Rev. Mat. Iberoam. (online first)
DOI 10.4171/RMI/1159

Dynamical aspects of the generalized Schrödinger problem via Otto calculus – A heuristic point of view

Ivan Gentil, Christian Léonard and Luigia Ripani

This article is dedicated to the memory of Kazumasa Kuwada.

Abstract. The defining equation

$$(*) \qquad \dot{\omega}_t = -F'(\omega_t)$$

of a gradient flow is *kinetic* in essence. This article explores some *dynamical* (rather than kinetic) features of gradient flows (i) by embedding equation $(*)$ into the family of slowed down gradient flow equations: $\dot{\omega}^\varepsilon_t = -\varepsilon F'(\omega^\varepsilon_t)$, where $\varepsilon > 0$, and (ii) by considering the *accelerations* $\ddot{\omega}^\varepsilon_t$. We shall focus on Wasserstein gradient flows. Our approach is mainly heuristic. It relies on Otto calculus.

A special formulation of the Schrödinger problem consists in minimizing some action on the Wasserstein space of probability measures on a Riemannian manifold subject to fixed initial and final data. We extend this action minimization problem by replacing the usual entropy, underlying the Schrödinger problem, with a general function on the Wasserstein space. The corresponding minimal cost approaches the squared Wasserstein distance when the fluctuation parameter ε tends to zero.

We show heuristically that the solutions satisfy some Newton equation, extending a recent result of Conforti. The connection with Wasserstein gradient flows is established and various inequalities, including evolutional variational inequalities and contraction inequalities under a curvature-dimension condition, are derived with a heuristic point of view. As a rigorous result we prove a new and general contraction inequality for the Schrödinger problem under a Ricci lower bound on a smooth and compact Riemannian manifold.

Mathematics Subject Classification (2010): Primary 37A50; Secondary 49-XX.
Keywords: Schrödinger problem, Wasserstein distance, Otto calculus, Newton equation, curvature-dimension conditions.

Contents

1. Introduction

The defining equation

$$\dot{\omega}_t = -F'(\omega_t), \tag{1.1}$$

of a gradient flow is *kinetic* in essence. This article explores some *dynamical* (rather than kinetic) features of gradient flows (i) by embedding equation (1.1) into the family of slowed down gradient flow equations: $\dot{\omega}^{\varepsilon}_t = -\varepsilon F'(\omega^{\varepsilon}_t)$, where $\varepsilon > 0$, and (ii) by considering the *accelerations* $\ddot{\omega}^{\varepsilon}_t$. We shall focus on Wasserstein gradient flows, i.e., gradient flows with respect to the Wasserstein metric on a space of probability measures. Our approach in this article is mainly heuristic, using Otto calculus.

Otto calculus is a powerful tool to understand the geometry of the Wasserstein space on a Riemannian manifold N. It offers a heuristic for considering the space $\mathcal{P}_2(N)$ of probability measures with finite second moments on the manifold, see (3.1), as an infinite dimensional Riemannian manifold, allowing one to address natural conjectures. Roughly speaking, Otto calculus is twofold. Firstly, it leads to the definition of the Wasserstein metric on $\mathcal{P}_2(N)$ whose corresponding squared Wasserstein distance between two probability measures μ and ν is given by the Benamou–Brenier formula

$$W_2^2(\mu,\nu) = \inf_{((\mu_s),(v_s))} \int_{[0,1]\times N} |v_s(x)|_x^2 \, \mu_s(dx)\, ds$$

where the infimum is taken over all $(\mu_s, v_s)_{0\le s\le 1}$ such that (μ_s) is a trajectory in $\mathcal{P}_2(N)$, starting from μ and arriving at ν and (v_s) is its velocity field, meaning that the transport equation $\partial_s \mu + \mathrm{div}\,(\mu v) = 0$ is satisfied.

Denoting the squared length of the velocity $\dot{\mu} = v$ by

$$|\dot{\mu}_s|^2_{\mu_s} := \inf\Big\{ \int_N |v|^2 \, d\mu_s ; v : \partial_s\mu + \mathrm{div}\,(\mu_s v) = 0 \Big\}, \tag{1.2}$$

we obtain the Riemannian distance like formula

$$W_2^2(\mu,\nu) = \inf_{(\mu_s)} \int_0^1 |\dot{\mu}_s|^2_{\mu_s}\, ds. \tag{1.3}$$

This provides us with natural definitions on $\mathcal{P}_2(N)$ of geodesics, gradients, Hessians and so on. We call the squared distance W_2^2 the *Wasserstein cost.*

Secondly, it appears that several PDEs whose solutions $(\mu_t)_{t\geq 0}$ are flows of probability measures, are gradient flows with respect to the Wasserstein metric, of some function $\mathcal{F}$:

$$\dot\mu_t = -\ \mathrm{grad}_{\mu_t} \mathcal{F}, \tag{1.4}$$

where the velocity $\dot\mu_t$ and the gradient of $\mathcal{F}$ are understood with respect to the Wasserstein metric. For instance it is well known since [34] that the heat equation is the gradient flow with respect to the Wasserstein metric of the usual entropy

$$\mathcal{E}nt(\mu) := \int_N \mu \log \mu \, d\mathrm{vol}\,. \tag{1.5}$$

We introduce a cost function which is a perturbed version of the Wasserstein cost W_2^2. It is defined for any regular function $\mathcal{F}$ on the set of probability measures, any $\varepsilon \geq 0$ and any probability measures μ, ν on the manifold, by

$$\mathcal{A}^\varepsilon_{\mathcal{F}}(\mu,\nu) := \inf_{(\mu_s)} \int_0^1 \Big(\frac{1}{2} |\dot\mu_s|^2_{\mu_s} + \frac{\varepsilon^2}{2} |\, \mathrm{grad}_{\mu_s} \mathcal{F}|^2_{\mu_s} \Big)\, ds,$$

where as in (1.3) the infimum runs through all paths $(\mu_s)_{0\leq s\leq 1}$ in $\mathcal{P}_2(N)$ from μ to ν. Remark that this family of cost functions embeds the Wasserstein cost $W_2^2 = \mathcal{A}^{\varepsilon=0}_{\mathcal{F}}$ as a specific limiting case, see (1.3). This paper investigates basic properties of $\mathcal{A}^\varepsilon_{\mathcal{F}}$ and of its minimizers which are called $\varepsilon\mathcal{F}$-interpolations. It also provides heuristic results which extend to $\mathcal{A}^\varepsilon_{\mathcal{F}}$ several known theorems about the optimal transport cost W_2^2 and the convexity properties of $\mathcal{F}$. The main motivation for introducing $\mathcal{A}^\varepsilon_{\mathcal{F}}$ is that the gradient flow solving

$$\dot\mu_t = -\varepsilon \ \mathrm{grad}_{\mu_t} \mathcal{F}, \tag{1.6}$$

is naturally associated to the Lagrangian $|\dot\mu|^2_\mu/2 + \varepsilon^2 |\,\mathrm{grad}_\mu \mathcal{F}|^2_\mu/2$. Indeed, any solution of (1.6) and any $\varepsilon\mathcal{F}$-interpolation satisfy the same Newton equation

$$\ddot\mu_s = \frac{\varepsilon^2}{2} \mathrm{grad}_{\mu_s} |\, \mathrm{grad}_{\mu_s} \mathcal{F}|^2_{\mu_s}$$

where $\ddot\mu$ denotes the acceleration with respect to the Wasserstein metric and a Wasserstein version of the Levi-Civita connection. When $\varepsilon = 1$, this is the equation of motion of the gradient flow (1.4), while when $\varepsilon = 0$, this is the equation of the free motion in the Wasserstein space characterizing McCann's displacement interpolations.

Let us quote some of our results.

- We denote the solution of (1.4) with initial state μ_0 by the semigroup notation: $\mu_t = \mathsf{S}^{\mathcal{F}}_t(\mu_0)$. The cost $\mathcal{A}^\varepsilon_{\mathcal{F}}$ satisfies the same contraction inequalities along the gradient flow $(\mathsf{S}^{\mathcal{F}}_t)$ as the one satisfied by the Wasserstein cost. The simplest contraction result states that, under a nonnegative Ricci curvature condition,
$$\mathcal{A}^\varepsilon_{\mathcal{F}}(\mathsf{S}^{\mathcal{F}}_t \mu, \mathsf{S}^{\mathcal{F}}_t \nu) \leq \mathcal{A}^\varepsilon_{\mathcal{F}}(\nu,\mu),$$
for any $t \geq 0$ and any probability measures μ, ν. This extends the well-known result by von Renesse and Sturm for the Wasserstein cost [47].

- Newton's equation satisfied by the $\varepsilon\mathcal{F}$-interpolations allows us to prove convexity properties of $\mathcal{F}$ along $\varepsilon\mathcal{F}$-interpolations. For instance, when the Ricci curvature is nonnegative, $\mathcal{F}$ is convex along the $\varepsilon\mathcal{F}$-interpolations. This generalizes McCann's result about the Wasserstein cost [41].

It is remarkable that, similarly to the Wasserstein cost, $\mathcal{A}^\varepsilon_{\mathcal{F}}$ behaves pretty well in presence of the Bakry–Émery curvature-dimension condition, see Section 4.

Schrödinger problem. A particular and fundamental case is when $\mathcal{F}$ is the standard entropy $\mathcal{E}nt$, see (1.5). The associated cost $\mathcal{A}^\varepsilon_{\mathcal{E}nt}$ is related to the Schrödinger problem by a Benamou–Brenier formula, see Section 5. It is identical up to an additive constant to an entropy minimization problem on the path space. The Newton equation satisfied by the $\varepsilon\mathcal{E}nt$-interpolations is a recent result by Conforti [17]. It is a keystone of this field.

Unlike the remainder of this article, our results about the Schrödinger problem are rigorous. In particular we prove that the entropic cost satisfies a general contraction inequality for the heat semigroup $(\mathsf{S}^{\mathcal{E}nt}_t)_{t\geq 0}$ under the assumption that the Ricci curvature is bounded from below by ρ in an n-dimensional Riemannian manifold: for any $t \geq 0$ and any probability measures μ, ν,

$$\mathcal{A}^\varepsilon_{\mathcal{E}nt}(\mathsf{S}^{\mathcal{E}nt}_t\mu, \mathsf{S}^{\mathcal{E}nt}_t\nu) \leq e^{-2\rho t}\mathcal{A}^\varepsilon_{\mathcal{E}nt}(\mu,\nu) - \frac{1}{n}\int_0^t e^{-2\rho(t-u)}(\mathcal{E}nt(\mathsf{S}^{\mathcal{E}nt}_u\mu) - \mathcal{E}nt(\mathsf{S}^{\mathcal{E}nt}_u\nu))^2 du.$$

The paper is organized as follows. In next Section 2 we treat the finite dimensional case where the state space is $\mathbb{R}^n$ equipped with the Euclidean metric. In this case, we are able to do explicitly all the computations using classical differential calculus. In Section 3 we treat the infinite dimensional case using the Otto calculus: we start recalling in a simple way the Otto calculus, then a heuristic derivation of the Newton equation is presented. Convexity properties are explored at Section 4. Finally in Section 5, the special case of the Schrödinger problem is investigated.

We have to mention again that, except for Section 5, all the results in this paper are heuristic, even if we believe that there is a way to prove them rigorously. Heuristic results are denoted with quotation marks. Some parts of the paper are related to other mathematical domains such as Euler equations and mean field games; we tried to extract references to known results from the large related literature. In order to propose a comprehensive document, we do not provide the detailed proofs in the finite dimensional case at Section 2.

2. Warm up in $\mathbb{R}^n$

In this section, $F\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth function ($\mathcal{C}^\infty$) with its first derivative (gradient) and second derivative (Hessian) denoted by F' and F''. All the results about gradient flows which are stated below are well known, see for instance [19].

2.1. Gradient flows in $\mathbb{R}^n$

The equation of a gradient flow: $[0,\infty) \ni t \mapsto \omega_t \in \mathbb{R}^n$ is

$$\dot{\omega}_t = -F'(\omega_t), \quad t \geq 0, \tag{2.1}$$

where $\dot{\omega}_t$ is the time derivative at time t of the path ω. This evolution equation makes sense in a Riemannian manifold if it is replaced by $\dot{\omega}_t = -\operatorname{grad}_{\omega_t} F$.

Remark 2.1. When F is ρ-convex for some $\rho \in \mathbb{R}$, that is

$$F'' \geq \rho \, \mathrm{Id} \tag{2.2}$$

in the sense of quadratic forms, there exists a unique solution of (2.1) for any initial state.

Definition 2.2 (Semigroup)**.** For any $x \in \mathbb{R}^n$, we denote

$$\mathsf{S}_t(x) := \omega_t^x, \quad t \geq 0, x \in \mathbb{R}^n,$$

where $(\omega_t^x)_{t\geq 0}$ is the solution of (2.1) starting form x.

From Remark 2.1, it follows that $(\mathsf{S}_t)_{t\geq 0}$ defines a semigroup, i.e., $\mathsf{S}_{s+t}(x) = \mathsf{S}_t\big(\mathsf{S}_s(x)\big)$ and $\mathsf{S}_0(x) = x$, for all $s, t \geq 0$, and all $x \in \mathbb{R}^n$. This semigroup is called the gradient flow of F (with respect to the Euclidean metric).

For any t, $x \mapsto \mathsf{S}_t(x)$ is continuously differentiable.

Equilibrium state. Any critical point $\hat{x}$ of F is an equilibrium of (2.1) since $F'(\hat{x}) = 0$ implies that $\dot{\omega}_t = 0$ for all t as soon as $\omega_0 = \hat{x}$. If $F''(\hat{x}) > 0$ (in the sense of quadratic forms), then $\hat{x}$ is a stable equilibrium, while when $F''(\hat{x}) < 0$, it is unstable. See Figure 1 for an illustration in dimension one. In the multidimensional case, $F''(\hat{x})$ may admit both stable and unstable directions.

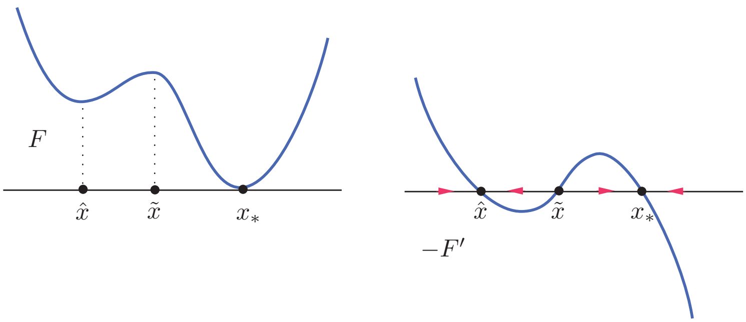

FIGURE 1. Graphical representations of F and $-F'$.

Two metric formulations of gradient flows. The evolution equation (2.1) only makes sense in presence of a differential geometric structure. It is worth rephrasing (2.1) in terms allowing a natural extension to a metric setting. The main idea is to express everything with the scalar quantities $|\dot\omega_t|$ and $|F'(\omega_t)|$ which admit metric analogues.

Proposition 2.3. *The gradient flow equation* (2.1) *is equivalent to*

$$\frac{1}{2}\int_s^t \left(|\dot\omega_r|^2 + |F'(\omega_r)|^2\right) dr \le F(\omega_s) - F(\omega_t), \quad \forall\, 0 \le s \le t.$$

In such case, the curve ω is said to be of maximal slope *with respect to F.*

Theorem 2.4 (Evolution variational inequality)**.**

(a) *The function F is ρ-convex if and only if*

$$\frac{d}{dt}\frac{1}{2}|y - \mathsf{S}_t(x)|^2 + \frac{\rho}{2}|y - \mathsf{S}_t(x)|^2 \le F(y) - F(\mathsf{S}_t(x)), \quad \forall t \ge 0, \forall x, y. \tag{2.3}$$

(b) *If ω is a $\mathcal{C}^1$ path satisfying*

$$\frac{d}{dt}\frac{1}{2}|y - \omega_t|^2 + \frac{\rho}{2}|y - \omega_t|^2 \le F(y) - F(\omega_t), \quad \forall t \ge 0, \forall y, \tag{2.4}$$

then it solves (2.1).

(c) *The function F is ρ-convex if and only if there exists a semigroup (T_t) such that for any x, $\mathsf{T}_t(x)$ is t-differentiable and*

$$\frac{d}{dt}_{|t=0^+}\frac{1}{2}|y - \mathsf{T}_t(x)|^2 + \frac{\rho}{2}|y - x|^2 \le F(y) - F(x), \quad \forall x, y. \tag{2.5}$$

In this case, $\mathsf{T} = \mathsf{S}$.

The evolution variational inequality (2.4) (EVI in short) is a key inequality for extending the notion of gradient flow to general spaces. For instance, it leads to the definition of a gradient flow in a geodesic space, see Definition 23.7 in [55]. This is a reason why many research papers focus on EVI.

2.2. The cost A_F^ε and the εF-interpolations in $\mathbb{R}^n$

In this section we see that gradient flows are special solutions of some Hamilton evolution equations and the related action minimizing problem is considered.

Free energy and the Fisher information. To draw an analogy with the infinite dimensional setting to be explored later on, where $\mathbb{R}^n$ will be replaced by the state space $\mathcal{P}_2(N)$ consisting of all probability measures with a finite second moment on some configuration space described by a Riemannian manifold N, $\mathbb{R}^n$ should be interpreted as the state space (not to be confused with the configuration space) and the function F as the "free energy" of the system. With this analogy in mind, we define the "Fisher information" I by

$$I := |F'|^2,$$

which appears as minus the free energy production along the gradient flow (2.1) since for any $t \geq 0$,

$$\frac{d}{dt}F(\omega_t) = F'(\omega_t) \cdot \dot{\omega}_t = -|F'(\omega_t)|^2 = -I(\omega_t). \tag{2.6}$$

As $I \geq 0$, this implies that $t \mapsto F(\omega_t)$ decreases as time passes. In mathematical terms: F is a Lyapunov function of the system, while with a statistical point of view, this property is an avatar of the second principle of thermodynamics.

Gradient flows as solutions of a Newton equation. Let us introduce a parameter $\varepsilon \geq 0$. For any $x \in \mathbb{R}^n$, the path

$$\omega_t^\varepsilon := \mathsf{S}_{\varepsilon t}(x)$$

satisfies the evolution equation

$$\dot{\omega}_t^\varepsilon = -\varepsilon F'(\omega_t^\varepsilon), \quad t \geq 0.$$

When ε is small, this corresponds to a *slowing down* of (2.1). The acceleration of its solution is given by $\ddot{\omega}_t^\varepsilon = -\varepsilon F''(\omega_t^\varepsilon) \cdot \dot{\omega}_t^\varepsilon = \varepsilon^2 F''F'(\omega_t^\varepsilon)$. Hence ω^ε satisfies the Newton equation

$$\ddot{\omega}^\varepsilon = -U^{\varepsilon\prime}(\omega^\varepsilon) \tag{2.7}$$

with the scalar potential

$$U^\varepsilon := -\frac{\varepsilon^2}{2} I = -\frac{\varepsilon^2}{2} |F'|^2.$$

A graphical representation of $U^{\varepsilon=1} = -|F'|^2/2$ corresponding to the free energy of Figure 1 is given at Figure 2.

