Regularity of the Schrödinger cost.

Gauthier Clerc

Regularity properties of the Schrödinger cost

Gauthier Clerc*

January 3, 2022

Abstract

The Schrödinger problem is an entropy minimisation problem on the space of probability measures. Its optimal value is a cost between two probability measures. In this article we investigate some regularity properties of this cost: continuity with respect to the marginals and time derivative of the cost along probability measures valued curves.

1 Introduction

The Schrödinger problem was formulated by Schrödinger himself in the articles [Sch31, Sch32] in the thirties. The modern approach of this problem has been mainly developed in the two seminal papers [Fol88] and [Lé14]. The discovery in [Mik04] that the Monge-Kantorovitch problem is recovered as the short time limit of the Schrödinger problem has triggered intense research activities in the last decade. This interest is due to the fact that adding an entropic penalty in the Monge-Kantorovitch problem leads to major computationnal advantages using the Sinkhorn algorithm (see for instance [PC19]). The Schrödinger problem can also be a fruitful tool to prove some functional inequalities (see [GGI20b], [ICLL20], etc.).
The problem is, observing the empirical distribution of a cloud of Brownian particles at time $t = 0$ and $t = 1$, to find the distribution at time $0 < s < 1$ of the cloud. In modern language, this is an entropy minimization problem. The relative entropy of two measures is loosely defined by

$$H(p|r) := \begin{cases} \int \log\left(\frac{dp}{dr}\right) dp \text{ if } p \ll r, \\ +\infty \text{ else.} \end{cases}$$

We leave the precise definition of the relative entropy to the main body of the paper. Given two probability measures μ and ν on a Riemannian manifold N equipped with a generator L of reversible measure m, the Schrödinger cost is defined as

$$\mathrm{Sch}(\mu, \nu) := \inf H(\gamma|R_{01}).$$

Here the infimum is taken over every probability measures on $N \times N$ with μ and ν as marginals and R_{01} is the joint law of initial and final position of the unique diffusion measure with generator

*Institut Camille Jordan, Umr Cnrs 5208, Université Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, F-69622 Villeurbanne cedex. Clerc@math.univ-lyon1.fr

L on $C([0,1],N)$ starting from m. Independently proven in different papers (see [CGP16, GLR17, GLR20, GT20]), the Benamou-Brenier-Schrödinger formula state that

$$\mathrm{Sch}(\mu,\nu) = \frac{\mathcal{C}(\mu,\nu)}{4} + \frac{H(\mu|m) + H(\nu|m)}{2}.$$

Here $\mathcal{C}(\mu,\nu)$ is the entropic cost given by

$$\mathcal{C}(\mu,\nu) := \inf \int_0^1 \left(\|v_s\|^2_{L^2(\mu_s)} + \left\| \nabla \log \frac{d\mu_s}{dm} \right\|^2_{L^2(\mu_s)} \right) ds, \tag{1}$$

and the infimum is taken over every $(\mu_s, v_s)_{0\leq s\leq 1}$ such that $(\mu_s)_{0\leq s\leq 1}$ is an absolutely continuous path with respect to the Wasserstein distance which connects μ to ν and satisfies in a weak sense for every $s \in [0,1]$

$$\begin{cases} v_s \in L^2(\mu_s), \\ \partial_s \mu_s = -\nabla \cdot (\mu_s v_s). \end{cases}$$

In this paper we investigate regularity properties of the functions $(\mu,\nu) \mapsto \mathrm{Sch}(\mu,\nu)$ and $(\mu,\nu) \mapsto \mathcal{C}(\mu,\nu)$. To my knowledge, regularity properties of the Schrödinger cost as function over probability measures haven't been investigate yet, but the stability of optimizer has been investigate in [Tam17] and more recently in [PME21] and [JP19]. We give an overview of the main contributions of this paper, leaving precise statements to other sections.

- In Section 3 we investigate continuity properties of the cost function $\mathcal{C}$. In Theorem 3.1 we show that
$$\lim_{k\to\infty} \mathcal{C}(\mu_k,\nu_k) = \mathcal{C}(\mu,\nu)$$
if $W_2(\mu_k,\mu) \underset{k\to\infty}{\to} 0$ (resp. ν_k to ν) with additional hypothesis about the entropy and the Fisher information along the sequences.
- In Section 4 we provide a few applications of the preceding continuity properties. The main result of this section is that using the continuity properties of Sch and $\mathcal{C}$ we are able to show that the Benamou-Brenier-Schrödinger formula (1) is valid assuming that both measures have finite entropy, finite Fisher information and locally bounded densities. Up to my knowledge this is a new result up to my knowledge, because no compactness assumptions are needed for the two marginals.
- In Section 5 we investigate the question of the derivability of the functions $t \mapsto \mathrm{Sch}(\mu_t,\nu_t)$ and $t \mapsto \mathcal{C}(\mu_t,\nu_t)$, where $(\mu_t)_{t\geqslant 0}$ and $(\nu_t)_{t\geqslant 0}$ are some curves on the Wasserstein space. These results extend the existing ones for the Wasserstein distance, see [AGS08, Theorem 8.4.7] and [Vil09, Theorem 23.9]. We prove that the derivative of the entropic cost is given for almost every t, by
$$\frac{d}{dt}\mathcal{C}(\mu_t,\nu_t) = \langle \dot{\mu}^t_s|_{s=1}, \dot{\nu}_t \rangle_{L^2(\nu_t)} - \langle \dot{\mu}^t_s|_{s=0}, \dot{\mu}_t \rangle_{L^2(\mu_t)},$$
where $(\mu^t_s)_{s\in[0,1]}$ is the minimizer of the problem (1) from μ_t to ν_t and $\dot{\mu}^t_s$ is the velocity of the path $s \mapsto \mu^t_s$ defined in other section. Such minimizers are called entropic interpolations. Note that this is exactly the formula which holds for the Wasserstein distance, replacing the Wasserstein geodesics by the entropic interpolations. For technical reasons we prove this formula in the case where $N = \mathbb{R}^n$ and L is the classical Laplacian operator.

2 Setting of our work

2.1 Markov semigroups

Let $(N, \mathfrak{g})$ be a smooth, connected and complete Riemannian manifold. We denote dx the Riemannian measure and $\langle \cdot, \cdot \rangle$ the Riemannian metric (we omit $\mathfrak{g}$ for simplicity). Let ∇ denote the gradient operator associated to $(N, \mathfrak{g})$ and $\nabla \cdot$ be the associated divergence in order to have for every smooth function f and vector field ζ

$$\int \langle \nabla f(x), \zeta(x) \rangle dx = - \int f(x) \nabla \cdot \zeta(x) dx.$$

Hence the Laplace-Beltrami operator can be defined as $\Delta = \nabla \cdot \nabla$. We consider a differential generator $L := \Delta - \langle \nabla, \nabla W \rangle$ for some smooth function $W : N \to \mathbb{R}$. We define the carré du Champ operator for every smooth functions f and g by

$$\Gamma(f, g) := \frac{1}{2} \left(L(fg) - fLg - gLf \right).$$

Under our current hypotheses we have $\Gamma(f) := \Gamma(f, f) = |\nabla f|^2$, which is the length of ∇f with to the Riemannian metric $\mathfrak{g}$. Let $Z := \int e^{-W} dx$, then if $Z < \infty$ the reversible probability measure associated with L is given by

$$dm := \frac{e^{-W}}{Z} dx.$$

If $Z = \infty$, the reversible measure associated with L is $dm := e^{-W} dx$ of infinite mass. Following the work of [BE85] we define the iterated carré du champ operator given by

$$\Gamma_2(f, g) = \frac{1}{2} \left(L\Gamma(f, g) - \Gamma(Lf, g) - \Gamma(f, Lg) \right),$$

for any smooth functions f and g and we denote $\Gamma_2(f) := \Gamma_2(f, f)$. We say that the operator L verifies the $CD(\rho, n)$ curvature-dimension condition with $\rho \in \mathbb{R}$ and $n \in (0, \infty]$ if for every smooth function f

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

For instance, $\mathbb{R}^n$ endowed with the classical Laplacian operator verify the $CD(0, n)$ curvature-dimension condition. With the Ornstein-Uhlenbeck operator, $\mathbb{R}^n$ verify the $CD(1, \infty)$ curvature-dimension condition. More generally a Riemannian manifold of dimension $n \in \mathbb{N}$ and with a Ricci tensor bounded from below by $\rho \in \mathbb{R}$ endowed with his Laplace-Beltrami operator verify the $CD(\rho, n)$ curvature-dimension condition. We assume that L is the generator of a Markov semigroup $(P_t)_{t \geqslant 0}$, this is for example the case when a $CD(\rho, \infty)$ curvature-dimension condition holds for some $\rho \in \mathbb{R}$. For every $f \in L^2(m)$ the family $(P_t f)_{t \geqslant 0}$ is defined as the unique solution of the Cauchy system

$$\begin{cases} \partial_t u = Lu, \\ u(\cdot, 0) = f(\cdot). \end{cases}$$

Under the $CD(\rho, \infty)$ curvature-dimension condition this Markov semigroup admit a probability kernel $p_t(x, dy)$ with density $p_t(x, y)$, that is for every $t \geqslant 0$ and $f \in L^2(m)$

$$\forall x \in N, \ P_t f(x) = \int f(y) p_t(x, dy) = \int f(y) p_t(x, y) dm(y),$$

for the existence of the kernel see [Gri09, Theorem 7.7]. We also define the dual semigroup $(P_t^*)_{t\geqslant 0}$ which acts on probability measures. Given a probability measure μ the family $(P_t^*\mu)_{t\geqslant 0}$ is given by the following equation

$$\int f dP_t^*\mu = \int P_t f d\mu,$$

for every $t \geqslant 0$ and every test function f. When $\mu \ll m$, we have $\frac{dP_t^*\mu}{dm} = P_t\left(\frac{d\mu}{dm}e^W\right)$. The function $(t,x) \mapsto \frac{dP_t^*\mu}{dx}(x)$ is a solution of the following Fokker-Planck type equation

$$\partial_t \nu_t = L^*\nu_t := \Delta\nu_t + \nabla\cdot(\nu_t \nabla W), \tag{2}$$

with initial value $\frac{d\mu}{dx}$. Here L^* is the dual operator of L in $L^2(dx)$.

