian measures with $m \in [(n-1)/n, 1) \cap (1/2, 1)$ as well as

$$\lim_{i\to\infty} (\det V_i)^{(1-m)/2} \Lambda_i^{-1} = \infty,$$

where $\Lambda_i$ is the largest eigenvalue of $V_i$. Then we have $\lim_{i\to\infty} \alpha_{(\mathbb{R}^n, N(v_i,V_i))}(r) = 0$ for all $r > 0$.

Under the additional assumption that $\omega(M) < \infty$, we further obtain the *$m$-normal concentration*. We first prove a computational lemma for later use.

**Lemma 6.4** (i) *For any $m \in (1/2, 1)$ and $a, r > 0$, we have*

$$\exp_m\big(-(ar-1)^2+1\big) \le (2m-1)^{1/(m-1)} \exp_m\bigg(-\frac{a^2}{2}r^2\bigg).$$

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

$$\exp_m\big((ar-1)^2-1\big) \ge \bigg(\frac{2}{m}-1\bigg)^{1/(m-1)} \exp_m\bigg(\frac{a^2}{2}r^2\bigg).$$

*Proof.* (i) We just calculate

$$\begin{aligned}
&\exp_m\big(-(ar-1)^2+1\big) \le \exp_m\bigg(-\frac{a^2}{2}r^2+2\bigg)\\
&= \bigg\{1+(m-1)\bigg(-\frac{a^2}{2}r^2+2\bigg)\bigg\}^{1/(m-1)}\\
&= (2m-1)^{1/(m-1)}\bigg\{1+(m-1)\bigg(-\frac{a^2}{2(2m-1)}r^2\bigg)\bigg\}^{1/(m-1)}\\
&\le (2m-1)^{1/(m-1)} \exp_m\bigg(-\frac{a^2}{2}r^2\bigg).
\end{aligned}$$

(ii) We similarly find

$$\begin{aligned}
&\exp_m\big((ar-1)^2-1\big) \ge \exp_m\bigg[\bigg(1-\frac{m}{2}\bigg)a^2r^2-\frac{2}{m}\bigg]\\
&= \bigg\{\bigg(\frac{2}{m}-1\bigg)+(m-1)\bigg(1-\frac{m}{2}\bigg)a^2r^2\bigg\}^{1/(m-1)}\\
&= \bigg(\frac{2}{m}-1\bigg)^{1/(m-1)}\bigg\{1+\frac{m}{2}(m-1)a^2r^2\bigg\}^{1/(m-1)}\\
&\ge \bigg(\frac{2}{m}-1\bigg)^{1/(m-1)} \exp_m\bigg(\frac{a^2}{2}r^2\bigg).
\end{aligned}$$

Note that the hypothesis $m \in (1,2)$ ensures that

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

$\square$

**Corollary 6.5 ($m$-normal concentration)** *Assume that $m \in [(n-1)/n, 1) \cap (1/2, 1)$, $\nu(M) = 1$, $\omega(M) < \infty$, $\mathrm{Ric}_N \geq 0$ and $\mathrm{Hess}\,\Psi \geq K$ holds for some $K > 0$. Then we have*

$$\alpha_{(M,\nu)}(r) \leq \frac{(2m-1)^{1/(m-1)}}{2} \exp_m \left( -\frac{mK}{4\omega(M)^{1-m}} r^2 \right)$$

*for all $r > 0$.*

*Proof.* We deduce from the Hölder inequality that

$$\int_B \sigma^m \, d\omega \leq \left( \int_B \sigma \, d\omega \right)^m \omega(B)^{1-m} = b^m \omega(B)^{1-m} \leq b^m \omega(M)^{1-m}$$

and, similarly, $G_m \leq \omega(M)^{1-m}$. In particular, $r^2 > 8\omega(M)^{1-m}/mK$ (otherwise the assertion is clear) implies $r^2 > 8G_m/mK$. Therefore we deduce from (6.3) that

$$\left( \sqrt{\frac{mK}{2}} r - \omega(M)^{(1-m)/2} \right)^2 - \omega(M)^{1-m} \leq -\omega(M)^{1-m} \ln_m(2b),$$

and hence

$$\alpha_{(M,\nu)}(r) \leq \frac{1}{2} \exp_m \left[ -\left( \omega(M)^{(m-1)/2} \sqrt{\frac{mK}{2}} r - 1 \right)^2 + 1 \right].$$

Then Lemma 6.4(i) completes the proof. $\square$

**Remark 6.6** Note that $\exp_m(-cr^2)$ is a polynomial of $r$, so that the $m$-normal concentration is weaker than the exponential concentration (i.e., $\alpha(r) \leq Ce^{-cr}$). This is natural and the most we can expect, because the $m$-Gaussian measures have only the polynomial decay for instance (Example 2.5).

For $m > 1$, Lemma 2.4(iii) ensures that $\mathrm{supp}\,\nu$ is bounded and $G_c(\nu) < \infty$ for all $c > 0$. Then the proof of Theorem 6.1(i) is applicable to $m \in (1,2]$ and gives (6.2) for all $r > 0$ and $\theta \in [0,1)$. Furthermore, under another condition that $\sigma$ is bounded, we again obtain the $m$-normal concentration.

**Proposition 6.7 ($m > 1$ case)** *Let $(M,\omega)$ satisfy $\mathrm{Ric}_N \geq 0$ and $m \in (1,2]$.*

(i) *Assume that $\nu(M) = 1$ and $\mathrm{Hess}\,\Psi \geq K$ holds for some $K > 0$. Then we have (6.2) for all $r > 0$ and $\theta \in [0,1)$.*

(ii) *If in addition $m < 2$ and $\|\sigma\|_\infty < \infty$, then we have*

$$\alpha_{(M,\nu)}(r)^{-1} \geq \left( \frac{2}{m} - 1 \right)^{1/(m-1)} \exp_m \left( \frac{mK\|\sigma\|_\infty^{1-m}}{4} r^2 \right)$$

*for all $r > 0$.*

*Proof.* (i) This is completely the same as Theorem 6.1(i), since $1/2 \ge m^{1/(1-m)}$ holds for $m \in (1, 2]$.

