数理解析研究所講究録
第 1896 巻 2014 年 98-114

Some evolution equations as Wasserstein gradient flows

Asuka Takatsu (takatsu@math.nagoya-u.ac.jp)

Graduate School of Mathematics, Nagoya University

Abstract

In the workshop, I demonstrated that a certain evolution equation on a weighted Riemannian manifold can be considered as a Wasserstein gradient flow (the talk was based on [7], where we used the notions of the information geometry). In this note, I discuss the usefulness of the information geometry in the Wasserstein geometry, especially its gradient flow structure.

1 Introduction

In [7], under appropriate conditions, we regard the evolution equation

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

on a weighted Riemannian manifold (M,ω) as the gradient flow of the functional E_φ^Ψ on the Wasserstein space $(\mathcal{P}_2(M), W_2)$. Here $M = (M,g)$ is a Riemannian manifold and $\omega = e^{-f}\,\mathrm{vol}_g$ is a positive measure on M, where $f \in C^\infty(M)$ and vol_g is the Riemannian volume measure on (M,g). The weighted divergence div_ω is defined for a vector field X on M by $\mathrm{div}_\omega(X) := \mathrm{div}(X) - g(X,\nabla f)$, where ∇ is the gradient and div is the divergence on (M,g), respectively. In the right-hand side of the evolution equation, φ is a continuous, non-decreasing, positive function on $(0,\infty)$ and Ψ is a function on M. The Wasserstein space $(\mathcal{P}_2(M), W_2)$ is a pair of the space $\mathcal{P}_2(M)$ of probability measures on (M,g) having finite second moment with its distance function W_2 which has its root in the optimal transport theory. The functional E_φ^Ψ on $\mathcal{P}_2(M)$ is the summation of the internal energy E_φ generating by $f_\varphi(r) := \int_0^r\int_1^t 1/\varphi(s)dsdt$ and the potential energy E^Ψ generating by Ψ. To be precise, for $\mu = \rho\omega \in \mathcal{P}_2(M)$, these energies are respectively defined by

$$E_\varphi(\mu) = \int_M f_\varphi(\rho)d\omega, \qquad E^\Psi(\mu) = \int_M \Psi d\mu.$$

It may be said that this interoperation of the evolution equation as a gradient flow is obtained by generalizing the case of the heat equation which is the case of $\varphi(s) = s$ via the information geometry. We expect that the use of the information geometry shall shed new light on the analysis of evolution equations. In this note, we explain the notions of the information geometry and its role in the Wasserstein gradient flow structure. For the emphasis on the usefulness of the information geometry, we discuss only the Euclidean case. Also some arguments are not rigorous for the sake of simplicity.

2 Wasserstein geometry

In this section, we first give the definition and some basic properties of the Wasserstein geometry, then review the Wasserstein gradient flow. The Wasserstein geometry is a metric geometry on the space of probability measures over a complete, separable, metric space. However, in this note, we restrict our attention to absolutely continuous probability measures on $\mathbb{R}^d$ with respect to the d-dimensional Lebesgue measure, and we moreover identify such a probability measure with its density function.

2.1 Basic properties

Let $\mathcal{P}^d$ be the space of non-negative, integrable functions on $\mathbb{R}^d$ having unit mass and finite second moment, that is,

$$\mathcal{P}^d := \left\{ \rho \in L^1(\mathbb{R}^d) \;\middle|\; \rho(x) \geq 0 \text{ a.e. } x \in \mathbb{R}^d, \int_{\mathbb{R}^d} \rho(x)dx = 1, \int_{\mathbb{R}^d} |x|^2 \rho(x)dx < \infty \right\}.$$

In this note, the integrability on $\mathbb{R}^d$ is with respect to the d-dimensional Lebesgue measure if not otherwise specified.

The (L^2-) *Wasserstein distance* between $\rho, \sigma \in \mathcal{P}^d$ is defined as

$$W_2(\rho, \sigma) = \inf_T \left(\int_{\mathbb{R}^d} |x - T(x)|^2 \rho(x)\, dx \right)^{1/2}, \tag{2.1}$$

where T runs over all measurable maps on $\mathbb{R}^d \to \mathbb{R}^d$ pushing ρ forward to σ. We say that a measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ *pushes* ρ *forward to* σ, denoted by $T_\sharp \rho = \sigma$, if

$$\int_{\mathbb{R}^d} \xi(x)\sigma(x)dx = \int_{\mathbb{R}^d} \xi(T(x))\rho(x)dx$$

holds for any non-negative function ξ on $\mathbb{R}^d$. For any $\rho, \sigma \in \mathcal{P}^d$, there exist a unique minimizer of the variational problem (2.1).

Theorem 2.1 ([2]) *Given $\rho, \sigma \in \mathcal{P}^d$, there exist a measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $T_\sharp \rho = \sigma$ and*

$$W_2(\rho, \sigma) = \left(\int_{\mathbb{R}^d} |x - T(x)|^2 \rho(x)\, dx \right)^{1/2}.$$

In addition, this map T is uniquely determined ρ-almost everywhere.

A minimizer T of the variational problem (2.1) is called an *optimal transport* between ρ and σ. Thus the variational problem (2.1) is solved, and moreover, W_2 is indeed a distance function on $\mathcal{P}^d$.

Theorem 2.2 ([14, Theorem 7.3]) *The pair $(\mathcal{P}^d, W_2)$ is a metric space.*

It is known that any two points $\rho, \sigma \in \mathcal{P}^d$ are joined by a unique length minimizing curve with respect to the Wasserstein distance function. The unique curve is generating by the optimal transport T between them. To be precise, set $T_t(x) := (1-t)x + tT(x)$ and $\rho_t := T_{t\sharp}\rho$. Then the curve $\{\rho_t\}_{t\in[0,1]} \subset \mathcal{P}^d$ is a unique length minimizing curve from ρ to σ with respect to the Wasserstein distance function.