2.2 Wasserstein space and absolutely continuous curves

The set $\mathcal{P}_2(N)$ of probability measures on N with finite second order moment can be endowed with the Wasserstein distance given for every $\mu,\nu \in \mathcal{P}_2(N)$ by,

$$W_2^2(\mu,\nu) := \inf \sqrt{\int d^2(x,y) d\pi(x,y)},$$

where the infimum is running over all $\pi \in \mathcal{P}(N\times N)$ with μ and ν as marginals and d is the Riemannian distance on (N, $\mathfrak{g}$). Recall that a path $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}_2(N)$ is absolutely continuous with to the Wasserstein distance W_2 if and only if

$$|\dot{\mu}_t| := \lim_{s\to t} \frac{W_2(\mu_s,\mu_t)}{|t-s|} \in L^1([0,1]).$$

In this case, there exists a unique vector field $(V_t)_{t\in[0,1]}$ such that $V_t \in L^2(\mu_t)$ and $|\dot{\mu}_t| = \|V_t\|_{L^2(\mu_t)}$. Furthermore this vector field can be characterized as the solution of the continuity equation

$$\partial_t \mu_t = -\nabla\cdot(V_t\mu_t)$$

with minimal norm in $L^2(\mu_t)$. We denote $\dot{\mu}_t = V_t$, and $(\dot{\mu}_t)_{t\in[0,1]}$ is called the velocity vector field of $(\mu_t)_{t\in[0,1]}$ or the velocity for short. Sometimes we also use the notation $\mathbf{dt}\mu_t = \dot{\mu}_t$.
In the famous paper [BB00] Benamou and Brenier showed that the Wasserstein distance admits a dynamical formulation

$$W_2^2(\mu,\nu) = \inf \int_0^1 \|\dot{\mu}_t\|^2_{L^2(\mu_t)} dt, \tag{3}$$

where the infimum is running over all absolutely continuous paths which connect μ to ν in $\mathcal{P}_2(N)$. In his article [Ott01], Felix Otto gave birth to a theory which allowed us to consider $(\mathcal{P}_2(N), W_2)$, heuristically at least, as an infinite dimensionnal Riemannian manifold. This theory was baptised "Otto calculus" later by Cédric Villani. For every $\mu \in \mathcal{P}_2(N)$ the tangent space of $\mathcal{P}_2(N)$ at μ can be defined as

$$T_\mu\mathcal{P}_2(N) := \overline{\{\nabla\varphi :\ \varphi \in C_c^\infty(N)\}}^{L^2(\mu)},$$

and the Riemannian metric is induced by the scalar product $\langle\cdot,\cdot\rangle_{L^2(\mu)}$, see for instance [Gig12, Section 1.4] or [GLR17, Section 3.2].

As in the Riemannian case, the acceleration of a curve can be defined as the covariant derivative of the veolcity field along the curve itself. If $(\mu_t)_{t\in[0,1]}$ is an absolutely continuous curve in $\mathcal{P}_2(N)$ and $(v_t)_{t\in[0,1]}$ is a vector field along $(\mu_t)_{t\in[0,1]}$, for every $t \in [0,1]$ we denote by $\mathbf{D_t}v_t$ the covariant derivative of v_t along $(\mu_t)_{t\in[0,1]}$ defined in [GLR17, Section 3.3]. It turns out that in the case where the velocity field of $(\mu_t)_{t\in[0,1]}$ has the form $(\nabla\varphi_t)_{t\in[0,1]}$ then the acceleration of $(\mu_t)_{t\in[0,1]}$ is given by

$$\forall t \in [0,1],\ \ddot{\mu}_t := \mathbf{D_t}\dot{\mu}_t = \nabla\left(\frac{d}{dt}\varphi_t + \frac{1}{2}\Gamma(\varphi_t)\right),$$

see [GLR17, Section 3.3]. Covariant derivative and acceleration can be defined in more general framework, see [Gig12, Section 5.1].

2.3 Schrödinger problem

Here we introduce the Schrödinger problem by his modern definition, following the two seminal papers [Lé14] and [Fol88]. The first object of interest is the relative entropy of two measures. The relative entropy of a probability measure p with to a measure r is loosely defined by

$$H(p|r) := \int \log\left(\frac{dp}{dr}\right) dp, \tag{4}$$

if $p \ll r$ and $+\infty$ elsewise. This definition is meaningful when r is a probability measure but not necessarily when r is unbounded. Assuming that r is σ-finite, there exists a function $\mathcal{W} : M \to [1,\infty)$ such that $z_{\mathcal{W}} := \int e^{-\mathcal{W}} dr < \infty$. Hence we can define a probability measure $r_{\mathcal{W}} := z_{\mathcal{W}}^{-1} e^{-\mathcal{W}} r$ and for every measure p such that $\int \mathcal{W} dp < \infty$

$$H(p|r) := H(p|r_{\mathcal{W}}) - \int \mathcal{W} dp - \log(z_{\mathcal{W}}),$$

where $H(p|r_{\mathcal{W}})$ is defined by the equation (4).
For $\mu, \nu \in \mathcal{P}(N)$ we define the Schrödinger cost from μ to ν by

$$\mathrm{Sch}(\mu,\nu) := \inf\left\{H(\gamma|R_{01}) :\ \gamma \in \mathcal{P}(N\times N),\ \gamma_0 = \mu,\ \gamma_1 = \nu\right\},$$

where R_{01} is the joint law of the initial and final position of the Markov process associated with L starting from m, which is given by

$$dR_{01}(x,y) = p_1(x,y)dm(x)dm(y).$$

To ensure the existence and unicity of minimizer, more hypothesis are needed. Namely we assume that there exists two non-negative mesurable functions $A, B : N \to \mathbb{R}$ such that

(i) $p_1(x,y) \geqslant e^{-A(x)-A(y)}$ uniformly in $x, y \in N$;

(ii) $\int e^{-B(x)-B(y)} p_1(x,y)\, m(dx)m(dy) < \infty$;

(iii) $\int (A+B)\, d\mu, \int (A+B)\, d\nu < \infty$;

(iv) $-\infty < H(\mu|m), H(\nu|m) < \infty$.

We define the set

$$\mathcal{P}_2^*(N) := \left\{\mu \in \mathcal{P}_2(N) :\ -\infty < H(\mu|m) < \infty,\ \int (A+B)\, d\mu < \infty\right\}.$$

If $\mu, \nu \in \mathcal{P}_2^*(N)$, it is proven that the Schrödinger cost $\mathrm{Sch}(\mu,\nu)$ is finite and admits a unique minimizer which takes the form

$$d\gamma = f \otimes g dR_{01}, \tag{5}$$

for two mesurable non-negative functions f and g, see [Tam17, Proposition 4.1.5]. Another fundamental result about the Schrödinger problem is an analogous formula to (3) for the Schrödinger cost.

Theorem 2.1 (Benamou-Brenier-Schrödinger formula) *Let $\mu, \nu \in \mathcal{P}_2^*(N)$ be two probability measures compactly supported and with bounded densities with to m. Then the following formula holds*

$$\mathrm{Sch}(\mu,\nu) = \frac{\mathcal{C}(\mu,\nu)}{4} + \frac{\mathcal{F}(\mu)+\mathcal{F}(\nu)}{2}, \tag{6}$$

where $\mathcal{C}(\mu,\nu)$ is the entropic cost between μ and ν given by

$$\mathcal{C}(\mu,\nu) := \inf \int_0^1 \left(\|\dot{\mu}_s\|^2_{L^2(\mu_s)} + \|\nabla \log(\mu_s)\|^2_{L^2(\mu_s)} \right) ds.$$