(ii) In (6.3) (with $m > 1$), we observe $\int_B \sigma^m \, d\omega \le b\|\sigma\|_\infty^{m-1}$ and $G_m \le \|\sigma\|_\infty^{m-1}$. Note also that $r^2 > 8\|\sigma\|_\infty^{m-1}/mK$ (otherwise the assertion is obvious) ensures $r^2 > 8G_m/mK$. These yield

$$\left(\sqrt{\frac{mK}{2}}r - \|\sigma\|_\infty^{(m-1)/2}\right)^2 - \|\sigma\|_\infty^{m-1} \le -b^{1-m}\|\sigma\|_\infty^{m-1}\ln_m(2b) \le \|\sigma\|_\infty^{m-1}\ln_m(b^{-1}).$$

Hence we have

$$\alpha_{(M,\nu)}(r)^{-1} \ge \exp_m\left[\left(\|\sigma\|_\infty^{(1-m)/2}\sqrt{\frac{mK}{2}}r - 1\right)^2 - 1\right],$$

and Lemma 6.4(ii) completes the proof. $\square$

Note that we obtained the estimate of the form $\alpha(r) \le C\exp_m(-cr^2)$ for $m < 1$, while $\alpha(r) \le C\{\exp_m(cr^2)\}^{-1}$ for $m > 1$. This is in a sense natural because the domain of $\exp_m$ is $(-\infty, 1/(1-m))$ for $m < 1$ and $[-1/(m-1), \infty)$ for $m > 1$.

# 7 Gradient flow of $H_m$

In this section, we show that the gradient flow of the $m$-relative entropy produces a weak solution to the porous medium equation. 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 explaining this coincidence (see, e.g., [JKO], [AGK] and [Vi2, Chapter 23]), among them, here we follow the rather 'metric' approach in [Oh1]. To do this, we start with a review of the geometric structure of the Wasserstein space and the general theory of gradient flows in it in accordance with the strategy in [Oh1] (see also [GO]). Throughout the section, $(M, g)$ is assumed to be compact, so that $\mathcal{P}^2(M) = \mathcal{P}(M)$ and $\sigma \in L^m(M, \omega)$.

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

We briefly review the geometric structure of $(\mathcal{P}(M), W_2)$. 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 ([St1, Proposition 2.10], [LV2, Theorem A.8]). In the case where $(M, g)$ is not nonnegatively curved, although $(\mathcal{P}(M), W_2)$ does not admit any lower curvature bound ([St1, Proposition 2.10]), we can show the following (see also [Oh1, Theorem 3.6]).

**Theorem 7.1** ([Gi, Theorem 3.4, Remark 3.5]) *Given $\mu \in \mathcal{P}(M)$ and unit speed geodesics $\alpha, \beta : [0, \delta) \longrightarrow \mathcal{P}(M)$ with $\alpha(0) = \beta(0) = \mu$, the joint limit*

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

*exists.*

Theorem 7.1 means that $(\mathcal{P}(M), W_2)$ looks like a Riemannian space (rather than a Finsler space), and we can investigate its infinitesimal structure 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 $\mu$. Given $\alpha, \beta \in \Sigma'_\mu[\mathcal{P}(M)]$, Theorem 7.1 verifies that the *angle*

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

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 *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.,

$$C_\mu[\mathcal{P}(M)] := \big(\Sigma_\mu[\mathcal{P}(M)] \times [0, \infty)\big) / \big(\Sigma_\mu[\mathcal{P}(M)] \times \{0\}\big),$$
$$\sigma_\mu\big((\alpha, s), (\beta, t)\big) := \sqrt{s^2 + t^2 - 2st \cos \angle_\mu(\alpha, \beta)}.$$

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

**Definition 7.2 (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 $\dot{\xi}(t) = \mathbf{v}$.

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

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

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

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

Define the *absolute gradient* (called the *local slope* in [AGS]) of $f$ at $\mu \in \mathcal{P}^*(M)$ by

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

Note that $-D_\mu f(\alpha) \le |\nabla_{\!-} f|(\mu)$ for any $\alpha \in \Sigma_\mu[\mathcal{P}(M)]$.

**Lemma 7.3** ([Oh1, Lemma 4.2]) *For each $\mu \in \mathcal{P}^*(M)$ with $0 < |\nabla_{\!-} f|(\mu) < \infty$, there exists unique $\alpha \in \Sigma_\mu[\mathcal{P}^*(M)]$ satisfying $D_\mu f(\alpha) = -|\nabla_{\!-} f|(\mu)$.*

Using $\alpha$ in the above lemma, we define the *negative gradient vector* of $f$ at $\mu$ as

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

In the case of $|\nabla_{\!-} f|(\mu) = 0$, we simply define $\nabla_{\!-} f(\mu)$ as the origin of $C_\mu[\mathcal{P}(M)]$.

**Definition 7.4 (Gradient curves)** A continuous curve $\xi : [0, l) \longrightarrow \mathcal{P}^*(M)$ which is locally Lipschitz on $(0, l)$ is called a *gradient curve* of $f$ if $|\nabla_{\!-} f|(\xi(t)) < \infty$ holds for all $t \in (0, \infty)$ and if it is right differentiable with $\dot{\xi}(t) = \nabla_{\!-} f(\xi(t))$ at all $t \in (0, l)$. We say that a gradient curve $\xi$ is *complete* if it is defined on entire $[0, \infty)$.

**Theorem 7.5** ([Oh1, Theorem 5.11, Corollary 6.3], [GO, Theorem 4.2])

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

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

$$W_2\big(\xi(t), \zeta(t)\big) \le e^{-tK} W_2\big(\xi(0), \zeta(0)\big) \tag{7.1}$$

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

Therefore the *gradient flow* $G : [0, \infty) \times \mathcal{P}^*(M) \longrightarrow \mathcal{P}^*(M)$ of $f$, given as $G(t, \mu) = \xi(t)$ in Theorem 7.5(i), is uniquely determined and extended to the closure $G : [0, \infty) \times \overline{\mathcal{P}^*(M)} \longrightarrow \overline{\mathcal{P}^*(M)}$ continuously.

## 7.3 $m$-relative entropy and the porous medium equation

We recall basic notions of calculus on weighted Riemannian manifolds $(M, \omega)$ with $\omega = e^{-\psi} \operatorname{vol}_g$. For a $C^1$-vector filed $V$ on $M$, we define the *weighted divergence* as

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

where $\operatorname{div} V$ denotes the usual divergence of $V$ for $(M, \operatorname{vol}_g)$. Note that, for any $f \in C^1(M)$,

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

For $f \in C^2(M)$, the *weighted Laplacian* is defined by

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

Then it is an established fact that the gradient flow of the relative entropy (or the *free energy*)

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

produces a solution to the associated *heat equation* (or the *Fokker-Planck equation*)

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

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, and [OS, Section 7] for the Finsler case.

Here we see that the same technique gives a weak solution to the *porous medium equation*

$$\frac{\partial \rho}{\partial t} = \frac{1}{m}\Delta^{\omega}(\rho^m) + \mathrm{div}_{\omega}(\rho \nabla \Psi) \tag{7.2}$$

as gradient flow of the $m$-relative entropy $H_m(\cdot|\nu)$, as was demonstrated by Otto [Ot] for the Tsallis entropy as well as $H_m(\cdot|N_m(0, cI_n))$ with respect to the $m$-Gaussian measures $N_m(0, cI_n)$ on $(\mathbb{R}^n, dx)$. Recall that $\nu = \exp_m(-\Psi)\omega$.