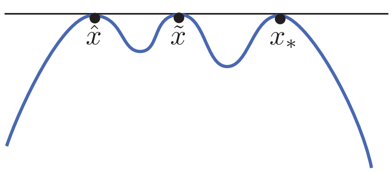

FIGURE 2. Graphical representation of $U = -|F'|^2/2$.

εF-interpolations. It is tempting to investigate the dynamical properties of the trajectories solving the corresponding Hamilton minimization problem. It is associated with the Lagrangian

$$L^\varepsilon(q,v) = \frac{|v|^2}{2} - U^\varepsilon(q) = \frac{|v|^2}{2} + \frac{\varepsilon^2}{2}|F'(q)|^2. \tag{2.8}$$

The associated Hamilton minimization principle is expressed below at (2.9).

Definition 2.5 (εF-cost, εF-interpolations)**.** For any $x, y \in \mathbb{R}^n$ and $\varepsilon \geq 0$, we define

$$A_F^\varepsilon(x,y) = \inf_{\omega \in \Omega^{xy}} \int_0^1 \left(\frac{1}{2}|\dot\omega_s|^2 + \frac{\varepsilon^2}{2}|F'(\omega_s)|^2 \right) ds, \tag{2.9}$$

where the infimum runs through all the subset Ω^{xy} of all paths starting from x and arriving at y. We call $A_F^\varepsilon(x,y)$ the εF-cost between x and y, and any minimizer $\omega^{\varepsilon,xy}$ of (2.9) is called an εF-interpolation between x and y.

- Remark that $A_F^\varepsilon = A_{\varepsilon F}^1$. Therefore, the above definitions of the εF-cost and εF-interpolation could also be called (ε, F)-cost and (ε, F)-interpolation.
- For any $x, y \in \mathbb{R}^n$, $A_F^\varepsilon(x,y) \geq 0$ but unless $\varepsilon = 0$, it is not a squared distance on $\mathbb{R}^n$. For instance whenever $F'(x) \neq 0$, we see that $A_F^\varepsilon(x,x) > 0$.

A key remark about this article is the following.

Remark 2.6 (Gradient flows are εF-interpolations)**.** Suppose that for any $\varepsilon > 0$ and all x, y there exists a unique εF-interpolation $\omega^{\varepsilon,xy}$. Then, for any x, the εF-interpolation between x and $y = \mathsf{S}_\varepsilon x$ is

$$\omega_s^{\varepsilon, x\mathsf{S}_\varepsilon x} = \mathsf{S}_{\varepsilon s} x, \quad 0 \leq s \leq 1.$$

In other words, the εF-interpolation matches with the gradient flow when the endpoint is well appropriate. The reason is clear from the uniqueness of the minimizer and the fact that the map $s \mapsto \mathsf{S}_{\varepsilon s} x$ satisfies Newton's equation (2.7).

The Hamiltonian corresponding to the Lagrangian (2.8) is

$$H^\varepsilon(q,p) = \frac{|p|^2}{2} + U^\varepsilon(q) = \frac{|p|^2}{2} - \frac{\varepsilon^2}{2}|F'(q)|^2,$$

and the equation of motion of any minimizer of (2.9), i.e., any εF-interpolation ω is given by the Hamilton system of equations: $\begin{cases} \dot\omega_s &= p_s, \\ \dot p_s &= -U^{\varepsilon\prime}(\omega_s), \end{cases} \quad 0 \leq s \leq 1.$

Proposition 2.7 (Properties of the εF-interpolations)**.**

(a) *For any $x, y \in \mathbb{R}^n$ and any $\varepsilon \geq 0$, the minimization problem (2.9) admits at least a solution $\omega^{\varepsilon,xy}$.*

(b) *Furthermore, $\lim_{\varepsilon \to 0} \omega^{\varepsilon,xy} = \omega^{xy}$ pointwise, where ω^{xy} is the constant speed geodesic from x to y.*

(c) *Any εF-interpolation is $\mathcal{C}^2$ and satisfies the Newton equation*

$$\ddot{\omega}_s^{\varepsilon,xy} = \frac{\varepsilon^2}{2} I'(\omega_s^{\varepsilon,xy}) = \varepsilon^2 F'' F'(\omega_s^{\varepsilon,xy}), \tag{2.10}$$

which is also (2.7).

(d) *Along any εF-interpolation ω, the Hamiltonian is conserved as a function of time*:

$$H(\omega_s, \dot{\omega}_s) = \frac{|\dot{\omega}_s|^2}{2} - \frac{\varepsilon^2}{2} |F'(\omega_s)|^2 = H(\omega_0, \dot{\omega}_0), \quad \forall 0 \leq s \leq 1.$$

- As already noticed at Remark 2.6, the path $(\mathsf{S}_{\varepsilon s}(x))_{0\leq s\leq 1}$ is an εF-interpolation between x and $\mathsf{S}_\varepsilon(x)$ satisfying Newton's equation (2.10) and one immediately sees with the very definition (2.1) of its evolution that the conserved value of the Hamiltonian along it is zero.
- When $\varepsilon = 0$, the εF-interpolation is the standard constant speed geodesic between x and y. For any small $\varepsilon > 0$, one can think of the εF-interpolation $\omega^{\varepsilon,xy}$ as a small perturbation of the geodesic.

Dual formulation of the cost A_F^ε.

Let $h\colon \mathbb{R}^n \mapsto \mathbb{R}$ be a Lipschitz function and let $\varepsilon > 0$. We define the Hamilton–Jacobi semigroup, for any $t \geq 0$ and $y \in \mathbb{R}^n$, by

$$Q_t^{\varepsilon F} h(y) = \inf_{\omega:\omega_t = y} \Big\{ h(\omega_0) + \int_0^t \frac{|\dot{\omega}_s|^2}{2} + \frac{\varepsilon^2}{2} |F'(\omega_s)|^2 \, ds \Big\}, \tag{2.11}$$

where the infimum is running over all $\mathcal{C}^1$ path ω such that $\omega_t = y$. The function $U\colon (t,y) \mapsto Q_t^{\varepsilon F} h(y)$, satisfies, in the sense of viscosity solutions, the Hamilton–Jacobi equation

$$\begin{cases} \partial_t U(t,y) + \frac{1}{2} |U'(t,y)|^2 = \frac{\varepsilon^2}{2} |F'(y)|^2; \\ U(0,\cdot) = h(\cdot). \end{cases} \tag{2.12}$$

Minimizers of (2.11) are solutions of the system

$$\begin{cases} \ddot{\omega}_s = \frac{\varepsilon^2}{2} [|F'|^2]'(\omega_s), \quad 0 \leq s \leq t, \\ \dot{\omega}_0 = h'(\omega_0), \quad \omega_t = y. \end{cases} \tag{2.13}$$

Proposition 2.8 (Dual formulation of A_F^ε)**.** *For any $x, y \in \mathbb{R}^n$,*

$$A_F^\varepsilon(x,y) = \sup_h \{ Q_1^{\epsilon F} h(y) - h(x) \} = \sup_h \{ Q_1^{\epsilon F} h(x) - h(y) \},$$

where the supremum runs through all regular enough functions h.

Heuristic proof. For any smooth path $(\omega_t)_{0\le t\le 1}$ between $\omega_0 = x$ and $\omega_1 = y$, from the definition of $Q_1^{\varepsilon F}$,

$$Q_1^{\varepsilon F} h(y) - h(x) \le \int_0^1 \Big(\frac{|\dot\omega_t|^2}{2} + \frac{\varepsilon^2}{2}|F'(\omega_t)|^2\Big)\,dt,$$

then

$$\sup_h \{Q_1^{\varepsilon} h(y) - h(x)\} \le A_F^{\varepsilon}(x,y).$$

Now, let γ be a minimizer of (2.11) with $t=1$ and choose h such that $h(z) = \dot\gamma_0 \cdot z$ in a large enough family of test functions. By (2.13), we know that γ satisfies Newton equation. Hence it is an ϵF-interpolation, that is a minimizer of the action between $\gamma_0 = x$ at time $t=0$ and $\gamma_1 = y$ at time $t=1$. In other words,

$$Q_1^{\varepsilon F} h(y) - h(x) = \int_0^1 \Big(\frac{|\dot\gamma_t|^2}{2} + \frac{\varepsilon^2}{2}|F'(\gamma_t)|^2\Big)\,dt = A_F^{\varepsilon}(x,y).$$

This proves the first equality: $A_F^{\varepsilon}(x,y) = \sup\{Q_1^{\varepsilon} h(y) - h(x)\}$.

As the action appearing at formula (2.9) is invariant with respect to time reversal, we see that $A_F^{\varepsilon}(x,y) = A_F^{\varepsilon}(y,x)$. It follows immediately that $A_F^{\varepsilon}(x,y) = \sup\{Q_1^{\varepsilon} h(x) - h(y)\}$. □

Alternate formulations of the cost A_F^{ε}. Since for any $x,y\in\mathbb{R}^n$ and any path ω from x to y, we have

$$\begin{aligned}\int_0^1 \big[|\dot\omega_s|^2 + \varepsilon^2|F'(\omega_s)|^2\big]\,ds &= \int_0^1 \big[|\dot\omega_s + \varepsilon F'(\omega_s)|^2 - 2\varepsilon F'(\omega_s)\cdot\dot\omega_s\big]\,ds\\ &= \int_0^1 |\dot\omega_s + \varepsilon F'(\omega_s)|^2\,ds - 2\varepsilon(F(y)-F(x)),\end{aligned}$$

we obtain the forward expression of the cost

$$A_F^{\varepsilon}(x,y) = \frac12 \inf_{\omega\in\Omega^{xy}} \Big\{\int_0^1 |\dot\omega_s + \varepsilon F'(\omega_s)|^2\,ds\Big\} - \varepsilon(F(y)-F(x)).$$

With the same way of reasoning, its backward formulation is

$$A_F^{\varepsilon}(x,y) = \frac12 \inf_{\omega\in\Omega^{xy}} \Big\{\int_0^1 |\dot\omega_s - \varepsilon F'(\omega_s)|^2\,ds\Big\} + \varepsilon(F(y)-F(x)). \tag{2.14}$$

It follows that its symmetric expression is

$$A_F^{\varepsilon}(x,y) = \frac14 \inf_{\omega\in\Omega^{xy}} \Big\{\int_0^1 |\dot\omega_s + \varepsilon F'(\omega_s)|^2 ds\Big\} + \frac14 \inf_{\omega\in\Omega^{xy}} \Big\{\int_0^1 |\dot\omega_s - \varepsilon F'(\omega_s)|^2 ds\Big\}.$$

Example. Let us treat the simple case corresponding to $F(x) = |x|^2/2$, $x\in\mathbb{R}^n$, where computations are simple and explicit.

- For any $t\ge 0$ and $x\in\mathbb{R}^n$, $\mathsf{S}_t(x) = e^{-t}x$.

- For any $x, y \in \mathbb{R}^n$, the εF-interpolation between x and y is given by

$$\omega_s^\varepsilon = \mathsf{S}_{\varepsilon s}(\alpha) + \mathsf{S}_{\varepsilon(1-s)}(\beta), \quad 0 \leq s \leq 1,$$

  where $\alpha = \dfrac{x - ye^{-\varepsilon}}{1 - e^{-2\varepsilon}}$, $\beta = \dfrac{y - xe^{-\varepsilon}}{1 - e^{-2\varepsilon}}$.

- For any $x \in \mathbb{R}^n$, $A_F^\varepsilon(x, x) = \varepsilon \dfrac{1 - e^{-\varepsilon}}{1 + e^{-\varepsilon}} |x|^2$.

- Moreover, the corresponding Hamilton–Jacobi equation (2.12) takes the form

$$\begin{cases} \partial_t U(t, y) + \frac{1}{2}|U'(t, y)|^2 = \varepsilon^2 |y|^2/2, & t > 0, \\ U(0, x) = h(x), \end{cases}$$

  and has an explicit solution:

$$Q_t^{\varepsilon F} h(y) = Q^0_{1-e^{-2\varepsilon t}} f(e^{-\varepsilon t} y)/(2\varepsilon) + \varepsilon |y|^2/2$$

  with $f(x) = h(x) - \varepsilon |x|^2/2$.

2.3. Convexity properties of the cost A_F^ε

Definition 2.9 ((ρ, n)-convexity)**.** Let $\rho \in \mathbb{R}$ and $\mathsf{n} \in (0, \infty]$. We say that the twice differentiable function F on $\mathbb{R}^n$ is (ρ, n)-convex if

$$F'' \geq \rho \,\mathrm{Id} + F' \otimes F'/\mathsf{n}. \tag{2.15}$$

Example. Let us give some examples where $n = 1$, $\mathsf{n} > 0$ and (2.15) is an equality.

- The map $x \mapsto -\mathsf{n} \log x$ is $(0, \mathsf{n})$-convex on $(0, \infty)$.
- When $\rho > 0$, the map $x \mapsto -\mathsf{n} \log \cos(x\sqrt{\rho/\mathsf{n}})$ is (ρ, n)-convex on the interval $(-\pi/2\sqrt{\mathsf{n}/\rho}, \pi/2\sqrt{\mathsf{n}/\rho})$.
- When $\rho < 0$, the map $x \mapsto -\mathsf{n} \log \sinh(x\sqrt{-\rho/\mathsf{n}})$ is (ρ, n)-convex on the interval $(0, \infty)$.

Contraction inequality under a convexity assumption.

Proposition 2.10 (Contraction of the gradient flow)**.** *Let us assume that F is (ρ, n)-convex. Then, for all $t \geq 0$, $\varepsilon \geq 0$ and $x, y \in \mathbb{R}^n$,*

$$A_F^\varepsilon(\mathsf{S}_t(x), \mathsf{S}_t(y)) \leq e^{-\rho t} A_F^\varepsilon(x, y) - \frac{1}{\mathsf{n}} \int_0^t e^{-2\rho(t-u)} [F(\mathsf{S}_u(x)) - F(\mathsf{S}_u(y))]^2 \, du. \tag{2.16}$$

Proof. Let $(\omega_s)_{0 \leq s \leq 1}$ be any smooth path between x and y. For all $t \geq 0$, the composed path $[\mathsf{S}_t(\omega_s)]_{0 \leq s \leq 1}$ is a path between $\mathsf{S}_t(x)$ and $\mathsf{S}_t(y)$ and it follows from the definition of A_F^ε that the first inequality in the subsequent chain, where L^ε is defined at (2.8), is satisfied:

$$\begin{aligned}
A_F^\varepsilon(\mathsf{S}_t(x), \mathsf{S}_t(y)) &\le \int_0^1 L^\varepsilon(\mathsf{S}_t(\omega_s), \partial_s \mathsf{S}_t(\omega_s))\, ds \\
&\le e^{-2\rho t} \int_0^1 L^\varepsilon(\omega_s, \dot\omega_s)\, ds - \frac{1}{\mathsf{n}} \int_0^t \int_0^1 e^{-2\rho(t-u)} [F'(\mathsf{S}_u(\omega_s)) \cdot \mathsf{S}_u'(\omega_s)\dot\omega_s]^2\, ds\, du \\
&\le e^{-2\rho t} \int_0^1 L^\varepsilon(\omega_s, \dot\omega_s)\, ds - \frac{1}{\mathsf{n}} \int_0^t e^{-2\rho(t-u)} [F(\mathsf{S}_u(y)) - F(\mathsf{S}_u(x))]^2\, du.
\end{aligned}$$

The second inequality is a consequence of Lemma 2.11 below, and the last inequality is implied by Jensen's inequality. This concludes the proof of the proposition. □

Lemma 2.11. *Let $\rho \in \mathbb{R}$ and $\mathsf{n} \in (0, \infty)$. The following assertions are equivalent.*

(i) *The function F is (ρ, n)-convex.*

(ii) *For any differentiable path $(\omega_s)_{0 \le s \le 1}$, any $\varepsilon \ge 0$ and any $t \ge 0$, $0 \le s \le 1$,*

(2.17)
$$L^\varepsilon(\mathsf{S}_t(\omega_s), \partial_s \mathsf{S}_t(\omega_s)) \le e^{-2\rho t} L^\varepsilon(\omega_s, \dot\omega_s) - \frac{1}{\mathsf{n}} \int_0^t e^{-2\rho(t-u)} [F'(\mathsf{S}_u(\omega_s)) \cdot \mathsf{S}_u'(\omega_s)\dot\omega_s]^2\, du.$$

Proof. Let us prove that (i) implies (ii). We have $\partial_s \mathsf{S}_t(\omega_s) = \mathsf{S}_t'(\omega_s)\dot\omega_s$, and for simplicity, we denote $x = \omega_s$ and $v = \dot\omega_s$. We shall use: $\partial_t \mathsf{S}_t(x) = -F'(\mathsf{S}_t(x))$, $\partial_t \mathsf{S}_t'(x) = -F''(\mathsf{S}_t(x))\mathsf{S}_t'(x)$ and $\partial_t F'(\mathsf{S}_t(x)) = -F''(\mathsf{S}_t(x))F'(\mathsf{S}_t(x))$. Let us set

$$\Lambda(t) := L^\varepsilon(\mathsf{S}_t(\omega_s), \partial_s \mathsf{S}_t(\omega_s)) = \frac{1}{2}|\mathsf{S}_t'(x)v|^2 + \frac{\varepsilon^2}{2}|F'(\mathsf{S}_t(x))|^2.$$

Its derivative is

$$\begin{aligned}
\partial_t \Lambda(t) &= -\mathsf{S}_t'(x)v \cdot F''(\mathsf{S}_t(x))\, \mathsf{S}_t'(x)v - \varepsilon^2 F'(\mathsf{S}_t(x)) \cdot F''(\mathsf{S}_t(x))F'(\mathsf{S}_t(x)) \\
&\overset{\text{(i)}}{\le} -2\rho\Lambda(t) - \frac{1}{\mathsf{n}}[F'(\mathsf{S}_t(x))\, \mathsf{S}_t'(x)v]^2 - \frac{\varepsilon^2}{\mathsf{n}}|F'(\mathsf{S}_t(x))|^4 \\
&\le -2\rho\Lambda(t) - \frac{1}{\mathsf{n}}[F'(\mathsf{S}_t(x))\, \mathsf{S}_t'(x)v]^2,
\end{aligned}$$

which implies (2.17).