We also mention the relation of convergences in $\mathcal{P}^d$ with respect to the Wasserstein distance function and the weak topology. For a sequence $\{\rho_n\}_{n\in\mathbb{N}} \subset \mathcal{P}^d$ and $\rho_\infty \in \mathcal{P}^d$, we say that $\{\rho_n\}_{n\in\mathbb{N}}$ *weakly converges to* ρ_∞ as $n \to \infty$ if it holds for any bounded continuous function ξ on $\mathbb{R}^d$ that

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \xi(x)\rho_n(x)dx = \int_{\mathbb{R}^d} \xi(x)\rho_\infty(x)dx.$$

Proposition 2.3 ([14, Theorem 7.12]) *For a sequence $\{\rho_n\}_{n\in\mathbb{N}} \subset \mathcal{P}^d$ and $\rho_\infty \in \mathcal{P}^d$, the following two conditions* (1) *and* (2) *are equivalent to each other:*

(1) $\lim_{n\to\infty} W_2(\rho_n, \rho_\infty) = 0$.

(2) $\{\rho_n\}_{n\in\mathbb{N}}$ *weakly converges to ρ_∞ as $n \to \infty$ and we have*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |x|^2\rho_n(x)dx = \int_{\mathbb{R}^d} |x|^2\rho_\infty(x)dx.$$

We remark that $(\mathcal{P}^d, W_2)$ is not complete. For example, let us consider the fundamental solution

$$g_t(x) := (4\pi t)^{-d/2} \exp\left(-\frac{|x|^2}{4t}\right)$$

of the heat equation. As $t \to 0$, g_t converges to the Dirac measure supported at the origin with respect to the Wasserstein distance function.

2.2 Gradient flow

As we mentioned in the previous subsection, any two points in $\mathcal{P}^d$ are joined by a unique length minimizing curve with respect to the Wasserstein distance function. This enables us to define the gradient of a functional on $(\mathcal{P}^d, W_2)$ via directional derivative, and to discuss its Wasserstein gradient flow. For a functional F on $\mathcal{P}^d$ and $\rho \in \mathcal{P}^d$, a curve $\{\rho_t\}_{t\in[0,l)} \subset \mathcal{P}^d$ is the *Wasserstein gradient flow* of F starting at ρ if we have

$$\frac{\partial}{\partial t}\rho_t = -\operatorname{grad} F(\rho_t)$$

for any $t \in (0, l)$ and $\rho_0 = \rho$, where $\operatorname{grad} F$ stands for the Wasserstein gradient of F. For example, see [9] for the details.

Let us first formulate tangent vectors at $\rho \in \mathcal{P}^d$, that is, the velocities of curves $\{\rho_t\}_{t\in(-\varepsilon)}$ with $\rho_0 = \rho$ at $t = 0$. For a curve $\{\rho_t\}_{t\in(-\varepsilon,\varepsilon)} \subset \mathcal{P}^d$ with $\rho_0 = \rho$, there exists a unique (up to additive constant) function ϕ on $\mathbb{R}^d$ satisfying

$$\int_{\mathbb{R}^d} \langle \nabla\xi(x), \nabla\phi(x)\rangle \rho(x)\, dx = \left.\frac{d}{dt}\right|_{t=0} \int_{\mathbb{R}^d} \xi(x)\rho_t(x)\, dx$$

for all smooth functions ξ on $\mathbb{R}^d$ with compact support, where ∇ is the gradient on $\mathbb{R}^d$. This can be interpreted as that ϕ is a solution of the elliptic equation of the form

$$-\operatorname{div}(\rho\nabla\phi) = \left.\frac{\partial\rho_t}{\partial t}\right|_{t=0}, \tag{2.2}$$

where div is the divergence on $\mathbb{R}^d$. Conversely, for a suitable function ϕ on $\mathbb{R}^d$ and $\rho \in \mathcal{P}^d$, there exists a unique curve $\{\rho_t\}_{t\in(-\varepsilon,\varepsilon)}$ satisfying (2.2) and $\rho_0 = \rho$. This yields that the velocity $\dot{\rho}_0$ of $\{\rho_t\}_{t\in(-\varepsilon,\varepsilon)}$ at $t = 0$, namely the tangent vector at ρ, can be considered as $-\operatorname{div}(\rho\nabla\phi)$ with the solution ϕ of (2.2). We thus identify the tangent space $T_\rho\mathcal{P}^d$ with the metric completion of the space defined by

$$\{v := -\operatorname{div}(\rho\nabla\phi) \mid \phi : \text{suitable function on } \mathbb{R}^d\}$$

with respect to the norm $\|\cdot\|_\rho$ induced by the scalar product $\langle\cdot,\cdot\rangle_\rho$ which is defined for $v_1 := -\operatorname{div}(\rho\nabla\phi_1)$ and $v_2 := -\operatorname{div}(\rho\nabla\phi_2)$ by

$$\langle v_1, v_2\rangle_\rho := \int_{\mathbb{R}^d} \langle\nabla\phi_1(x), \nabla\phi_2(x)\rangle\, \rho(x)\, dx.$$

By Benamou–Brenier formula, the 'Riemannian' distance function of $(\mathcal{P}^d, \langle\cdot,\cdot\rangle_*)$ coincides with the Wasserstein distance function W_2.

Theorem 2.4 ([1, Theorem 4.1]) *For any $\rho_0, \rho_1 \in \mathcal{P}^d$ with suitable conditions, we have*

$$W_2(\rho_0, \rho_1)^2$$
$$= \inf\left\{\int_0^1 \|\dot{\rho}_t\|_{\rho_t}^2 dt \;\middle|\; \{\rho_t\}_{t\in[0,1]} \subset \mathcal{P}^d \text{ is a curve from } \rho_0 \text{ to } \rho_1 \text{ with the velocity } \dot{\rho}_t \text{ at } t\right\}.$$

Using this expression, let us explain the gradient of internal energies and potential energies on $(\mathcal{P}^d, W_2)$.