**Theorem 7.6 (Gradient flow of $H_m$)** *Let $(M, g)$ be compact, $m \in ((n-1)/n, 1) \cup (1, 2]$ and $\Psi$ be Lipschitz. If a curve $(\mu_t)_{t \in [0,\infty)} \subset \mathcal{P}_{\mathrm{ac}}(M_0, \omega)$ is a gradient curve of $H_m(\cdot|\nu)$, then its density function $\rho_t$ is a weak solution to the porous medium equation* (7.2). *To be precise,*

$$\int_M \phi_{t_1}\, d\mu_{t_1} - \int_M \phi_{t_0}\, d\mu_{t_0} = \int_{t_0}^{t_1} \int_M \left\{ \frac{\partial \phi_t}{\partial t} + \frac{1}{m}\rho_t^{m-1}\Delta^{\omega}\phi_t - \langle \nabla \phi_t, \nabla \Psi \rangle \right\} d\mu_t\, dt \tag{7.3}$$

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

*Proof.* Fix $t \in (0, \infty)$ and, given small $\delta > 0$, choose $\mu^{\delta} \in \mathcal{P}(M)$ as a minimizer of the function

$$\mu \longmapsto H_m(\mu|\nu) + \frac{W_2(\mu, \mu_t)^2}{2\delta}.$$

We postpone the proof of the following technical claim until the end of the section. We remark that the hypothesis $m \le 2$ comes into play only in the proof of Claim 7.7(iii) (see Lemma 7.11).

**Claim 7.7** (i) *Such $\mu^{\delta}$ indeed exists and is absolutely continuous.*

(ii) *We have*

$$\lim_{\delta \to 0} \frac{W_2(\mu^{\delta}, \mu_t)^2}{2\delta} = 0, \qquad \lim_{\delta \to 0} H_m(\mu^{\delta}|\nu) = H_m(\mu_t|\nu).$$

*In particular, $\mu^{\delta}$ converges to $\mu_t$ weakly.*

(iii) *Moreover, by putting $\mu^{\delta} = \rho^{\delta}\omega$, $(\rho^{\delta})^m$ converges to $\rho_t^m$ in $L^1(M, \omega)$.*

Take a Lipschitz function $\varphi : M \longrightarrow \mathbb{R}$ such that $\mathcal{T}(x) := \exp_x(\nabla \varphi(x))$ gives the optimal transport from $\mu^{\delta}$ to $\mu_t$. We consider the transport $\mu^{\delta}_{\varepsilon} := (\mathcal{F}_{\varepsilon})_{\sharp}\mu^{\delta}$ in another direction for small $\varepsilon > 0$, where $\mathcal{F}_{\varepsilon}(x) := \exp_x(\varepsilon \nabla \phi_t(x))$. It immediately follows from the choice of $\mu^{\delta}$ that

$$H_m(\mu^{\delta}_{\varepsilon}|\nu) + \frac{W_2(\mu^{\delta}_{\varepsilon}, \mu_t)^2}{2\delta} \ge H_m(\mu^{\delta}|\nu) + \frac{W_2(\mu^{\delta}, \mu_t)^2}{2\delta}. \tag{7.4}$$

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

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \frac{W_2(\mu_\varepsilon^\delta, \mu_t)^2 - W_2(\mu^\delta, \mu_t)^2}{\varepsilon} \\
&\le \limsup_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_M \left\{ d\big(\mathcal{F}_\varepsilon(x), \mathcal{T}(x)\big)^2 - d\big(x, \mathcal{T}(x)\big)^2 \right\} d\mu^\delta(x) \\
&= -\int_M 2\langle \nabla\phi_t, \nabla\varphi\rangle \, d\mu^\delta.
\end{aligned}$$

We used the first variation formula for the distance $d$ in the last line. Thanks to the compactness of $M$, there is a constant $C > 0$ such that

$$\phi_t\big(\mathcal{T}(x)\big) \le \phi_t(x) + \langle \nabla\phi_t(x), \nabla\varphi(x)\rangle + Cd\big(x, \mathcal{T}(x)\big)^2.$$

Thus we obtain, by virtue of Claim 7.7(ii),

$$\begin{aligned}
&\liminf_{\delta\to 0} \frac{1}{2\delta} \limsup_{\varepsilon\to 0} \frac{W_2(\mu_\varepsilon^\delta, \mu_t)^2 - W_2(\mu^\delta, \mu_t)^2}{\varepsilon} \le -\limsup_{\delta\to 0} \frac{1}{\delta} \int_M \langle \nabla\phi_t, \nabla\varphi\rangle \, d\mu^\delta \\
&\le \liminf_{\delta\to 0} \frac{1}{\delta} \left[ \int_M \{\phi_t - \phi_t(\mathcal{T})\} \, d\mu^\delta + CW_2(\mu^\delta, \mu_t)^2 \right] \\
&= \liminf_{\delta\to 0} \frac{1}{\delta} \left\{ \int_M \phi_t \, d\mu^\delta - \int_M \phi_t \, d\mu_t \right\}.
\end{aligned}$$

Next we calculate the difference of entropies in (7.4). We put $\mu^\delta = \rho^\delta\omega$, $\mu_\varepsilon^\delta = \rho_\varepsilon^\delta\omega$ and $\mathbf{J}_\varepsilon^\omega := e^{\psi - \psi(\mathcal{F}_\varepsilon)} \det(D\mathcal{F}_\varepsilon)$. Then we obtain from the Jacobian equation $\rho_\varepsilon^\delta(\mathcal{F}_\varepsilon)\mathbf{J}_\varepsilon^\omega = \rho^\delta$ that

$$\begin{aligned}
&H_m(\mu_\varepsilon^\delta|\nu) - \frac{1}{m} \int_M \sigma^m \, d\omega = \frac{1}{m(m-1)} \int_M \{(\rho_\varepsilon^\delta)^m - m\rho_\varepsilon^\delta \sigma^{m-1}\} \, d\omega \\
&= \frac{1}{m(m-1)} \int_M \{\rho_\varepsilon^\delta(\mathcal{F}_\varepsilon)^{m-1} - m\sigma(\mathcal{F}_\varepsilon)^{m-1}\} \rho_\varepsilon^\delta(\mathcal{F}_\varepsilon) \mathbf{J}_\varepsilon^\omega \, d\omega \\
&= \frac{1}{m(m-1)} \int_M \left\{ \left( \frac{\rho^\delta}{\mathbf{J}_\varepsilon^\omega} \right)^{m-1} - m\sigma(\mathcal{F}_\varepsilon)^{m-1} \right\} d\mu^\delta.
\end{aligned}$$