Let us now assume (ii) and show that it implies (i). Since (2.17) is an equality at time $t = 0$, fixing $s = 0$, $\omega_0 = x$ and $\dot\omega_0 = v$, the first order Taylor expansion of (2.17) in t around $t = 0$ implies that for any $v, x \in \mathbb{R}^n$,

$$v \cdot (F''(x) - \rho\,\mathrm{Id})v - \frac{1}{\mathsf{n}}(F'(x) \cdot v)^2 + \varepsilon^2 F'(x) \cdot (F''(x) - \rho\,\mathrm{Id})F'(x) \ge 0. \tag{2.18}$$

Suppose ad absurdum that (i) is false. Then (2.15) fails and there exist $x_o, v_o \in \mathbb{R}^n$ such that

$$v_o \cdot (F''(x_o) - \rho\,\mathrm{Id})v_o - \frac{1}{\mathsf{n}}(F'(x) \cdot v_o)^2 < 0.$$

But taking $x = x_o$, $v = \lambda v_o$ in (2.18) and sending λ to infinity leads to a contradiction. □

Actually we believe that the contraction inequality (2.16) is equivalent to the (ρ, n)-convexity of the function F as in the case when $\varepsilon = 0$.

Convexity properties along εF-interpolations. Let us introduce the notation

$$\theta_a(s) := \frac{1 - e^{-2as}}{1 - e^{-2a}}.$$

Note that $\lim_{a \to 0} \theta_a(s) = s$.

Proposition 2.12 (Convexity under the (ρ, ∞)-condition)**.** *Let F be a (ρ, ∞)-convex function with $\rho \in \mathbb{R}$. Then any εF-interpolation ω satisfies, for all $0 \le s \le 1$,*

$$\begin{aligned} F(\omega_s) \le{}& \theta_{\rho\varepsilon}(1-s) F(\omega_0) + \theta_{\rho\varepsilon}(s) F(\omega_1) \\ &- \frac{1 - e^{-2\rho\varepsilon}}{2\varepsilon}\, \theta_{\rho\varepsilon}(s)\, \theta_{\rho\varepsilon}(1-s)\, [A_F^\varepsilon(\omega_0, \omega_1) + \varepsilon F(\omega_0) + \varepsilon F(\omega_1)]. \end{aligned} \tag{2.19}$$

Proof. We start following the smart proof of Conforti (Theorem 1.4 in [17]). Let $(\omega_s)_{0 \le s \le 1}$ be an εF-interpolation, and let $\overleftarrow{h}$ and $\vec{h}$ be two functions on $[0,1]$ such that for any $s \in [0,1]$,

$$F(\omega_s) = \vec{h}(s) - \overleftarrow{h}(s)$$

and

$$\vec{h}'(s) = \frac{1}{4\varepsilon} |\dot{\omega}_s + \varepsilon F'(\omega_s)|^2, \qquad \overleftarrow{h}'(s) = \frac{1}{4\varepsilon} |\dot{\omega}_s - \varepsilon F'(\omega_s)|^2.$$

This is possible since $\frac{d}{ds} F(\omega_s) = F'(\omega_s) \cdot \dot{\omega}_s = \vec{h}'(s) - \overleftarrow{h}'(s)$.

Then, for any $s \in [0,1]$, using the Newton equation (2.10) satisfied by ω and the (ρ, ∞)-convexity (2.15) of F, we obtain

$$\vec{h}''(s) = \frac{1}{2} F''(\varepsilon F' + \dot{\omega}_s, \varepsilon F' + \dot{\omega}_s) \ge 2\rho\,\varepsilon\, \vec{h}'(s).$$

Similarly we have $\overleftarrow{h}''(s) \le -2\rho\varepsilon \overleftarrow{h}'(s)$. We know by Lemma 4.1 in [17] that these inequalities imply

$$\vec{h}(s) \le \vec{h}(1) - \theta_{\rho\varepsilon}(1-s)[\vec{h}(1) - \vec{h}(0)], \qquad \overleftarrow{h}(s) \ge \overleftarrow{h}(0) + \theta_{\rho\varepsilon}(s)[\overleftarrow{h}(1) - \overleftarrow{h}(0)].$$

Arranging the terms in $F(\omega_s) = \vec{h}(s) - \overleftarrow{h}(s)$, we see that

$$F(\omega_s) \le \theta_{\rho\varepsilon}(1-s) F(\omega_0) + \theta_{\rho\varepsilon}(s) F(\omega_1) - (1 - e^{-2\rho\varepsilon}) \theta_{\rho\varepsilon}(s) \theta_{\rho\varepsilon}(1-s) [\vec{h}(1) - \overleftarrow{h}(0)].$$

Now the proof differs from Conforti's one. By the definitions of $\vec{h}$ and $\overleftarrow{h}$, and using the backward formulation (2.14) of the cost A_F^ε, we obtain

$$2\varepsilon[\vec{h}(1) - \overleftarrow{h}(0)] = \int_0^1 \frac{1}{2} |\dot{\omega}_s - \varepsilon F'(\omega_s)|^2 ds + 2\varepsilon F(\omega_1) = A_F^\varepsilon(x,y) + \varepsilon F(\omega_0) + \varepsilon F(\omega_1),$$

which gives us the desired inequality (2.19). □

Proposition 2.12 implies a Talagrand-type inequality, i.e., a comparison between a cost function and an entropy, see [17].

Corollary 2.13 (Talagrand-type inequality for the cost A_F^ε)**.** *Assume that F is (ρ,∞)-convex with $\rho>0$ and that it is normalized by $\inf F=0$. Then for any $\varepsilon>0$ and $x,y\in\mathbb{R}^n$,*

$$A_F^\varepsilon(x,y)\le\varepsilon\,\frac{1+e^{-\rho\varepsilon}}{1-e^{-\rho\varepsilon}}\,(F(x)+F(y)). \tag{2.20}$$

In particular, if $\inf F=F(x^*)=0$ at x^*, then for any $y\in\mathbb{R}^n$,

$$A_F^\varepsilon(x^*,y)\le\varepsilon\,\frac{1+e^{-2\rho\varepsilon}}{1-e^{-2\rho\varepsilon}}\,F(y). \tag{2.21}$$

Proof. Both (2.20) and (2.21) are direct consequence of (2.19), the first at time $s=1/2$ with $F(\omega_{1/2})\ge0$, and the latter at time $s=0$ with $\inf F=F(x^*)=0$. □

Letting ε tend to zero in (2.21), we see that $F(x)\ge\rho|x-x^*|^2/2$, which is the optimal inequality.

Costa's lemma states the concavity of the exponential entropy along the heat semigroup. Here is an analogous result.

Proposition 2.14 (Convexity under the $(0,\mathsf{n})$-condition)**.** *Let us assume that F is $(0,\mathsf{n})$-convex. Then, for any $\varepsilon\ge0$ and any εF-interpolation ω, the function*

$$[0,1]\ni s\mapsto e^{-F(\omega_s)/\mathsf{n}}$$

is concave.

Proof. Differentiating twice $\Lambda(s):=\exp\{-F(\omega_s)/\mathsf{n}\}$, we get $\Lambda'(s)=-\Lambda(s)F'(\omega)\cdot\dot\omega_s/\mathsf{n}$, and

$$\begin{aligned}\Lambda''(s)&=-\frac{1}{\mathsf{n}}\Lambda'(s)F'(\omega_s)\cdot\dot\omega_s-\frac{1}{\mathsf{n}}\Lambda(s)F''(\omega_s)(\dot\omega_s,\dot\omega_s)-\frac{1}{\mathsf{n}}\Lambda(s)F'(\omega_s)\cdot\ddot\omega_s\\&=-\frac{1}{\mathsf{n}}\Lambda(s)\Big[-\frac{1}{\mathsf{n}}|F'(\omega_s)\dot\omega_s|^2+\underbrace{F''(\omega_s)(\dot\omega_s,\dot\omega_s)}_{\ge\frac{1}{\mathsf{n}}(F'(\omega_s)\cdot\dot\omega_s)^2}+\underbrace{\varepsilon^2F''(\omega_s)(F'(\omega_s),F'(\omega_s))}_{\ge\frac{\varepsilon^2}{\mathsf{n}}|F'(\omega_s)|^4}\Big]\\&\le0,\end{aligned}$$

which is the desired result. We used (2.10) at second equality, and the $(0,\mathsf{n})$-convexity of F at last inequality. □

As a direct consequence of Propositions 2.12 and 2.14, we obtain the following.

Corollary 2.15. *Let ω be an εF-interpolation with $\varepsilon>0$.*

(a) *If F is (ρ,∞)-convex with $\rho\in\mathbb{R}$, then*

$$-\frac{d^+}{ds}F(\omega_s)\Big|_{s=1}+\rho A_F^\varepsilon(\omega_0,\omega_1)\le\frac{\rho\,\varepsilon(1+e^{-2\rho\varepsilon})}{1-e^{-2\rho\,\varepsilon}}\,[F(\omega_0)-F(\omega_1)].$$

(b) *If F is $(0,\mathsf{n})$-convex with $\mathsf{n}>0$, then*

$$-\frac{d^+}{ds}F(\omega_s)\Big|_{s=1}\le\mathsf{n}\,[1-e^{-(F(\omega_0)-F(\omega_1))/\mathsf{n}}]. \tag{2.22}$$

In order to prepare the proof of the analogue of EVI at Proposition 2.17, we need the next result.

Proposition 2.16 (Derivative formula)**.** *For any $x, y \in \mathbb{R}^n$,*

$$\frac{d}{dt}^+\Big|_{t=0} A_F^\varepsilon(\mathsf{S}_t(x), y) \le -\frac{d}{ds}\Big|_{s=1} F(\omega_s^{yx}),$$

where ω^{yx} is any εF-interpolation from y to x.

Proof. Let ω be an εF-interpolation from y to x, (we drop the superscript yx for simplicity). Then for any $t \ge 0$, $(\eta_{s,t})_{0\le s\le 1} = (\mathsf{S}_{st}(\omega_s))_{0\le s\le 1}$ is a path from y to $\mathsf{S}_t(x)$ and by definition of A_F^ε, we have

$$A_F^\varepsilon(\mathsf{S}_t(x), y) \le \int_0^1 \left[\frac{1}{2}|\partial_s \eta_{s,t}|^2 + \frac{\varepsilon^2}{2}|F'(\eta_{s,t})|^2\right] ds.$$

Since at $t = 0$ this is an equality, we see that

$$\begin{aligned}\frac{d}{dt}^+\Big|_{t=0} & A_F^\varepsilon(\mathsf{S}_t(x), y) \\ &\le \int_0^1 \big[-F'(\omega_s)\cdot\dot\omega_s - sF''(\omega_s)(\dot\omega_s,\dot\omega_s) - s\varepsilon^2 F''(\omega_s)(F'(\omega_s),F'(\omega_s))\big]\, ds.\end{aligned}$$

Differentiating $H(s) := F(\omega_s)$, we obtain $H'(s) = F'(\omega_s)\cdot\dot\omega_s$ and $H''(s) = F''(\omega_s)(\dot\omega_s,\dot\omega_s) + F'(\omega_s)\cdot\ddot\omega_s$. With the Newton equation (2.10): $\ddot\omega_s = \varepsilon^2 F''F'(\omega_s)$, we arrive at

$$\frac{d}{dt}^+\Big|_{t=0} A_F^\varepsilon(\mathsf{S}_t(x), y) \le \int_0^1 [-H'(s) - sH''(s)]\, ds = -H'(1),$$

which is the announced result. □

This specific method of applying the gradient flow $\mathsf{S}_{st}(\omega_s)$ to a path ω has been successfully used in [19].

One derives immediately from Corollary 2.15 and Proposition 2.16 the following.

Proposition 2.17 (EVI under (ρ,∞) or $(0,\mathsf{n})$-convexity)**.**

(a) *Assume that F is (ρ,∞)-convex. Then, for any $x, y \in \mathbb{R}^n$,*

$$\frac{d}{dt}^+\Big|_{t=0} A_F^\varepsilon(\mathsf{S}_t(x), y) + \rho A_F^\varepsilon(x,y) \le \frac{\rho\,\varepsilon(1+e^{-2\rho\,\varepsilon})}{1-e^{-2\rho\,\varepsilon}}\,[F(y)-F(x)]. \tag{2.23}$$

(b) *Assume that F is $(0,\mathsf{n})$-convex. Then, for any $x, y \in \mathbb{R}^n$,*

$$\frac{d}{dt}^+\Big|_{t=0} A_F^\varepsilon(\mathsf{S}_t(x), y) \le \mathsf{n}\,[1 - e^{-(F(y)-F(x))/\mathsf{n}}].$$

These inequalities should be compared with the EVI formulation (2.5) which holds in the case where $\varepsilon = 0$ and $\mathsf{n} = \infty$. Note that one recovers (2.5) by letting ε tend to zero in (2.23), or by sending n to infinity in the last inequality.

3. Newton's equation in the Wasserstein space

This section is dedicated to the infinite dimension case where the states live in the Wasserstein space on a Riemannian manifold. This is done by relying on the results of previous section in $\mathbb{R}^n$ as an analogical guideline, and by using Otto's heuristic. Only non-rigorous statements and "proofs" are presented in this section. For the sake of completeness, we start recalling Otto calculus which was introduced in the seminal paper [44]. This heuristic theory is known for long and well explained at many places [55], [29], [2]. Our contribution is the introduction of the Newton equation, a key point of our computations.

3.1. The setting

The configuration space is a Riemannian manifold (N, g) equipped with a Riemannian measure vol and the state space to be considered later on is the Wasserstein space

$$M := \mathcal{P}_2(N) \tag{3.1}$$

of all probability measures μ on N such that $\int_N d^2(x_o, \cdot)\, d\mu < \infty$, where d stands for the Riemannian distance.

Carré du champ. The gradient in (N, g) is denoted by ∇ and the divergence by $\nabla\cdot$. Functions and vector fields are assumed to be smooth enough for allowing all the computations. In particular, for any vector field $\vec{A}$ and function f with a compact support, the following integration by parts formula holds

$$\int f\, \nabla\cdot \vec{A}\, d\mathrm{vol} = -\int \nabla f \cdot \vec{A}\, d\mathrm{vol},$$

using the inner product on (N, g). The carré du champ Γ in N is defined for any functions f, g on N by

$$\Gamma(f,g) : x \mapsto \nabla f(x) \cdot \nabla g(x), \quad x \in N.$$

As usual we write $\Gamma(f,g) = \nabla f \cdot \nabla g$, and state $\Gamma(f) = \nabla f \cdot \nabla f = |\nabla f|^2$. Denoting $\Delta_\mathrm{g} = \Delta = \nabla \cdot \nabla$ (we drop the subscript g for simplicity), the Laplace–Beltrami operator on N, we have

$$\int \Gamma(f,g)\, d\mathrm{vol} = -\int f \Delta g\, d\mathrm{vol}\,.$$

Iteration of Γ, hidden connection. The iterated carré du champ operator in (N, g) introduced by Bakry and Émery in [4] (see also [5]) is defined for any f by

$$\Gamma_2(f) = \frac{1}{2}\Delta\Gamma(f) - \Gamma(f, \Delta f). \tag{3.2}$$

It happens to be the left-hand side of the Bochner–Lichnerowicz identity,

$$\Gamma_2(f) = ||\nabla^2 f||^2_{\mathrm{HS}} + \mathrm{Ric}_\mathrm{g}(\nabla f, \nabla f), \tag{3.3}$$

where $||\nabla^2 f||^2_{\mathrm{HS}}$ is the Hilbert–Schmidt norm of the Hessian $\nabla^2 f$ of f and $\mathrm{Ric}_{\mathrm{g}}$ is the Ricci tensor of (N, g).

For any functions $f, g, h \colon N \mapsto \mathbb{R}$, the evaluation of the Hessian of f applied to two gradients $\nabla g, \nabla h$ only depends on the carré du champ Γ:

$$\nabla^2 f(\nabla g, \nabla h) = \frac{1}{2}\big[\Gamma(\Gamma(f,g),h) + \Gamma(\Gamma(f,h),g) - \Gamma(\Gamma(g,h),f)\big] \tag{3.4}$$

and of course $2\nabla^2 f(\nabla f, \nabla f) = \Gamma(\Gamma(f), f)$.

The Levi-Civita connection is implicitly invoked in the Bochner–Lichnerowicz identity, in particular for computing the Hessian of f, and it is hidden in (3.4). This is commented on at [5], p. 158.

3.2. Some PDEs are Wasserstein gradient flows

We present basic notions which are useful to visualize heuristically the Wasserstein space $M = \mathcal{P}_2(N)$ as an infinite Riemannian manifold.

The same notation is used for the probability measure μ and its density $d\mu/d\mathrm{vol}$.

Velocity, tangent space and Wasserstein metric. If (μ_t) is a path in M, then $\partial_t \mu_t$ satisfies $\int_N \partial_t \mu_t \, d\mathrm{vol} = 0$. In other words the tangent space at some $\mu \in M$ is a subset of functions f satisfying $\int_N f \, d\mathrm{vol} = 0$.

Let us explore another representation of the tangent space. Again, let (μ_t) be a path in M. Then for each t, there exists a unique (up to some identification) map $\Phi_t \colon N \mapsto \mathbb{R}$ such that the continuity equation

$$\partial_t \mu_t = -\nabla \cdot (\mu_t \nabla \Phi_t) \tag{3.5}$$

holds. As a hint for the proof of this statement, denoting $\mathsf{a}_\mu(\nabla f) := -\nabla \cdot (\mu \nabla f)$, we see that $\mathsf{a}_\mu(\nabla f) = v$ is an elliptic equation in f and $\nabla \Phi_t = \mathsf{a}^{-1}_{\mu_t}(\partial_t \mu_t)$ exists because $v = \partial_t \mu_t$ satisfies $\int \partial_t \mu_t \, d\mathrm{vol} = 0$. This allows one to identify the velocity $\partial_t \mu_t$ with the gradient vector field $\nabla \Phi_t$, via the mapping

$$\dot{\mu}_t := \nabla \Phi_t = \mathsf{a}^{-1}_{\mu_t}(\partial_t \mu_t). \tag{3.6}$$

Within Otto's heuristics, the tangent space of M at μ is represented by the set of gradients

$$\mathrm{T}_\mu M = \{\nabla \Phi, \ \Phi : N \mapsto \mathbb{R}\}.$$

To be rigorous it is necessary to define $\mathrm{T}_\mu M$ as the closure in $L^2(N, \mu)$ of $\{\nabla \Phi, \ \Phi \in \mathcal{C}^\infty_c(N)\}$. *We call $\dot{\mu}_t \in \mathrm{T}_\mu M$ the Wasserstein velocity to distinguish it from the standard velocity $\partial_t \mu_t$.*

The continuity equation (3.5) is a keystone of the theory. It is valid for instance when N is a smooth compact Riemannian manifold and the density $\mu(t, x)$ is smooth and positive. This extends to more general settings, as explained in [3], Chapter 8.

Definition 3.1 (Wasserstein metric)**.** The inner product on $\mathrm{T}_\mu M$ is defined for any $\nabla\Phi, \nabla\Psi$ by

$$\langle \nabla\Phi, \nabla\Psi\rangle_\mu = \int \nabla\Phi \cdot \nabla\Psi \, d\mu = \int \Gamma(\Phi, \Psi)\, d\mu.$$

Hence the speed of a path (μ_t) in M is: $|\dot\mu_t|_{\mu_t} = \left(\int |\mathsf{a}_{\mu_t}^{-1}(\partial_t\mu_t)|^2 d\mu_t\right)^{1/2}$.

