ANNALES
DE LA FACULTÉ
DES SCIENCES
DE
TOULOUSE
Mathématiques

Ivan Gentil, Christian Léonard and Luigia Ripani

About the analogy between optimal transport and minimal entropy

Tome XXVI, nº 3 (2017), p. 569-600.

<http://afst.cedram.org/item?id=AFST_2017_6_26_3_569_0>

Annales de la faculté des sciences de Toulouse Volume XXVI, n° 3, 2017
pp. 569-600

About the analogy between optimal transport and minimal entropy (*)

Ivan Gentil (1), Christian Léonard (2) and Luigia Ripani (3)

Abstract. — We describe some analogy between optimal transport and the Schrödinger problem where the transport cost is replaced by an entropic cost with a reference path measure. A dual Kantorovich type formulation and a Benamou–Brenier type representation formula of the entropic cost are derived, as well as contraction inequalities with respect to the entropic cost. This analogy is also illustrated with some numerical examples where the reference path measure is given by the Brownian motion or the Ornstein–Uhlenbeck process.

Our point of view is measure theoretical, rather than based on stochastic optimal control, and the relative entropy with respect to path measures plays a prominent role.

Résumé. — Nous décrivons des analogies entre le transport optimal et le problème de Schrödinger lorsque le coût du transport est remplacé par un coût entropique avec une mesure de référence sur les trajectoires. Une formule duale de Kantorovich, une formulation de type Benamou–Brenier du coût entropique sont démontrées, ainsi que des inégalités de contraction par rapport au coût entropique. Cette analogie est aussi illustrée par des exemples numériques où la mesure de référence sur les trajectoires est donnée par le mouvement Brownien ou bien le processus d'Ornstein–Uhlenbeck.

(*) Reçu le 4 novembre 2015, accepté le 13 mai 2016.

Keywords: Schrödinger problem, entropic interpolation, Wasserstein distance, Kantorovich duality.

(1) Institut Camille Jordan, UMR CNRS 5208. Université Claude Bernard. Lyon, France — gentil@math.univ-lyon1.fr

(2) MODALX, Université Paris Nanterre, UFR SEGMI. Nanterre, France — christian.leonard@u-paris10.fr

(3) Institut Camille Jordan, UMR CNRS 5208. Université Claude Bernard. Lyon, France — ripani@math.univ-lyon1.fr

Authors partially supported by the French ANR projects: GeMeCoD and ANR-12-BS01-0019 STAB. The third author is supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

Article proposé par Fabrice Baudoin.

Notre approche s'appuie sur la théorie de la mesure, plutôt que sur le contrôle optimal stochastique, et l'entropie relative joue un rôle fondamental.

1. Introduction

In this article, some analogy between optimal transport and the Schrödinger problem is investigated. A Kantorovich type dual equality, a Benamou–Brenier type representation of the entropic cost and contraction inequalities with respect to the entropic cost are derived when the transport cost is replaced by an entropic one. This analogy is also illustrated with some numerical examples.

Our point of view is measure theoretical rather than based on stochastic optimal control as is done in the recent literature; the relative entropy with respect to path measures plays a prominent role.

Before explaining the Schrödinger problem which is associated to an entropy minimization, we first introduce the Wasserstein quadratic transport cost W_2^2 and its main properties. For simplicity, our results are stated in $\mathbb{R}^n$ rather than in a general Polish space. Let us note that properties of the quadratic transport cost can be found in the monumental work by C. Villani [21, 22]. In particular its square root W_2 is a (pseudo-)distance on the space of probability measures which is called Wasserstein distance. It has been intensively studied and has many interesting applications. For instance it is an efficient tool for proving convergence to equilibrium of evolution equations, concentration inequalities for measures or stochastic processes and it allows to define curvature in metric measure spaces, see the textbook [22] for these applications and more.