Here the infimum is running over every absolutely continuous path $(\mu_s)_{s\in[0,1]}$ which connects μ to ν in $\mathcal{P}_2(N)$ and $\mathcal{F}$ is defined as

$$\mathcal{F}(\mu) := H(\mu|m).$$

Different versions of this theorem have been obtained under various hypothesis, see [CGP16, GLR17, GLR20, GT20].
The functional $\mathcal{F} : \mathcal{P}_2(N) \to [0,\infty]$ is central on this work. Its gradient can be identified by the equation $\frac{d}{dt}\mathcal{F}(\mu_t) = \langle \mathbf{grad}_{\mu_t}\mathcal{F}, \dot{\mu}_t\rangle_{\mu_t}$ and is given for every $\mu \in \mathcal{P}_2(N)$ with smooth density against m by

$$\mathbf{grad}_\mu \mathcal{F} := \nabla \log\left(\frac{d\mu}{dm}\right).$$

Those definitions allowed us to see the Fokker-Planck type equation (2) as the gradient flow equation of $\mathcal{F}$. Indeed every solution $(\nu_t)_{t\geqslant 0}$ of this equation verify

$$\dot{\nu}_t = -\nabla\left(\log\frac{\nu_t}{dx} + W\right) = -\nabla \log\left(\frac{d\nu_t}{dm}\right) = -\mathbf{grad}_{\nu_t}\mathcal{F},$$

see [GLR20, Section 3.2]. With Otto calculus, we can also introduce the notions of Hessian and covariant derivative. A great fact is that the Hessian of $\mathcal{F}$ can be expressed in term of Γ_2, indeed

$$\forall \mu \in \mathcal{P}_2(N),\ \forall\ \nabla\varphi, \nabla\psi \in T_\mu\mathcal{P}_2(N),\ \mathbf{Hess}_\mu\mathcal{F}(\nabla\varphi,\nabla\psi) = \int \Gamma_2(\nabla\varphi,\nabla\psi)d\mu,$$

see [GLR17, Section 3.3]. The quantity $\mathcal{I}(\mu) := \left\|\nabla \log\frac{d\mu}{dm}\right\|^2_{L^2(\mu)}$ which appears in the previous definition is central in this work, it is called the Fisher information. According to the Otto calculus formalism, the Fisher information admits the nice interpretation,

$$\mathcal{I}(\mu) := \|\mathbf{grad}_\mu\mathcal{F}\|^2_{L^2(\mu)}.$$

Minimizers of the entropic cost $\mathcal{C}(\mu,\nu)$ are called entropic interpolations and take the form

$$\mu_t = P_t f P_{1-t} g dm,$$

where f and g are the two positive functions which appears in the equation (5). Due to this particular structure, velocity and acceleration of entropic interpolations can be explicitly computed. It holds that for every $t \in [0,1]$

$$\dot{\mu}_t = \nabla \left(\log P_{1-t} g - \log P_t f \right).$$

But the most important fact, is that entropic interpolations are solutions of the following Newton equation

$$\ddot{\mu}_t = \nabla \frac{d}{dt} \log \mu_t + \nabla^2 \log \mu_t \dot{\mu}_t,$$

which can be rewrite in the Otto calculus formalism as

$$\ddot{\mu}_t = \mathbf{Hess}_{\mu_t} \mathcal{F} \, \mathbf{grad}_{\mu_t} \mathcal{F}. \tag{7}$$

This equation was first derived in [Con19, Theorem 1.2], see also [GLR20, Sec 3.3, Propositon 3.5].

2.4 Flow maps

In this subsection we follow [Gig12, Sec 2.1]. We need this result only in the euclidean framework, hence in this subsection we take $N = \mathbb{R}^n$ for simplicity. A crucial ingredient of the proof of the Theorem 5.2 is, given a path $(\mu_t)_{t\in[0,1]}$, the existence of a family of maps $(T_{t\to s})_{t,s\in[0,1]}$ such that for every $s,t \in [0,1]$

$$\frac{d}{ds} T_{t\to s} = \dot{\mu}_s \circ T_{t\to s}$$

and

$$T_{t\to s} \# \mu_t = \mu_s.$$

These maps are called the flow maps associated with $(\mu_s)_{s\in[0,1]}$. The existence of such maps can be garanted by some regularity assumptions on the path. Before the statement we recall the definition of the Lipschitz constant of a vector field proposed by Gigli in [Gig12].

Definition 2.2 (Lipschitz constant of a vector field) *For every smooth compactly supported vector field ζ on $\mathbb{R}^n$ we define*

$$L(\zeta) := \sup_{x,y\in\mathbb{R}^n} \frac{|\zeta(x) - \zeta(y)|}{|x-y|}$$

Then for every $\mu \in \mathcal{P}_2(N)$ and every $v \in T_\mu \mathcal{P}_2(\mathbb{R}^n)$ we define

$$\mathcal{L}(v) := \inf \varliminf_{n\to\infty} L(\zeta_n),$$

where the infimum is taken over sequences $(\zeta_n)_{n\in\mathbb{N}}$ of smooth compactly supported vector fields which converges to v in $L^2(\mu)$ when $n \to \infty$.

Note that in the case where v is smooth and compactly supported $\mathcal{L}(v)$ is the Lipschitz constant of v.

Theorem 2.3 (Cauchy Lipschitz on manifolds, [Gig12, Theorem 2.6]) *Let $(\mu_t)_{t\in[0,1]} \subset \mathcal{P}_2(\mathbb{R}^n)$ be an absolutely continuous path such that*

$$\int_0^1 \mathcal{L}(\dot{\mu}_t)dt < \infty, \ \int_0^1 \|\dot{\mu}_t\|_{L^2(\mu_t)}dt < \infty.$$

Then there exists a family of maps $(T_{t\to s})_{t,s\in[0,1]}$ such that

$$\begin{cases} T_{t\to s} : \mathrm{supp}(\mu_t) \to \mathrm{supp}(\mu_s), & \forall t,s \in [0,1], \\ T_{t\to t}(x) = x, & \forall x \in \mathrm{supp}(\mu_t),\ t \in [0,1], \\ \frac{d}{dr} T_{t\to r_{|r=s}} = \dot{\mu}_s \circ T_{t\to s}, & \forall t \in [0,1],\ a.e - s \in [0,1]. \end{cases}$$

and the map $x \mapsto T_{t\to s}(x)$ is Lipschitz for every $s,t \in [0,1]$. Furthermore for every $s,t,r \in [0,1]$ and $x \in \mathrm{supp}(\mu_t)$

$$T_{r\to s} \circ T_{t\to r}(x) = T_{t\to s}(x),$$

and

$$T_{t\to s}\#\mu_t = \mu_s.$$

2.5 hypothesis about the heat kernel

Here is a summary of all hypothesis needed in all the paper.

(H1) The $CD(\rho,\infty)$ curvature-dimension condition holds for some $\rho \in \mathbb{R}$.

(H2) hypothesis (i) to (iv) in Section 2.3.

The first hypothesis (H1) is needed to defined Markov semigroups as introduced in [BGL14]. The second hypothesis (H2) is needed to ensure existence and unicity of minimizers of the Schrödinger problem. For instance those hypothesis hold true when $N = \mathbb{R}^n$ is equipped with the classical Laplacian operator or the Ornstein-Ulhenbeck operator, or when N is compact.

3 Continuity of the entropic cost

Here we are interested in the continuity of the function $(\mu,\nu) \mapsto \mathcal{C}(\mu,\nu)$ where $\mathcal{C}(\mu,\nu)$ is defined as an infimum over all absolutely continuous paths connecting μ to ν.

Theorem 3.1 (Continuity of the entropic cost) *Let $\mu,\nu \in \mathcal{P}_2^*(N)$ and $(\mu_k)_{k\in\mathbb{N}}, (\nu_k)_{k\in\mathbb{N}} \subset \mathcal{P}_2^*(N)$ be two sequences such that μ_k converges toward μ with to the Wasserstein distance (resp.. ν_k toward ν). We also assume that for every $k \in \mathbb{N}$ there exists an entropic interpolation from μ_k to ν_k (resp. from μ to ν) and*

$$\sup\{\mathcal{I}(\mu_k), \mathcal{I}(\nu_k);\ k \in \mathbb{N}\} < +\infty$$

and

$$\sup\{\mathcal{F}(\mu_k), \mathcal{F}(\nu_k);\ k \in \mathbb{N}\} < +\infty.$$

Then

$$\mathcal{C}(\mu_k,\nu_k) \underset{k\to\infty}{\to} \mathcal{C}(\mu,\nu).$$

Proof

◁ To begin we will show that

$$\overline{\lim_{k\to\infty}}\,\mathcal{C}(\mu_k,\nu_k)\leq\mathcal{C}(\mu,\nu).$$

To do so let us consider some particular path from μ_k to ν_k. For every $k\in\mathbb{N}$, $\varepsilon\in(0,1/2)$ and $\delta\in(0,\varepsilon/2)$, we define a path $\eta^{k,\varepsilon,\delta}$ from μ_k to μ given by

$$\eta_t^{k,\varepsilon,\delta}=\begin{cases}P_t^*(\mu_k), & t\in[0,\varepsilon/2-\delta),\\ P^*_{\varepsilon/2-\delta}(\gamma_t), & t\in[\varepsilon/2-\delta,\varepsilon/2+\delta],\\ P^*_{\varepsilon-t}(\mu), & t\in(\varepsilon/2+\delta,\varepsilon],\end{cases}$$

where for all $t\in(\varepsilon/2-\delta,\varepsilon/2+\delta)$ we define $\gamma_t=\alpha_{\frac{t-(\varepsilon/2-\delta)}{2\delta}}$ and $(\alpha_t)_{t\in[0,1]}$ is a Wasserstein constant speed geodesic from μ_k to μ. We also define a path $(\tilde{\eta_t}^{k,\varepsilon,\delta})_{t\in[0,\varepsilon]}$ in the exact same way, but changing μ_k in ν and μ in ν_k, that is a path from ν to ν_k.
We denote by $(\mu_t)_{t\in[0,1]}$ the entropic interpolation from μ to ν. Then for every $0<\varepsilon<1/2$, $k\in\mathbb{N}$ and $\delta\in(0,\varepsilon/2)$ we define a path $\left(\zeta_t^{k,\varepsilon,\delta}\right)_{t\in[0,1]}$ by

$$\zeta_t^{k,\varepsilon,\delta}=\begin{cases}\eta_t^{k,\varepsilon,\delta}, & t\in[0,\varepsilon),\\ \mu_{\frac{t-\varepsilon}{1-2\varepsilon}}, & t\in[\varepsilon,1-\varepsilon],\\ \tilde{\eta}^{k,\varepsilon,\delta}_{t-(1-\varepsilon)}, & t\in(1-\varepsilon,1].\end{cases}$$