Thus we have

$$\begin{aligned}
&H_m(\mu^\delta|\nu) - H_m(\mu_\varepsilon^\delta|\nu) \\
&= \frac{1}{m(m-1)} \int_M \Big[ (\rho^\delta)^{m-1} \{1 - (\mathbf{J}_\varepsilon^\omega)^{1-m}\} - m\{\sigma^{m-1} - \sigma(\mathcal{F}_\varepsilon)^{m-1}\} \Big] \, d\mu^\delta.
\end{aligned}$$

Note that

$$\lim_{\varepsilon\to 0} \frac{\mathbf{J}_\varepsilon^\omega - 1}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{e^{\psi - \psi(\mathcal{F}_\varepsilon)} \det(D\mathcal{F}_\varepsilon) - 1}{\varepsilon} = \Delta\phi_t - \langle \nabla\phi_t, \nabla\psi\rangle = \Delta^\omega \phi_t.$$

Hence we obtain, together with Claim 7.7(iii),

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\frac{H_m(\mu^\delta|\nu)-H_m(\mu^\delta_\varepsilon|\nu)}{\varepsilon}\\
&=\int_M\left\{\frac{1}{m}(\rho^\delta)^{m-1}\Delta^\omega\phi_t+\frac{1}{m-1}\langle\nabla\phi_t,\nabla(\sigma^{m-1})\rangle\right\}d\mu^\delta \qquad (7.5)\\
&\to\int_M\left\{\frac{1}{m}\rho_t^{m-1}\Delta^\omega\phi_t+\frac{1}{m-1}\langle\nabla\phi_t,\nabla(\sigma^{m-1})\rangle\right\}d\mu_t
\end{aligned}$$

as $\delta$ tends to zero.

These together imply, as $\sigma^{m-1}=1-(m-1)\Psi$,

$$\liminf_{\delta\to 0}\frac{1}{\delta}\left\{\int_M\phi_t\,d\mu^\delta-\int_M\phi_t\,d\mu_t\right\}\ge\int_M\left\{\frac{1}{m}\rho_t^{m-1}\Delta^\omega\phi_t-\langle\nabla\phi_t,\nabla\Psi\rangle\right\}d\mu_t.$$

Moreover, equality holds since we can change $\phi$ into $-\phi$. Recall from [GO, (4)] (see also [Oh1, Lemma 6.4]) that

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

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

$$\begin{aligned}
&\lim_{\delta\to 0}\frac{1}{\delta}\left\{\int_M\phi_{t+\delta}\,d\mu_{t+\delta}-\int_M\phi_t\,d\mu_t\right\}\\
&=\lim_{\delta\to 0}\frac{1}{\delta}\left\{\int_M(\phi_{t+\delta}-\phi_t)\,d\mu_{t+\delta}+\int_M\phi_t\,d\mu_{t+\delta}-\int_M\phi_t\,d\mu_t\right\}\\
&=\int_M\left\{\frac{\partial\phi_t}{\partial t}+\frac{1}{m}\rho_t^{m-1}\Delta^\omega\phi_t-\langle\nabla\phi_t,\nabla\Psi\rangle\right\}d\mu_t
\end{aligned}$$

as desired. $\square$

Recall from Theorem 4.1 that the entropy $H_m(\cdot|\nu)$ is $K$-convex if (and only if) $\mathrm{Ric}_N\ge 0$ and $\mathrm{Hess}\,\Psi\ge K$. Combining this with Theorems 7.5, 7.6, we obtain the following.

**Corollary 7.8** *The weak solution to the porous medium equation constructed in Theorem* 7.6 *enjoys the contraction property* (7.1) *under the assumptions* $\mathrm{Ric}_N\ge 0$ *and* $\mathrm{Hess}\,\Psi\ge K$.

The argument in the proof of Theorem 7.6 also shows that the absolute gradient of $H_m(\cdot|\nu)$ at $\mu$ coincides with the square root of the $m$-relative Fisher information introduced in (5.2), now for general $m$. Compare this with Theorem 5.2.

**Proposition 7.9** *Take* $m\in[(n-1)/n,1)\cup(1,\infty)$ *and* $\mu=\rho\omega\in\mathcal{P}_{\mathrm{ac}}(M,\omega)$ *such that* $\rho$ *is Lipschitz. For any* $(d^2/2)$*-convex function* $\varphi:M\longrightarrow\mathbb{R}$ *and the corresponding transport* $\mu_t:=(\mathcal{T}_t)_\sharp\mu$ *with* $\mathcal{T}_t(x):=\exp_x(t\nabla\varphi(x))$, $t\ge 0$, *it holds that*

$$\lim_{t\to 0}\frac{H_m(\mu_t|\nu)-H_m(\mu|\nu)}{t}=\frac{1}{m-1}\int_M\langle\nabla(\rho^{m-1}-\sigma^{m-1}),\nabla\varphi\rangle\,d\mu.$$

*In particular, we have $|\nabla_{-}[H_m(\cdot|\nu)]|(\mu) = \sqrt{I_m(\mu|\nu)}$ and, if $|\nabla_{-}[H_m(\cdot|\nu)]|(\mu) < \infty$, then the negative gradient vector $\nabla_{-}[H_m(\cdot|\nu)](\mu)$ is achieved by*

$$\nabla\varphi = -\nabla\biggl(\frac{\rho^{m-1} - \sigma^{m-1}}{m-1}\biggr).$$

*Proof.* Recall that $\varphi$ is twice differentiable a.e., and that $\mu_t$ is absolutely continuous for $t < 1$ ([Vi2, Theorem 8.7]). Using the calculation deriving (7.5), we obtain

$$\begin{aligned}
&\lim_{t\to 0} \frac{H_m(\mu|\nu) - H_m(\mu_t|\nu)}{t} \\
&= \int_M \biggl\{\frac{1}{m}\rho^{m-1}\Delta^{\omega}\varphi + \frac{1}{m-1}\langle\nabla\varphi, \nabla(\sigma^{m-1})\rangle\biggr\}\,d\mu \\
&= -\int_M \biggl\{\frac{1}{m}\langle\nabla(\rho^m), \nabla\varphi\rangle - \frac{\rho}{m-1}\langle\nabla\varphi, \nabla(\sigma^{m-1})\rangle\biggr\}\,d\omega \\
&= -\frac{1}{m-1}\int_M \langle\nabla(\rho^{m-1} - \sigma^{m-1}), \nabla\varphi\rangle\,d\mu.
\end{aligned}$$

As any geodesic with respect to $W_2$ is realized in this way (Theorem 2.7), we have $|\nabla_{-}[H_m(\cdot|\nu)]|(\mu) = \sqrt{I_m(\mu|\nu)}$ and

$$\nabla_{-}[H_m(\cdot|\nu)](\mu) = -\nabla\biggl(\frac{\rho^{m-1} - \sigma^{m-1}}{m-1}\biggr).$$

$\square$

**Remark 7.10** The family of $m$-Gaussian measures (Example 2.5) again has a role to play here. On the unweighted Euclidean space $(\mathbb{R}^n, dx)$, it is known by [OW, Proposition 5] that a solution to the porous medium equation starting from an $m$-Gaussian measure will keep being $m$-Gaussian. An explicit expression of such solutions is given in [Ta2].