The Wasserstein quadratic cost W_2^2 and the Monge–Kantorovich problem

Let $\mathcal{P}(\mathbb{R}^n)$ be the set of all probability measures on $\mathbb{R}^n$. We denote its subset of probability measures with a second moment by $\mathcal{P}_2(\mathbb{R}^n) = \{\mu \in \mathcal{P}(\mathbb{R}^n); \int |x|^2\,\mu(\mathrm{d}x) < \infty\}$. For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, the Wasserstein quadratic cost is

$$W_2^2(\mu_0, \mu_1) = \inf_\pi \int_{\mathbb{R}^n\times\mathbb{R}^n} |y-x|^2\,\pi(\mathrm{d}x\,\mathrm{d}y), \tag{1.1}$$

where the infimum is running over all the couplings π of μ_0 and μ_1, namely, all the probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ_0 and μ_1, that is for any bounded measurable functions φ and ψ on $\mathbb{R}^n$,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (\varphi(x) + \psi(y))\, \pi(\mathrm{d}x\, \mathrm{d}y) = \int_{\mathbb{R}^n} \varphi\, \mathrm{d}\mu_0 + \int_{\mathbb{R}^n} \psi\, \mathrm{d}\mu_1. \tag{1.2}$$

In restriction to $\mathcal{P}_2(\mathbb{R}^n)$, the pseudo-distance W_2 becomes a genuine distance. The *Monge–Kantorovich problem* with a quadratic cost function, consists in finding the optimal couplings π that minimize (1.1).

The entropic cost $\mathcal{A}^R$ and the Schrödinger problem

Let us fix some reference nonnegative measure R on the path space $\Omega = C([0,1], \mathbb{R}^n)$ and denote R_{01} the measure on $\mathbb{R}^n \times \mathbb{R}^n$. It describes the joint law of the initial position X_0 and the final position X_1 of a random process on $\mathbb{R}^n$ whose law is R. This means that

$$R_{01} = (X_0, X_1)_\# R$$

is the push-forward of R by the mapping (X_0, X_1). Recall that the push-forward of a measure α on the space A by the measurable mapping $f : \mathsf{A} \to \mathsf{B}$ is defined by

$$f_\# \alpha(\mathrm{d}b) = \alpha(f^{-1}(\mathrm{d}b)), \qquad \mathrm{d}b \subset \mathsf{B},$$

in other words, for any positive function H,

$$\int H d(f_\# \alpha) = \int H(f) \mathrm{d}\alpha.$$

For any probability measures μ_0 and μ_1 on $\mathbb{R}^n$, the entropic cost $\mathcal{A}^R(\mu_0, \mu_1)$ of (μ_0, μ_1) is defined by

$$\mathcal{A}^R(\mu_0, \mu_1) = \inf_\pi H(\pi \,|\, R_{01})$$

where $H(\pi | R_{01}) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \log(\mathrm{d}\pi / \mathrm{d}R_{01})\, d\pi$ is the relative entropy of π with respect to R_{01} and π runs through all the couplings of μ_0 and μ_1. The *Schrödinger problem* consists in finding the unique optimal entropic plan π that minimizes the above infimum.

In this article, we choose R as the reversible Kolmogorov continuous Markov process specified by the generator $\frac{1}{2}(\Delta - \nabla V \cdot \nabla)$ and the initial reversing measure $e^{-V(x)}\, \mathrm{d}x$.

Aim of the paper

Below in this introductory section, we are going to focus onto four main features of the quadratic transport cost W_2^2. Namely:

- the Kantorovich dual formulation of W_2^2;
- the Benamou–Brenier dynamical formulation of W_2^2;
- the displacement interpolations, that is the W_2-geodesics in $\mathcal{P}_2(\mathbb{R}^n)$;
- the contraction of the heat equation with respect to W_2^2.

The goal of this article is to recover analogous properties for $\mathcal{A}^R$ instead of W_2^2, by replacing the Monge–Kantorovich problem with the Schrödinger problem.

Several aspects of the quadratic Wasserstein cost

Let us provide some detail about these four properties.