This is an absolutely continuous path which connects μ_k to ν_k, hence, by the very definition of the cost $\mathcal{C}$ we have

$$\mathcal{C}(\mu_k,\nu_k)\leq\int_0^1\left\|\dot{\zeta}_t^{k,\varepsilon,\delta}\right\|^2_{L^2(\zeta_t^{k,\varepsilon,\delta})}+\mathcal{I}(\zeta_t^{k,\varepsilon,\delta})dt.$$

Due to the hypothesis (H1) we can apply the local logarithmic Sobolev inequalities stated in [BGL14, Theorem 5.5.2] and the ρ-convexity of the entropy (see [Vil09, Corollary 17.19]) to find

$$2\int_0^{\varepsilon/2-\delta}\mathcal{I}(P_t^*\mu_k)dt+2\int_0^{\varepsilon/2-\delta}\mathcal{I}(P_t^*\mu)dt\leq\frac{1-e^{-\rho\varepsilon}}{\rho}\left(\mathcal{I}(\mu)+\mathcal{I}(\mu_k)\right),$$

and

$$\int_{\varepsilon/2-\delta}^{\varepsilon/2+\delta}\mathcal{I}(P^*_{\varepsilon/2-\delta}\gamma_t)dt\leq\frac{4\delta\rho}{e^{\rho(\varepsilon-2\delta)}-1}\int_0^1\mathcal{F}(\alpha_t)dt\leq\frac{2\delta\rho}{e^{\rho(\varepsilon-2\delta)}-1}\left(\mathcal{F}(\mu)+\mathcal{F}(\mu_k)-\frac{\rho}{2}W_2^2(\mu,\mu_k)\right).$$

Here for $t\in(\varepsilon/2-\delta,\varepsilon/2+\delta)$ we denote $\mathbf{dt}P_{\varepsilon/2-\delta}\gamma_t=\dot{P}_{\varepsilon/2-\delta}\gamma_t$ the velocity field of the path $(P_{\varepsilon/2-\delta}\gamma_s)_{s\in(\varepsilon/2-\delta,\varepsilon/2+\delta)}$ at time t. We need to estimate

$$\int_{\varepsilon/2-\delta}^{\varepsilon/2+\delta}\left\|\mathbf{dt}P^*_{\varepsilon/2-\delta}\gamma_t\right\|_{L^2(P^*_{\varepsilon/2-\delta}\gamma_t)}dt.$$

Using [AGS08, Theorem 8.3.1], for every $t\in(\varepsilon/2-\delta,\varepsilon/2+\delta)$ we have

$$\left\|\mathbf{dt}P^*_{\varepsilon/2-\delta}\gamma_t\right\|_{L^2(P^*_{\varepsilon/2-\delta}(\gamma_t))}=\lim_{u\to t}\frac{W_2(P^*_{\varepsilon/2-\delta}\gamma_t,P^*_{\varepsilon/2-\delta}\gamma_u)}{|t-u|}.$$

Finally, using the $CD(\rho,\infty)$ contraction property [BGL14, Theorem 9.7.2] we obtain

$$\left\|\mathbf{dt}P^*_{\varepsilon/2-\delta}\gamma_t\right\|_{L^2(P^*_{\varepsilon/2-\delta}(\gamma_t))}\leq e^{-\rho(\varepsilon/2-\delta)}\frac{W_2(\mu_k,\mu)}{2\delta}.$$

We have shown

$$\int_0^\varepsilon \left\|\dot{\zeta}_t^{n,\varepsilon}\right\|^2_{L^2(\zeta_t^{n,\varepsilon})} + \mathcal{I}\left(\zeta_t^{n,\varepsilon}\right) dt \leq \frac{1-e^{-\rho\varepsilon}}{\rho}\left(\mathcal{I}(\mu)+\mathcal{I}(\mu_k)\right) + \frac{2\delta\rho}{e^{\rho(\varepsilon-2\delta)}-1}\left(\mathcal{F}(\mu)+\mathcal{F}(\mu_k)-\frac{\rho}{2}W_2^2(\mu,\mu_k)\right) + e^{-\rho(\varepsilon-2\delta)}\frac{W_2^2(\mu_k,\mu)}{4\delta^2}.$$

A similar estimate hold for the integral from $1-\varepsilon$ to 1 and we obtain

$$\mathcal{C}(\mu_k,\nu_k) \leq \frac{1-e^{-\rho\varepsilon}}{\rho}\left(\mathcal{I}(\mu)+\mathcal{I}(\mu_k)+\mathcal{I}(\nu)+\mathcal{I}(\nu_k)\right)+ \frac{2\delta\rho}{e^{\rho(\varepsilon-2\delta)}-1}\left(\mathcal{F}(\mu)+\mathcal{F}(\mu_k)+\mathcal{F}(\nu)+\mathcal{F}(\nu_k)-\frac{\rho}{2}\left(W_2^2(\mu,\mu_k)+W_2^2(\nu,\nu_k)\right)\right) + e^{-\rho(\varepsilon-2\delta)}\frac{W_2^2(\mu_k,\mu)+W_2^2(\nu_k,\nu)}{4\delta^2} + \int_0^1 \frac{1}{1-2\varepsilon}\left\|\dot{\mu}_t\right\|^2_{L^2(\mu_t)} + (1-2\varepsilon)\mathcal{I}(\mu_t)dt.$$

Finally, letting in this order k tend to ∞, δ tend to 0, and ε tend to 0 we obtain the desired inequality.
To obtain the lim inf inequality, we consider the same path but swapping the role of μ_k and μ (resp. ν_k and ν) and using the fact that $1-2\varepsilon < \frac{1}{1-2\varepsilon}$, we obtain for every $k \in \mathbb{N}$, $\varepsilon \in (0,1/2)$ and $\delta \in (0,\varepsilon)$

$$\mathcal{C}(\mu,\nu) \leq \frac{1-e^{-\rho\varepsilon}}{\rho}\left(\mathcal{I}(\mu)+\mathcal{I}(\mu_k)+\mathcal{I}(\nu)+\mathcal{I}(\nu_k)\right) + \frac{2\delta\rho}{e^{\rho(\varepsilon-2\delta)}-1}\left(\mathcal{F}(\mu)+\mathcal{F}(\mu_k)+\mathcal{F}(\nu)+\mathcal{F}(\nu_k)-\frac{\rho}{2}\left(W_2^2(\mu,\mu_k)+W_2^2(\nu,\nu_k)\right)\right) + e^{-\rho(\varepsilon-2\delta)}\frac{W_2^2(\mu_k,\mu)+W_2^2(\nu_k,\nu)}{4\delta^2} + \frac{1}{1-2\varepsilon}\mathcal{C}(\mu_k,\nu_k).$$

Letting k tends to ∞, δ tends to 0 and ε tend to zero we obtain

$$\mathcal{C}(\mu,\nu) \leq \varliminf_{k\to\infty} \mathcal{C}(\mu_k,\nu_k).$$

▷

4 Extension of some properties to the non compactly supported case

4.1 Benamou-Brenier-Schrödinger formula

As mentionned before, the Benamou-Brenier-Schrödinger formula has been obtained under various hypothesis. Here we show that the result hold true in the case where both measures are not compactly supported but assuming that they have finite fisher information, finite entropy and locally bounded densities, using continuity properties of the cost proved before and existing results. Recall that, in the existing litterature, this formula is proved assuming that the two measures have bounded supports and densities, see [GT20, Theorem 4.3].

Proposition 4.1 (Benamou-Brenier-Schrödinger formula) *Let $\mu, \nu \in \mathcal{P}_2^*(N)$ be two measures with locally bounded densities with respect to m such that $\mathcal{I}(\mu), \mathcal{I}(\nu) < \infty$. Furthermore, assume that there exists an entropic interpolation from μ to ν. Then*

$$\mathrm{Sch}(\mu, \nu) = \frac{\mathcal{C}(\mu, \nu)}{4} + \frac{\mathcal{F}(\mu) + \mathcal{F}(\nu)}{2}.$$

Notice that, the hypothesis of existence of entropic interpolations is not so restrictive. Indeed if $N = \mathbb{R}^n$, entropic interpolations always exists for measures in $\mathcal{P}_2^*(N)$, see [Lé14, Proposition 4.1].

Proof

$\triangleleft$ Let $x \in N$, for every $n \in \mathbb{N}$, we define

$$\mu_n = \alpha_n \mathbb{1}_{B(x,n)} \frac{d\mu}{dm} dm,$$

where α_n is a constant renormalization. Analogously we can define a sequence $(\nu_n)_{n\in\mathbb{N}}$ which converges to ν when $n \to \infty$. As μ_n and ν_n are compactly supported, we can apply the Benamou-Brenier-Schrödinger formula, namely

$$\mathrm{Sch}(\mu_n, \nu_n) = \frac{\mathcal{C}(\mu_n, \nu_n)}{4} + \frac{\mathcal{F}(\mu_n) + \mathcal{F}(\nu_n)}{2}. \tag{8}$$

It can be easily shown that $W_2(\mu_n, \mu) \underset{n\to\infty}{\to} 0$, $\mathcal{I}(\mu_n) \underset{n\to\infty}{\to} \mathcal{I}(\mu)$, and $\mathcal{F}(\mu_n) \underset{n\to\infty}{\to} \mathcal{F}(\mu)$ (resp. ν_n and ν). Hence by the theorem 3.1 the right-hand side of (8) converges toward $\frac{\mathcal{C}(\mu,\nu)}{4} + \frac{\mathcal{F}(\mu)+\mathcal{F}(\nu)}{2}$ when $n \to \infty$.