## 7.4 Proof of Claim 7.7

(i) The existence follows from, as usual, the compactness of $\mathcal{P}(M)$ and the lower semicontinuity of $H_m(\cdot|\nu)$ (Lemma 3.4). The absolute continuity is obvious for $m > 1$.

For $m < 1$, decompose $\mu^\delta$ into absolutely continuous and singular parts $\mu^\delta = \rho\omega + \mu^s$ and suppose $\mu^s(M) > 0$. We modify $\mu^\delta$ into $\hat{\mu}_r \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ as

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

for small $r > 0$. Then we find

$$\begin{aligned}
\int_M \sigma^{m-1}\,d\hat{\mu}_r &\le \int_M \sigma^{m-1}\,d\mu^\delta + \int_M \biggl|\sigma(y)^{m-1} - \frac{1}{\omega(B(y,r))}\int_{B(y,r)} \sigma^{m-1}\,d\omega\biggr| d\mu^s(y) \\
&\le \int_M \sigma^{m-1}\,d\mu^\delta + \Bigl\{\sup_M |\nabla(\sigma^{m-1})| \cdot \mu^s(M)\Bigr\} r.
\end{aligned}$$

Given an optimal coupling $\pi = \pi_1 + \pi_2$ of $\mu^\delta$ and $\mu_t$ 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 $\mu_t$. Hence we observe

$$\begin{aligned}
W_2(\hat{\mu}_r, \mu_t)^2 &\le \int_{M\times M} d(x,z)^2 d\pi_1(x,z) + \int_{M\times M} \{d(y,z)+r\}^2\, d\pi_2(y,z) \\
&\le \int_{M\times M} d(x,z)^2 d\pi(x,z) + \{2 \operatorname{diam} M + r\} r \pi_2(M\times M) \\
&\le W_2(\mu^\delta, \mu_t)^2 + \{3 \operatorname{diam} M \cdot \mu^s(M)\} r.
\end{aligned}$$

Finally, it follows from the Hölder inequality that

$$\begin{aligned}
&\int_M \hat{\rho}_r^m \, d\omega = \int_M \left[ \int_M \left\{ \frac{\rho(x)}{\mu^s(M)} + \frac{\chi_{B(y,r)}(x)}{\omega(B(y,r))} \right\} d\mu^s(y) \right]^m d\omega(x) \\
&\ge \mu^s(M)^{m-1} \int_M \left[ \int_M \left\{ \frac{\rho(x)}{\mu^s(M)} + \frac{\chi_{B(y,r)}(x)}{\omega(B(y,r))} \right\}^m d\mu^s(y) \right] d\omega(x) \\
&\ge \mu^s(M)^{m-1} \int_M \left\{ \int_{M\setminus B(y,r)} \frac{\rho^m}{\mu^s(M)^m}\, d\omega + \int_{B(y,r)} \frac{1}{\omega(B(y,r))^m}\, d\omega \right\} d\mu^s(y) \\
&= \int_M \rho^m \, d\omega - \mu^s(M)^{-1} \int_M \left( \int_{B(y,r)} \rho^m \, d\omega \right) d\mu^s(y) \\
&\quad + \mu^s(M)^{m-1} \int_M \omega\big(B(y,r)\big)^{1-m} \, d\mu^s(y).
\end{aligned}$$

As $M$ is compact, we find

$$\mu^s(M)^{m-1} \int_M \omega\big(B(y,r)\big)^{1-m}\, d\mu^s(y) \ge \mu^s(M)^m C_1(\omega, m) r^{n(1-m)},$$

and

$$\begin{aligned}
\int_{B(y,r)} \rho^m \, d\omega &= \int_{B(y,r)} (\rho\sigma^{m-1})^m \sigma^{m(1-m)} \, d\omega \\
&\le \left( \int_{B(y,r)} \rho\sigma^{m-1} \, d\omega \right)^m \left( \int_{B(y,r)} \sigma^m \, d\omega \right)^{1-m} \\
&\le \left( \int_{B(y,r)} \rho\sigma^{m-1} \, d\omega \right)^m C_2(\omega, \sigma, m) r^{n(1-m)}.
\end{aligned}$$

Since $\lim_{r\to 0} \sup_{y\in M} \int_{B(y,r)} \rho\sigma^{m-1}\, d\omega = 0$, these imply

$$\int_M \hat{\rho}_r^m \, d\omega \ge \int_M \rho^m \, d\omega + C_1(\omega, m) \mu^s(M)^m r^{n(1-m)}.$$

Combining these, we conclude that

$$\begin{aligned}&H_m(\hat{\mu}_r|\nu) + \frac{W_2(\hat{\mu}_r,\mu_t)^2}{2\delta} - H_m(\mu^\delta|\nu) - \frac{W_2(\mu^\delta,\mu_t)^2}{2\delta}\\&\le -C_3(\omega,m)\mu^s(M)^m r^{n(1-m)} + C_4(M,\sigma,m,\delta)\mu^s(M)r,\end{aligned}$$

where $C_3, C_4 > 0$. Then $n(1-m) < 1$ and $\mu^s(M) > 0$ yield that

$$H_m(\hat{\mu}_r|\nu) + \frac{W_2(\hat{\mu}_r,\mu_t)^2}{2\delta} < H_m(\mu^\delta|\nu) + \frac{W_2(\mu^\delta,\mu_t)^2}{2\delta}$$

holds for small $r > 0$. This contradicts the construction of $\mu^\delta$, therefore we obtain $\mu^s(M) = 0$.