Kantorovich dual formulation of W_2^2

The following duality result was proved by Kantorovich in [8]. For any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$W_2^2(\mu_0,\mu_1) = \sup_{\psi} \left\{ \int_{\mathbb{R}^n} \psi \, \mathrm{d}\mu_1 - \int_{\mathbb{R}^n} Q\psi \, \mathrm{d}\mu_0 \right\}, \tag{1.3}$$

where the supremum runs over all bounded continuous function ψ and

$$Q\psi(x) = \sup_{y\in\mathbb{R}^n} \left\{ \psi(y) - |x-y|^2 \right\}, \quad x \in \mathbb{R}^n.$$

It is often expressed in the equivalent form,

$$W_2^2(\mu_0,\mu_1) = \sup_{\varphi} \left\{ \int_{\mathbb{R}^n} \widetilde{Q}\varphi \, \mathrm{d}\mu_1 - \int_{\mathbb{R}^n} \varphi \, \mathrm{d}\mu_0 \right\}, \tag{1.4}$$

where the supremum runs over all bounded continuous function φ and

$$\widetilde{Q}\varphi(y) = \inf_{x\in\mathbb{R}^n} \left\{ \varphi(x) + |y-x|^2 \right\}, \quad y \in \mathbb{R}^n.$$

The map $Q\psi$ is called the sup-convolution of ψ and its defining identity is sometimes referred to as the Hopf–Lax formula.

Benamou–Brenier formulation of W_2^2

The Wasserstein cost admits a dynamical formulation: the so-called Benamou–Brenier formulation which was proposed in [5]. It states that for any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$W_2^2(\mu_0, \mu_1) = \inf_{(\nu,v)} \int_{\mathbb{R}^n \times [0,1]} |v_t|^2 \, \mathrm{d}\nu_t \, \mathrm{d}t, \tag{1.5}$$

where the infimum runs over all paths $(\nu_t, v_t)_{t \in [0,1]}$ where $\nu_t \in \mathcal{P}(\mathbb{R}^n)$ and $v_t(x) \in \mathbb{R}^n$ are such that ν_t is absolutely continuous with respect to time in the sense of [1, Ch. 1] for all $0 \leqslant t \leqslant 1$, $\nu_0 = \mu_0$, $\nu_1 = \mu_1$ and

$$\partial_t \nu_t + \nabla \cdot (\nu_t v_t) = 0, \quad 0 \leqslant t \leqslant 1.$$

In this equation which is understood in a weak sense, $\nabla \cdot$ stands for the standard divergence of a vector field in $\mathbb{R}^n$ and ν_t is identified with its density with respect to Lebesgue measure. This general result is proved in [1, Ch. 8]. A proof under the additional requirement that μ_0, μ_1 have compact supports is available in [21, Thm. 8.1].

Displacement interpolations

The metric space $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ is geodesic. This means that for any probability measure $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, there exists a path $(\mu_t)_{t \in [0,1]}$ in $\mathcal{P}_2(\mathbb{R}^n)$ such that for any $s, t \in [0,1]$,

$$W_2(\mu_s, \mu_t) = |t - s| W_2(\mu_0, \mu_1).$$

Such a path is a constant speed geodesic in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, see [1, Ch. 7]. Moreover when ν is absolutely continuous with respect to the Lebesgue measure, there exists a convex function ψ on $\mathbb{R}^n$ such that for any $t \in [0,1]$, the geodesic is given by

$$\mu_t = ((1-t) \operatorname{Id} + t \nabla \psi)_{\#} \mu_0. \tag{1.6}$$

This interpolation is called the McCann displacement interpolation in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$, see [21, Ch. 5].

Contractions in Wasserstein distance

Contraction in Wasserstein distance is a way to define the curvature of the underlying space or of the reference Markov operator. In its general formulation, the von Renesse–Sturm theorem tells that the heat equation in a smooth, complete and connected Riemannian manifold satisfies a contraction property with respect to the Wasserstein distance if and only if the Ricci curvature is bounded from below, see [18]. In the context of the present article where Kolmogorov semigroups on $\mathbb{R}^n$ are considered, two main contraction results will be explained with more details in Section 6.