For the left hand-side, note that by the space restriction property of the Schrödinger cost [Tam17, Proposition 4.2.2], for every $n \in \mathbb{N}$ the optimal transport plan for the Schrödinger problem from μ_n to ν_n is given fro every probability set A of $N \times N$ by

$$\gamma_n(A) := \frac{\gamma(A \cap B(x,n)^2)}{\mu(B(x,n))\nu(B(x,n))},$$

where γ is the optimal transport plan for the Schrödinger problem from μ to ν. Hence $\mathrm{Sch}(\mu_n, \nu_n) = H(\gamma_n | R_{01}) \underset{n\to\infty}{\to} H(\gamma | R_{01}) = \mathrm{Sch}(\mu, \nu)$, and the result is proved. $\triangleright$

4.2 Longtime properties of the entropic cost

The entropic cost $\mathcal{C}(\mu, \nu)$ can be defined with more generality using a parameter $T > 0$. For $\mu, \nu \in \mathcal{P}_2(N)$ and $T > 0$ we define

$$C_T(\mu, \nu) := \inf \int_0^T \|\dot{\mu}_t\|_{L^2(\mu_t)} + \mathcal{I}(\mu_t) dt.$$

In [GGI20a, Theorem 3.6] and [Con19, Theorem 1.4], estimates are provided for high values of T, but only in the case where both measures are compactly supported and smooth. Using the Proposition 3.1 we are able to extend these estimates to the non-compactly supported and non-smooth case. The following lemma will be very useful, it is proved in [ICLL20, Lemma 3.1].

Lemma 4.2 (Approximation by compactly supported measures) *Let $\mu \in \mathcal{P}_2(N)$ be a probability measure such that $\mathcal{F}(\mu) < \infty$ and $\mathcal{I}(\mu) < \infty$. Then there exists a sequence $(\mu_k)_{k\in\mathbb{N}} \subset \mathcal{P}_2(N)$ such that*

(i) $\mathcal{F}(\mu_k) \underset{k\to\infty}{\to} \mathcal{F}(\mu)$, $\mathcal{I}(\mu_k) \underset{k\to\infty}{\to} \mathcal{I}(\mu)$ *and* $W_2(\mu_k, \mu) \underset{k\to\infty}{\to} 0$.

(ii) $\frac{d\mu_k}{dm} \in C_c^\infty(N)$ *for every* $k \in \mathbb{N}$.

Using this lemma and the Theorem 3.1 we can easily extend the estimates provided in [Con19, Theorem 1.4] and [GGI20a, Theorem 3.6]. Note that in [Con19] the author has already extended the estimate which holds under the $CD(\rho, \infty)$ curvature-dimension condition to the non-compact case, but we believe this is a pertinent example to illustrate the utility of Proposition 3.1. The validity of the $CD(0, n)$ estimate for non-compactly supported measures is a new result up to my knowledge.

Corollary 4.3 (Talagrand type inequality for the entropic cost) *Let $\mu, \nu \in \mathcal{P}_2(N)$ be two probability measures with finite entropy and Fisher information. Assume that there exists an entropic interpolation from μ to ν. Then if the $CD(\rho, \infty)$ curvature-dimension condition holds for some $\rho > 0$*

$$C_T(\mu, \nu) \leq 2 \inf_{t\in(0,T)} \left\{ \frac{1 + e^{-2\rho t}}{1 - e^{-2\rho t}} \mathcal{F}(\mu) + \frac{1 + e^{-2\rho(T-t)}}{1 - e^{-2\rho(T-t)}} \mathcal{F}(\nu) \right\}.$$

If the $CD(0, n)$ curvature-dimension condition holds for some $n > 0$ then

$$C_T(\mu, \nu) \leq C_1(\mu, \nu) + 2n \log(T).$$

These estimates are very useful, for instance they are fundamental to show the longtime convergence of entropic interpolations, see [GGI20a].

5 Derivability of the Schrödinger cost

In this section, we take $N = \mathbb{R}^n$ for some $n \in \mathbb{N}$ and $L = \Delta$ is the classical Laplacian operator. In this case the heat semigroup $(P_t)_{t\geqslant 0}$ is given by the following density

$$\forall x, y \in \mathbb{R}^n,\ t > 0,\ p_t(x, y) = \frac{1}{(4\pi t)^{n/2}} e^{-\frac{|x-y|^2}{4t}},$$

and the reversible measure m is the Lebesgue measure. Notice that in this case, the funtions A and B wich appears in hypothesis (i) to (iv) in Section 2.3 can be chosen as

$$\forall x \in \mathbb{R}^n,\ A(x) = B(x) := |x|^2.$$

Hence in this case

$$\mathcal{P}_2^*(\mathbb{R}^n) = \{\mu \in \mathcal{P}_2(\mathbb{R}^n) :\ -\infty < \mathcal{F}(\mu) < \infty\}.$$

A natural question is the following: given a probability measure ν can we find a formula for the derivative of the function $t \mapsto \mathcal{C}(\mu_t, \nu)$ where $(\mu_t)_{t\in[0,1]}$ is a smooth curve in $\mathcal{P}_2(N)$? From a formal point view, we can easily find an answer. Here we use the notation $\mathbf{dt}\mu_s^t$ (resp. $\mathbf{ds}\mu_s^t$) for the velocity of a given path $(\mu_s^t)_{(s,t)\in[0,1]\times[0,1]}$ wrt to t (resp. wrt s), to avoid confusion

between the two variables. For every $t \in [0,1]$, let $(\mu_s^t)_{s\in[0,1]}$ be the entropic interpolation from μ_t to ν, then

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\mathcal{C}(\mu_t,\nu) &= \frac{d}{dt}\int_0^1 \|\mathbf{ds}\mu_s^t\|^2_{L^2(\mu_s^t)} + \|\mathbf{grad}_{\mu_s^t}\mathcal{F}\|^2_{L^2(\mu_s^t)}ds \\ &= \int_0^1 \langle \mathbf{D_t ds}\mu_s^t, \mathbf{ds}\mu_s^t\rangle_{L^2(\mu_s^t)} + \mathbf{Hess}_{\mu_s^t}\mathcal{F}(\mathbf{ds}\mu_s^t, \mathbf{grad}_{\mu_s^t}\mathcal{F})ds \\ &= \int_0^1 \langle \mathbf{D_s dt}\mu_s^t, \mathbf{ds}\mu_s^t\rangle_{L^2(\mu_s^t)} + \mathbf{Hess}_{\mu_s^t}\mathcal{F}(\mathbf{ds}\mu_s^t, \mathbf{grad}_{\mu_s^t}\mathcal{F})ds.\end{aligned}$$

Here we have used [GLR20, Lemma 20] to invert the derivatives. Noticing that

$$\langle \mathbf{D_s dt}\mu_s^t, \mathbf{ds}\mu_s^t\rangle_{L^2(\mu_s^t)} = \frac{d}{ds}\langle \mathbf{dt}\mu_s^t, \mathbf{ds}\mu_s^t\rangle_{L^2(\mu_s^t)} - \langle \mathbf{dt}\mu_s^t, \mathbf{Dsds}\mu_s^t\rangle_{L^2(\mu_s^t)}$$

and using the Newton equation (7) we have

$$\frac{1}{2}\frac{d}{dt}\mathcal{C}(\mu_t,\nu) = \int_0^1 \frac{d}{ds}\langle \mathbf{dt}\mu_s^t, \mathbf{ds}\mu_s^t\rangle_{L^2(\mu_s^t)}ds = -\langle \dot{\mu}_t, \mathbf{ds}\mu_s^t|_{s=0}\rangle_{L^2(\mu_t)}. \tag{9}$$

Unfortunately we do not see how to turn this proof into a rigorous one.

From another point of view, we can try to derive the static formulation of the Schrödinger problem. Once again, we can easily guess a formula from a heuristic point of view. Indeed, let $(\mu_t)_{t\in[0,1]}$ be a smooth curve in $\mathcal{P}_2(N)$. For every $t \in [0,1]$ we denote by $\gamma_t = f^t \otimes g^t dR_{01}$ the optimal transport plan for the Schrödinger problem from μ_t to ν. Then

$$\begin{aligned}\frac{d}{dt}\mathrm{Sch}(\mu_t,\nu) &= \frac{d}{dt}H(\gamma_t|R_{01}) \\ &= \langle \dot{\gamma}_t, \nabla \log \gamma_t\rangle_{L^2(\gamma_t)}.\end{aligned}$$

Using the fact that γ_t is a transport plan from μ_t to ν it can be easily shown that $\langle \dot{\gamma}_t, \nabla \log \gamma_t\rangle_{\gamma_t} = \langle \dot{\mu}_t, \nabla \log f^t\rangle_{\mu_t}$. Hence we obtain

$$\frac{d}{dt}\mathrm{Sch}(\mu_t,\nu) = \langle \dot{\mu}_t, \nabla \log f^t\rangle_{L^2(\mu_t)}.$$

Note that this is equivalent to the equation (9) thanks to the Benamou-Brenier-Schrödinger formula. This proof is not rigorous because we don't have the regularity properties needed for γ_t. To prove our results, we follow the idea of Villani in [Vil09, Theorem 23.9] where he computes the derivative of the Wasserstein distance along curves. Before the statement of our main theorem a technical lemma is needed. This lemma is an easy corollary of the proof of [Tam17, Theorem 4.2.3].