Comparing with (1.2), we see that minimizing velocity fields satisfying the continuity equation (3.5) are gradient fields.

As a definition, the Wasserstein cost for transporting μ onto ν is

$$W_2^2(\mu,\nu) := \inf_\pi \int_{N\times N} d^2(x,y)\, d\pi,$$

where d is the Riemannian distance on N and π runs through the set of all couplings between μ and ν. The Benamou–Brenier theorem [7] states that

$$\inf_{(\xi,v)} \int_{[0,1]\times N} |v_t|^2 \, d\xi_t \, dt = W_2^2(\mu,\nu), \tag{3.7}$$

where the infimum runs through the set of all (ξ, v) such that for all t, $\xi_t \in \mathcal{P}_2(N)$ and v_t is a time dependent vector field satisfying the continuity equation $\partial_t\xi + \nabla\cdot(\xi v) = 0$ and the endpoint constraint $\xi_0 = \mu, \xi_1 = \nu$. This means that the Riemannian distance associated to the Wasserstein metric is W_2, recall (1.3). The proof of the analogous result in a metric space can be found in [3].

Other metrics. The standard velocity $\partial_t\mu_t$ lives in the space

$$\mathcal{T}_\mu := \{\partial\nu : \int \partial\nu \, d\mathrm{vol} = 0\}$$

equipped with the inner product of $L^2(\mathrm{vol})$, while the Wasserstein velocity $\dot\mu_t$ lives in $\mathrm{T}_\mu M$ equipped with the above Wasserstein inner product. Both of them represents the same object, they are linked by (3.6), but they give rise to distinct gradient flows.

Although the present article focuses on the Wasserstein metric, let us give some examples of alternate metrics based on the tangent space $\mathcal{T}_\mu$.

1) The first example is described in [21], Section 9.6. For any $\mu \in M$ and two perturbations $\partial\nu, \partial\nu'$ around μ, the inner product is the standard scalar product in $L^2(\mathrm{vol})$:

$$\langle \partial\nu, \partial\nu'\rangle_{\mu,1} = \int \partial\nu\, \partial\nu' \, d\mathrm{vol}\,. \tag{3.8}$$

2) The second example is rather similar. For any $\mu \in M$ and $\partial\nu, \partial\nu' \in \mathcal{T}_\mu$,

$$\langle \partial\nu, \partial\nu'\rangle_{\mu,2} = \int \nabla\Delta^{-1}(\partial\nu)\cdot\nabla\Delta^{-1}(\partial\nu')\, d\mathrm{vol} = -\int \Delta^{-1}(\partial\nu)\partial\nu' \, d\mathrm{vol}\,. \tag{3.9}$$

This metric is explained in [44], Section 1.2.

Both inner products $\langle \cdot, \cdot \rangle_{\mu,1}$ and $\langle \cdot, \cdot \rangle_{\mu,2}$ do not depend on μ and one can show that the geodesic $(\mu_s)_{0\le s\le 1}$ between μ and ν satisfies: $\partial^2_{ss}\mu_s = 0$. It follows that (μ_s) is an affine interpolation, meaning that the mass is not transported but teleported. The geometric content of these metrics is poorer than the Wasserstein metric one.

3) For any $\mu \in M$ and two perturbations $\partial\nu, \partial\nu' \in \mathcal{T}_\mu$ around μ,

$$\langle \partial\nu, \partial\nu' \rangle_{\mu,3} = \int \frac{\nabla\Delta^{-1}(\partial\nu)\cdot\nabla\Delta^{-1}(\partial\nu')}{\mu}\, d\mathrm{vol}\,. \tag{3.10}$$

This is the Markov transportation metric which is defined and used in [10].

The heat equation

$$\partial_t u = \Delta u \tag{3.11}$$

is the gradient flow of $\mathcal{F}_1(\mu) = \int |\nabla\mu|^2\, d\mathrm{vol}$ with respect to the metric $\langle \cdot, \cdot \rangle_{\mu,1}$ defined at (3.8) and also of $\mathcal{F}_2(\mu) = \int \mu^2/2\, d\mathrm{vol}$ with respect to the metric $\langle \cdot, \cdot \rangle_{\mu,2}$ defined at (3.9), cf. [21]. The heat equation is also the gradient flow of the standard entropy defined below at (3.15), with respect to the Markov transportation metric defined in (3.10). We shall see in a moment at (3.17) that this equation is also the gradient flow of the same entropy with respect to the Wasserstein metric.

Wasserstein gradient flow. Let us go back to the Wasserstein metric and define the gradient with respect to this metric of a function $\mathcal{F} \colon M \mapsto \mathbb{R}$. If $\mathcal{F}$ is differentiable at μ, there exists a gradient field $\nabla\Phi$ such that for any path (μ_t) satisfying $\dot\mu_t = \nabla\Psi_t$ and $\mu_{t_o} = \mu$, we have

$$\frac{d}{dt}\mathcal{F}(\mu_t)\Big|_{t=t_o} = \langle \nabla\Phi, \nabla\Psi_{t_o} \rangle_\mu.$$

As a definition, the gradient of $\mathcal{F}$ at μ is

$$\mathrm{grad}_\mu \mathcal{F} := \nabla\Phi \in \mathrm{T}_\mu M.$$

Similarly to the finite dimensional setting, one can define a gradient flow in M with respect to the Wasserstein metric.

Definition 3.2 (Wasserstein gradient flow)**.** A path $(\mu_t)_{t\ge 0}$ in M is a Wasserstein gradient flow of a function $\mathcal{F}$ if for any $t \ge 0$,

$$\dot\mu_t = -\,\mathrm{grad}_{\mu_t} \mathcal{F}.$$

We denote by $\mu_t = \mathsf{S}_t^{\mathcal{F}}\mu$, the solution of the above gradient flow equation starting from $\mu \in M$.

Examples of Wasserstein gradient flows. Consider the following interesting type of functions:

$$\mathcal{F}(\mu) = \int f(\mu)\, d\mathrm{vol}, \tag{3.12}$$

where $f\colon \mathbb{R}\mapsto\mathbb{R}$ and the Wasserstein gradient flow equation

$$\dot\mu_t = -\operatorname{grad}_{\mu_t}\mathcal{F}. \tag{3.13}$$

With a path (μ_t) satisfying $\dot\mu_t = \nabla\Psi_t$, we obtain

$$\begin{aligned}\frac{d}{dt}\mathcal{F}(\mu_t) &= \int f'(\mu_t)\partial_t\mu_t\, d\mathrm{vol} = -\int f'(\mu_t)\nabla\cdot(\mu_t\nabla\Psi_t)\, d\mathrm{vol}\\ &= \int \nabla[f'(\mu_t)]\cdot\nabla\Psi_t\, d\mu_t = \langle\nabla[f'(\mu_t)],\nabla\Psi_t\rangle_{\mu_t}.\end{aligned}$$

Therefore

$$\operatorname{grad}_\mu\mathcal{F} = \nabla[f'(\mu)]\in \mathrm{T}_\mu M. \tag{3.14}$$

We mention three standard examples of PDEs whose solutions are Wasserstein gradient flows.

(1) The standard entropy, that is the relative entropy with respect to the volume measure, is defined for any $\mu\in M$ by

$$\mathcal{E}nt(\mu) := \int \mu\log\mu\, d\mathrm{vol}\,. \tag{3.15}$$

Let us show that *the Wasserstein gradient flow of the entropy is the heat equation.*

The function $\mathcal{F}=\mathcal{E}nt$ corresponds to $f(a)=a\log a$ and its Wasserstein gradient (3.14) is

$$\operatorname{grad}_\mu\mathcal{E}nt = \nabla\log\mu. \tag{3.16}$$

With (3.5), we see that the gradient flow equation (3.13) writes as

$$\partial_t\mu_t = \nabla\cdot(\mu_t\nabla\log\mu_t) = \Delta\mu_t, \tag{3.17}$$

which is the heat equation (3.11). This was discovered in the seminal paper [34].

This example can be generalized by considering the Fokker–Planck equation with a potential V:

$$\partial_t\mu_t = \Delta\mu_t + \nabla\cdot(\mu_t\nabla V) = \nabla\cdot(\mu_t\nabla(\log\mu_t + V)), \tag{3.18}$$

whose solution is the Wasserstein gradient flow of

$$\mathcal{F}(\mu) = \mathcal{E}nt(\mu) + \int V\, d\mu.$$

(2) The Rényi entropy of order $p>0$, with $p\neq 1$, is

$$\mathcal{R}_p(\mu) := \frac{1}{p-1}\int \mu^p\, d\mathrm{vol}\,. \tag{3.19}$$

Let us show that *the Wasserstein gradient flow equation* (3.13) *of the Rényi entropy is the porous media equation* (*or fast diffusion equation*)

$$\partial_t\mu_t = \Delta\mu_t^p.$$

The function $\mathcal{F} = \mathcal{R}_p$ corresponds to $f(a) = a^p/(p-1)$ and its Wasserstein gradient (3.14) is

$$\operatorname{grad}_\mu \mathcal{R}_p = \frac{p}{p-1} \nabla \mu^{p-1}. \tag{3.20}$$

With (3.5), we see that the gradient flow equation writes as $\partial_t \mu_t = \nabla \cdot \big(\mu_t \frac{p}{p-1} \nabla \mu_t^{p-1}\big) = \Delta \mu_t^p$. This was proved rigorously in [44].

(3) The granular media equation in $\mathbb{R}^n$ is

$$\partial_t \mu_t = \nabla \cdot (\mu_t \nabla(\log \mu_t + V + W * \mu_t)), \tag{3.21}$$

where $V\colon \mathbb{R}^n \mapsto \mathbb{R}$ is a confinement potential and $W\colon \mathbb{R}^n \mapsto \mathbb{R}$ is an interaction potential. We assume that $W(-x) = W(x)$ and $W * \mu_t(x) = \int W(x-y)\mu_t(x)\, d\mathrm{vol}$ is the usual convolution in $\mathbb{R}^n$. We directly derive from this equation

$$\dot{\mu}_t = -\nabla(\log \mu_t + V + W * \mu_t).$$

Let us introduce

$$\mathcal{F}(\mu) = \int \Big(\log \mu + V + \frac{1}{2} W * \mu\Big)\, d\mu.$$

By the same computation, we obtain

$$\operatorname{grad}_\mu \mathcal{F} = \nabla(\log \mu + V + W * \mu).$$

So the granular media equation (3.21) is the Wasserstein gradient flow equation of the above functional $\mathcal{F}$. This result is derived and used in [14].

Many other functionals are also studied. For instance the gradient flow of the modified Fisher information,

$$\mathcal{F}(\mu) = \frac{1}{2} \int |\nabla \log \mu|^2 \mu + \int V \mu,$$

is used in [28] to prove existence of fourth order evolutional equation.

3.3. Second order derivative in M

Covariant derivative of a gradient field. Let us show that the covariant derivative $D_t \nabla \Phi_t$ of the vector field $\nabla \Phi_t$ in the tangent space $\mathrm{T}_{\mu_t} M$, along a path $(\mu_t)_{t \geq 0}$ with velocity $\dot{\mu}_t = \nabla \Theta_t$, is

$$D_t \nabla \Phi_t = \mathsf{Proj}_{\mu_t}\big(\nabla \partial_t \Phi_t + \nabla^2 \Phi_t \nabla \Theta_t\big), \tag{3.22}$$

where Proj_{μ_t} is the orthogonal projection on the tangent space $\mathrm{T}_{\mu_t} M$. To see this, we compute the time-derivative of $\langle \nabla \Phi_t, \nabla \Psi_t \rangle_{\mu_t} = \int \Gamma(\Phi_t, \Psi_t)\, d\mu_t$, where $\nabla \Phi_t$ and $\nabla \Psi_t$ belong to $\mathrm{T}_{\mu_t} M$:

$$\frac{d}{dt} \langle \nabla \Phi_t, \nabla \Psi_t \rangle_{\mu_t} = \int \big[\Gamma(\partial_t \Phi_t, \Psi_t) + \Gamma(\Phi_t, \partial_t \Psi_t) + \Gamma(\Gamma(\Phi_t, \Psi_t), \Theta_t)\big]\, d\mu_t.$$

But from (3.4),

$$\Gamma(\Gamma(\Phi_t, \Psi_t), \Theta_t) = \nabla^2\Psi_t(\nabla\Phi_t, \nabla\Theta_t) + \nabla^2\Phi_t(\nabla\Psi_t, \nabla\Theta_t).$$

Hence

$$\begin{aligned}&\frac{d}{dt}\langle\nabla\Phi_t, \nabla\Psi_t\rangle_{\mu_t}\\ &\quad = \int \big(\nabla\partial_t\Phi_t + \nabla^2\Phi_t\nabla\Theta_t\big)\cdot\nabla\Psi_t\, d\mu_t + \int \big(\nabla\partial_t\Psi_t + \nabla^2\Psi_t\nabla\Theta_t\big)\cdot\nabla\Phi_t\, d\mu_t.\end{aligned}$$

As the covariant derivative must obey the chain rule:

$$\frac{d}{dt}\langle\nabla\Phi_t, \nabla\Psi_t\rangle_{\mu_t} = \langle D_t\nabla\Phi_t, \nabla\Psi_t\rangle_{\mu_t} + \langle\nabla\Phi_t, D_t\nabla\Psi_t\rangle_{\mu_t}, \tag{3.23}$$

we have shown (3.22). Introducing the convective derivative

$$\mathsf{D}_{\dot\mu_t} := \partial_t + \dot\mu_t\cdot\nabla$$

used in fluid mechanics, (3.22) writes as

$$D_t\nabla\Phi_t = \mathsf{Proj}_{\mu_t}\mathsf{D}_{\dot\mu_t}\nabla\Phi_t. \tag{3.24}$$

Acceleration. The acceleration $\ddot\mu_t = D_t\dot\mu_t$ of the path (μ) is the covariant derivative of $\dot\mu_t = \nabla\Phi_t$. It is given by (3.22) and (3.24) with $\nabla\Phi = \nabla\Theta$:

$$\ddot\mu_t = \nabla\Big(\partial_t\Phi_t + \frac{1}{2}\Gamma(\Phi_t)\Big) = \mathsf{D}_{\dot\mu_t}\dot\mu_t \in \mathrm{T}_{\mu_t}M. \tag{3.25}$$

It follows from (3.25) that any geodesic (μ_t) in the Wasserstein space satisfies

$$\begin{cases} \partial_t\mu_t = -\nabla\cdot(\mu_t\nabla\Phi_t),\\ \partial_t\Phi_t + \dfrac{1}{2}\Gamma(\Phi_t) = 0, \end{cases} \tag{3.26}$$

see for instance [55]. Such a geodesic is sometimes called a displacement interpolation or a McCann geodesic.

Hessian. Let $(\mu_t)_{t\geq 0}$ be a Wasserstein geodesic and let $\mathcal{F}$ be a function on M. With (3.24) and $\ddot\mu_t = 0$, we see that the Wasserstein Hessian of $\mathcal{F}$ at $\mu_t \in M$, applied to $\dot\mu_t$ is

$$\begin{aligned}\mathrm{Hess}_{\mu_t}\mathcal{F}(\dot\mu_t, \dot\mu_t) &= \frac{d^2}{dt^2}\mathcal{F}(\mu_t) = \frac{d}{dt}\langle \mathrm{grad}_{\mu_t}\mathcal{F}, \dot\mu_t\rangle_{\mu_t}\\ &= \langle D_t\,\mathrm{grad}_{\mu_t}\mathcal{F}, \dot\mu_t\rangle_{\mu_t} = \langle \mathsf{D}_{\dot\mu_t}\,\mathrm{grad}_{\mu_t}\mathcal{F}, \dot\mu_t\rangle_{\mu_t}.\end{aligned}$$

Actually, we need some additional information about $\mathcal{F}$ and its gradient to give a more explicit expression of the Hessian of $\mathcal{F}$. Let us have a look at the important case where

$$\mathcal{F}(\mu) = \int f(\mu)\,\mathrm{dvol}.$$

In the subsequent lines, (μ_t) is a generic path with velocity $\dot\mu = \nabla\Phi$.

Since $D_t \operatorname{grad}_{\mu_t} \mathcal{F} = \nabla(-f''(\mu_t)\nabla\cdot(\mu_t \nabla \Phi_t)) + \mathsf{Proj}_{\mu_t} \nabla^2(f'(\mu_t))\nabla\Phi_t$, we see that

$$\begin{aligned}\operatorname{Hess}_\mu \mathcal{F}(\dot\mu,\dot\mu)\\ &= \int \Big(-\Gamma\big[f''(\mu)\nabla\cdot(\mu\nabla\Phi),\Phi\big] + \Gamma\big[\Phi,\Gamma(f'(\mu),\Phi)\big] - \frac{1}{2}\Gamma\big[f'(\mu),\Gamma(\Phi)\big]\Big)\, d\mu,\end{aligned}$$

where we used

$$\nabla^2[f'(\mu)](\nabla\Phi,\nabla\Phi) = \Gamma\big(\Gamma(f'(\mu),\Phi),\Phi\big) - \frac{1}{2}\Gamma\big(\Gamma(\Phi),f'(\mu)\big).$$

Integrating by parts, we obtain

$$\operatorname{Hess}_\mu \mathcal{F}(\dot\mu,\dot\mu) = \int \big(\big[\mu^2 f''(\mu) - \mu f'(\mu) + f(\mu)\big](\Delta\Phi)^2 + [\mu f'(\mu) - f(\mu)]\Gamma_2(\Phi)\big)\, d\mathrm{vol}\,.$$

These computations appear in [45], p. 9, see also [55], p. 425. The Hessian of the entropy is

$$\operatorname{Hess}_\mu \mathcal{E}nt(\dot\mu,\dot\mu) = \int \Gamma_2(\Phi)\, d\mu. \tag{3.27}$$

In the case of a gradient flow where $\nabla\Phi = -\operatorname{grad}_\mu \mathcal{E}nt = -\nabla\log\mu$, we obtain

$$\operatorname{Hess}_\mu \mathcal{E}nt(\operatorname{grad}_\mu \mathcal{E}nt, \operatorname{grad}_\mu \mathcal{E}nt) = \int \Gamma_2(\log\mu)\, d\mu. \tag{3.28}$$

The Hessian of the Rényi entropy (3.19), corresponding to $f(a) = a^p/(p-1)$, is

$$\operatorname{Hess}_\mu \mathcal{R}_p(\dot\mu,\dot\mu) = \int \big((p-1)(\Delta\Phi)^2 + \Gamma_2(\Phi)\big)\, \mu^{p-1}\, d\mu. \tag{3.29}$$

The Schwarz theorem. Next result is a Schwarz theorem for a path in the Wasserstein space depending both on the time parameters s and t. For the special purpose of next statement, we introduce the notation

$$\mathsf{d}_t\mu_t := \dot\mu_t$$

for the expression (3.6) of the dot derivative.