Organization of the paper

The setting of the present work and notation are introduced in Section 2. The entropic cost $\mathcal{A}^R$ is defined with more detail in Section 3 together with the related notion of entropic interpolation, an analogue of the displacement interpolation. A dual Kantorovich type formulation and a Benamou–Brenier type formulation of the entropic cost are derived respectively at Sections 4 and 5. Section 6 is dedicated to the contraction properties of the heat flow with respect to the entropic cost. In the last Section 7, we give some examples of entropic interpolations between Gaussian distributions when the reference path measure is given by the Brownian motion or the Ornstein–Uhlenbeck process.

Literature

The Benamou–Brenier formulation of the entropic cost which is stated at Corollary 5.3 was proved recently by Chen, Georgiou and Pavon in [7] (in a slightly less general setting) without any mention to optimal transport (in this respect Corollary 5.6 relating the entropic and Wasserstein costs is new). Although our proof of Corollary 5.3 is close to their proof, we think it is worth including it in the present article to emphasize the analogies between displacement and entropic interpolations. In addition, we also provide a time asymmetric version of this formulation in Theorem 5.1.

Both [16] and [7] share the same stochastic optimal control viewpoint. This differs from the entropic approach of the present paper.

Let us notice that Theorem 6.1 is a new result: it provides contraction inequalities with respect to the entropic cost. Moreover, examples and comparison proposed at the end of the paper, with respect two different kernels (Gaussian and Ornstein–Uhlenbeck) are new.

2. The reference path measure

We make precise the reference path measure R to which the entropic cost $\mathcal{A}^R$ is associated. Although more general reversible path measures R would be all right to define a well-suited entropic cost, we prefer to consider the specific class of Kolmogorov Markov measures. This choice is motivated by the fact that, as presented in [12], the Monge Kantorovich problem is the limit of a sequence of entropy minimization problems, when a proper fluctuation parameter tends to zero. The Kolmogorov Markov measures, as reference measures in the Schrödinger problem, admit as a limit case the Monge Kantorovich problem with quadratic cost function, namely the Wasserstein distance.

Notation

For any measurable set Y, we denote respectively by $\mathcal{P}(Y)$ and $\mathcal{M}(Y)$ the set of all the probability measures and all positive σ-finite measures on Y. The *relative entropy* of a probability measure $p \in \mathcal{P}(Y)$ with respect to a positive measure $r \in \mathcal{M}(Y)$ is loosely defined by

$$H(p|r) := \begin{cases} \int_Y \log(\mathrm{d}p/\mathrm{d}r)\mathrm{d}p \in (-\infty, \infty], & \text{if } p \ll r, \\ \infty, & \text{otherwise.} \end{cases}$$

For some assumptions on the reference measure r that guarantee the above integral to be meaningful and bounded from below, see after the regularity hypothesis (Reg2) at page 579. For a rigorous definition and some properties of the relative entropy with respect to an unbounded measure see [14]. The state space $\mathbb{R}^n$ is equipped with its Borel σ-field and the path space Ω with the canonical σ-field $\sigma(X_t; 0 \leqslant t \leqslant 1)$ generated by the canonical process

$$X_t(\omega) := \omega_t \in \mathbb{R}^n, \quad \omega = (\omega_s)_{0 \leqslant s \leqslant 1} \in \Omega,\ 0 \leqslant t \leqslant 1.$$

For any path measure $Q \in \mathcal{M}(\Omega)$ and any $0 \leqslant t \leqslant 1$, we denote

$$Q_t(\,\cdot\,) := Q(X_t \in \cdot\,) = (X_t)_\# Q \in \mathcal{M}(\mathbb{R}^n),$$

the push-forward of Q by X_t. When Q is a probability measure, Q_t is the law of the random location X_t at time t when the law of the whole trajectory is Q.