Lemma 5.1 *Let $(\mu_k)_{k\in\mathbb{N}}, (\nu_k)_{k\in\mathbb{N}} \subset \mathcal{P}_2^*(\mathbb{R}^n)$ and $\mu,\nu \in \mathcal{P}_2^*(\mathbb{R}^n)$ such that μ_k converges toward μ with respect to the Wasserstein distance when $k \to \infty$ (resp. ν_k to ν). For every $k \in \mathbb{N}$, we denote by $\gamma_k = f^k \otimes g^k dR_{01}$ the optimal transport plan for the Schrödinger problem from μ_k to ν_k and $\gamma = f \otimes g dR_{01}$ the optimal transport plan for the Schrödinger problem from μ to ν. Assume that $(\frac{d\mu_k}{dm})_{k\in\mathbb{N}}$ and $(\frac{d\nu_k}{dm})_{k\in\mathbb{N}}$ are uniformly bounded in compact sets. Then for every compact set $K \subset N$, up to extraction $(f^k)_{k\in N}$ and $(g^k)_{k\in N}$ are uniformly bounded in $L^\infty(K,m)$. Furthermore*

$$f^k \underset{k\to\infty}{\overset{*}{\rightharpoonup}} f, \; g^k \underset{k\to\infty}{\overset{*}{\rightharpoonup}} g,$$

where the weak star convergence is understood in $L^\infty(K,m)$.

In addition to this lemma, the following fact is central in our proof. Given two probability measures p, r on $\mathbb{R}^n$ and a smooth enough function $\varphi : \mathbb{R}^n \to \mathbb{R}^n$, we have

$$\frac{d\varphi\#p}{dm} = \frac{\frac{dp}{dm}}{|\det J_\varphi|} \circ \varphi^{-1},$$

where $|\det J_\varphi|$ is the Jacobian determinant of φ. We often refer to this result as the Monge-Ampère equation or the Jacobian equation, see [Vil09, Theorem 11.1] or [AGS08, Lemma 5.5.3]. Using this equation, we obtain

$$H(\varphi\#p|r) = H(p|r) - \int \log|\det J_\varphi| dp + \int \left(\log \frac{dr}{dm} - \log \frac{dr}{dm} \circ \varphi\right) dp, \tag{10}$$

where $|\det J_\varphi|$ is the Jacobian determinant of φ. Given a curve $(\mu_t)_t \subset \mathcal{P}_2(N)$ and a measure $\nu \in \mathcal{P}_2(N)$, the idea of the following proof is to apply equation (10) with $r = R_{01}$, $p = \gamma_t$ is the optimal transport plan for the Schrödinger problem from μ_t to ν and $\varphi = T_{t\to s} \times \mathrm{Id}$ to bound from above $\mathrm{Sch}(\mu_s, \nu)$, and then let $s \to t$.

Theorem 5.2 (Derivative of the Schrödinger cost) *Let $(\nu_t)_{t\in(t_1,t_2)} \subset \mathcal{P}_2^*(\mathbb{R}^n)$ and $(\mu_t)_{t\in(t_1,t_2)} \subset \mathcal{P}_2^*(\mathbb{R}^n)$ be two absolutely continuous curves, for some $(t_1, t_2) \subset \mathbb{R}$. Furthermore assume that*

(i) For every $t \in (t_1, t_2)$ the measure μ_t has smooth bounded density against m.

(ii) There exists a constant $C > 0$ such that for every $t \in (t_1, t_2)$ and $x \in \mathbb{R}^n$ we have $|\dot\mu_t(x)| \le C(1 + |x|)$.

(iii) The sequence $\left(\frac{d\mu_t}{dm}\right)_{t\in(t_1,t_2)}$ and $\left(\frac{d\nu_t}{dm}\right)_{t\in(t_1,t_2)}$ are uniformly bounded in compact sets.

(iv) The functions $t \mapsto \mathcal{F}(\mu_t)$ and $t \mapsto \mathcal{F}(\nu_t)$ are derivable and $\frac{d}{dt}\mathcal{F}(\mu_t) = \langle \nabla \log \frac{d\mu_t}{dm}, \dot\mu_t\rangle_{L^2(\mu_t)}$ (resp. $\mathcal{F}(\nu_t)$).

(v) $\int_{t_1}^{t_2} \mathcal{L}(\dot\mu_t)dt < \infty$ and $\int_{t_1}^{t_2} \|\dot\mu_t\|_{L^2(\mu_t)}dt < \infty$, where $\mathcal{L}(\dot\mu_t)$ is defined at definition 2.2.

(vi) For every $t \in (t_1, t_2)$ the functions f^t and g^t are in $L^1(m)$, where (f^t, g^t) is the unique solution in $L^\infty(m) \times L^\infty(m)$ of the Schrödinger system

$$\begin{cases} \frac{d\mu_t}{dm} = f^t P_1 g^t, \\ \frac{d\nu_t}{dm} = g^t P_1 f^t. \end{cases}$$

Then the application $t \mapsto \mathrm{Sch}(\mu_t, \nu)$ is differentiable almost everywhere and we have for almost every $t \in (t_1, t_2)$

$$\frac{d}{dt}\mathrm{Sch}(\mu_t, \nu_t) = \langle \dot\mu_t, \nabla \log f^t\rangle_{L^2(\mu_t)} + \langle \dot\nu_t, \nabla \log g^t\rangle_{L^2(\nu_t)}.$$

Furthermore for almost every $t \in (t_1, t_2)$ this equality can be rewritten as

$$\frac{d}{dt}\mathcal{C}(\mu_t, \nu_t) = \langle \dot\nu_t, \mathbf{ds}\mu^t_{s|s=1}\rangle_{L^2(\nu_t)} - \langle \dot\mu_t, \mathbf{ds}\mu^t_{s|s=0}\rangle_{L^2(\mu_t)},$$

where $(\mu_s^t)_{s\in[0,1]}$ is the entropic interpolation from μ_t to ν_t.

Proof
◁ To begin we want to show

$$\frac{d}{dt}\mathrm{Sch}(\mu_t,\nu) = \langle \dot{\mu}_t, \nabla \log f^t\rangle_{L^2(\nu_t)},$$

for every $\nu \in \mathcal{P}_2^*(\mathbb{R}^n)$ such that $\frac{d\nu}{dm} \in L^\infty(m)$.

For every $t \in [0,1]$, γ_t denotes the optimal transport plan in the Schrödinger problem from μ_t to ν. Let $t \in [0,1]$ be fixed. Then for every s small enough by the very definition of the cost $\mathrm{Sch}(\mu_{t+s},\nu) \leq H((T_{t\to t+s}\times \mathrm{Id})\#\gamma_t|R_{01})$ where $(T_{t_1\to t_2})_{t_1,t_2\in[0,1]}$ are the flow maps associated to $(\mu_s)_{s\in[0,1]}$ defined in the subsection 2.4. Applying the equation (10) with $r = R_{01}$, $p = \gamma_t$ and $\varphi = T_{t\to t+s}\times \mathrm{Id}$ we obtain

$$H((T_{t\to t+s}\times \mathrm{Id})\#\gamma_t|R_{01}) = H(\gamma_t|R_{01}) + \int \log p_1 d\gamma_t - \int \log p_1 d(T_{t\to t+s}\times \mathrm{Id})\#\gamma_t \\ - \int \log|\det J_{T_{t\to t+s}}(x)| d\mu_t(x).$$

As noticed in [Vil09, Eq (23.11)], by the hypothesis (*ii*) there exists a constant C such that for every $y \in \mathbb{R}^n$ ans $s_1, s_2 \in [0,1]$

$$\begin{cases} |T_{s_1\to s_2}(x)| \leq C(1+|x|) \\ |x - T_{s_1\to s_2}(x)| \leq C|s_1 - s_2|(1+|x|). \end{cases} \tag{11}$$

For every $x, y \in \mathbb{R}^n$, we have $\log p_1(T_{t\to t+s}x, y) = -\frac{|T_{t\to t+s}x - y|^2}{4} - \frac{n}{2}\log(4\pi)$ and

$$\begin{aligned} \left|\frac{1}{2}\frac{d}{ds}|T_{t\to t+s}(x) - y|^2\right| &= |\langle \dot{\mu}_{t+s} \circ T_{t\to t+s}(x), T_{t\to t+s}(x) - y\rangle| \\ &\leq \frac{|\dot{\mu}_{t+s}\circ T_{t\to t+s}(x)|^2}{2} + \frac{|T_{t\to t+s}(x) - y|^2}{2} \\ &\leq C(1+|x|^2+|y|^2) \in L^1(\gamma_t), \end{aligned}$$

for some constant $C > 0$. Hence we can differentiate over the integral at time $s = 0$ to find

$$\int \log p_1 \; d(T_{t\to t+s}\times \mathrm{Id})\#\gamma_t = \int \log p_1 d\gamma_t + s\int \langle \dot{\mu}_t(x), \nabla_x \log p_1(x,y)\rangle d\gamma_t(x,y) + o(s). \tag{12}$$