"Lemma" 3.3 (Commutation of the dot derivatives)**.** *For any map $(\mu_{s,t})_{s,t\ge 0}$ from $[0,\infty)^2$ to M,*

$$\mathsf{d}_s\mathsf{d}_t\mu_{s,t} = \mathsf{d}_t\mathsf{d}_s\mu_{s,t} \in \mathrm{T}_{\mu_{s,t}}M, \tag{3.30}$$

where $\mathsf{d}_s\mathsf{d}_t\mu_{s,t}$ is the covariant derivative in s of the vector field $\mathsf{d}_t\mu_{s,t}$.

This will be used later during the heuristic proof of Proposition 3.5.

Heuristic proof of "Lemma" 3.3. There exist two functions $\Phi_{s,t}, \Psi_{s,t}$ such that $\mathsf{d}_s\mu_{s,t} = \nabla\Phi_{s,t}$ and $\mathsf{d}_t\mu_{s,t} = \nabla\Psi_{s,t}$. By (3.6), this means

$$\partial_s\mu_{s,t} = -\nabla\cdot(\mu_{s,t}\nabla\Phi_{s,t}) \quad \text{and} \quad \partial_t\mu_{s,t} = -\nabla\cdot(\mu_{s,t}\nabla\Psi_{s,t}).$$

From (3.22),

$$\mathsf{d}_s\mathsf{d}_t\mu_{s,t} = D_s\nabla\Psi_{s,t} = \mathsf{P}_{\mu_{s,t}}\big(\nabla\partial_s\Psi_{s,t} + \nabla^2\Psi_{s,t}\nabla\Phi_{s,t}\big)$$
$$\mathsf{d}_t\mathsf{d}_s\mu_{s,t} = D_t\nabla\Phi_{s,t} = \mathsf{P}_{\mu_{s,t}}\big(\nabla\partial_t\Phi_{s,t} + \nabla^2\Phi_{s,t}\nabla\Psi_{s,t}\big).$$

On the other hand, by the standard Schwarz theorem: $\partial_s\partial_t\mu_{s,t} = \partial_t\partial_s\mu_{s,t} =: \partial^2_{st}\mu_{s,t}$,

$$\begin{aligned}\partial^2_{st}\mu_{s,t} &= -\nabla\cdot(-\nabla\cdot(\mu_{s,t}\nabla\Psi_{s,t})\nabla\Phi_{s,t} + \mu_{s,t}\nabla\partial_t\Phi_{s,t})\\ &= -\nabla\cdot(-\nabla\cdot(\mu_{s,t}\nabla\Phi_{s,t})\nabla\Psi_{s,t} + \mu_{s,t}\nabla\partial_s\Psi_{s,t}).\end{aligned} \tag{3.31}$$

It is enough to prove that for any $\nabla\chi \in \mathrm{T}_{\mu_{s,t}}M$,

$$\langle \mathsf{d}_s\mathsf{d}_t\mu_{s,t} - \mathsf{d}_t\mathsf{d}_s\mu_{s,t}, \nabla\chi\rangle_{\mu_{s,t}} = 0. \tag{3.32}$$

Dropping the subscript s,t and using (3.31),

$$\begin{aligned}\langle(\mathsf{d}_s\mathsf{d}_t\mu - \mathsf{d}_t\mathsf{d}_s\mu, \nabla\chi\rangle_\mu &= \int \mu\left[\nabla\partial_s\Psi + \nabla^2\Psi\nabla\Phi - \nabla\partial_t\Phi - \nabla^2\Phi\nabla\Psi\right]\cdot\nabla\chi\, d\mathrm{vol}\\ &= \int \big[\nabla^2\Psi(\nabla\Phi,\nabla\chi) - \nabla^2\Phi(\nabla\Psi,\nabla\chi) - \Delta\Psi\Gamma(\Phi,\chi) - \Gamma(\log\mu,\Psi)\Gamma(\Phi,\chi)\\ &\quad + \Delta\Phi\Gamma(\Psi,\chi) + \Gamma(\log\mu,\Phi)\Gamma(\Psi,\chi)\big]\, d\mu.\end{aligned}$$

From (3.4) we see that

$$\int\left[\nabla^2\Psi(\nabla\Phi,\nabla\chi) - \nabla^2\Phi(\nabla\Psi,\nabla\chi)\right]d\mu = \int\left[\Gamma(\Phi,\Gamma(\Psi,\chi)) - \Gamma(\Psi,\Gamma(\Phi,\chi))\right]d\mu.$$

Integrating by parts,

$$\begin{aligned}\int \Delta\Phi\Gamma(\Psi,\chi)\, d\mu &= -\int \Gamma(\Phi,\mu\Gamma(\Psi,\chi))\, d\mathrm{vol}\\ &= -\int\left[\Gamma(\Phi,\Gamma(\Psi,\chi)) + \Gamma(\Psi,\chi)\Gamma(\Phi,\log\mu)\right]d\mu\end{aligned}$$

and something similar for $\int \Delta\Psi\Gamma(\Phi,\chi)\, d\mu$. Adding all these quantities, we obtain (3.32). □

3.4. $\varepsilon\mathcal{F}$-interpolations solve Newton's equation

In this section we investigate the analogue in the Wasserstein space of Proposition 2.7.

The carré du champ operator on M is defined, for any functions $\mathcal{F},\mathcal{G}\colon M\mapsto\mathbb{R}$, by

$$\mathbf{\Gamma}(\mathcal{F},\mathcal{G})(\mu) := \langle \operatorname{grad}_\mu\mathcal{F}, \operatorname{grad}_\mu\mathcal{G}\rangle_\mu, \quad \mu\in M, \tag{3.33}$$

with $\mathbf{\Gamma}(\mathcal{F},\mathcal{F})(\mu) = \mathbf{\Gamma}(\mathcal{F})(\mu) = |\operatorname{grad}_\mu\mathcal{F}|^2_\mu \geq 0$.

For instance, if $\mathcal{F}(\mu) = \int f(\mu)dx$ and $\mathcal{G}(\mu) = \int g(\mu)dx$ for some real functions f and g, then

$$\mathbf{\Gamma}(\mathcal{F}, \mathcal{G})(\mu) = \int \nabla f'(\mu) \cdot \nabla g'(\mu)\, d\mu = \int \Gamma(f'(\mu), g'(\mu))\, d\mu = \int f''(\mu) g''(\mu) \Gamma(\mu)\, d\mu,$$

where we have used the chain rule in (N, g).

In analogy with the Definition 2.5 of the εF-cost and εF-interpolations, we introduce the following.

"Definition" 3.4 ($\varepsilon\mathcal{F}$-cost, $\varepsilon\mathcal{F}$-interpolations)**.** Let $\mathcal{F}$ be a (regular enough) function on M. For any $\varepsilon > 0$ and $\mu, \nu \in M$, we define the $\varepsilon\mathcal{F}$-cost between μ and ν by

$$\mathcal{A}^{\varepsilon}_{\mathcal{F}}(\mu, \nu) = \inf_{(\mu)} \int_0^1 \Big(\frac{1}{2} |\dot{\mu}_s|^2_{\mu_s} + \frac{\varepsilon^2}{2} \mathbf{\Gamma}(\mathcal{F})(\mu_s) \Big)\, ds, \tag{3.34}$$

where $|\dot{\mu}_s|^2_{\mu_s} = \langle \dot{\mu}_s, \dot{\mu}_s \rangle_{\mu_s}$, and the infimum runs over all paths $(\mu_s)_{0 \le s \le 1}$ in M such that $\mu_0 = \mu$ and $\mu_1 = \nu$.

Minimizers are called *$\varepsilon\mathcal{F}$-interpolations* and are denoted by $(\mu^{\varepsilon,\mu\nu}_s)_{0 \le s \le 1}$ (for simplicity we sometimes omit ε, μ and ν).

- When $\varepsilon = 0$, we recover the formula (1.3).
- Since $\mathcal{A}^{\varepsilon}_{\mathcal{F}} \ge \frac{1}{2} W_2^2$, $\mathcal{A}^{\varepsilon}_{\mathcal{F}}$ appears as an approximation from above of the squared Wasserstein distance.

The $\varepsilon\mathcal{F}$-cost is the action associated with the Lagrangian

$$\mathcal{L}^{\varepsilon}(\mu, \dot{\mu}) = |\dot{\mu}|^2_{\mu}/2 + \varepsilon^2\, \mathbf{\Gamma}(\mathcal{F})(\mu)/2$$

corresponding to the scalar potential

$$\mathcal{U}^{\varepsilon} := -\frac{\varepsilon^2}{2} \mathbf{\Gamma}(\mathcal{F}).$$

"Proposition" 3.5 (Interpolations solve Newton's equation)**.** *Any $\varepsilon\mathcal{F}$-interpolation μ satisfies the Newton equation*

$$\ddot{\mu} = \frac{\varepsilon^2}{2} \operatorname{grad}_{\mu} \mathbf{\Gamma}(\mathcal{F}). \tag{3.35}$$

Heuristic proof. Let μ be an $\varepsilon\mathcal{F}$-interpolation and let us take any perturbation $(\mu_{s,t})_{s \in [0,1], t \in \mathbb{R}}$ of μ verifying: $\mu_{s,0} = \mu_s$ for any $s \in [0,1]$, $\mu_{0,t} = \mu_0$, $\mu_{1,t} = \mu_1$ for any t, and $\overset{\odot}{\mu}_{0,0} = \overset{\odot}{\mu}_{1,0} = 0$, where the Wasserstein t-velocity of $\mu_{s,t}$ is denoted by $\overset{\odot}{\mu}_{s,t}$, while its s-velocity is denoted as usual by $\dot{\mu}_{s,t}$. Let us differentiate the action $\Lambda(t) := \int_0^1 \big(|\dot{\mu}_{s,t}|^2_{\mu_{s,t}}/2 - \varepsilon^2 \mathcal{U}(\mu_{s,t}) \big)\, ds$, $t \in \mathbb{R}$, of the t-perturbation:

$$\Lambda'(t) = \int_0^1 \big(\langle \dot{\mu}_{s,t}, \dot{\overset{\odot}{\mu}}_{s,t} \rangle_{\mu_{s,t}} - \varepsilon^2 \langle \operatorname{grad}_{\mu_{s,t}} \mathcal{U}, \overset{\odot}{\mu}_{s,t} \rangle_{\mu_{s,t}} \big)\, ds.$$

Let us integrate by parts $\int_0^1 \langle \dot\mu_{s,t}, \dot{\overset{\odot}{\mu}}_{s,t}\rangle_{\mu_{s,t}}\, ds$. By the chain rule (3.23),

$$\frac{d}{ds}\langle \dot\mu_{s,t}, \overset{\odot}{\mu}_{s,t}\rangle_{\mu_{s,t}} = \langle \ddot\mu_{s,t}, \overset{\odot}{\mu}_{s,t}\rangle_{\mu_{s,t}} + \langle \dot\mu_{s,t}, \dot{\overset{\odot}{\mu}}_{s,t}\rangle_{\mu_{s,t}},$$

and by Lemma 3.3,

$$\int_0^1 \langle \dot\mu_{s,t}, \dot{\overset{\odot}{\mu}}_{s,t}\rangle_{\mu_{s,t}} ds = -\int_0^1 \langle \ddot\mu_{s,t}, \overset{\odot}{\mu}_{s,t}\rangle_{\mu_{s,t}} ds + \langle \dot\mu_{1,t}, \overset{\odot}{\mu}_{1,t}\rangle_{\mu_{1,t}} - \langle \dot\mu_{0,t}, \overset{\odot}{\mu}_{0,t}\rangle_{\mu_{0,t}}.$$

Taking into account the boundary conditions $\overset{\odot}{\mu}_{0,0}=\overset{\odot}{\mu}_{1,0}=0$, we obtain at $t=0$,

$$\Lambda'(0) = -\int_0^1 \left\langle (\ddot\mu_s + \varepsilon^2 \operatorname{grad}_{\mu_s} \mathcal{U}), \overset{\odot}{\mu}_{s,0} \right\rangle_{\mu_s} ds.$$

Since $(\mu_s)_{0\le s\le 1}$ is a minimizer, $\Lambda'(0)\ge 0$. As $(\overset{\odot}{\mu}_{s,0})$ can be chosen arbitrarily on any interval $s\in[\delta, 1-\delta]$ with $\delta>0$, this shows that $\Lambda'(0)=0$, showing that the Newton equation (3.35) holds. □

From Newton's equation (3.35) and the expression (3.25) of the acceleration $\ddot\mu$, one deduces the subsequent result.

"Corollary" 3.6 (Equations of motion of the interpolations)**.** *The $\varepsilon\mathcal{F}$-interpolation between μ and ν satisfies the system of equations*

$$\begin{cases} \partial_s\mu_s = -\nabla\cdot(\mu_s\nabla\Phi_s), \\ \nabla\partial_s\Phi_s + \nabla\Gamma(\Phi_s)/2 = \dfrac{\varepsilon^2}{2}\operatorname{grad}_{\mu_s}\mathbf{\Gamma}(\mathcal{F}), \\ \mu_0=\mu, \quad \mu_1=\nu. \end{cases} \tag{3.36}$$

When $\varepsilon=0$, one recovers the system (3.26) satisfied by the McCann geodesics. We also believe that (3.36) implies regularity properties of the $\varepsilon\mathcal{F}$-interpolation, for instance when $\mathcal{F}$ is the Rényi entropy.

In $\mathbb{R}^n$, the system (3.36) appears as the Euler equation where the initial and the final configuration are prescribed,

$$\begin{cases} \partial_s\mu_s = -\nabla\cdot(\mu_s\nabla\Phi_s), \\ \partial_s(\mu_s\nabla\Phi_s) + \nabla\cdot(\mu_s\nabla\Phi_s\otimes\nabla\Phi_s) = \dfrac{\varepsilon^2}{2}\mu_s\operatorname{grad}_{\mu_s}\mathbf{\Gamma}(\mathcal{F}), \\ \mu_0=\mu, \quad \mu_1=\nu, \end{cases}$$

where $\nabla\cdot$ is the divergence applied to each column. This system is a particular case of Eq. (17) in [24], where the solution also minimizes some action.

Examples. Let us treat two examples.

(1) In the case where $\mathcal{F}=\mathcal{E}nt$ is the usual entropy, the system (3.36) becomes

$$\begin{cases} \partial_s\mu_s = -\nabla\cdot(\mu_s\nabla\Phi_s), \\ \partial_s\Phi_s + \Gamma(\Phi_s)/2 = \varepsilon^2\left[\Gamma(\log\mu_s)/2 - \Delta\mu_s/\mu_s\right], \\ \mu_0=\mu, \quad \mu_1=\nu, \end{cases} \tag{3.37}$$

This fundamental example is related to the Schrödinger problem. It will be developed carefully at Section 5.

(2) In the case where $\mathcal{F} = \mathcal{R}_p$ is the Rényi entropy (3.19), the system (3.36) becomes

$$\begin{cases} \partial_s \mu_s = -\nabla \cdot (\mu_s \nabla \Phi_s), \\ \partial_s \Phi_s + \dfrac{1}{2}\Gamma(\Phi_s) = \dfrac{\varepsilon^2 p^2}{2}\left((3-2p)\,\mu_s^{2p-4}\,\Gamma(\mu_s) - 2\mu_s^{2p-3}\Delta\mu_s\right), \\ \mu_0 = \mu, \quad \mu_1 = \nu. \end{cases}$$

With $p = 1$, we are are back to the entropy $\mathcal{E}nt$. When $p \neq 1$, we don't know how to solve this system.

In the case where $\mathcal{F}(\mu) = \int f(\mu)\, d\mathrm{vol}$, we have

$$\mathrm{grad}_\mu \mathcal{F} = \nabla[f'(\mu)] = f''(\mu)\nabla\mu, \quad \text{and} \quad \mathbf{\Gamma}(\mathcal{F})(\mu) = \int f''(\mu)^2\,\Gamma(\mu)\, d\mu.$$

Moreover, after some computations, we obtain

$$\mathrm{grad}_\mu \mathbf{\Gamma}(\mathcal{F}) = -\nabla\left[2f''(\mu)^2\mu\Delta\mu + (2f'''(\mu)f''(\mu)\mu + f''(\mu)^2)\Gamma(\mu)\right].$$

Dual formulation of the cost $\mathcal{A}^\varepsilon_{\mathcal{F}}$.

Let $\mathcal{H}$ be a functional on M and let $\varepsilon > 0$. We define the Hamilton–Jacobi semigroup for any $t \geq 0$ and $\nu \in M$ by, ,

$$\mathcal{Q}_t^{\varepsilon\mathcal{F}}\mathcal{H}(\nu) = \inf\left\{\mathcal{H}(\mu_0) + \int_0^t \frac{|\dot\mu_s|^2}{2} + \frac{\varepsilon^2}{2}\mathbf{\Gamma}(\mathcal{F})(\mu_s)\, ds\right\}, \tag{3.38}$$

where the infimum runs over all path $(\mu_s)_{0\leq s\leq t}$ such that $\mu_t = \nu$. The function $\mathcal{U}\colon (t,\nu) \mapsto \mathcal{Q}_t^{\varepsilon\mathcal{F}}\mathcal{H}(\nu)$ satisfies the Hamilton–Jacobi equation

$$\begin{cases} \partial_t \mathcal{U}(t,\nu) + \dfrac{1}{2}\mathbf{\Gamma}(\mathcal{U})(t,\nu) = \dfrac{\varepsilon^2}{2}\mathbf{\Gamma}(\mathcal{F})(\nu), \\ \mathcal{U}(0,\cdot) = \mathcal{H}(\cdot). \end{cases} \tag{3.39}$$

Even if the definition of the function $\mathcal{Q}_t^{\varepsilon\mathcal{F}}\mathcal{H}$ is formal, following the work initiated by Gangbo–Nguyen–Tudorascu [24] (see also [1], [32] and also [25] for their applications to the mean field games), the functional $\mathcal{U}$ satisfies the Hamilton–Jacobi equation in the Wasserstein space in the sense of viscosity solutions.

Minimizers of (3.38) are solutions of the system

$$\begin{cases} \ddot\mu_s = \dfrac{\varepsilon^2}{2}\,\mathrm{grad}_{\mu_t}\mathbf{\Gamma}(\mathcal{F}), \quad 0 \leq s \leq t, \\ \dot\mu_0 = \mathrm{grad}_{\mu_0}\mathcal{H}, \quad \mu_t = \nu. \end{cases} \tag{3.40}$$

"Proposition" 3.7 (Dual formulation of $\mathcal{A}^\varepsilon_{\mathcal{F}}$)**.** *For any $\mu, \nu \in M$,*

$$\mathcal{A}^\varepsilon_{\mathcal{F}}(\mu,\nu) = \sup_{\mathcal{H}}\{\mathcal{Q}_1^{\varepsilon\mathcal{F}}\mathcal{H}(\mu) - \mathcal{H}(\nu)\},$$

where the supremum runs through all functionals $\mathcal{H}$ on M.