We first consider the internal energy E_f generating by $f \in C[0,\infty) \cap C^2(0,\infty)$, which is defined for $\rho \in \mathcal{P}^d$ by

$$E_f(\rho) := \int_{\mathbb{R}^d} f(\rho(x))dx.$$

For any curve $\{\rho_t\}_{t\in(-\varepsilon,\varepsilon)} \subset \mathcal{P}^d$ with $\rho_0 = \rho$ and the velocity $\dot{\rho}_0 = -\operatorname{div}(\rho\nabla\phi)$, we directly compute

$$\begin{aligned}
\left.\frac{d}{dt}E_f(\rho_t)\right|_{t=0} &= \int_{\mathbb{R}^d}\left[\left.\frac{\partial}{\partial t}f(\rho_t)\right|_{t=0}\right]dx = \int_{\mathbb{R}^d}\left[f'(\rho)\cdot\left.\frac{\partial}{\partial t}\rho_t\right|_{t=0}\right]dx \\
&= -\int_{\mathbb{R}^d}[f'(\rho)\operatorname{div}(\rho\nabla\phi)]\,dx = \int_{\mathbb{R}^d}\langle\nabla f'(\rho), \nabla\phi\rangle\,\rho\,dx \\
&= \left\langle -\operatorname{div}(\rho\nabla f'(\rho)), \dot{\rho}_0\right\rangle_\rho.
\end{aligned}$$

On the other hand, the Riemannian calculus gives

$$\left.\frac{d}{dt}E_f(\rho_t)\right|_{t=0} = \operatorname{Diff}E_f|_\rho(\dot{\rho}_0) = \langle\operatorname{grad} E_f, \dot{\rho}_0\rangle_\rho,$$

where $\mathrm{Diff} E_f$ is the differential map of E_f on $\mathcal{P}^d$. Since the tangent vector $\dot{\rho}_0 \in T_\rho \mathcal{P}^d$ is arbitrary, we have

$$\operatorname{grad} E_f|_\rho = -\operatorname{div}\left(\rho \nabla f'(\rho)\right).$$

We next consider the potential energy E^Ψ generating by $\Psi \in C^2(\mathbb{R}^d)$, which is defined for $\rho \in \mathcal{P}^d$ by

$$E^\Psi(\rho) := \int_{\mathbb{R}^d} \Psi(x)\rho(x)dx.$$

Similarly, for any curve $\{\rho_t\}_{t\in(-\varepsilon,\varepsilon)} \subset \mathcal{P}^d$ with $\rho_0 = \rho$ and the velocity $\dot{\rho}_0 = -\operatorname{div}(\rho\nabla\phi)$, we find that

$$\begin{aligned} \left.\frac{d}{dt}E^\Psi(\rho_t)\right|_{t=0} &= \int_{\mathbb{R}^d}\left[\Psi\cdot\left.\frac{\partial}{\partial t}\rho_t\right|_{t=0}\right]dx = -\int_{\mathbb{R}^d}\left[\Psi\operatorname{div}(\rho\nabla\phi)\right]dx = \int_{\mathbb{R}^d}\langle\nabla\Psi,\nabla\phi\rangle\,\rho\,dx \\ &= \left\langle -\operatorname{div}\left(\rho\nabla\Psi\right), \dot{\rho}_0\right\rangle_\rho, \end{aligned}$$

and

$$\left.\frac{d}{dt}E^\Psi(\rho_t)\right|_{t=0} = \mathrm{Diff}E^\Psi|_\rho(\dot{\rho}_0) = \left\langle \operatorname{grad} E^\Psi, \dot{\rho}_0\right\rangle_\rho,$$

which implies

$$\operatorname{grad} E^\Psi|_\rho = -\operatorname{div}\left(\rho\nabla\Psi\right).$$

In this way, we find that, for a Wasserstein gradient flow $\{\rho_t\}_{t\in[0,l)}$ of $E_f^\Psi := E_f + E^\Psi$,

$$\frac{\partial}{\partial t}\rho_t = \operatorname{div}\left(\rho_t\nabla f'(\rho_t) + \rho_t\nabla\Psi\right)$$

holds for any $t \in (0,l)$.

We next discuss the convexity of a functional which plays an important role in an asymptotic analysis in its gradient flow since if a functional is convex, then its gradient flow has a contraction property (for instance, see [7] and references therein). To do this, we define the lower bound of the Hessian of functionals on $(\mathcal{P}^d, W_2)$. Recall that, for $\Psi \in C^2(\mathbb{R}^d)$ and $K \in \mathbb{R}$, the Hessian of Ψ is bounded below by K, namely

$$\mathrm{Hess}_x\,\Psi(v,v) \geq K|v|^2$$

holds for any $x, v \in \mathbb{R}^d$ if and only if we have

$$\Psi((1-t)x+ty) \leq (1-t)\Psi(x) + t\Psi(y) - \frac{K}{2}t(1-t)|x-y|^2$$

for any $x, y \in \mathbb{R}^d$ and $t \in [0,1]$ (see [14, §2.1.3], for instance).

Definition 2.5 Given $K \in \mathbb{R}$, we say that a functional $F : \mathcal{P}^d \to (-\infty, \infty]$ is *displacement K-convexity*, if, for any length minimizing curve $\{\rho_t\}_{t\in[0,1]} \subset \mathcal{P}^d$ with constant speed,

$$F(\rho_t) \leq (1-t)F(\rho_0) + tF(\rho_1) - \frac{K}{2}(1-t)tW_2(\rho_0,\rho_1)^2$$

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

As for the displacement convexity of internal energies and potential energies, the following criteria are known.

Theorem 2.6 ([6]) (1) *Let f be a positive, convex function on $(0,\infty)$. Assume that f is C^2 on $(0,\infty)$ and satisfies $\lim_{r\downarrow 0} f(r)=0$. If moreover the function defined by*

$$r \mapsto \frac{rf'(r)-f(r)}{r^{1-\frac{1}{d}}}$$

is non-decreasing on $(0,\infty)$, then the internal energy E_f generating by f is displacement 0-convex on $\mathcal{P}^d$.
(2) *For $\Psi \in C^2(\mathbb{R}^d)$, if the Hessian of Ψ is bounded below by $K \in \mathbb{R}$, then the potential energy E^Ψ generating by Ψ is displacement K-convex on $\mathcal{P}^d$.*

For the internal energy E_f generating by f, $\psi_f(r) := rf'(r) - f(r)$ is called the *pressure function* of f. As mentioned in [10], the Wasserstein gradient flow of E_f is written as

$$\frac{\partial}{\partial t}\rho(t,x) = \Delta(\psi_f(\rho(t,x))).$$

3 Example

In this section, we see the evolution equation of the form

$$\frac{\partial}{\partial t}\rho = \frac{1}{2-q}\Delta(\rho^{2-q}) + \operatorname{div}(\rho\nabla\Psi)$$

on $\mathbb{R}^d$, where $d \geq 2$, $q \in (0,(d+1)/d)$ and $\Psi \in C^2(\mathbb{R}^d)$. If Ψ is a constant function, namely without drift, the evolution equation is called the the *fast diffusion equation* for $q>1$, the *porous medium equation* for $q<1$, and the *heat equation* for $q=1$. We remark that, in [5], the heat equation is regarded as a Wasserstein gradient flow, where they used a time-discrete iterative variational scheme. On the other hand, in [8], the fast diffusion equation, the porous medium equation, and the heat equation are interpreted as Wasserstein gradient flows by using the Riemannian structure of the Wasserstein space, where the interpretation of the Riemannian structure differs from one given in Section 2.