The Kolmogorov Markov measure R and its semigroup

Most of our results can be stated in the general setting of a Polish state space. For the sake of simplicity, the setting of the present paper is particularized. The state space is $\mathbb{R}^n$ and the reference path measure R is the Markov path measure associated with the generator

$$\frac{1}{2}(\Delta - \nabla V \cdot \nabla) \tag{2.1}$$

and the corresponding reversible measure

$$m = e^{-V}\,\mathrm{Leb}$$

as its initial measure, where Leb is the Lebesgue measure. It is assumed that the potential V is a $\mathcal{C}^2$ function on $\mathbb{R}^n$ such that the martingale problem associated with the generator (2.1) on the domain $\mathcal{C}^2$ and the initial measure m admits a unique solution $R \in \mathcal{M}(\Omega)$. This is the case for instance when the following hypothesis are satisfied.

Existence hypothesis (Exi)

There exists some constant $c > 0$ such that one of the following assumptions holds true:

(i) $\lim_{|x|\to\infty} V(x) = +\infty$ and $\inf\{|\nabla V|^2 - \Delta V/2\} > -\infty$, or
(ii) $-x \cdot \nabla V(x) \leqslant c(1 + |x|^2)$, for all $x \in \mathbb{R}^n$.

See [19, Thm. 2.2.19] for the existence result under the assumptions (i) or (ii). For any initial condition $X_0 = x \in \mathbb{R}^n$, the path measure $R_x := R(\,\cdot \mid X_0 = x) \in \mathcal{P}(\Omega)$ is the law of the weak solution of the stochastic differential equation

$$\mathrm{d}X_t = -\nabla V(X_t)/2\ \mathrm{d}t + \mathrm{d}W_x(t), \quad 0 \leqslant t \leqslant 1, \tag{2.2}$$

where W_x is an R_x-Brownian motion. The Kolmogorov Markov measure is

$$R(\,\cdot\,) = \int_{\mathbb{R}^n} R_x(\,\cdot\,)\, m(\mathrm{d}x) \in \mathcal{M}(\Omega).$$

Recall that $m = e^{-V}\,\mathrm{Leb}$ is not necessarily a probability measure. The Markov semigroup associated to R is defined for any bounded measurable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ and any $t \geqslant 0$, by

$$T_t f(x) = E_{R_x} f(X_t), \quad x \in \mathbb{R}^n.$$

It is reversible with reversing measure m as defined in [3].

Regularity hypothesis (Reg1)

We also assume for simplicity that $(T_t)_{t\geqslant 0}$ admits for any $t>0$, a *density kernel* with respect to m, a probability density $p_t(x,y)$ such that

$$T_t f(x) = \int_{\mathbb{R}^n} f(y) p_t(x,y)\, m(\mathrm{d}y). \tag{2.3}$$

For instance, when $V(x)=|x|^2/2$, then R is the path measure associated to the Ornstein–Uhlenbeck process with the Gaussian measure as its reversing measure. When $V=0$, we recover the Brownian motion with Lebesgue measure as its reversing measure. Examples of Kolmogorov semigroups admitting a density kernel can be found for instance in [20, Ch. 3] or [2, Cor. 4.2]. This semigroup is fixed once for all.

Properties of the path measure R

The measure R is our *reference* path measure and it satisfies the following properties.

(a) It is Markov, that is for any $t\in[0,1]$, $R(X_{[t,1]}\in\cdot\mid X_{[0,t]}) = R(X_{[t,1]}\in\cdot\mid X_t)$. See [14] for the definition of the conditional law for unbounded measures since R is not necessarily a probability measure.

(b) It is reversible. This means that for all $0\leqslant T\leqslant 1$, the restriction $R_{[0,T]}$ of R to the sigma-field $\sigma(X_{[0,T]})$ generated by $X_{[0,T]}=(X_t)_{0\leqslant t\leqslant T}$, is invariant with respect to time-reversal, that is $[(X_{T-t})_{0\leqslant t\leqslant T}]_\# R_{[0,T]} = R_{[0,T]}$.