Heuristic proof. It is the analogue of the proof of Proposition 2.8. □

“Proposition” 3.8 (Conserved quantity)**.** *Let* $(\mu_s)_{0\le s\le 1}$ *be an* $\varepsilon\mathcal{F}$*-interpolation. Then the map*

$$s \mapsto |\dot\mu_s|^2_{\mu_s} - \varepsilon^2\, \mathbf{\Gamma}\mathcal{F}(\mu_s)$$

is constant on $[0,1]$.

The result is obvious from the Newton equation (3.35) satisfied by the $\varepsilon\mathcal{F}$-interpolations $(\mu_s)_{0\le s\le 1}$.

4. Inequalities based upon $\mathcal{A}^\varepsilon_\mathcal{F}$

We go on using Section 2 as a guideline. To this aim we need to introduce the curvature-dimension property in the Wasserstein space.

4.1. (ρ,n)-convexity in the Wasserstein space

Definition 4.1 ((ρ,n)-convexity)**.** A function $\mathcal{F}$ on M is (ρ,n)-convex, where $\rho\in\mathbb{R}$ and $\mathsf{n}\in(0,\infty]$, if

$$\mathrm{Hess}_\mu\, \mathcal{F}(\dot\mu,\dot\mu) \ge \rho\,|\dot\mu|^2_\mu + \frac{1}{\mathsf{n}}\langle \mathrm{grad}_\mu\, \mathcal{F},\dot\mu\rangle^2_\mu, \tag{4.1}$$

for any $\mu\in M$ and $\dot\mu\in T_\mu M$.

This analogue of Definition 2.9 was introduced in [20], Section 2.1.

Example 4.2. Let us look at the examples of the standard and Rényi entropies on the n-dimensional manifold (N,g).

(1) Suppose that $\mathrm{Ric}_\mathrm{g}\ge \rho\,\mathrm{Id}$ for some real ρ. Applying this inequality and the Cauchy–Schwarz inequality: $||\nabla^2\Phi||^2_{\mathrm{HS}}\ge(\Delta\Phi)^2/n$, to the expression (3.3) of Γ_2, we see that the Bakry–Émery curvature-dimension condition holds with $\mathsf{n}=n$, i.e., for any function f,

$$\Gamma_2(f)\ge\rho\Gamma(f)+\frac{1}{n}(\Delta f)^2.$$

By (3.27), for any $\dot\mu=\nabla\Phi\in \mathrm{T}_\mu M$,

$$\begin{aligned}\mathrm{Hess}_\mu\,\mathcal{E}nt(\dot\mu,\dot\mu) &= \int\Gamma_2(\Phi)\,d\mu \ge \rho\int\Gamma(\Phi)\,d\mu+\frac{1}{n}\int(\Delta\Phi)^2\,d\mu\\ &\ge \rho\,|\dot\mu|^2_\mu+\frac{1}{n}\Big(\int\Delta\Phi\,d\mu\Big)^2 = \rho\,|\dot\mu|^2_\mu+\frac{1}{n}\langle\mathrm{grad}_\mu\,\mathcal{E}nt,\dot\mu\rangle^2_\mu,\end{aligned}$$

where we used $\mathrm{grad}_\mu\,\mathcal{E}nt=\nabla\log\mu$, see (3.16).

In other words, $\mathcal{E}nt$ is (ρ,n)-convex.

(2) The Rényi entropy $\mathcal{R}_p$, defined at (3.19), is convex whenever $\mathrm{Ric}_\mathrm{g}\ge 0$ and $p\ge 1-1/n$, $p\ne 1$. Indeed, it follows from (3.29) that

$$\mathrm{Hess}_\mu\,\mathcal{R}_p(\dot\mu,\dot\mu)\ge(p-1+1/n)\int(\Delta\Phi)^2\mu^{p-1}\,d\mu\ge 0,$$

where $\dot\mu=\nabla\Phi$.

4.2. Contraction inequality

Next proposition is the infinite dimensional analogue of Proposition 2.10.

"Proposition" 4.3 (Contraction under (ρ, n)-convexity)**.** *Suppose that $\mathcal{F}$ is (ρ, n)-convex. Then for any $\mu, \nu \in M$ and $t \geq 0$,*

$$\mathcal{A}^{\varepsilon}_{\mathcal{F}}(\mathsf{S}^{\mathcal{F}}_t \mu, \mathsf{S}^{\mathcal{F}}_t \nu) \leq e^{-2\rho t} \mathcal{A}^{\varepsilon}_{\mathcal{F}}(\mu, \nu) - \frac{1}{\mathsf{n}} \int_0^t e^{-2\rho(t-u)} (\mathcal{F}(\mathsf{S}^{\mathcal{F}}_u \mu) - \mathcal{F}(\mathsf{S}^{\mathcal{F}}_u \nu))^2 du. \tag{4.2}$$

Heuristic proof. The proof follows the line of Proposition 2.10. The Wasserstein setting analogue of the implication (i)⇒ (ii) of Lemma 2.11, which is used in the proof of Proposition 2.10, still holds true. Its relies on the commutation of the dot derivatives stated at "Lemma" 3.3. □

This result is proved and stated rigorously at Theorem 5.5 in the context of a compact and smooth Riemannian manifold. It gives a better contraction inequality than Theorem 6.1 in [27].

In the case of $\varepsilon = 0$, we recover the dimensional contraction proved in [10], [26], and [12]. See also [11]. When $\varepsilon > 0$ and $\rho = 0$, an improved contraction inequality is also given in [48] for the usual entropy. The equivalence between contraction inequality for the Wasserstein distance ($\varepsilon = 0$) with $\mathsf{n} = \infty$ and a lower bound on the Ricci curvature was first proved in [47]. See also [36], [6] for a dimensional contraction with two different times.

4.3. Convexity properties along $\varepsilon\mathcal{F}$-interpolations

Two kinds of convexity properties can be explored for the cost $\mathcal{A}^{\varepsilon}_{\mathcal{F}}$. When $\mathcal{F} = \mathcal{E}nt$ is the usual entropy, the first convexity property has been introduced by Conforti under a (ρ, ∞)-convexity assumption and the second one by the third author under a $(0, \mathsf{n})$-convexity assumption.

Next result extends in our context Conforti's convexity inequality (Theorem 1.4 in [17]). Let us recall our notation $\theta_a(s) := (1 - e^{-2as})/(1 - e^{-2a})$. Note that $\lim_{a \to 0} \theta_a(s) = s$.

"Proposition" 4.4 (Convexity under the (ρ, ∞)-condition)**.** *Let $\mathcal{F}$ be a (ρ, ∞)-convex function with $\rho \in \mathbb{R}$. Then any $\varepsilon\mathcal{F}$-interpolation $(\mu_s)_{0 \leq s \leq 1}$ satisfies*

$$\begin{aligned} \mathcal{F}(\mu_s) \leq{} & \theta_{\rho\varepsilon}(1-s)\, \mathcal{F}(\mu_0) + \theta_{\rho\varepsilon}(s)\, \mathcal{F}(\mu_1) \\ & - \frac{1 - e^{-2\rho\varepsilon}}{2\varepsilon}\, \theta_{\rho\varepsilon}(s)\, \theta_{\rho\varepsilon}(1-s)\, [\mathcal{A}^{\varepsilon}_{\mathcal{F}}(\mu_0, \mu_1) + \varepsilon \mathcal{F}(\mu_0) + \varepsilon \mathcal{F}(\mu_1)]. \end{aligned} \tag{4.3}$$

Heuristic proof. Exact analogue of Proposition 2.12's proof. □

When $\rho = 0$ the inequality (4.3) simply implies that the map $s \mapsto \mathcal{F}(\mu_s)$ is convex: a result obtained by the second author for the usual entropy in [39]. When $\varepsilon = 0$, we recover the usual convexity of $\mathcal{F}$ along McCann interpolations: the starting point of the Lott–Sturm–Villani theory [51], [40],

$$\mathcal{F}(\mu_s) \leq (1-s)\mathcal{F}(\mu_0) + s\mathcal{F}(\mu_1) - 2\rho s(1-s) W_2^2(\mu_0, \mu_1), \quad \forall 0 \leq s \leq 1.$$

From the above inequality, when $\rho > 0$, we can deduce a Talagrand inequality relating the Wasserstein distance with the entropy. Indeed, taking $s = 1/2$, one obtains

$$W_2^2(\mu_0, \mu_1) \le \frac{1}{\rho}\,(\mathcal{F}(\mu_0) + \mathcal{F}(\mu_1)).$$

The same property holds for the cost $\mathcal{A}^\varepsilon_{\mathcal{F}}$ as proposed in Corollary 1.2 of [17].

"Corollary" 4.5 (Talagrand inequality for the cost $\mathcal{A}^\varepsilon_{\mathcal{F}}$)**.** *Assume that $\mathcal{F}$ is (ρ, ∞)-convex with $\rho > 0$ and that it is normalized by $\inf \mathcal{F} = 0$. Then for any $\varepsilon > 0$ and $\mu, \nu \in M$,*

$$\mathcal{A}^\varepsilon_{\mathcal{F}}(\mu, \nu) \le \varepsilon\, \frac{1 + e^{-\rho\varepsilon}}{1 - e^{-\rho\varepsilon}}\,(\mathcal{F}(\mu) + \mathcal{F}(\nu)).$$

In particular, if $m \in M$ minimizes $\mathcal{F} : \inf \mathcal{F} = \mathcal{F}(m) = 0$, then for any μ,

$$\mathcal{A}^\varepsilon_{\mathcal{F}}(\mu, m) \le \varepsilon\, \frac{1 + e^{-2\rho\varepsilon}}{1 - e^{-2\rho\varepsilon}}\, \mathcal{F}(\mu).$$

This is the exact analogue of Proposition 2.13.

Next result is a generalization of a result proved for the usual entropy by the third author [48].

"Proposition" 4.6 (Convexity under the $(0, \mathsf{n})$-condition)**.** *Suppose that $\mathcal{F}$ is $(0, \mathsf{n})$-convex with $\mathsf{n} > 0$. Then, for any $\mathcal{F}$-interpolation $(\mu_s)_{0 \le s \le 1}$, the map*

$$[0, 1] \ni s \mapsto \exp\left(-\mathcal{F}(\mu_s)/\mathsf{n}\right), \tag{4.4}$$

is concave.

Heuristic proof. Analogous to the proof of Proposition 2.14. □

"Proposition" 4.7 (EVI inequality under (ρ, ∞) or $(0, \mathsf{n})$-convexity)**.**

(a) *Assume that $\mathcal{F}$ is (ρ, ∞)-convex. Then, for any $\mu, \nu \in M$,*

$$\frac{d}{dt}^{+}\Big|_{t=0} \mathcal{A}^\varepsilon_{\mathcal{F}}(\mathsf{S}^{\mathcal{F}}_t \mu, \nu) + \rho\, \mathcal{A}^\varepsilon_{\mathcal{F}}(\mu, \nu) \le \frac{\rho\, \varepsilon (1 + e^{-2\rho\,\varepsilon})}{1 - e^{-2\rho\,\varepsilon}}\, [\mathcal{F}(\nu) - \mathcal{F}(\mu)]. \tag{4.5}$$

(b) *Assume that $\mathcal{F}$ is $(0, \mathsf{n})$-convex. Then, for any $\mu, \nu \in M$,*

$$\frac{d}{dt}^{+}\Big|_{t=0} \mathcal{A}^\varepsilon_{\mathcal{F}}(\mathsf{S}^{\mathcal{F}}_t \mu, \nu) \le \mathsf{n}\left[1 - e^{-(\mathcal{F}(\nu) - \mathcal{F}(\mu))/\mathsf{n}}\right].$$

Heuristic proof. It follows exactly the line of proof of Proposition 2.17. The analogues of the preliminary results Corollary 2.15 and Proposition 2.16 are derived by means of the commutation of the dot derivatives obtained at "Lemma" 3.3. □

When $\varepsilon = 0$, we recover from (4.5), the EVI inequality under the (ρ, ∞)-convexity of $\mathcal{F}$,

$$\frac{d}{dt}^{+}\Big|_{t=0} \frac{1}{2} W_2^2(\mathsf{S}_t^{\mathcal{F}}\mu, \nu) + \frac{\rho}{2} W_2^2(\mu, \nu) \leq \mathcal{F}(\nu) - \mathcal{F}(\mu).$$

This inequality is commented on for instance in [3].

Under the $(0, \mathsf{n})$-convexity of $\mathcal{F}$, the inequality has the same form

$$\frac{d}{dt}^{+}\Big|_{t=0} \frac{1}{2} W_2^2(\mathsf{S}_t^{\mathcal{F}}\mu, \nu) \leq \mathsf{n}\left(1 - e^{-(\mathcal{F}(\nu) - \mathcal{F}(\mu))/\mathsf{n}}\right).$$

This was proved in [20].

5. Link with the Schrödinger problem

5.1. Entropic cost

This section is dedicated to the specific setting where

$$\mathcal{F}(\mu) = \mathcal{E}nt(\mu) = \int \mu \log \mu \, d\mathrm{vol}\,.$$

We depart from the heuristic style of previous sections and provide rigorous results. This requires some regularity hypotheses: (N, g) is assumed to a compact, connected and smooth Riemannian manifold.

The Riemannian measure vol is normalized to be a probability measure. Since all the measures considered in this section are smooth and absolutely continuous with respect to vol, we identify measures and densities. The heat semigroup is denoted by $(\mathsf{P}_t)_{t \geq 0}$ and defined for any smooth function f by $\mathsf{P}_t f(x) = u(t, x)$ where u solves $\begin{cases} \partial_t u = \Delta u, \\ u(0, \cdot) = f, \end{cases} \quad t \geq 0.$

We already saw that the heat equation is the Wasserstein gradient flow of $\mathcal{E}nt$, see (3.17). We also know that

$$\mathrm{grad}_\mu \mathcal{E}nt = \nabla \log \mu, \quad \mathbf{\Gamma}(\mathcal{E}nt)(\mu) = \int \left|\nabla \log \mu\right|^2 d\mu,$$

and we expect that any $\varepsilon\mathcal{E}nt$-interpolation (μ) solves the Newton equation (3.35):

$$\ddot{\mu} = \frac{\varepsilon^2}{2} \mathrm{grad}_\mu \mathbf{\Gamma}(\mathcal{E}nt). \tag{5.1}$$

The gradient flow of the usual entropy is denoted by $\mathsf{S}_t^{\mathcal{E}nt}$ and for any $\mu \in \mathcal{P}(M)$,

$$\mathsf{S}_t^{\mathcal{E}nt}(\mu) = \mathsf{P}_t\Big(\frac{d\mu}{d\mathrm{vol}}\Big)\,\mathrm{vol},$$

where $\mathsf{P}_t = \exp(t\Delta)$ is the usual heat semigroup.

The rigorous definition of the cost $\mathcal{A}^{\varepsilon}_{\mathcal{E}nt}$ on the space (N, g) is the following.

Definition 5.1 (Entropic cost)**.** For any positive and smooth densities $\mu, \nu \in M$,

$$\mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mu,\nu) = \inf_{(\mu_t,\nabla\Phi_t)_{0\le t\le 1}} \Big\{ \int_0^1 \Big(\frac{1}{2}|\nabla\Phi_t|^2_{\mu_t} + \frac{\varepsilon^2}{2}\int |\nabla \log \mu_t|^2\, d\mu_t \Big)\, dt \Big\},$$

where the infimum runs through all smooth and positive paths $(\mu_t)_{0\le t\le 1}$ and smooth gradient vector fields $(\nabla\Phi_t)_{0\le t\le 1}$ such that

$$\begin{cases} \partial_t \mu_t = -\nabla\cdot(\mu_t\nabla\Phi_t), \\ \mu_0 = \mu, \quad \mu_1 = \nu. \end{cases} \tag{5.2}$$

Another equivalent formulation is

$$\mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mu,\nu) = \inf_{(\mu_t)} \Big\{ \int_0^1 \Big(\frac{1}{2}|\dot\mu_t|^2_{\mu_t} + \frac{\varepsilon^2}{2}\int \big|\nabla \log \mu_t\big|^2\, d\mu_t \Big)\, dt \Big\}, \tag{5.3}$$

where the infimum is taken over all smooth $(\mu_t)_{0\le t\le 1}$ such that $\mu_0=\mu$ and $\mu_1=\nu$.

5.2. The Schrödinger problem

Let us recall the Schrödinger problem as explained for instance in the survey [38]. For any two probability measures q,r on some measure space, the *relative entropy* of q with respect to r is

$$H(\mathsf{q}|\mathsf{r}) = \begin{cases} \displaystyle\int \log\frac{d\mathsf{q}}{d\mathsf{r}}\, d\mathsf{q}, & \text{if } \mathsf{q}\ll\mathsf{r}, \\ +\infty, & \text{otherwise.} \end{cases}$$

The usual entropy is $\mathcal{E}nt(\mu) = H(\mu|\,\mathrm{vol})$, $\mu\in\mathcal{P}(N)$. The set of all probability measures on the path space $\Omega = \mathcal{C}([0,1],N)$ is denoted by $\mathcal{P}(\Omega)$. For any $\varepsilon>0$,

$$R^{\varepsilon}\in\mathcal{P}(\Omega)$$

is the law of the reversible Brownian motion with generator $\varepsilon\Delta$ and initial measure vol. Note that this motion has the same law as $\sqrt{2\varepsilon}$ times a standard reversible Brownian motion. The joint law of the initial and final positions of this process is noted

$$R^{\varepsilon}_{01}(dx\,dy)\in\mathcal{P}(N\times N).$$

For any couple of measures $\mu,\nu\in M=\mathcal{P}(N)$ such that $H(\mu|\,\mathrm{vol}) = \mathcal{E}nt(\mu) < +\infty$ and $H(\nu|\,\mathrm{vol}) = \mathcal{E}nt(\nu)<+\infty$, the Schrödinger problem is

$$\text{minimize} \quad \varepsilon H(Q|R^{\varepsilon}) \quad \text{subject to} \quad Q\in\mathcal{P}(\Omega): Q_0=\mu, Q_1=\nu, \tag{5.4}$$

where for any $0\le s\le 1$, $Q_s = (X_s)_{\#}Q\in\mathcal{P}(N)$ denotes the s-marginal of Q, i.e., the law of the position X_s at time s of the canonical process with law Q. Its value

$$\mathrm{Sch}^{\varepsilon}(\mu,\nu) := \inf\,(5.4)$$

is called the Schrödinger cost, its solution Q the Schrödinger bridge, and the time-marginal flow $(Q_s)_{0\le s\le 1}$ the entropic interpolation between μ and ν.