3.1 Heat equation

Take $f(r) := r\log(r)$. We then have

$$\psi_f(r) := rf'(r) - f(r) = r,$$

and the function

$$r \mapsto \frac{\psi_f(r)}{r^{1-1/d}} = r^{1/d}$$

is trivially non-decreasing on $(0,\infty)$. Thus the internal energy

$$E_f(\rho) := \int_{\mathbb{R}^d} \rho\log(\rho)dx$$

is displacement 0-convex. In this case, E_f is called the *Boltzmann entropy* with negative sign. Recall that a minimizer of E_f on $\mathcal{P}^d$ with the mean and the covariance constraints is a Gaussian measure, which is characterized by the exponential function. Needless to say, the exponential function $\exp(t)$ is a solution of the ordinary differential equation

$$\frac{d}{dt}y(t) = y(t), \quad y(0) = 1.$$

A typical example of Gaussian densities is the fundamental solution

$$(4\pi t)^{-\frac{d}{2}} \exp\left(-\frac{|x|^2}{4t}\right)$$

of the heat equation

$$\frac{\partial}{\partial t}u(t,x) = \Delta u(x,t) \tag{3.1}$$

on $\mathbb{R}^d$.

More generally, for any function $\Psi \in C^2(\mathbb{R}^d)$ whose Hessian is bounded below by $K > 0$, there exists $c \in \mathbb{R}$ such that $\sigma := \exp(-\Psi + c) \in \mathcal{P}^d$ and

$$\inf_{\rho \in \mathcal{P}^d} E_f^\Psi(\rho) \geq E_f^\Psi(\sigma)$$

holds, where we set $E_f^\Psi := E_f + E^\Psi$.

The functional H_f^Ψ on $\mathcal{P}^d$ defined by

$$H_f^\Psi(\rho) := E_f^\Psi(\rho) - E_f^\Psi(\sigma) = \int_{\mathbb{R}^d} \rho \log\left(\frac{\rho}{\sigma}\right) dx$$

is called the *relative entropy* of ρ with respect to σ. The non-negativity of H_f^Ψ also follows from the convexity of f since we have

$$H_f^\Psi(\rho) = \int_{\mathbb{R}^d} [f(\rho) - f(\sigma) - f'(\sigma)(\rho - \sigma)]\, dx.$$

The relative entropy H_f^Ψ is displacement K-convex and its Wasserstein gradient flow is a solution of the Fokker–Planck equation given by

$$\frac{\partial}{\partial t}\rho(t,x) = \Delta\rho + \operatorname{div}(\rho\nabla\Psi). \tag{3.2}$$

Formally, H_f^Ψ is decreasing along its Wasserstein gradient flow ρ_t in time $t > 0$ and any Wasserstein gradient flow of H_f^Ψ asymptotically approaches to σ.

If we take $\Psi(x) = |x|^2/2$, then $c = -\log(2\pi)^{d/2}$ and σ is the Lebesgue density of the standard Gaussian measure. In this case, for a solution $\rho(t,x)$ of the Fokker–Planck equation (3.2), the function

$$u(t,x) := (1+2t)^{-d/2} \cdot \rho\left(\frac{1}{2}\log(1+2t), \frac{x}{\sqrt{1+2t}}\right)$$

is a solution of the heat equation (3.1). This time-dependent scaling is well-known, however recently, a different time-dependent scaling is used to analyze asymptotic behavior of certain evolution equations (see [3] and Section 5 below).

3.2 Porous Medium Equation/Fast Diffusion Equation

Fix $q \in (0,1) \cup (1,2)$ and set the function f_q on $[0,\infty)$ by

$$f_q(r) := \frac{r^{2-q} - (2-q)r}{(2-q)(1-q)}.$$

Note that f_q is convex and $f_q(r) \to r \log r$ as $q \to 1$. The direct computation yields

$$\psi_{f_q}(r) := r f_q'(r) - f_q(r) = \frac{r^{2-q}}{2-q}$$

and the function

$$r \mapsto \frac{\psi_{f_q}(r)}{r^{1-\frac{1}{d}}} = \frac{r^{1-q+\frac{1}{d}}}{2-q}$$

is non-decreasing on $(0,\infty)$ if $q \le (d+1)/d$. We remark that E_{f_q} is related to the $(2-q)$-*Tsallis entropy* with negative sign (see [13]).

For $q \in (0,1)\cup(1,(d+4)/(d+2))$, a minimizer of E_{f_q} on $\mathcal{P}^d$ under the mean and the covariance constraints is called the q-Gaussian measure and characterized the q-exponential function $\exp_q$ given by

$$\exp_q(t) := [1 + (1-q)t]_+^{1/(1-q)},$$

where we set $[t]_+ := \max\{t, 0\}$ and by convention $0^a := \infty$ for $a < 0$. The assumption $q < (d+2)/(d+4)$ ensures the finiteness of the second moment of q-Gaussian measures. Note that $\exp_q(t) \to \exp(t)$ as $q \to 1$. The q-exponential function is a solution of the ordinary differential equation given by

$$\frac{d}{dt} y(t) = y(t)^q, \quad y(0) = 1.$$

A typical example of q-Gaussian densities is the self-similar solution

$$ct^{-\frac{d}{d(1-q)+2}} \cdot \exp_q\left(-\lambda |x|^2 / t^{\frac{2}{d(1-q)+2}}\right)$$

of the following evolution equation

$$\frac{\partial}{\partial t} u(t,x) = \frac{1}{2-q} \Delta \left(u(x,t)^{2-q}\right) \tag{3.3}$$

on $\mathbb{R}^d$, where $c, \lambda \in \mathbb{R}$ are constants depending on q and d (for instance, see [11]).