Any reversible measure R is stationary, i.e. $R_t=m$, for all $0\leqslant t\leqslant 1$ for some $m\in\mathcal{M}(\mathbb{R}^n)$. This measure m is called the reversing measure of R and is often interpreted as an equilibrium of the dynamics specified by the kernel $(R_x; x\in\mathbb{R}^n)$. One says for short that R is m-reversible.

3. Entropic cost and entropic interpolations

We define the Schrödinger problem, the entropic cost and the entropic interpolation which are respectively the analogues of the Monge–Kantorovich problem, the Wasserstein cost and the displacement interpolation that were briefly described in the introduction.

Let us state the definition of the entropic cost associated with R.

Definition 3.1 (Entropic cost). — *Consider the projection*

$$R_{01} := (X_0, X_1)_{\#} R \in \mathcal{M}(\mathbb{R}^n \times \mathbb{R}^n)$$

of R onto the endpoint space $\mathbb{R}^n \times \mathbb{R}^n$. For any $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$,

$$\mathcal{A}^R(\mu_0, \mu_1) = \inf\{H(\pi \,|\, R_{01});\ \pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n) : \pi_0 = \mu_0, \pi_1 = \mu_1\} \in (-\infty, \infty]$$

is the R-entropic cost of (μ_0, μ_1).

This definition is related to a static Schrödinger problem. It also admits a dynamical formulation.

Definition 3.2 (Dynamical formulation of the Schrödinger problem). *The Schrödinger problem associated to R, μ_0 and μ_1 consists in finding the minimizer $\hat{P}$ of the relative entropy $H(\,\cdot\,|\,R)$ among all the probability path measures $P \in \mathcal{P}(\Omega)$ with prescribed initial and final marginals $P_0 = \mu_0$ and $P_1 = \mu_1$,*

$$H(\hat{P}|R) = \min\{H(P|R),\ P \in \mathcal{P}(\Omega), P_0 = \mu_0, P_1 = \mu_1\}. \tag{3.1}$$

It is easily seen that its minimal value is the entropic cost,

$$\mathcal{A}^R(\mu_0, \mu_1) = \inf\{H(P \,|\, R);\ P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} \in (-\infty, \infty], \tag{3.2}$$

see for instance [15, Lem. 2.4].

In the rest of this work, the entropic cost will always be associated with the fixed reference measure $R \in \mathcal{M}(\Omega)$, therefore without ambiguity we drop the index and denote $\mathcal{A}^R = \mathcal{A}$.

Remarks 3.3. —

(1) First of all, when R is not a probability measure, the relative entropy might take some negative values and even the value $-\infty$. However, because of the decrease of information by push-forward mappings, we have

$$H(P|R) \geqslant \max(H(\mu_0 \,|\, m), H(\mu_1 \,|\, m)),$$

see [14, Thm. 2.4] for instance. Hence $H(P \,|\, R)$ is well defined in $(-\infty, \infty]$ whenever $H(\mu_0 \,|\, m) > -\infty$ or $H(\mu_1 \,|\, m) > -\infty$. This will always be assumed.

(2) Even the nonnegative quantity $\mathcal{A}(\mu_0, \mu_1) - \max(H(\mu_0 \,|\, m), H(\mu_1 \,|\, m)) \geqslant 0$ cannot be the square of a distance such as the Wasserstein cost W_2^2. As a matter of fact, considering the special situation where $\mu_0 = \mu_1 = \mu$, we have $\mathcal{A}(\mu, \mu) \geqslant H(\mu \,|\, m) > 0$ as soon as μ differs from m. This is a consequence of Theorem 5.1 below.