The Schrödinger problem admits an equivalent static formulation:

$$\mathrm{Sch}^{\varepsilon}(\mu,\nu)=\inf\{\varepsilon H(\pi|R^{\varepsilon}_{01});\ \pi\in\mathcal{P}(N\times N):\pi_0=\mu,\pi_1=\nu\},$$

where $\pi_0,\pi_1\in\mathcal{P}(N)$ are the marginals of π. It was proved by Beurling [9] and Föllmer [23], see also [38], that for any probability measures $\mu,\nu\in\mathcal{P}(N)$ with finite entropy and finite second order moments, the minimum of (5.4) is achieved by a unique probability measure, whose shape is characterized by the product form

$$P=f(X_0)\,g(X_1)R^{\varepsilon}\in\mathcal{P}(\Omega), \tag{5.5}$$

where f and g are nonnegative functions (the ε-dependence is omitted on the functions f and g). The logarithm of these functions play the same role as the Kantorovich potentials in the optimal transportation theory.

The time-marginals of the minimizer P define a flow of probability measures $(\mu_s)_{0\le s\le 1}$ called the *entropic ε-interpolation* between μ and ν. It follows from the Markov property of the Brownian motion and (5.5) that $(\mu_s)_{0\le s\le 1}$ takes the particular form

$$\mu_s=\mathsf{P}_{\varepsilon s}f\,\mathsf{P}_{\varepsilon(1-s)}g\ \mathrm{vol}, \tag{5.6}$$

where in the present context $\mathsf{P}_t=\exp(t\Delta)$ is the heat semigroup.

It is proved in [42], see also [37] for a more general result, that

$$\lim_{\varepsilon\to 0}\mathrm{Sch}^{\varepsilon}(\mu,\nu)=W_2^2(\mu,\nu)/4.$$

Next lemma establishes a connection between the Schrödinger problem and previous sections.

Lemma 5.2 (Entropic interpolation solves (3.37))**.** *Let $(\mu_s)_{0\le s\le 1}$ be a path of probability measures on N such that*

$$\mu_s=\mathsf{P}_{\varepsilon s}f\mathsf{P}_{\varepsilon(1-s)}g\ \ \mathrm{vol}\,.$$

where f,g are smooth and positive functions on N. Then $(\mu_s)_{0\le s\le 1}$ solves (3.37), i.e.,

$$\begin{cases}\partial_s\mu_s=-\nabla\cdot(\mu_s\nabla\Phi_s),\\ \partial_s\Phi_s+\Gamma(\Phi_s)/2=\varepsilon^2(\Gamma(\log\mu_s)/2-\Delta\mu_s/\mu_s),\end{cases} \tag{5.7}$$

with $\Phi_s=\varepsilon\log\mathsf{P}_{\varepsilon(1-s)}g-\varepsilon\log\mathsf{P}_{\varepsilon s}f$.

This system has an explicit solution given by $\mu_s=\exp\big((\varphi_s+\psi_s)/\varepsilon\big)$ where $\varphi_s:=\varepsilon\log\mathsf{P}_{\varepsilon s}f$ and $\psi_s:=\varepsilon\log\mathsf{P}_{\varepsilon(1-s)}g$. This gives $\Phi_s=\psi_s-\varphi_s$ and φ_s, ψ_s satisfy the Hamilton–Jacobi–Bellman equations

$$\partial_s\varphi_s=\varepsilon\Delta\varphi_s+\Gamma(\varphi_s),\quad \partial_s\psi_s=-\varepsilon\Delta\psi_s-\Gamma(\psi_s). \tag{5.8}$$

A heuristic presentation of this computation was proposed in Section 3.4.

In $\mathbb{R}^n$, the system (5.7) appears as the pressureless Euler equation:

$$\begin{cases} \partial_s \mu_s = -\nabla \cdot (\mu_s \nabla \Phi_s), \\ \partial_s(\mu_s \nabla \Phi_s) + \nabla \cdot (\mu_s \nabla \Phi_s \otimes \nabla \Phi_s) = -\dfrac{\varepsilon^2}{2}\, \mu_s\, \nabla\Big(\dfrac{\Delta\sqrt{\mu_s}}{\sqrt{\mu_s}}\Big), \\ \mu_0 = \mu, \quad \mu_1 = \nu. \end{cases}$$

This system is a particular case of Equation (1.1) in [22]. Let us emphasize that this specific system admits the entropic interpolation as an explicit smooth solution; this was unnoticed in [22].

Newton equation. Recall that the acceleration is given at (3.25) by $\ddot{\mu}_t = \nabla\big(\partial_t \Phi_t + \frac{1}{2}\Gamma(\Phi_t)\big) \in \mathrm{T}_{\mu_t} M$. Together with (5.7), this leads us to the Newton equation (5.1):

$$\ddot{\mu} = \varepsilon^2\, \nabla(\Gamma(\log \mu)/2 - \Delta\mu/\mu), \tag{5.9}$$

which was recently derived by Conforti in [17].

After the seminal works of Schrödinger in 1931 [49], [50], different aspects of entropic interpolations were investigated by Bernstein [8], and almost thirty years later by Beurling [9]. During the 70's, Jamison rediscovered this theory [33], and this also happened independently during the 80's to Zambrini [56], who initiated the theory of Euclidean quantum theory [18], [16]. The works by Jamison and Zambrini opened the way to the study of second order stochastic differential equations in order to derive the Newton equations for the Schrödinger bridges in terms of random paths, see for instance the papers by Krener, Thieullen and Zambrini [53], [35], [54].

The equation (5.9) in the Wasserstein space has a different nature than the above mentioned Newton equations in the path space. It sheds a new light on the dynamics of the entropic interpolations.

5.3. Benamou–Brenier and Kantorovich formulas

We show that the Schrödinger cost Sch^ε is equal, up to an additive constant, to the entropic cost $\mathcal{A}^\varepsilon_{\mathcal{E}nt}$, using the Benamou–Brenier–Schrödinger formula. This was proved in [15] and in Corollary 5.3 of [27] in $\mathbb{R}^n$ with a Kolmogorov generator, and in a more general case, like a $\mathrm{RCD}^*(K,N)$ space, in Theorem 5.4.3 of [52], for instance in the compact Riemannian manifold (N, g).

Theorem 5.3 (Benamou–Brenier–Schrödinger formula, see [15], [27], [52], [31])**.** *For any couple of positive and smooth probability measures μ, ν on N,*

$$\mathrm{Sch}^\varepsilon(\mu,\nu) = \mathcal{A}^\varepsilon_{\mathcal{E}nt}(\mu,\nu)/2 + \varepsilon\,[\mathcal{E}nt(\mu) + \mathcal{E}nt(\nu)]/2. \tag{5.10}$$

Proof. Let us recall, in our context, the proof proposed in Theorem 5.4.3 of [52]. Let $(\mu_s)_{0\le s\le 1}$ be a path from μ to ν with velocity $\nabla\Psi_s$, and let $(\nu_s)_{0\le s\le 1}$ be the entropic interpolation between μ and ν that is the path given by (5.6) for some functions f and g. If $(\nabla\Phi_s)_{0\le s\le 1}$ denotes its velocity, then from Lemma 5.2, the couple $(\nu_s, \nabla\Phi_s)_{0\le s\le 1}$ satisfies (5.7). We have

$$\frac{d}{ds}\int \Phi_s\, d\mu_s = \int \partial_s\Phi_s\, d\mu_s + \int \Phi_s\partial_s\mu_s\ \mathrm{dvol} = \int \partial_s\Phi_s\, d\mu_s + \int \Gamma(\Phi_s,\Psi_s)\, d\mu_s,$$

and from (5.7) and an integration by parts,

$$\begin{aligned}\frac{d}{ds}\int \Phi_s\, d\mu_s &= \int \Big[-\frac{1}{2}\Gamma(\Phi_s) + \frac{\varepsilon^2}{2}\Big(\Gamma(\log\nu_s) - 2\,\frac{\Delta\nu_s}{\nu_s}\Big) + \Gamma(\Phi_s,\Psi_s)\Big]\, d\mu_s \\ &= \int \Big[-\frac{1}{2}\Gamma(\Phi_s) + \Gamma(\Phi_s,\Psi_s) + \frac{\varepsilon^2}{2}\big(2\,\Gamma(\log\mu_s,\log\nu_s) - \Gamma(\log\nu_s)\big)\Big]\, d\mu_s.\end{aligned}$$

By Proposition 4.1.5 in [52], the functions f and g are positive and smooth. Applying the Cauchy–Schwarz inequality to Γ, we obtain

$$\frac{d}{ds}\int \Phi_s\, d\mu_s \le \int \Big[\frac{1}{2}\Gamma(\Psi_s) + \frac{\varepsilon^2}{2}\Gamma(\log\mu_s)\Big]\, d\mu_s,$$

with equality if $\Psi_s = \Phi_s$ (which implies that $\mu_s = \nu_s$). Hence,

$$\int \Phi_1\, d\nu - \int \Phi_0\, d\mu \le \int_0^1\int \Big[\frac{1}{2}\Gamma(\Psi_s) + \frac{\varepsilon^2}{2}\Gamma(\log\mu_s)\Big]\, d\mu_s ds, \tag{5.11}$$

and taking the minimum over all paths $(\mu_s)_{0\le s\le 1}$, we have obtained

$$\int \Phi_1\, d\nu - \int \Phi_0\, d\mu \le \mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mu,\nu),$$

But when $\mu_s = \nu_s$, (5.11) is an equality. This implies that

$$\int \Phi_1\, d\nu - \int \Phi_0\, d\mu = \mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mu,\nu).$$

Since (ν_s) is optimal, we know by (5.6) that $\nu_s = \mathsf{P}_{\varepsilon s} f \mathsf{P}_{\varepsilon(1-s)} g$ and $\Phi_s = \varepsilon\log\mathsf{P}_{\varepsilon(1-s)}g - \varepsilon\log\mathsf{P}_{\varepsilon s}f$. Hence,

$$\begin{aligned}\int \Phi_1\, d\nu - \int \Phi_0\, d\mu &= 2\varepsilon\Big(\int f\,\mathsf{P}_\varepsilon g\log f\,\mathrm{dvol} + \int g\,\mathsf{P}_\varepsilon f\log g\,\mathrm{dvol}\Big) - \varepsilon\,(\mathcal{E}nt(\mu) + \mathcal{E}nt(\nu)) \\ &= 2\varepsilon\, H(P|R^{\varepsilon}_{01}) - \varepsilon\,(\mathcal{E}nt(\mu) + \mathcal{E}nt(\nu)),\end{aligned}$$

where P is given in (5.5). Then

$$\int \Phi_1\, d\nu - \int \Phi_0\, d\mu = 2\,\mathrm{Sch}^{\varepsilon}(\mu,\nu) - \varepsilon\,(\mathcal{E}\mathrm{nt}(\mu) + \mathcal{E}\mathrm{nt}(\nu)),$$

and the Benamou–Brenier–Schrödinger formula (5.10) is proved. □

This result tells us that the action minimization problem (5.3) and the Schrödinger problem (5.4) are equivalent. In particular, the minimizers coincide and satisfy Newton's equation (5.1). This was proved rigorously in Theorem 1.3 of [17] in the same setting. A similar formal computation is done in [46] for the Schrödinger equation (not the problem).

We propose now a new proof of the dual formulation of the Schrödinger problem. The proof is given in [43] by using stochastic control, and [27], Theorem 4.1, in a more general context. The one proposed here is simpler. Let us note that independently and at the same moment, the result was also proved in a $\mathrm{RCD}^*(\rho, \mathsf{n})$ space in [31].

Theorem 5.4 (Kantorovich–Schrödinger dual formulation, see [43], [27], [31])**.** *For any couple of positive and smooth probability measures μ, ν on N,*

$$\varepsilon \sup_{h\in\mathcal{C}(N)} \Big\{ \int \log h\, d\nu - \int \log \mathsf{P}_\varepsilon h\, d\mu \Big\} = \frac{1}{2}\mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mu,\nu) + \frac{\varepsilon}{2}(\mathcal{E}nt(\nu) - \mathcal{E}nt(\mu)). \tag{5.12}$$

Proof. For any $\varepsilon\mathcal{E}nt$-interpolation $(\nu_s)_{0\le s\le 1}$ between μ and ν, and with velocity $(\nabla\Phi_s)_{0\le s\le 1}$, and any smooth and positive function h on N,

$$\begin{aligned}
\int \log h\, d\nu - \int \log \mathsf{P}_\varepsilon h\, d\mu &= \int_0^1 \frac{d}{ds}\int \log \mathsf{P}_{\varepsilon(1-s)} h\, d\nu_s\, ds\\
&= \int_0^1 \int \Big(-\varepsilon \frac{\Delta \mathsf{P}_{\varepsilon(1-s)}h}{\mathsf{P}_{\varepsilon(1-s)}h} + \Gamma(\log \mathsf{P}_{\varepsilon(1-s)}h, \Phi_s)\Big)\, d\nu_s\, ds\\
&= \int_0^1 \Big(\int \varepsilon\, \Gamma\Big(\frac{\nu_s}{\mathsf{P}_{\varepsilon(1-s)}h}, \mathsf{P}_{\varepsilon(1-s)}h\Big)\, d\mathrm{vol} + \int \Gamma(\log \mathsf{P}_{\varepsilon(1-s)}h, \Phi_s)\, d\nu_s\Big)\, ds\\
&= \int_0^1 \int \big(\Gamma(\varepsilon \log \nu_s + \Phi_s, \log \mathsf{P}_{\varepsilon(1-s)}h) - \varepsilon\, \Gamma(\log \mathsf{P}_{\varepsilon(1-s)}h)\big)\, d\nu_s\, ds.
\end{aligned}$$

Now, because the $\varepsilon\mathcal{E}nt$-interpolation writes as $\nu_s = \mathsf{P}_{\varepsilon s} f \mathsf{P}_{\varepsilon(1-s)} g$ vol for some positive smooth and bounded functions f and g (see [52] again), and $\Phi_s = \varepsilon \log \mathsf{P}_{\varepsilon(1-s)} g - \varepsilon \log \mathsf{P}_{\varepsilon s} f$, we have: $\varepsilon \log \nu_s + \Phi_s = 2\varepsilon \log \mathsf{P}_{\varepsilon(1-s)} g$. This leads to the following inequality:

$$\begin{aligned}
&\int \log h\, d\nu - \int \log \mathsf{P}_\varepsilon h\, d\mu\\
&\qquad = \varepsilon \int_0^1 \int \big(2\,\Gamma(\log \mathsf{P}_{\varepsilon(1-s)} g, \log \mathsf{P}_{\varepsilon(1-s)} h) - \Gamma(\log \mathsf{P}_{\varepsilon(1-s)} h)\big)\, d\nu_s\, ds\\
&\qquad \le \varepsilon \int_0^1 \int \Gamma(\log \mathsf{P}_{\varepsilon(1-s)} g)\, d\nu_s\, ds = \int \log g\, d\nu - \int \log \mathsf{P}_\varepsilon g\, d\mu,
\end{aligned}$$

with equality when $h = g$. In other words,

$$\sup_{h\in\mathcal{C}(N)} \Big\{ \int \log h\, d\nu - \int \log \mathsf{P}_\varepsilon h\, d\mu \Big\} = \int \log g\, d\nu - \int \log \mathsf{P}_\varepsilon g\, d\mu.$$

Moreover,

$$\begin{aligned}
\int \log g\, d\nu - \int \log \mathsf{P}_\varepsilon g\, d\mu &= \int (\log g) g \mathsf{P}_\varepsilon f\, d\mathrm{vol} - \int (\log \mathsf{P}_\varepsilon g) f \mathsf{P}_\varepsilon g\, d\mathrm{vol}\\
&= \int (\log g) g \mathsf{P}_\varepsilon f\, d\mathrm{vol} + \int (\log f) f \mathsf{P}_\varepsilon g\, d\mathrm{vol} - \int (\log f \mathsf{P}_\varepsilon g) f \mathsf{P}_\varepsilon g\, d\mathrm{vol}\\
&= H(P|R^\varepsilon) - \mathcal{E}nt(\mu),
\end{aligned}$$

where P is the optimal probability measure given in (5.5). Taking equation (5.10) into account, this concludes the proof. □

Let us see how the dual formulation proposed at Proposition 3.7 is in fact the same as the one in (5.12). Let us define for any $t \geq 0$ and $\mu \in M$,

$$\mathcal{U}(t,\mu) := \int Q_t^\varepsilon f \, d\mu + \varepsilon \, \mathcal{E}nt(\mu), \tag{5.13}$$

where $Q_t^\varepsilon f = -2\varepsilon \log \mathsf{P}_{\varepsilon t} \exp(-f/2\varepsilon)$ is the solution of the Hamilton–Jacobi–Bellman equation in N,

$$\partial_t u + \frac{1}{2}\Gamma(u) = \varepsilon \Delta u.$$

Then, from a direct computation, the map $\mathcal{U}$ solves the Hamilton–Jacobi equation in the Wasserstein space:

$$\partial_t \mathcal{U}(t,\mu) + \frac{1}{2}\mathbf{\Gamma}(\mathcal{U})(t,\nu) = \frac{\varepsilon^2}{2}\mathbf{\Gamma}(\mathcal{E}nt)(\mu), \tag{5.14}$$

starting from the initial condition $\mu \mapsto \int f \, d\mu + \varepsilon \mathcal{E}nt(\mu)$.

By Proposition 3.7, we have

$$\sup_f \Big\{ \int Q_t^\varepsilon f \, d\mu + \varepsilon \, \mathcal{E}nt(\mu) - \int f \, d\nu - \varepsilon \, \mathcal{E}nt(\nu) \Big\} = \mathcal{A}^\varepsilon_{\mathcal{E}nt}(\mu,\nu).$$

On the other hand, from the definition of $Q_t^\varepsilon f$,

$$\begin{aligned} \sup_f \Big\{ \int Q_t^\varepsilon f \, d\mu + \varepsilon \mathcal{E}nt(\mu) - \int f \, d\nu - \varepsilon \, \mathcal{E}nt(\nu) \Big\} \\ = \varepsilon \sup_h \Big\{ \int \log h \, d\nu - \log P_\varepsilon h \, d\mu \Big\} + \frac{\varepsilon}{2}\left(\mathcal{E}nt(\mu) - \mathcal{E}nt(\nu)\right), \end{aligned}$$

hence we recover formally the identity (5.12). It is interesting to notice that the Hamilton–Jacobi equation (5.14) has a "smooth" solution given by (5.13). Let us note that the dual formulation of the Schrödinger problem has also been computed recently as a particular case in [13].

5.4. Contraction inequality

We are now able to give a new and rigorous result.