In the rest of this subsection, we always assume $q \in (0,1) \cup (1,(d+1)/d)$ and $d \ge 2$, which guarantees the displacement 0-convexity of E_{f_q} on $\mathcal{P}^d$ and the finiteness of the second moment of q-Gaussian measures on $\mathbb{R}^d$. For any function $\Psi \in C^2(\mathbb{R}^d)$ whose Hessian is bounded below by $K > 0$, there exists $c \in \mathbb{R}$ such that $\sigma := \exp_q(-\Psi + c) \in \mathcal{P}^d$ and

$$E_{f_q}^{\Psi}(\rho) \ge E_{f_q}^{\Psi}(\sigma)$$

holds for any $\rho \in \mathcal{P}^d$ whose support is contained in the support of σ, where we set $E_{f_q}^{\Psi} := E_{f_q} + E^{\Psi}$. The functional $H_{f_q}^{\Psi}$ on $\mathcal{P}^d$ defined by

$$H_{f_q}^{\Psi}(\rho) := E_{f_q}^{\Psi}(\rho) - E_{f_q}^{\Psi}(\sigma)$$

is called the *q-relative entropy* of ρ with respect to σ. In the case of $\rho \in \mathcal{P}^d$ whose support is contained in the support of σ, the non-negativity of $H^{\Psi}_{f_q}$ also follows from the convexity of f_q since we have

$$H^{\Psi}_{f_q}(\rho) = \int_{\mathbb{R}^d} \left[f_q(\rho) - f_q(\sigma) - f_q'(\sigma)(\rho - \sigma)\right] dx.$$

The q-relative entropy $H^{\Psi}_{f_q}$ is displacement K-convex and its Wasserstein gradient flow is a solution of the following evolution equation:

$$\frac{\partial}{\partial t}\rho(t,x) = \frac{1}{2-q}\Delta(\rho^{2-q}) + \operatorname{div}(\rho\nabla\Psi). \tag{3.4}$$

Formally, $H^{\Psi}_{f_q}$ is decreasing along its Wasserstein gradient flow ρ_t in time $t > 0$ and any Wasserstein gradient flow of $H^{\Psi}_{f_q}$ asymptotically approaches to σ.

If we take $\Psi(x) = |x|^2/2$, then σ is the Lebesgue density of a q-Gaussian measure. Moreover, for a solution $\rho(t,x)$ of the evolution equation (3.4), the function

$$u(t,x) := \left(1 + \frac{t}{\alpha}\right)^{-d\alpha} \cdot \rho\left(\alpha \log\left(1 + \frac{t}{\alpha}\right), x \Big/ \left(1 + \frac{t}{\alpha}\right)^{\alpha}\right) \tag{3.5}$$

is a solution of the evolution equation (3.3), where $\alpha = \alpha(q,d) := 1/(d(1-q)+2)$. When $q \to 1$, the evolution equation (3.3) and its self-similar solution recover the heat equation (3.1) and its fundamental solution, respectively. Since $\alpha(1,d) = 1/2$, this time-dependent scaling (3.5) can be extended to the case of $q = 1$.

4 Information geometry

In the previous section, we discus a certain evolution equation from the viewpoint of the Wasserstein gradient flow of the functional E^{Ψ}_f consisting of the internal energy E_f generating by f and the potential energy E^{Ψ} generating by Ψ, where f satisfies the condition in Theorem 2.6(1) and the Hessian of Ψ is bounded below by $K > 0$. Although there are many choice of f, in this section, we introduce the method to generalize $f(r) = r\log(r)$, which is the internal density of the Boltzmann energy, by using the information geometry associated to φ. We refer to [7] and references therein for the details.

In this section, a function $\varphi : (0,\infty) \to (0,\infty)$ is always assumed to be continuos, non-decreasing, positive function with

$$\varphi(0) := \lim_{s \downarrow 0} \varphi(s) = 0, \quad \varphi(1) = 1.$$

Define the *φ-logarithmic function* by

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

for $t \in (0,\infty)$. Since the function $\ln_\varphi$ is clearly increasing, there exists its inverse function on $\ln_\varphi((0,\infty))$. We extend the inverse function to the whole of $\mathbb{R}^d$ by

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

where we set

$$l_\varphi := \inf_{t>0} \ln_\varphi(t) = \lim_{t\downarrow 0} \ln_\varphi(t), \qquad L_\varphi := \sup_{t>0} \ln_\varphi(t) = \lim_{t\uparrow\infty} \ln_\varphi(t).$$

We call $\exp_\varphi$ the *φ-exponential function.* Note that $\exp_\varphi$ is a solution of the ordinary differential equation given by

$$\frac{d}{dt} y(t) = \varphi(y(t)), \quad y(0) = 1.$$

We define a kind of the differentiable coefficient of φ as

$$\theta_\varphi := \sup_{s>0} \left\{ \frac{s}{\varphi(s)} \cdot \limsup_{\varepsilon\downarrow 0} \frac{\varphi(s+\varepsilon) - \varphi(s)}{\varepsilon} \right\} \ge 0.$$

If $\theta_\varphi < 2$, then the function on $(0,\infty)$ defined as

$$f_\varphi(r) := \int_0^r \ln_\varphi(t) dt$$

is well-defined (see [7, Lemma 2.8]). The function f_φ is clearly convex and $f_\varphi(0) = 0$. Set

$$\psi_\varphi(r) := r f_\varphi'(r) - f_\varphi(r) = \int_0^r \int_t^r \frac{1}{\varphi(s)} ds dt = \int_0^r \frac{s}{\varphi(s)} ds$$

as the pressure function of f_φ.

Proposition 4.1 ([7, Theorem 3.5]) *If $\theta_\varphi \le q < 2$, then*

$$r \mapsto \frac{\psi_\varphi(r)}{r^{1-(q-1)}}$$

is non-decreasing on $r \in (0,\infty)$.

This yields that, for φ satisfying

$$\theta_\varphi - 1 \le \frac{1}{d},$$

the internal energy $E_\varphi := E_{f_\varphi}$ is displacement 0-convex on $\mathcal{P}^d$.