(3) A good news about $\mathcal{A}$ is that since R is reversible, it is symmetric: $\mathcal{A}(\mu,\nu)=\mathcal{A}(\nu,\mu)$. To see this, let us denote $X_t^*=X_{1-t}, 0\leqslant t\leqslant 1$, and $Q^*:=(X^*)_\# Q$ the time reversal of any $Q\in\mathcal{M}(\Omega)$. As X^* is one-one, we have $H(P\,|\,R)=H(P^*\,|\,R^*)$ and since we assume that $R^*=R$, we see that

$$H(P\,|\,R)=H(P^*\,|\,R),\quad \forall P\in\mathcal{P}(\Omega). \tag{3.3}$$

Hence, if P solves (3.1) with $(\mu_0,\mu_1)=(\mu,\nu)$, then $X^*{}_\# P$ solves (3.1) with $(\mu_0,\mu_1)=(\nu,\mu)$ and these Schrödinger problems share the same value.

Existence of a minimizer. Entropic interpolation

We recall some general results from [15, Thm. 2.12] about the solution of the dynamical Schrödinger problem (3.1). Let us denote by $p(x,y)$ the probability density introduced in (2.3), at time $t=1$, so that

$$R_{01}(\mathrm{d}x\,\mathrm{d}y)=m(\mathrm{d}x)p(x,y)m(\mathrm{d}y).$$

In order for (3.1) to admit a unique solution, it is enough that it satisfies the following hypothesis:

Regularity hypothesis (Reg2)

(i) $p(x,y)\geqslant e^{-A(x)-A(y)}$ for some nonnegative measurable function A on $\mathbb{R}^n$;
(ii) $\int_{\mathbb{R}^n\times\mathbb{R}^n} e^{-B(x)-B(y)}p(x,y)\,m(\mathrm{d}x)m(\mathrm{d}y)<\infty$ for some nonnegative measurable function B on $\mathbb{R}^n$;
(iii) $\int_{\mathbb{R}^n}(A+B)\,\mathrm{d}\mu_0, \int_{\mathbb{R}^n}(A+B)\,\mathrm{d}\mu_1<\infty$ where A appears at (i) and B appears at (ii);
(iv) $-\infty<H(\mu_0\,|\,m), H(\mu_1\,|\,m)<\infty$.

Assumptions (ii)–(iii) are useful to define rigorously $H(\mu_0\,|\,m)$ and $H(\mu_1\,|\,m)$. Under these assumptions the entropic cost $\mathcal{A}(\mu_0,\mu_1)$ is finite and the minimizer P of the Schrödinger problem (3.1) is characterized, in addition to the marginal constraints $P_0=\mu_0, P_1=\mu_1$, by the product formula

$$P=f_0(X_0)g_1(X_1)\,R \tag{3.4}$$

for some measurable functions f_0 and g_1 on $\mathbb{R}^n$. The uniqueness of the solution is a direct consequence of the fact that (3.1) is a *strictly* convex minimization problem.

Definition 3.4 (Entropic interpolation). — *The* R-entropic interpolation *between* μ_0 *and* μ_1 *is defined as the marginal flow of the minimizer* P *of* (3.1)*, that is* $\mu_t := P_t \in \mathcal{P}(\mathbb{R}^n), 0 \leqslant t \leqslant 1$.

Proposition 3.5. — *Under the hypotheses (Exi), (Reg1) and (Reg2), the* R*-entropic interpolation between* μ_0 *and* μ_1 *is characterized by*

$$\mu_t = e^{\varphi_t + \psi_t}\, m, \quad 0 \leqslant t \leqslant 1, \tag{3.5}$$

where

$$\varphi_t = \log T_t f_0, \quad \psi_t = \log T_{1-t} g_1, \quad 0 \leqslant t \leqslant 1, \tag{3.6}$$

and the measurable functions f_0, g_1 *solve the following system*

$$\frac{\mathrm{d}\mu_0}{\mathrm{d}m} = f_0 T_1 g_1, \quad \frac{\mathrm{d}\mu_1}{\mathrm{d}m} = g_1 T_1 f_0. \tag{3.7}$$

The system (3.7) is often called the Schrödinger system. It simply expresses the marginal constraints. Its solutions (f_0, g_1) are precisely the functions that appear in the identity (3.4). Actually it is difficult or impossible to solve explicitly the system (3.7). However, in Section 7, we will see some particular examples in the Gaussian setting where the system admits an explicit solution. Some numerical algorithms have been proposed recently in [6].