Notice that thanks to the Monge-Ampère equation we have

$$\int \log|\det J_{T_{t\to t+s}}| d\mu_t = \mathcal{F}(\mu_t) - \mathcal{F}(\mu_{t+s}) = -s\langle \nabla \log \frac{d\mu_t}{dm}, \dot{\mu}_t\rangle_{L^2(\mu_t)} + o(s).$$

Combining this with the equation (12), we have

$$\mathrm{Sch}(\mu_{t+s},\nu) \leq \mathrm{Sch}(\mu_t,\nu) - s\left(\int \langle \nabla_x \log p_1(x,y), \dot{\mu}_t(x)\rangle d\gamma_t(x,y) - \langle \nabla \log \frac{d\mu_t}{dm}, \dot{\mu}_t\rangle_{L^2(\mu_t)}\right) + o(s).$$

Observe that using the hypothesis (*vi*) we have

$$\begin{aligned}\int \langle \nabla_x \log p_1(x,y), \dot{\mu}_t(x)\rangle d\gamma_t(x,y) &= \int\int \langle \nabla_x p_1(x,y), \dot{\mu}_t(x)\rangle f^t(x) g^t(y) dm(x) dm(y)\\ &= \int \langle \nabla_x \int p_1(x,y) g^t(y) dm(y), \dot{\mu}_t(x)\rangle f^t(x) dm(x)\\ &= \int \langle \nabla P_1 g^t(x), \dot{\mu}_t(x)\rangle f^t(x) dm(x)\\ &= \int \langle \nabla \log \frac{d\mu_t}{dm}(x), \dot{\mu}_t(x)\rangle d\mu_t(x) - \int \langle \nabla \log f^t(x), \dot{\mu}_t(x)\rangle d\mu_t(x).\end{aligned}$$

Hence we obtain

$$\overline{\lim_{s\to 0}} \frac{\mathrm{Sch}(\mu_{t+s},\nu) - \mathrm{Sch}(\mu_t,\nu)}{s} \leq \langle \dot{\mu}_t, \nabla \log f^t\rangle_{L^2(\mu_t)}.$$

For the reverse inequality we use the same kind of estimates. By definition we have $\mathrm{Sch}(\mu_t,\nu) \leq H((T_{t+s\to t}\times Id)\#\gamma_{t+s}|R_{01})$. Applying equation (10) we have

$$\begin{aligned}H((T_{t+s\to t}\times Id)\#\gamma_{t+s}|R_{01}) = H(\gamma_{t+s}|R_{01}) &- \int \log|\det J_{T_{t+s\to t}}| d\mu_{t+s}\\ &- \int \left(\frac{|T_{t+s\to t}x - y|^2 - |x-y|^2}{4}\right) d\gamma_{t+s}.\end{aligned}$$

As already noticed we have $\int \log|\det J_{T_{t+s\to t}}| d\mu_{t+s} = \mathcal{F}(\mu_{t+s}) - \mathcal{F}(\mu_t) = s\langle \dot{\mu}_t, \nabla \log \mu_t\rangle_{L^2(\mu_t)} + o(s)$. Now we have to deal with a more complicated term. We want to show that

$$\int \left(\frac{|T_{t+s\to t}x - y|^2 - |x-y|^2}{4}\right) d\gamma_{t+s}(x,y) = s\int \langle \nabla_x \log p_1(x,y), \dot{\mu}_t(x)\rangle d\gamma_t(x,y) + o(s).$$

Notice that using (11) we have for every $s>0$

$$\left||T_{t+s\to t}x - y|^2 - |x-y|^2\right| \leq Cs(1+|x|^2+|y|^2) \tag{13}$$

for some $C>0$.

For every $s\in\mathbb{R}$ small enough, we denote $v_s(x,y) = \frac{|T_{t+s\to t}x-y|^2-|x-y|^2}{s}$ and $v(x,y) = -2\langle x-y, \dot{\mu}_t(x)\rangle$. Of course for every $x,y\in N$, we have

$$v_s(x,y) \underset{s\to 0}{\to} v(x,y)$$

and by (13)

$$|v_s(x,y)| \leq P(x,y) := C(1+|x|^2+|y|^2).$$

Let χ_R be the product function $\chi_R = \mathbb{1}_{B(0,R)}\otimes \mathbb{1}_{B(0,R)}$. By the Lemma 5.1, for every $R>0$ there exists a sequence $(s_k^R)_{k\in\mathbb{N}}$ which tends to zero when k tend to ∞ such that the sequences $(f^{t+s_k^R})$, $(g^{t+s_k^R})$ are uniformly bounded in $L^\infty(B(0,R),m)$ and

$$f^{t+s_k^R} \underset{k\to\infty}{\overset{*}{\rightharpoonup}} f^t,\ g^{t+s_k^R} \underset{k\to\infty}{\overset{*}{\rightharpoonup}} g^t,$$

where the weak star convergence is understood in $L^\infty(K^R,m)$. Now for simplicity we denote $s_k^R = s_k$ and $K^R = B(0,R)$.

Note that

$$\int v_{s_k^R}(x,y)d\gamma_{t+s_k^R}(x,y) - \int v(x,y)d\gamma_t(x,y) = \int (1-\chi_R(x,y))v_{t+s_k^R}(x,y)d\gamma_{t+s_k^R}(x,y)$$
$$+ \int \chi_R(x,y)\left(v_{t+s_k^R}(x,y)f^{t+s_k^R}(x)g^{t+s_k^R}(y) - v(x,y)f^t(x)g^t(y)\right)dR_{01}(x,y)$$
$$+ \int (\chi_R(x,y)-1)v(x,y)d\gamma_t(x,y). \quad (14)$$

To obtain the desired estimate we are going to pass to the limsup in k, then let R tend to $+\infty$. The third term is independent of k and by the dominated convergence theorem it is immediate that it tend to 0 when $R \to \infty$. Things are trickier for the second term. Denote

$$\varphi_k(x) := \int v_{s_k^R}(x,y)g^{t+s_k^R}(y)p_1(x,y)dm(y)$$

and

$$\varphi(x) := \int v(x,y)g^t(y)p_1(x,y)dm(y).$$

Then

$$\left|\int \chi_R(x,y)v_{s_k^R}(x,y)d\gamma_{t+s_k^R}(x,y) - \int \chi_R(x,y)v(x,y)d\gamma(x,y)\right|$$
$$= \left|\int_K \varphi_k f^{t+s_k^R}dm - \int_{K^R}\varphi f^t dm\right|$$
$$\leq \left|\int_{K^R} f^{t+s_k^R}(\varphi^k - \varphi)dm\right| + \left|\int_{K^R}\left(f^{t+s_k^R} - f^t\right)\varphi dm\right|$$
$$\leq \sup_{n\in\mathbb{N}}\|f^{t+s_n}\|_{L^\infty(K^R,m)}\|\varphi_k - \varphi\|_{L^1(K^R,m)} + \left|\int_{K^R}\left(f^{t+s_k^R} - f^t\right)\varphi dm\right|.$$

The second term tends to zero thanks to the weak star convergence of $f^{t+s_k^R}$ toward f^t when $k \to \infty$. Furthermore the same kind of calculus gives for every $k \in \mathbb{N}$

$$|\varphi_k(x) - \varphi(x)| \leq \sup_{n\in\mathbb{N}}\|g^{t+s_n}\|_{L^\infty(K,m)}\|\left(v_k(x,\cdot) - v(x,\cdot)\right)p_1(x,\cdot)\|_{L^1(K,m)}$$
$$+ \int v(x,y)(g(x) - g^{t+s_k^R}(x))dR_{01}(x,y).$$

Again the second term tends to zero thanks to the weak star convergence of $g^{t+s_k^R}$. Using the upper bound

$$|\left(v_k(x,\cdot) - v(x,\cdot)\right)p_1(x,\cdot)| \leq 2P(x,\cdot)p_1(x,\cdot) \in L^1(K,m)$$

we have by the dominated convergence theorem

$$\|\left(v_k(x,\cdot) - v(x,\cdot)\right)p_1(x,\cdot)\|_{L^1(K,m)} \underset{k\to\infty}{\to} 0\ .$$

Hence $\varphi_k \underset{k\to\infty}{\to} \varphi$ pointwise. Noticing that for every $x \in K^R$, we have

$$|\varphi_k(x)| \leq \sup_{n\in\mathbb{N}}\|g^{t+s_k^R}\|_{L^\infty(K^R,m)}\int P(x,y)p_1(x,y)dm(y) \in L^1(K^R,m).$$

By the dominated convergence theorem,

$$\|\varphi_k - \varphi\|_{L^1(K^R,m)} \underset{k\to\infty}{\to} 0.$$

Thus the second term in (14) tends to zero when $k \to +\infty$. For the first term term, notice that for $R \geqslant 1$

$$\int (1-\chi_R) v_{t+s_k^R} d\gamma_{t+s_k^R} \leq 2C \int_{|x|\geqslant R} |x|^2 d\mu_{t+s_k^R}(x) + C \int_{|y|\geqslant R} |y|^2 d\nu(y),$$

thus it converges to zero, see [Vil09, Definition 6.8 and Theorem 6.9]. Hence for every $R > 0$ by letting k tends to $+\infty$ and R tends to $+\infty$ in

$$\frac{Sch(\mu_t,\nu) - Sch(\mu_{t+s_k^R},\nu)}{s_k^R} \leq -\frac{1}{s_k^R} \int \log|\det J_{T_{t+s_k^R\to t}}| d\mu_{t+s_k^R} + \int v_{s_k^R}(x,y) d\gamma_{t+s_k^R}(x,y)$$

we obtain

$$\overline{\lim_{s\to 0}} \frac{Sch(\mu_t,\nu) - Sch(\mu_{t+s},\nu)}{s} \leq \int \langle \nabla \log p_1(x,y), \dot\mu_t(x)\rangle d\gamma_t(x,y) + \langle \dot\mu_t, \nabla \log \frac{d\mu_t}{dm}\rangle_{\mu_t}$$