(ii) By the choice of $\mu^\delta$, we have

$$H_m(\mu^\delta|\nu) + \frac{W_2(\mu^\delta,\mu_t)^2}{2\delta} \le H_m(\mu_t|\nu)$$

which immediately implies $\lim_{\delta\to 0} W_2(\mu^\delta,\mu_t)^2 \le \lim_{\delta\to 0} 2\delta H_m(\mu_t|\nu) = 0$. Thus $\mu^\delta$ converges to $\mu_t$ weakly, and hence

$$\limsup_{\delta\to 0}\frac{W_2(\mu^\delta,\mu_t)^2}{2\delta} \le H_m(\mu_t|\nu) - \liminf_{\delta\to 0} H_m(\mu^\delta|\nu) \le 0$$

by the lower semi-continuity. This further yields

$$H_m(\mu_t|\nu) \le \liminf_{\delta\to 0} H_m(\mu^\delta|\nu) \le \limsup_{\delta\to 0} H_m(\mu^\delta|\nu) \le H_m(\mu_t|\nu).$$

(iii) This is a consequence of the following general lemma. $\diamondsuit$

**Lemma 7.11** *Assume $m \in [(n-1)/n, 1) \cup (1,2]$ and that 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 as well as $\lim_{i\to\infty} H_m(\mu_i|\nu) = H_m(\mu|\nu) < \infty$. Then, by setting $\mu_i = \rho_i\omega$ and $\mu = \rho\omega$, $\rho_i^m$ converges to $\rho^m$ in $L^1(M,\omega)$.*

*Proof.* Note that the convergence of $H_m(\mu_i|\nu)$ ensures $\lim_{i\to\infty}\int_M \rho_i^m\,d\omega = \int_M \rho^m\,d\omega$. We shall show the following:

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

Then we have, for $m < 1$,

$$\int_M |\rho_i^m - \rho^m|\,d\omega \le \int_M |\rho_i - \rho|^m\,d\omega \le \omega(M)^{1-m}\bigg(\int_M |\rho_i - \rho|\,d\omega\bigg)^m,$$

and

$$\begin{aligned}&\int_M |\rho_i - \rho|\,d\omega\\&\le \int_M \big[|\min\{\rho_i, C\} - \min\{\rho, C\}| + \max\{\rho_i - C, 0\} + \max\{\rho - C, 0\}\big]\,d\omega\\&\to 0\end{aligned}$$

as $i \to \infty$ and then $C \to \infty$. For $m \in (1, 2]$, we similarly find

$$\begin{aligned}\int_M |\rho_i^m - \rho^m| \, d\omega &\le m \int_M |\rho_i - \rho| (\rho_i + \rho)^{m-1} \, d\omega \\ &\le m \bigg( \int_M |\rho_i - \rho|^m \, d\omega \bigg)^{1/m} \bigg( \int_M (\rho_i + \rho)^m \, d\omega \bigg)^{(m-1)/m},\end{aligned}$$

and

$$\begin{aligned}&\int_M |\rho_i - \rho|^m \, d\omega \\ &\le 2^{m-1} \int_M \big[ |\min\{\rho_i, C\} - \min\{\rho, C\}|^m + \max\{\rho_i - C, 0\}^m + \max\{\rho - C, 0\}^m \big] \, d\omega \\ &\to 0\end{aligned}$$

as $i \to \infty$ and then $C \to \infty$.

To show $(*)$, we suppose that it is false. Then there are some constants $C, \varepsilon > 0$ and sequences $\{k_j\}_{j \in \mathbb{N}}, \{l_j\}_{j \in \mathbb{N}} \subset \mathbb{N}$ going to infinity such that

$$\| \min\{\rho_{k_j}, C\} - \min\{\rho_{l_j}, C\} \|_{L^2(M,\omega)} \ge \varepsilon$$

for all $j \in \mathbb{N}$. This implies, together with $d^2[t^m/m(m-1)]/dt^2 = t^{m-2}$ and $\tau^{m-2}(\tau - \varepsilon)^2 \ge (1-\varepsilon)^2$ for $\tau \ge 1 \ge \varepsilon$,

$$\begin{aligned}&\frac{1}{m(m-1)} \int_M \bigg( \frac{\rho_{k_j} + \rho_{l_j}}{2} \bigg)^m \, d\omega \\ &\le \int_M \bigg[ \frac{\rho_{k_j}^m + \rho_{l_j}^m}{2m(m-1)} - \frac{\max\{\rho_{k_j}, \rho_{l_j}\}^{m-2}}{8} |\rho_{k_j} - \rho_{l_j}|^2 \bigg] \, d\omega \\ &\le \int_M \frac{\rho_{k_j}^m + \rho_{l_j}^m}{2m(m-1)} \, d\omega - \frac{C^{m-2}}{8} \int_M |\min\{\rho_{k_j}, C\} - \min\{\rho_{l_j}, C\}|^2 \, d\omega \\ &\le \int_M \frac{\rho_{k_j}^m + \rho_{l_j}^m}{2m(m-1)} \, d\omega - \frac{C^{m-2}}{8} \varepsilon^2.\end{aligned}$$

This means that $\bar{\mu}_j := \{(\rho_{k_j} + \rho_{l_j})/2\}\omega$ satisfies

$$\limsup_{j \to \infty} H_m(\bar{\mu}_j|\nu) \le \lim_{i \to \infty} H_m(\mu_i|\nu) - \frac{C^{m-2}}{8}\varepsilon^2 = H_m(\mu|\nu) - \frac{C^{m-2}}{8}\varepsilon^2,$$

this contradicts the lower semi-continuity of $H_m(\cdot|\nu)$. □

# 8 Finsler case

We finally stress that most results in this article are extended to Finsler manifolds, according to the theory developed in [Oh2], [OS] (see also a survey [Oh3]). Briefly speaking, 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], [Sh] for the basics of Finsler geometry, and to [Oh2], [OS], [Oh3] for the details omitted in the following discussion.

A *Finsler manifold* $(M, F)$ will be a pair of an $n$-dimensional $C^\infty$-manifold $M$ and a $C^\infty$-Finsler structure $F : TM \longrightarrow [0, \infty)$ satisfying the following *regularity*, *positive homogeneity*, and *strong convexity* conditions:

(1) $F$ is $C^\infty$ on $TM \setminus 0$, where $0$ stands for the zero section;

(2) $F(\lambda v) = \lambda F(v)$ holds for all $v \in TM$ and $\lambda \geq 0$;

(3) In any local coordinate system $(x^i)_{i=1}^n$ of $U \subset M$ and the corresponding coordinate $v = \sum_i v^i (\partial/\partial x^i)_x$ of $T_x M$ with $x \in U$, the $n \times n$-matrix

$$\left( \frac{\partial^2 (F^2)}{\partial v^i \partial v^j}(v) \right)_{i,j=1}^n$$

is positive definite for all $v \in T_x M \setminus 0$ and $x \in U$.

Then the distance $d$, geodesics and the exponential map are defined in the same manner as Riemannian geometry, whereas $d$ is typically nonsymmetric (and not a distance in a precise sense) since $F$ is merely positively homogeneous. Nonetheless, $d$ satisfies the positivity and the triangle inequality.