Example 4.2 (1) The case of $\varphi(s) = s$ is the most important case. In this case, the φ-logarithmic (resp. φ-exponential) function is the usual logarithmic (resp. exponential) function and

$$l_\varphi = -\infty, \quad L_\varphi = \infty, \quad \theta_\varphi = 1.$$

The convex function f_φ and its pressure function ψ_φ are respectively given by

$$f_\varphi(r) = r\log(r) - r, \qquad \psi_\varphi(r) = r.$$

(2) Another important case is $\varphi_q(s) := s^q$ for $q \in (0,1) \cup (0,2)$, where the φ-logarithmic and the φ-exponential functions are power functions of the form

$$\ln_q(t) := \ln_{\varphi_q}(t) = \frac{t^{1-q} - 1}{1-q}, \quad \exp_q(\tau) := \exp_{\varphi_q}(\tau) = [1 + (1-q)\tau]_+^{\frac{1}{1-q}}.$$

Since $\ln_q(t) \to \log(t)$ and $\exp_q(\tau) \to \exp(\tau)$ hold as $q \to 1$, we denote $\ln_1(t) := \log(t)$ and $\exp_1(\tau) := \exp(\tau)$ for convenience. It is easy to check

$$l_q := l_{\varphi_q} = \begin{cases} -\dfrac{1}{1-q} & \text{if } q < 1, \\ -\infty & \text{if } q > 1, \end{cases} \qquad L_q := L_{\varphi_q} = \begin{cases} \infty & \text{if } q < 1, \\ -\dfrac{1}{1-q} & \text{if } q > 1, \end{cases} \qquad \theta_{\varphi_q} = q,$$

and

$$f_q(r) := f_{\varphi_q}(r) = \frac{r^{2-q} - (2-q)r}{(2-q)(1-q)}, \qquad \psi_q(r) := \psi_{\varphi_q}(r) = \frac{r^{2-q}}{2-q}.$$

It follows from [7, Lemma 2.10] with [11, Proposition 3.2] that, for any $\theta_\varphi - 1 < 2/(d+2)$ and $c > 0$, there exists $\lambda \in (l_\varphi, L_\varphi)$ such that $\exp_\varphi(\lambda - c|x|^2) \in \mathcal{P}^d$.

In the rest of this section, we assume that $\theta_\varphi - 1 < 1/d$ and $d \geq 2$. Then by Proposition 4.1, the internal energy E_φ generating by f_φ is displacement 0-convex. Recall that E_φ is a functional on $\mathcal{P}^d$ defined by

$$E_\varphi(\rho) := \int_{\mathbb{R}^d} f_\varphi(\rho(x))dx.$$

Fix any function $\Psi \in C^2(\mathbb{R}^d)$ whose Hessian is bounded below by $K > 0$ and

$$\inf_{x \in \mathbb{R}^d} \Psi \geq -L_{\theta_\varphi}.$$

Due to [7, Lemma 4.5], we may assume that $\sigma := \exp_\varphi(-\Psi) \in \mathcal{P}^d$ without loss of generality. Note that the support $\mathrm{supp}(\sigma)$ of σ coincides with the closure of $\Psi^{-1}(-L_\varphi, -l_\varphi)$ and σ is the unique minimize of

$$E_\varphi^\Psi(\rho) := E_\varphi(\rho) + E^\Psi(\rho) = \int_{\mathbb{R}^d} [f_\varphi(\rho(x)) + \Psi(x)\rho(x)]dx$$

on the convex subset $\mathcal{P}^d_{\Psi,\varphi}$ of $(\mathcal{P}^d, W_2)$ defined by

$$\mathcal{P}^d_{\Psi,\varphi} := \{\rho \in \mathcal{P}^d \mid \text{the support of } \rho \text{ is contained in the closure of } \Psi^{-1}(-L_\varphi, -l_\varphi)\ \}.$$

The minimality of $E_\varphi^\Psi(\sigma)$ follows from the strict convexity of f_φ and the fact $\tau = \ln_\varphi(\exp_\varphi(\tau)) = f'_\varphi(\exp(\tau))$ for $\tau \in (l_\varphi, L_\varphi)$. Precisely, we compute

$$\begin{aligned} E_\varphi^\Psi(\rho) - E_\varphi^\Psi(\sigma) &= \int_{\mathrm{supp}(\sigma)} [f_\varphi(\rho) - f_\varphi(\sigma) + \Psi(\rho - \sigma)]dx \\ &= \int_{\mathrm{supp}(\sigma)} [f_\varphi(\rho) - f_\varphi(\sigma) - f'_\varphi(\sigma)(\rho - \sigma)]dx \geq 0. \end{aligned}$$

This means that, under the mean and the covariance constraints (with support condition), the minimizer of E_φ is characterized by the φ-exponential function.

We mention the amount given by

$$D_\varphi(\rho_0|\rho_1) := \int_{\mathbb{R}^d} [f_\varphi(\rho_0) - f_\varphi(\rho_1) - f'_\varphi(\rho_1)(\rho_0 - \rho_1)]dx$$

is called the *divergence* in the information geometry, which behaves like the squared distance function. In this note, we call the functional on $\mathcal{P}^d_{\Psi,\varphi}$ defined by

$$H_\varphi^\Psi(\rho) := D_\varphi(\rho|\sigma) = E_\varphi^\Psi(\rho) - E_\varphi^\Psi(\sigma) \geq 0$$

the *φ-relative entropy* with respect to σ.

Remark 4.3 Take $\varphi(s) = s$ (resp. $\varphi_q(s) = s^q$), then H_φ^Ψ coincides with the classical relative entropy (resp. q-relative entropy).

Since the φ-relative entropy is displacement K-convex on $\mathcal{P}^d_{\Psi,\varphi}$, its Wasserstein gradient flow

$$\frac{\partial}{\partial t}\rho = \mathrm{div}\left(\rho \nabla f'_\varphi(\rho)\right) + \mathrm{div}\left(\rho \nabla \Psi\right) = \Delta(\psi_\varphi(\rho)) + \mathrm{div}\left(\rho \nabla \Psi\right)$$

may have the K-contraction property with respect to the Wasserstein distance function. In other words, it holds

$$W_2(\rho_t, \tilde{\rho}_t) \leq e^{-Kt} W_2(\rho_0, \tilde{\rho}_0)$$

for any solutions $\rho_t(x) = \rho(t,x)$, $\tilde{\rho}_t(x) = \tilde{\rho}(t,x)$ of the above evolution equation and $t > 0$. Moreover, H_φ^Ψ is decreasing along its Wasserstein gradient flow in time $t > 0$. In [7, Sections 8,9], we discuss it in the setting of a weighted Riemannian manifold and there are many researches in the setting of the Euclidean case (without notions of information theory). For example, see [4].