Theorem 5.5 (Contraction under (ρ, n)-convexity). *Suppose that (N, g) is a smooth, connected n-dimensional Riemannian manifold with $\mathrm{Ric}_g \geq \rho$ for some $\rho \in \mathbb{R}$. Then for any smooth and positive probability measures μ and ν, and for any $t \geq 0$,*

$$\mathcal{A}^\varepsilon_{\mathcal{E}nt}(\mathsf{S}_t^{\mathcal{E}nt}\mu, \mathsf{S}_t^{\mathcal{E}nt}\nu) \leq e^{-2\rho t}\mathcal{A}^\varepsilon_{\mathcal{E}nt}(\mu,\nu) - \frac{1}{n}\int_0^t e^{-2\rho(t-u)}\left(\mathcal{E}nt(\mathsf{S}_u^{\mathcal{E}nt}\mu) - \mathcal{E}nt(\mathsf{S}_u^{\mathcal{E}nt}\nu)\right)^2 du. \tag{5.15}$$

Proof. Let $(\mu_s)_{0 \leq s \leq 1}$ be any smooth and positive path between μ and ν satisfying

$$\partial_s \mu_s = -\nabla \cdot (\mu_s \nabla \Phi_s),$$

with a smooth velocity $\nabla\Phi_s$. Let us denote for any $t \geq 0$, $\mu_{s,t} := \mathsf{S}_t^{\mathcal{E}nt}(\mu_s)$. It is a path between $\mathsf{S}_t^{\mathcal{E}nt}(\mu)$ and $\mathsf{S}_t^{\mathcal{E}nt}(\nu)$. From the continuity equation,

$$\partial_s \mathsf{S}_t^{\mathcal{E}nt}(\mu_s) = -\mathsf{S}_t^{\mathcal{E}nt}(\nabla\cdot(\mu_s\nabla\Phi_s)) = -\nabla\cdot\Big(\mathsf{S}_t^{\mathcal{E}nt}(\mu_s)\,\frac{\mathsf{R}_t(\mu_s\nabla\Phi_s)}{\mathsf{S}_t^{\mathcal{E}nt}(\mu_s)}\Big),$$

where $(\mathsf{R}_t)_{t\geq 0}$ is the Hodge–de Rham semigroup on forms. The definition and properties of this semigroup can be found in this context in [26]. Of course the s-velocity $\dot\mu_{s,t}$ of $\mu_{s,t}$ is not equal to $\mathsf{R}_t(\mu_s\nabla\Phi_s)/\mathsf{S}_t^{\mathcal{E}nt}(\mu_s)$ since the latter is not a gradient. But some inequality remains valid:

$$|\dot\mu_{s,t}|^2_{\mu_{s,t}} \leq \int \Big|\frac{\mathsf{R}_t(\mu_s\nabla\Phi_s)}{\mathsf{S}_t^{\mathcal{E}nt}(\mu_s)}\Big|^2 \mathsf{S}_t^{\mathcal{E}nt}(\mu_s)\, d\mathrm{vol}\,.$$

Hence

$$\mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mathsf{P}_t f,\mathsf{P}_t g) \leq \int_0^1\int \Big[\frac{1}{2}\Big|\frac{\mathsf{R}_t(\mu_s\nabla\Phi_s)}{\mathsf{P}_t(\mu_s)}\Big|^2 + \frac{\varepsilon^2}{2}\big|\nabla\log\mathsf{P}_t(\mu_s)\big|^2\Big]\mathsf{P}_t(\mu_s)\, d\mathrm{vol}\, ds. \tag{5.16}$$

Now we follow Theorem 3.8 in [26]. Let t be fixed and for any $r \in [0,t]$,

$$\Lambda(r) := \mathsf{P}_r\bigg(\frac{|\mathsf{R}_{t-r}(\omega)|^2}{\mathsf{P}_{t-r}(g)}\bigg) + \varepsilon^2\mathsf{P}_r(\Gamma(\log\mathsf{P}_{t-r}g)\mathsf{P}_{t-r}g),$$

where ω is some given 1-form and g is some given smooth and positive function on N. First, from the usual computations on Markov semigroups by using Theorem 5.5.3 in [5] and the Ricci bound on (N,g),

$$\frac{\partial}{\partial_r}\,\mathsf{P}_r\big(\Gamma(\log\mathsf{P}_{t-r}g)\mathsf{P}_{t-r}g\big) = 2\mathsf{P}_r(\Gamma_2(\log\mathsf{P}_{t-r}g)\mathsf{P}_{t-r}g) \geq 2\rho\mathsf{P}_r(\Gamma(\log\mathsf{P}_{t-r}g)\mathsf{P}_{t-r}g).$$

As in the proof of Theorem 3.8 in [26], we obtain

$$\begin{aligned}\Lambda'(r) &\geq 2\rho\,\mathsf{P}_r\Big(\frac{|\mathsf{R}_{t-r}(\omega)|^2}{\mathsf{P}_{t-r}(g)}\Big) + 2\rho\,\varepsilon^2\,\mathsf{P}_r(\Gamma(\log\mathsf{P}_{t-r}g)\mathsf{P}_{t-r}g)\\ &\quad+\frac{2}{n}\,\frac{[\mathsf{P}_t(\nabla\cdot\omega) - \mathsf{P}_r(\nabla(\log\mathsf{P}_{t-r}g)\cdot\mathsf{R}_{t-r}\omega)]^2}{\mathsf{P}_r g}\\ &= 2\rho\Lambda(r) + \frac{2}{n}\,\frac{[\mathsf{P}_t(\nabla\cdot\omega) - \mathsf{P}_r(\nabla(\log\mathsf{P}_{t-r}g)\cdot\mathsf{R}_{t-r}\omega)]^2}{\mathsf{P}_r g}.\end{aligned}$$

Integrating over $[0,t]$,

$$\Lambda(0) \leq e^{-2\rho t}\Lambda(t) - \frac{2}{n}\int_0^t e^{-2\rho r}\,\frac{[\mathsf{P}_t(\nabla\cdot\omega) - \mathsf{P}_r(\nabla(\log\mathsf{P}_{t-r}g)\cdot\mathsf{R}_{t-r}\omega)]^2}{\mathsf{P}_r g}\,dr.$$

Let us choose $\omega = \mu_s\nabla\Phi_s$ and $g = \mu_s$. Integrate this inequality with respect to the probability measure $d\mathrm{vol}$, by (5.16)

$$\begin{aligned}&\mathcal{A}^{\varepsilon}_{\mathcal{E}nt}(\mathsf{P}_t f,\mathsf{P}_t g) \leq e^{-2\rho t}\frac{1}{2}\int_0^1\int\big[|\nabla\Phi_s|^2\,d\mu_s + \varepsilon^2\Gamma(\log\mu_s)\big]\,d\mu_s ds\\ &-\frac{1}{n}\int_0^t\int_0^1\int e^{-2\rho r}\,\frac{[\mathsf{P}_t(\nabla\cdot(\mu_s\nabla\Phi_s)) - \mathsf{P}_r(\nabla(\log\mathsf{P}_{t-r}\mu_s)\cdot\mathsf{R}_{t-r}(\mu_s\nabla\Phi_s))]^2}{\mathsf{P}_r\mu_s}\,d\mathrm{vol}\,ds\,dr.\end{aligned}$$

Applying twice the Cauchy–Schwarz inequality,

$$\int_0^t \int_0^1 \int e^{-2\rho r} \frac{\left[\mathsf{P}_t(\nabla \cdot (\mu_s \nabla \Phi_s)) - \mathsf{P}_r(\nabla(\log \mathsf{P}_{t-r}\mu_s) \cdot \mathsf{R}_{t-r}(\mu_s \nabla \Phi_s))\right]^2}{\mathsf{P}_r \mu_s} \, d\mathrm{vol}\, ds\, dr$$
$$\geq \int_0^t e^{-2\rho(t-r)} \Big(\int_0^1 \int \nabla(\log \mathsf{P}_r \mu_s) \cdot \mathsf{R}_r(\mu_s \nabla \Phi_s) \; d\mathrm{vol}\, ds \Big)^2 \, dr.$$

Now, since

$$\int \nabla(\log \mathsf{P}_r \mu_s) \cdot \mathsf{R}_r(\mu_s \nabla \Phi_s)\, d\mathrm{vol} = -\int \log(\mathsf{P}_r \mu_s) \mathsf{P}_r(\mu_s \nabla \Phi_s)\, d\mathrm{vol} = \frac{d}{ds} \mathcal{E}nt(\mathsf{P}_r \mu_s),$$

we have

$$\int_0^1 \int \nabla(\log \mathsf{P}_r \mu_s) \cdot \mathsf{R}_r(\mu_s \nabla \Phi_s)\, d\mathrm{vol}\, ds = \mathcal{E}nt(\mathsf{P}_r \nu) - \mathcal{E}nt(\mathsf{P}_r \mu).$$

Taking the infimum over all smooth and positive paths (μ_s) leads to the announced inequality. □

Remark 5.6. This new result is interesting since the first one, proved in [27], was not satisfactory at the light of this one. Indeed, we do not need to use a distortion of the time to obtain a contraction inequality. The Benamou–Brenier–Schrödinger formulation, namely the definition of the cost $\mathcal{A}^{\varepsilon}_{\mathcal{F}}$, appears as an efficient method to deal with the Schrödinger problem and many other inequalities.

Of course, this contraction inequality can be generalized to a more general operator $\Delta_g + V$ where V is a vector field in (N, g).

5.5. Conclusion

• The general cost $\mathcal{A}^{\varepsilon}_{\mathcal{F}}$ shares the same curvature properties than the Wasserstein distance, which is quite surprising for the Schrödinger problem as a minimisation problem of the relative entropy.

• We believe that the $\mathcal{F}$-interpolations are smooth as in the case of the Schrödinger problem. This gives a way used in [30] to reach the second order differentiation formula on the Wasserstein space.

References

[1] Ambrosio, L. and Feng, J.: On a class of first order Hamilton–Jacobi equations in metric spaces. *J. Differential Equations* **256** (2014), no. 7, 2194–2245.

[2] Ambrosio, L. and Gigli, N.: A user's guide to optimal transport. In *Modelling and optimisation of flows on networks,* 1–155. Lecture Notes in Math. 2062, Fond. CIME/CIME Found. Subser., Springer, Heidelberg, 2013.

[3] Ambrosio, L., Gigli N. and Savaré, G.: *Gradient flows in metric spaces and in the space of probability measures.* Second edition. Lectures in Mathematics, ETH Zürich, Birkhäuser Verlag, Basel, 2008.

[4] Bakry, D. and Emery, M.: Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84,* 177206. Lecture Notes in Math. 1123, Springer, Berlin, 1985.

[5] Bakry, D., Gentil, I. and Ledoux, M.: *Analysis and geometry of Markov diffusion operators.* Grundlehren der Mathematischen Wissenschaften 348, Springer, Cham, 2014.

[6] Bakry, D., Gentil, I. and Ledoux, M.: On Harnack inequalities and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **14** (2015), no. 3, 705–727.

[7] Benamou, J.-D. and Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** (2000), no. 3, 375–393.

[8] Bernstein, S.: Sur les liaisons entre les grandeurs aléatoires. In *Verhand. Internat. Math. Kongr., vol. 1, 288–309.* Zürich, 1932.

[9] Beurling, A.: An automorphism of product measures. *Ann. of Math. (2)* **72** (1960), 189–200.

[10] Bolley, F., Gentil, I. and Guillin, A.: Dimensional contraction via Markov transportation distance. *J. Lond. Math. Soc. (2)* **90** (2014), no. 1, 309–332.

[11] Bolley, F., Gentil, I. and Guillin, A.: Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp–Lieb inequalities. *Ann. Probab.* **46** (2018), no. 1, 261–301.

[12] Bolley, F., Gentil, I., Guillin, A. and Kuwada, K.: Equivalence between dimensional contractions in Wasserstein distance and the curvature-dimension condition. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **18** (2018), no. 3, 845–880.

[13] Bowles, M. and Ghoussoub, N.: A theory of transfers: Duality and convolution. Preprint, arXiv: `1804.08563`, 2018.

[14] Carrillo, J. A., McCann, R. J. and Villani, C.: Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Rev. Mat. Iberoamaricana* **19** (2003), no. 3, 971–1018.

[15] Chen, Y., Georgiou, T. T. and Pavon, M.: On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.* **169** (2016), no. 2, 671–691.

[16] Chung, K. L. and Zambrini, J.-C.: *Introduction to random time and quantum randomness.* World Scientific, 2003.

[17] Conforti, G.: Second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probab. Theory Related Fields* **174** (2019), no. 1-2, 1–47.

[18] Cruzeiro, A. B. and Zambrini, J.-C.: Malliavin calculus and Euclidean quantum mechanics. I. Functional calculus. *J. Funct. Anal.* **96** (1991), no. 1, 62–95.

[19] Daneri, S. and Savaré, G.: Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.* **40** (2008), no. 3, 1104–1122.

[20] Erbar, M., Kuwada,K. and Sturm, K.-T.: On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.* **201** (2015), no. 3, 993–1071.

[21] Evans, L. C.: *Partial differential equations.* Graduate Studies in Mathematics 19, American Mathematical Society, Providence, RI, 1998.

[22] Feng, J. and Nguyen, T.: Hamilton–Jacobi equations in space of measures associated with a system of conservation laws. *J. Math. Pures Appl. (9)* **97** (2012), no. 4, 318–390.

[23] Föllmer, H.: Random fields and diffusion processes. In *École d'Été de Probabilités de Saint-Flour XVXVII, 1985-87*, 101–203. Lecture Notes in Math. 1362, Springer, Berlin, 1988.

[24] Gangbo, W., Nguyen, T. and Tudorascu, A.: Hamilton–Jacobi equations in the Wasserstein space. *Methods Appl. Anal.* **15** (2008), no. 2, 155–184.

[25] Gangbo, W. and Swiech, A.: Existence of a solution to an equation arising from the theory of mean field games. *J. Differential Equations* **259** (2015), no. 11, 6573–6643.

[26] Gentil, I.: Dimensional contraction in Wasserstein distance for diffusion semigroups on a Riemannian manifold. *Potential Anal.* **42** (2015), no. 4, 861–873.

[27] Gentil, I., Léonard, C. and Ripani, L.: About the analogy between optimal transport and minimal entropy.. *Ann. Fac. Sci. Toulouse Math. (6)* **26** (2017), no. 3, 569–601.

[28] Gianazza, U., Savaré, G. and Toscani, G: The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.* **194** (2009), no. 1, 133–220.

[29] Gigli, N.: Second order analysis on $(\mathcal{P}_2(M), W_2)$. *Mem. Amer. Math. Soc.* **216** (2012), no. 1018, 154 pp.

[30] Gigli, N. and Tamanini, L.: Second order differentiation formula on compact $\mathrm{RCD}^*(K,N)$ spaces. To appear in *J. Eur. Math. Soc.*

[31] Gigli, N. and Tamanini, L.: Benamou–Brenier and Kantorovich duality formulas for the entropic cost on $\mathrm{RCD}^*(K,N)$ spaces. To appear in *Probab. Theory Relat. Fields.*, doi: 10.1007/s00440-019-00909-1, 2019.

[32] Hynd, R. and Kim, H. K.: Value functions in the Wasserstein spaces: finite time horizons. *J. Funct. Anal.* **269** (2015), no. 4, 968–997.

[33] Jamison, B.: The Markov processes of Schrödinger. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** (1975), no. 4, 323-331.

[34] Jordan, R., Kinderlehrer, D. and Otto, F.: The variational formulation of the Fokker–Planck equation. *SIAM J. Math. Anal.* **29** (1998), no. 1, 1–17.

[35] Krener, A.: Reciprocal diffusions in flat space. *Probab. Theory Related Fields* **107** (1997), no. 2, 243–281.

[36] Kuwada, K.: Space-time Wasserstein controls and Bakry–Ledoux type gradient estimates. *Calc. Var. Partial Differential Equations* **54** (2015), no. 1, 127–161.

[37] Léonard, C.: From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* **262** (2012), no. 4, 1879–1920.

[38] Léonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34** (2014), no. 4, 1533–1574.

[39] Léonard, C.: On the convexity of the entropy along entropic interpolations. In *Measure theory in non-smooth spaces*, 195–242. Partial Differential Equations and Measure Theory, De Gruyter Open, Warsaw, 2017.

[40] Lott, J. and Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)* **169** (2009), no. 3, 903–991.

[41] McCann, R. J.: Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** (1995), no. 2, 309–323.

[42] Mikami, T.: Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Related Fields* **129** (2004), no. 2, 245–260.

[43] Mikami, T. and Thieullen, M.: Duality theorem for the stochastic optimal control problem. *Stochastic Process. Appl.* **116** (2006), no. 12, 1815–1835.

[44] Otto, F.: The geometry of dissipative evolution equations: The porous medium equation. *Commun. Partial Differential Equations* **26** (2001), no. 1-2, 101–174.

[45] Otto, F. and Westdickenberg, M.: Eulerian calculus for the contraction in the Wasserstein distance. *SIAM J. Math. Anal.* **37** (2005), no. 4, 1227–1255.

[46] von Renesse, M.-K.: An optimal transport view of Schrödinger's equation. *Canad. Math. Bull.* **55** (2012), no. 4, 858–869.

[47] von Renesse, M.-K. and Sturm, K.-T.: Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure Appl. Math.* **58** (2005), no. 7, 923–940.

[48] Ripani, L.: Convexity and regularity properties for entropic interpolations. *J. Funct. Anal.* **277** (2019), no. 2, 368–391.

[49] Schrödinger, E.: Über die Umkehrung der Naturgesetze. Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math., 144–153, 1931.

[50] Schrödinger, E.: Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré* **2** (1932), no. 4, 269–310.

[51] Sturm, K.-T.: On the geometry of metric measure spaces. I and II. *Acta Math.* **196** (2006), no. 1, 65–177.

[52] Tamanini, L.: *Analysis and geometry of RCD spaces via the Schrödinger problem.* PhD Thesis, Pisa, 2017.

[53] Thieullen, M.: Second order stochastic differential equations and non-Gaussian reciprocal diffusions. *Probab. Theory Related Fields* **97** (1993), no. 1-2, 231–257.

[54] Thieullen, M. and Zambrini, J.-C.: Symmetries in the stochastic calculus of variations. *Probab. Theory Related Fields* **107** (1997), no. 3, 401–427.

[55] Villani, C.: *Optimal transport. Old and new.* Grundlehren der Mathematischen Wissenschaften 338, Springer-Verlag, Berlin, 2009.

[56] Zambrini, J.-C.: Variational processes and stochastic versions of mechanics. *J. Math. Phys.* **27** (1986), 2307–2330.

Received June 7, 2018. Published online January 10, 2020.

Ivan Gentil: Univ. Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, 69622 Villeurbanne, France.
E-mail: gentil@math.univ-lyon1.fr

Christian Léonard: Modal-X, Université Paris Nanterre, 92001 Nanterre, France.
E-mail: christian.leonard@math.cnrs.fr

Luigia Ripani: Univ. Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, 69622 Villeurbanne, France.
E-mail: luigia.ripani@u-cergy-pontoise.fr

This research was supported by the French ANR-17-CE40-0030 EFI project, Labex MILYON (ANR-10-LABX-0070) and Labex MME-DII (ANR11-LBX-0023-01).