On a Finsler manifold $(M, F)$, there is no constructive measure as good as the Riemannian volume measure in the Riemannian case, but we can consider an arbitrary positive $C^\infty$-measure $\omega$ on $M$ and associate it with the weighted Ricci curvature $\mathrm{Ric}_N$ ([Oh2]). This curvature turns out surprisingly useful, and the argument in [Oh2] is applicable to generalizing the whole results in Sections 4–6 to the Finsler setting. (We need a little trick only in Proposition 5.4, put $\mu = (1 - \varepsilon f)\sigma\omega$ when $m > 1$.)

**Theorem 8.1** *Let $(M, F)$ be a forward complete Finsler manifold and $\omega$ be a positive $C^\infty$-measure on $M$. Then the following results in this article hold true also for $(M, F, \omega)$ (with appropriate interpretations for the nonsymmetric distance, cf.* [Oh2]*):*

- *Theorem* 4.1;
- *Proposition* 5.1*, Theorem* 5.2*, Proposition* 5.4;
- *Theorem* 6.1*, Corollary* 6.5*, Proposition* 6.7*.*

As for Section 7, due to the lack of the analogue of Theorem 7.1, we can not directly follow the Riemannian argument. Nonetheless, we can follow the discussion in [OS] using a (formal) Finsler structure of the Wasserstein space, and obtain results corresponding to Theorem 7.6 and Proposition 7.9. The point is the usage of the structure of the underlying space $M$, while we did not explicitly use it in Subsections 7.1, 7.2. See [OS, Sections 6, 7] for further details.

Let $(M, F)$ be compact from now on. Due to Otto's idea [Ot, Section 4], we introduce a Finsler structure of $(\mathcal{P}(M), W_2)$ as follows. Given $\mu \in \mathcal{P}(M)$, we define the *tangent space* at $\mu$ by

$$T_\mu \mathcal{P} := \Big( \overline{\{\nabla\varphi \,|\, \varphi \in C^\infty(M)\}}, F_W(\mu, \cdot) \Big),$$
$$F_W(\mu, \nabla\varphi) := \bigg( \int_M F(\nabla\varphi)^2 \, d\mu \bigg)^{1/2},$$

where the *gradient vector* $\nabla\varphi(x) \in T_x M$ is the Legendre transform of the derivative $D\varphi(x) \in T_x^* M$, and the closure was taken with respect to $F_W(\mu, \cdot)$. We remark that the gradient $\nabla$ is a nonlinear operator (i.e., $\nabla(\varphi_1 + \varphi_2)(x) \neq \nabla\varphi_1(x) + \nabla\varphi_2(x)$), since the Legendre transform is nonlinear (unless $F|_{T_x M}$ is Riemannian).

Now, we take a locally Lipschitz curve $(\mu_t)_{t \in I} \subset \mathcal{P}(M)$ on an open interval $I \subset \mathbb{R}$. We can associate it with the *tangent vector field* $\dot{\mu}_t = \Phi(t, \cdot) \in T_{\mu_t} \mathcal{P}$, that is, $\Phi$ is a Borel vector field on $I \times M$ with $\Phi(t, x) \in T_x M$ and $F(\Phi) \in L^\infty_{\mathrm{loc}}(I \times M, d\mu_t dt)$ satisfying the *continuity equation* $\partial\mu_t/\partial t + \mathrm{div}(\Phi_t \mu_t) = 0$ in the weak sense that

$$\int_I \int_M \bigg\{ \frac{\partial\phi_t}{\partial t} + D\phi_t(\Phi_t) \bigg\} \, d\mu_t dt = 0 \tag{8.1}$$

holds for all $\phi \in C_c^\infty(I \times M)$ ([AGS, Theorem 8.3.1], [OS, Theorem 7.3]). Using these 'differentiable' structures, we can consider gradient curves in a way different from the 'metric' approach in Section 7.

**Definition 8.2** Given a function $f : \mathcal{P}(M) \longrightarrow (-\infty, \infty]$ and $\mu \in \mathcal{P}(M)$ with $f(\mu) < \infty$, we say that $f$ is *differentiable* at $\mu$ if there is $\Phi \in T_\mu \mathcal{P}$ such that

$$\lim_{t \downarrow 0} \frac{f((\mathcal{T}_t)_\sharp \mu) - f(\mu)}{t} = \int_M \mathcal{L}(\Phi)(\nabla\varphi) \, d\mu$$

holds for all $\varphi \in C^\infty(M)$, where $\mathcal{T}_t(x) := \exp_x(t\nabla\varphi)$ and $\mathcal{L} : T_x M \longrightarrow T_x^* M$ stands for the Legendre transform. Then we write $\nabla_W f(\mu) = \Phi$.

Then a gradient curve should be a solution to $\dot{\mu}_t = \nabla_W[-H_m(\cdot|\nu)](\mu_t)$. We first show that $\nabla_W[-H_m(\cdot|\nu)](\mu_t)$ is described by the Fisher information like Proposition 7.9.

**Proposition 8.3** *Take $\mu = \rho\omega \in \mathcal{P}_{\mathrm{ac}}(M, \omega)$ with $\rho^m \in H^1(M, \omega)$. If $(\rho^{m-1} - \sigma^{m-1})/(1 - m) \notin H^1(M, \mu)$, then $-H_m(\cdot|\nu)$ is not differentiable at $\mu$. If $(\rho^{m-1} - \sigma^{m-1})/(1 - m) \in H^1(M, \mu)$, then $-H_m(\cdot|\nu)$ is differentiable at $\mu$ and we have*

$$\nabla_W[-H_m(\cdot|\nu)](\mu) = \nabla\bigg( \frac{\rho^{m-1} - \sigma^{m-1}}{1 - m} \bigg) \in T_\mu \mathcal{P}.$$

*Proof.* Fix arbitrary $\varphi \in C^\infty(M)$ and put $\mathcal{T}_t(x) := \exp_x(t\nabla\varphi(x))$, $\mu_t = \rho_t\omega := (\mathcal{T}_t)_\sharp \mu$ for sufficiently small $t > 0$. Then the Jacobian equation $\rho = \rho_t(\mathcal{T}_t)\mathbf{J}_t^\omega$ holds $\mu$-a.e., where $\mathbf{J}_t^\omega(x)$ stands for the Jacobian of $D\mathcal{T}_t(x) : T_x M \longrightarrow T_{\mathcal{T}_t(x)} M$ with respect to $\omega$. Thus we obtain, as in the proof of Theorem 7.6,

$$H_m(\mu_t|\nu) = H_m(\mu|\nu) + \frac{1}{m(m-1)} \int_M \big[ \rho^{m-1}\{(\mathbf{J}_t^\omega)^{1-m} - 1\} + m\{\sigma^{m-1} - \sigma(\mathcal{T}_t)^{m-1}\} \big] \, d\mu.$$