We close this section with comments of the advantage obtained by using the information geometry. As mentioned before, the φ-relative entropy behaves as the squared distance function in the context of the information geometry. Then it is natural to compare the two 'distance' functions, the Wasserstein distance function W_2 and the square roof to the φ-relative entropy. Assume that φ satisfies the condition in Theorem 2.6(1) and the Hessian of $\Psi \in C^2(\mathbb{R}^d)$ is bounded below by $K > 0$, which guarantees the existence of a unique minimizer σ of E_φ^Ψ on $\mathcal{P}^d_{\Psi,\varphi}$. We then have

$$W_2(\rho, \sigma)^2 \leq \sqrt{\frac{2}{K} H_\varphi^\Psi(\rho)} \tag{4.1}$$

for any $\rho \in \mathcal{P}^d_{\Psi,\varphi}$ (for the proof, see [7, Section 6]). In the case of $\varphi(s) = s$, namely H^Ψ_φ is the classical relative entropy, the inequality (4.1) is the Talagrand inequality from which we derive the Gaussian concentration inequality for σ. In a similar way, the inequality (4.1) provides the q-Gaussian concentration inequality for σ, where q depends on φ. There are several researches in which the inequality (4.1) was proved without notions of the information geometry, however we may not find such a concentration inequality for the minimizer σ of E^Ψ_φ unless we use the information geometry.

As similar as the variant of Talagrand inequality, the displacement K-convexity of H^Ψ_φ provides a variant of logarithmic Sobolev inequality which compares the φ-relative entropy and the *φ-Fisher information* I^Ψ_φ defined for $\rho \in \mathcal{P}^d_{\Psi,\varphi}$ by

$$I^\Psi_\varphi(\rho) := \int_{\mathbb{R}^d} |\nabla[\ln_\varphi(\rho(x)) - \ln_\varphi(\sigma(x))]|^2 \rho(x)\, dx.$$

To be precise, we have

$$H^\Psi_\varphi(\rho) \leq \frac{1}{2K} I^\Psi_\varphi(\rho)$$

for any $\rho \in \mathcal{P}^d_{\Psi,\varphi}$. If we take $\varphi(s) = s$, we then find

$$I^\Psi_\varphi(\rho) = \int_{\mathbb{R}^d} \left|\nabla \log\left(\frac{\rho(x)}{\sigma(x)}\right)\right|^2 \rho(x)\, dx = 4\int_{\mathbb{R}^d} \left|\nabla\sqrt{\frac{\rho(x)}{\sigma(x)}}\right|^2 \sigma(x)\, dx,$$

that is I^Ψ_φ coincides with the classical Fisher information, and the φ-logarithmic Sobolev inequality recovers the classical logarithmic Sobolev inequality of the form

$$\int_{\mathbb{R}^d} \left(\frac{\rho}{\sigma}\right) \log\left(\frac{\rho}{\sigma}\right) \sigma dx \leq \frac{1}{2K}\int_{\mathbb{R}^d} \left|\nabla \log\left(\frac{\rho}{\sigma}\right)\right|^2 \rho\, dx.$$

Note that if the reference probability measure σ satisfies some convexity condition, then the classical Talagrand inequality and the classical logarithmic Sobolev inequality are equivalent to each other (see [9, Theorem 1]).

In this way, if we introduce the notions of the information geometry to a evolution equation which is realized as the Wasserstein gradient flow of a displacement K-convex functional for $K > 0$, we can easily find its entropy functional and describe the stationary solution. Moreover, we generalize functional inequalities and estimate the concentration function of the stationary solution.

5 Remarks on time-dependent scaling

This section is devoted to explain the time-dependent scaling given in [3] in terms of push-forward by dilations on $\mathbb{R}^d$. Continuously, let $\varphi : (0,\infty) \to (0,\infty)$ be a continuous, non-decreasing, positive function such that $\theta_\varphi - 1 < 1/d$ for some $d \in \mathbb{N}$ with $d \geq 2$. Set f_φ and ψ_φ as the internal energy density and its pressure function defined in Section 4. We discuss the time-dependent scaling of a Wasserstein gradient flow of the internal energy E_φ, that is a solution of the evolution equation given by

$$\frac{\partial}{\partial t}u(x,t) = \Delta(\psi_\varphi(u(x,t))). \tag{5.1}$$

Roughly speaking, this time-dependent scaling is the projection from $\mathcal{P}^d$ to the unit sphere $\mathbb{S}(\mathcal{P}^d)$ with center at the Dirac measure supported at the origin of $\mathbb{R}^d$ in the Wasserstein space, that is,

$$\mathbb{S}(\mathcal{P}^d) := \left\{ \rho \in \mathcal{P}^d \;\middle|\; \int_{\mathbb{R}^d} |x|^2 \rho(x)dx = 1 \right\}.$$

The key of the proof is that the dilation on $\mathbb{R}^d$ induces the dilation on $\mathcal{P}^d$ via the push-forward (see [12]).

Given any $s > 0$, the *dilation* $\delta[s]$ of scale s on $\mathbb{R}^d$ is a map from $\mathbb{R}^d$ to $\mathbb{R}^d$ defined by $\delta[s]x = sx$ for $x \in \mathbb{R}^d$. Similarly, for $s > 0$, we define the map $D[s] : \mathcal{P}^d \to \mathcal{P}^d$ by $D[s](\rho) = \delta[s]_\sharp \rho$ and call the *dilation* of scale s on $\mathcal{P}^d$. By the change of variables, we easily check that, for $\rho \in \mathcal{P}^d$, $u_s := D[s](\rho)$ satisfies

$$\rho(x) = s^d \cdot u_s(sx)$$

for ρ-almost every $x \in \mathbb{R}^d$, or equivalently

$$s^{-d} \cdot \rho(x/s) = u_s(x).$$

Remark 5.1 Using the dilations, we rewrite the time-dependent scaling (3.5) as

$$u_t := D\left[\left(1 + \frac{t}{\alpha}\right)^\alpha\right]\left(\rho_{\alpha \log(1+\frac{t}{\alpha})}\right),$$

where we denote $u_t(x) := u(t,x)$ and $\rho_{\alpha\log(1+\frac{t}{\alpha})}(x) := \rho(\alpha \log(1 + \frac{t}{\alpha}), x)$.