This is enough to conclude as in the previous case and obtain

$$\underline{\lim_{s\to 0}} \frac{\mathrm{Sch}(\mu_{t+s},\nu) - \mathrm{Sch}(\mu_t,\nu)}{s} \geqslant \langle \dot\mu_t, \nabla \log f^t\rangle_{\mu_t}.$$

This ends the case where $\nu_t = \nu$ is constant. Now we need to use a "doubling of variables "technique. Let $s, s', t \in [0,1]$ and $\gamma_{s,t}$ (resp. $\gamma_{s',t}$) be the optimal transport plan for the Schrödinger problem from μ_s (resp. μ'_s) to ν_t. Then, using the same tricks as before we have

$$H(\gamma_{s',t}|R_{01}) - H(\gamma_{s,t}|R_{01}) \leq \mathcal{F}(\mu_{s'}) - \mathcal{F}(\mu_s) + \frac{1}{4}\int \left(|x-y|^2 - |T_{s\to s'}x - y|^2\right) d\gamma_{s,t}(x,y).$$

Now using (13), the fact that $s \mapsto \mathcal{F}(\mu_s)$ is Lipschitz continuous and the fact that second order moment of of both curves are locally bounded, there exists a constant $C > 0$ such that

$$H(\gamma_{s',t}|R_{01}) - H(\gamma_{s,t}|R_{01}) \leq C|s - s'|.$$

By symmetry we can take absolute values in this inequality and it follows that the function $(s,t) \mapsto \mathrm{Sch}(\mu_s,\nu_t)$ is locally absolutely continuous in s uniformly in t (also absolutely continuous in t uniformly in s). Hence by [Vil09, Lemma 23.28] the desired result follow. ▷

Example 5.3 (Contraction property along Gaussian curves) *For $m, \sigma^2 > 0$ we denote by $\mathcal{N}(m,\sigma^2)$ the probability measure on $\mathbb{R}$ given by*

$$d\mathcal{N}(m,\sigma^2)(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

Considering the measures $\mu := \mathcal{N}(m_0,1)$ and $\nu := \mathcal{N}(m_1,1)$ it follows from [GGI20a, Sec A.2] that the curves $(P_t^*\mu)_{t\geqslant 0}$ and $(P_t^*\nu)_{t\geqslant 0}$ satisfies the hypothesis of Theorem 5.2. If we denote

$(\mu_s^t)_{s\in[0,1]}$ *the entropic interpolation from* $P_t^*\mu$ *to* $P_t^*\nu$ *applying Theorem 5.2 we obtain for almost every* $t > 0$

$$\begin{aligned}\frac{d}{dt}\mathcal{C}(P_t^*\mu, P_t^*\nu) &= -\langle \mathbf{grad}_{P_t^*\nu}\mathcal{E}nt, \mathbf{ds}\mu_s^t|_{s=1}\rangle_{L^2(\nu_t)} + \langle \mathbf{grad}_{P_t^*\mu}\mathcal{E}nt, \mathbf{ds}\mu_s^t|_{s=0}\rangle_{L^2(\mu_t)} \\ &= -\frac{d}{ds}\mathcal{E}nt(\mu_s^t)|_{s=1} + \frac{d}{ds}\mathcal{E}nt(\mu_s^t)|_{s=0} \\ &= -\int_0^1 \frac{d^2}{ds^2}\mathcal{E}nt(\mu_s^t)ds.\end{aligned} \tag{15}$$

Using the Newton equation (7) and the $CD(0,1)$ *curvature-dimension condition, for every* $t > 0$ *and* $s \in [0,1]$ *we have*

$$\begin{aligned}\frac{d^2}{ds^2}\mathcal{E}nt(\mu_s^t) &= \mathbf{Hess}_{\mu_s^t}\mathcal{E}nt(\mathbf{ds}\mu_s^t, \mathbf{d}s\mu_s^t) + \mathbf{Hess}_{\mu_s^t}\mathcal{E}nt(\mathbf{grad}_{\mu_s^t}\mathcal{E}nt, \mathbf{grad}_{\mu_s^t}\mathcal{E}nt) \\ &\geqslant \left(\left(\frac{d}{ds}\mathcal{E}nt(\mu_s^t)\right)^2 + |\mathbf{grad}_{\mu_s^t}\mathcal{E}nt(\mu_s^t)|^4_{L^2(\mu_s^t)}\right).\end{aligned}$$

Then by the Jensen inequality and neglecting the second term we obtain for every $t \geqslant 0$,

$$\int_0^1 \frac{d^2}{ds^2}\mathcal{E}nt(\mu_s^t)ds \geqslant \left(\mathcal{E}nt(P_t^*\mu) - \mathcal{E}nt(P_t^*\nu)\right)^2.$$

Using this and integrating the equality (15) we obtain for every $t \geqslant 0$,

$$\mathcal{C}(P_t^*\mu, P_t^*\nu) \leq \mathcal{C}(\mu,\nu) - \int_0^t \left(\mathcal{E}nt(P_u^*\mu) - \mathcal{E}nt(P_u^*\nu)\right)^2 du.$$

Hence we recover the $(0,1)$*-contraction property of the entropic cost proved in [GLR20, Theorem 37]. The result could, of course, be proven in* $\mathbf{R}^n$ *for* $n \geqslant 1$.

Acknowledgements This work was supported by the French ANR-17-CE40-0030 EFI project. I want to thanks warmly the anonymous referees for their work.

References

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures. 2nd ed.* Basel: Birkhäuser, 2nd ed. edition, 2008.

[BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[BE85] D. Bakry and M. Emery. Diffusions hypercontractives. *Sémin. de probabilités XIX, Univ. Strasbourg 1983/84, Proc., Lect. Notes Math. 1123, 177-206*, 1985.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators.* Cham: Springer, 2014.

[CGP16] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.*, 169(2):671–691, 2016.

[Con19] G. Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *Probab. Theory Relat. Fields*, 174(1-2):1–47, 2019.

[Fol88] H. Follmer. Random fields and diffusion processes. Calcul des probabilités, Éc. d'Été, Saint-Flour. 1985-87, Lect. Notes Math. 1362, 101-203, 1988.

[GGI20a] G.Clerc, G.Conforti, and I.Gentil. Long-time behaviour of entropic interpolations, 2020.

[GGI20b] G.Clerc, G.Conforti, and I.Gentil. On the variational interpretation of local logarithmic Sobolev inequalities. working paper or preprint, November 2020.

[Gig12] N. Gigli. Second order analysis on $(\mathcal{P}_2(M), W_2)$. *Mem. Am. Math. Soc.*, 1018:154, 2012.

[GLR17] I. Gentil, C. Léonard, and L. Ripani. About the analogy between optimal transport and minimal entropy. *Ann. Fac. Sci. Toulouse Math. (6)*, 26(3):569–601, 2017.

[GLR20] I. Gentil, C. Léonard, and L. Ripani. Dynamical aspects of the generalized schrödinger problem via otto calculus – a heuristic point of view. *Rev. Mat. Iberoam.*, 2020.

[Gri09] A. Grigor'yan. *Heat kernel and analysis on manifolds*, volume 47 of *AMS/IP Studies in Advanced Mathematics*. American Mathematical Society, Providence, RI; International Press, Boston, MA, 2009.

[GT20] N. Gigli and L. Tamanini. Benamou-Brenier and Kantorovich duality formulas for the entropic cost on $RCD^*(K, N)$ spaces . *Probab. Theory Relat. Fields*, 176(1-2):1–34, 2020.

[ICLL20] I.Gentil, C.Léonard, L.Ripani, and L.Tamanini. An entropic interpolation proof of the HWI inequality. *Stochastic Process. Appl.*, 130(2):907–923, 2020.

[JP19] J.Backhoff-Veraguas and P.Gudmund. Stability of martingale optimal transport and weak optimal transport. *arXiv preprint arXiv:1904.04171*, 2019.

[Lé14] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.

[Mik04] T. Mikami. Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probab. Theory Relat. Fields*, 129(2):245–260, 2004.

[Ott01] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Commun. Partial Differ. Equations*, 26(1-2):101–174, 2001.

[PC19] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[PME21] P.Ghosal, M.Nutz, and E.Bernton. Stability of entropic optimal transport and schrödinger bridges. *arXiv preprint arXiv:2106.03670*, 2021.

[Sch31] E. Schrödinger. Über die Umkehrung der Naturgesetze. *Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math.*, 144:144–153, 1931.

[Sch32] E. Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Ann. Inst. H. Poincaré*, 2:269–310, 1932.

[Tam17] L. Tamanini. Analysis and geometry of RCD spaces via the Schrödinger problem. *Phd Thesis*, 2017.

[Vil09] C. Villani. *Optimal transport. Old and new.* Berlin: Springer, 2009.