As the change of variables formula implies

$$\lim_{t\to 0}\int_M \frac{(\mathbf{J}^\omega_t)^{1-m}-1}{t}\rho^m\,d\omega = (1-m)\lim_{t\to 0}\int_M \frac{\mathbf{J}^\omega_t-1}{t}\rho^m\,d\omega$$
$$= (1-m)\lim_{t\to 0}\int_M \frac{\rho^m-\rho(\mathcal{T}_t)^m}{t}\mathbf{J}^\omega_t\,d\omega = (m-1)\int_M D(\rho^m)(\nabla\varphi)\,d\omega$$
$$= m\int_M D(\rho^{m-1})(\nabla\varphi)\,d\mu,$$

we obtain

$$\lim_{t\to 0}\frac{H_m(\mu|\nu)-H_m(\mu_t|\nu)}{t} = \int_M D\biggl(\frac{\rho^{m-1}-\sigma^{m-1}}{1-m}\biggr)(\nabla\varphi)\,d\mu.$$

This yields

$$\nabla_W[-H_m(\cdot|\nu)](\mu) = \nabla\biggl(\frac{\rho^{m-1}-\sigma^{m-1}}{1-m}\biggr)$$

provided that $(\rho^{m-1}-\sigma^{m-1})/(1-m)\in H^1(M,\mu)$. If $(\rho^{m-1}-\sigma^{m-1})/(1-m)\notin H^1(M,\mu)$, then we find

$$\limsup_{\tilde\mu\to\mu}\frac{H_m(\mu|\nu)-H_m(\tilde\mu|\nu)}{W_2(\mu,\tilde\mu)} = \infty$$

by approximating $\rho^{m-1}-\sigma^{m-1}$ with $\phi\in C^\infty(M)$ and choosing $\varphi=\phi/(1-m)$. Hence $H_m(\cdot|\nu)$ is not differentiable at $\mu$. $\square$

**Theorem 8.4** *Let $(\mu_t)_{t\geq 0}\subset\mathcal{P}_{\mathrm{ac}}(M,\omega)$ be a continuous curve that is locally Lipschitz on $(0,\infty)$, and assume that $\mu_t=\rho_t\omega$ satisfies $\rho_t^m\in H^1(M,\omega)$ as well as $(\rho_t^{m-1}-\sigma^{m-1})/(1-m)\in H^1(M,\mu_t)$ for a.e. $t\in(0,\infty)$. Then*

$$\dot\mu_t = \nabla_W[-H_m(\cdot|\nu)](\mu_t)$$

*holds at a.e. $t\in(0,\infty)$ if and only if $(\rho_t)_{t\geq 0}$ is a weak solution to the reverse porous medium equation of the form*

$$\frac{\partial\rho}{\partial t} = -\operatorname{div}_\omega\biggl[\rho\nabla\biggl(\frac{\rho^{m-1}-\sigma^{m-1}}{1-m}\biggr)\biggr]. \tag{8.2}$$

*Proof.* If $\dot\mu_t=\nabla_W[-H_m(\cdot|\nu)](\mu_t)$ holds, then Proposition 8.3 yields

$$\dot\mu_t = \nabla\biggl(\frac{\rho_t^{m-1}-\sigma^{m-1}}{1-m}\biggr).$$

Then it follows from the continuity equation (8.1) that

$$\int_0^\infty\int_M\frac{\partial\phi_t}{\partial t}\,d\mu_t dt = -\int_0^\infty\int_M D\phi_t\biggl[\nabla\biggl(\frac{\rho_t^{m-1}-\sigma^{m-1}}{1-m}\biggr)\biggr]\,d\mu_t dt$$

for all $\phi\in C_c^\infty([0,\infty)\times M)$. Therefore $\rho_t$ weakly solves (8.2).

Conversely, if $\rho_t$ is a weak solution to (8.2), then the same calculation implies that

$$\Phi_t = \nabla\left(\frac{\rho_t^{m-1} - \sigma^{m-1}}{1-m}\right)$$

satisfies the continuity equation (8.1). Therefore Proposition 8.3 shows $\dot{\mu}_t = \Phi_t = \nabla_W[-H_m(\cdot|\nu)](\mu_t)$. $\Box$

We meant by the 'reverse' porous medium equation the porous medium equation with respect to the *reverse* Finsler structure $\overleftarrow{F}(v) := F(-v)$. As the gradient vector for $\overleftarrow{F}$ is written by $\overleftarrow{\nabla} u = -\nabla(-u)$, (8.2) is indeed rewritten as

$$\frac{\partial \rho}{\partial t} = \mathrm{div}_\omega\left[\rho \overleftarrow{\nabla}\left(\frac{\rho^{m-1} - \sigma^{m-1}}{m-1}\right)\right] = \mathrm{div}_\omega\left[\rho \overleftarrow{\nabla}\left(\frac{\rho^{m-1}}{m-1} + \Psi\right)\right].$$

# 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.

[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.

[AGS] L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures, Birkhäuser Verlag, Basel, 2005.

[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.

[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.

[CMS] D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, A Riemannian interpolation inequality á la Borell, Brascamp and Lieb, Invent. Math. **146** (2001), 219–257.

[FG] A. Figalli and N. Gigli, Local semiconvexity of Kantorovich potentials on non-compact manifolds, To appear in ESAIM Control Optim. Calc. Var.

[Gi] N. Gigli, On the inverse implication of Brenier-McCann theorems and the structure of $(\mathscr{P}_2(M), W_2)$, Preprint (2009). Available at http://cvgmt.sns.it/people/gigli/

[GO] N. Gigli and S. Ohta, First variation formula in Wasserstein spaces over compact Alexandrov spaces, Preprint (2010). Available at http://www.math.kyoto-u.ac.jp/~sohta/

[JKO] R. Jordan, D. Kinderlehrer and F. Otto, The variational formulation of the Fokker-Planck equation, SIAM J. Math. Anal. **29** (1998), 1–17.

[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.

[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.

[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.

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

[Ot] F. Otto, The geometry of dissipative evolution equation: 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.

[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), 1–18.

[Sh] Z. Shen, Lectures on Finsler geometry, World Scientific Publishing Co., Singapore, 2001.

[St1] K.-T. Sturm, On the geometry of metric measure spaces, Acta Math. **196** (2006), 65–131.

[St2] K.-T. Sturm, On the geometry of metric measure spaces. II, Acta Math. **196** (2006), 133–177.

[Ta1] A. Takatsu, Wasserstein geometry of Gaussian measures, Preprint (2008). Available at http://www.math.tohoku.ac.jp/~sa6m21/

[Ta2] A. Takatsu, Wasserstein geometry of porous medium equation, Preprint (2010). Available at http://www.math.tohoku.ac.jp/~sa6m21/

[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.