We now see the scaling given in [3], which depends not only time but also initial data. They used the temperature (second moment) of solutions. For any $\rho \in \mathcal{P}^d$, define its *inverse temperature* $\beta[\rho]$ by

$$\beta(\rho) := \left(\int_{\mathbb{R}^d} \frac{|x|^2}{2} \rho(x)dx \right)^{-1}.$$

We then compute

$$\int_{\mathbb{R}^d} \frac{|x|^2}{2} D[\beta(\rho)^{\frac{1}{2}}](\rho)(x)dx = \int_{\mathbb{R}^d} \frac{|\beta(\rho)^{\frac{1}{2}} x|^2}{2} \rho(x)\, dx = 1,$$

which means $D[\beta(\rho)^{1/2}](\rho) \in \mathbb{S}(\mathcal{P}^d)$. In what follows, we assume that the set $\mathcal{S}^d$ given by

$$\left\{ u \in \mathbb{S}(\mathcal{P}^d) \mid \text{there exists a unique global Wasserstein gradient flow of } E_\varphi \text{ starting at } u \right\}$$

is not empty. For $t > 0$, we define the *renormalized flow map* $S[t]$ at time t from $\mathcal{S}^d$ to $\mathbb{S}(\mathcal{P}^d)$ by

$$S[t](u) := D[\beta(u_t)^{\frac{1}{2}}](u_t),$$

where $u_t(x) := u(t,x)$ is the global Wasserstein gradient flow of E_φ starting at u.

Remark 5.2 Even in the case of $\varphi(s) = s^q$ for $q \in (0, (d+1)/d)$, this scaling is generally different from the inverse of the time-dependent scaling (3.5). However, the both scaling for the self-similar solution are similar to each other. For example, if we consider the heat equation, which corresponds to the case of $\varphi(s) = s$, and

$$u(x) := \left(\frac{4\pi}{d}\right)^{-\frac{d}{2}} \exp\left(-\frac{d|x|^2}{4}\right).$$

Then the Wasserstein gradient flow $u_t(x) := u(t,x)$ of E_φ starting at u is the self-similar solution, that is

$$u_t(x) = \left(\frac{4\pi(1+dt)}{d}\right)^{-\frac{d}{2}} \exp\left(-\frac{d|x|^2}{4(1+dt)}\right)$$

and its inverse temperature is $\beta(u_t) = 1/(1+dt)$. Therefore we have

$$D[\beta(u_t)^{\frac{1}{2}}](u_t) \equiv u.$$

On the other hand, if we take

$$v(x) := (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{|x|^2}{2}\right),$$

which is the stationary solution of

$$\frac{\partial}{\partial t}\rho = \Delta\rho + \operatorname{div}(\rho x),$$

then the Wasserstein gradient flow $v_t(x) := v(t,x)$ of E_φ starting at v is given by

$$v_t(x) = (2\pi(1+2t))^{-\frac{d}{2}} \exp\left(-\frac{|x|^2}{2(1+2t)}\right).$$

Applying the inverse time-dependent scaling (3.5), we have

$$D[(1+2t)^{-\frac{1}{2}}](v_t) \equiv v.$$

Usually, the long time asymptotics of the evolution equation (5.1) cannot be characterized by self-similar solutions. However, in [3], they characterized a universal asymptotic profile by fixed points of $S[t]$. To do this, they assumed the following condition:

(NL2) there exists $c > 0$ and $m > (d-2)/d$ such that it holds for all $r > 0$ that

$$\psi_\varphi'(r) = \frac{r}{\varphi(r)} \geq c r^{m-1}.$$

Theorem 5.3 ([3, Theorem 2]) *Suppose (NL2). There exist $t_* > 0$ and a curve $\{v_t\}_{t>t_*} \subset \mathcal{S}^d$ such that, $S[t](v_t) = v_t$ for $t > t_*$ and it holds*

$$\lim_{t\to\infty} W_2(v_t, S[t](u)) = 0$$

for any $u \in \mathcal{S}^d$.

In the proof, they used an $L^1 - L^\infty$ regularizing property which is derived from the condition (NL2).

Theorem 5.4 ([3, Theorem 1]) *If we assume* (NL2), *then for any $u \in \mathcal{S}^d$, the global Wasserstein gradient flow $u_t(x) := u(t,x)$ of E_φ staring at u belongs to $L^\infty(\mathbb{R}^d)$. Moreover, there exists $C > 0$, which does not depend on u, such that*

$$\|u_t\|_\infty \leq Ct^{-\frac{d}{d(m-1)+2}}$$

holds for any $t > 0$.

It is thus important to find such an m in the condition (NL2). From the viewpoint of the information geometry, we may use $2 - \theta_\varphi$ instead of m since we have

$$\psi'_\varphi(r) = \frac{r}{\varphi(r)} \geq \frac{r^{1-\theta_\varphi}}{\varphi(1)}$$

for $r > 1$ according to the following property.

Lemma 5.5 ([7, Lemma 2.10]) *The function $r \mapsto r^{\theta_\varphi}/\varphi(r)$ is non-decreasing on $(0, \infty)$.*

Remark 5.6 In [3], they required another conditions on ψ_φ. However, under the assumption $\theta_\varphi - 1 < 1/d$ and $\mathcal{S}^d \neq \emptyset$, ψ_φ verifies such conditions.

Thus if we use the notions of the information geometry, we may give natural example of ψ_φ which satisfies the conditions in [3].

References

[1] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math.**84** (2000), 375–393.

[2] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math.**44** (1991), 375–417.

[3] J. A. Carrillo, M. Di Francesco and G. Toscani, *Intermediate asymptotics beyond homogeneity and self-similarity: long time behavior for $u_t = \Delta\phi(u)$*, Arch. Ration. Mech. Anal.**180** (2006), 127–149.

[4] J. A. Carrillo, R. J. McCann, and C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Ration. Mech. Anal. **179** (2006), 217–263.

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

[6] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math.**128** (1997), 153–179.

[7] S. Ohta and A. Takatsu, *Displacement convexity of generalized relative entropies. II*, Comm. Anal. Geom. **21** (2013), 687–785.

[8] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), 101–174.

[9] 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.

[10] F. Otto and M. Westdickenberg, *Eulerian calculus for the contraction in the Wasserstein distance*, SIAM J. Math. Anal. **37** (2005), 1227–1255 (electronic).

[11] A. Takatsu, *Behaviors of φ-exponential distributions in Wasserstein geometry and an evolution equation*, SIAM J. Math. Anal. **45** (2013), 2546–2556.

[12] A. Takatsu ahd T. Yokota, *Cone structure of L^2-Wasserstein spaces*, J. Topol. Anal. **04** (2012), 237–253.

[13] C. Tsallis, *Introduction to nonextensive statistical mechanics*, Springer, 2009.

[14] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.