In our setting where R is the Kolmogorov path measure defined in (2.1), the entropic interpolation μ_t admits a density $\mu_t(z) := \mathrm{d}\mu_t/\mathrm{d}z$ with respect to the Lebesgue measure. It is important to notice that, contrary to the McCann interpolation, the function $(t,x) \mapsto \mu_t(x)$ is smooth on $(t,x) \in]0,1[\times\mathbb{R}$ and solves the transport equation

$$\partial_t \mu_t + \nabla \cdot (\mu_t\, v^{cu}(t, \mu_t)) = 0 \tag{3.8}$$

with the initial condition μ_0 and where $v^{cu}(t, \mu_t, z) = \nabla\psi_t(z) - \nabla V(z)/2 + \nabla \log \mu_t(z)/2$ refers to the *current velocity* introduced by Nelson in [17, Ch. 11] (this will be recalled at Section 5). The current velocity is a smooth function and the ordinary differential equation

$$\dot{x}_t(x) = v^{cu}(t, x_t(x)), \quad x_0 = x$$

admits a unique solution for any initial position x, the solution of the continuity equation (3.8) admits the following push-forward expression:

$$\mu_t = (x_t)_\# \mu_0, \quad 0 \leqslant t \leqslant 1, \tag{3.9}$$

in analogy with the displacement interpolation given in (1.6).

Remark 3.6 (From the entropic cost to the Wasserstein cost). — The Wasserstein distance is a limit case of the entropic cost. We shall use this result to compare contraction properties in Section 6 and also to illustrate the examples in Section 7.

Let us consider the following dilatation in time with ratio $\varepsilon > 0$ of the reference path measure R: $R^\varepsilon := (X^\varepsilon)_\# R$ where $X^\varepsilon(t) := X_{\varepsilon t}$, $0 \leqslant t \leqslant 1$. It is shown in [12] that some renormalization of the entropic cost $\mathcal{A}^{R^\varepsilon}$ converges to the Wasserstein distance when ε goes to 0. Namely,

$$\lim_{\varepsilon \to 0} \varepsilon \mathcal{A}^{R^\varepsilon}(\mu_0, \mu_1) = W_2^2(\mu_0, \mu_1)/2. \tag{3.10}$$

Even better, when μ_0 and μ_1 are absolutely continuous, the entropic interpolation $(\mu_t^{R^\varepsilon})_{0 \leqslant t \leqslant 1}$ between μ_0 and μ_1 converges as ε tends to zero towards the McCann displacement interpolation $(\mu_t)_{0 \leqslant t \leqslant 1}$, see (1.6).

4. Kantorovich dual equality for the entropic cost

We derive the analogue of the Kantorovich dual equality (1.3) when the Wasserstein cost is replaced by the entropic cost.

Theorem 4.1 (Kantorovich dual equality for the entropic cost). — *Let V, μ_0 and μ_1 be such that the hypothesis (Exi), (Reg1) and (Reg2) stated in Section 3 are satisfied. We have*

$$\mathcal{A}(\mu_0, \mu_1) = H(\mu_0 \,|\, m) + \sup \left\{ \int_{\mathbb{R}^n} \psi \, \mathrm{d}\mu_1 - \int_{\mathbb{R}^n} \mathcal{Q}^R \psi \, \mathrm{d}\mu_0;\ \psi \in C_b(\mathbb{R}^n) \right\}$$

where for every $\psi \in C_b(\mathbb{R}^n)$,

$$\mathcal{Q}^R \psi(x) := \log E_{R^x} e^{\psi(X_1)} = \log T_1(e^\psi)(x),\ x \in \mathbb{R}^n.$$

This result was obtained by Mikami and Thieullen in [16] with an alternate statement and a different proof. The present proof is based on an abstract dual equality which is stated below at Lemma 4.2. Let us first describe the setting of this lemma.

Let $\mathbb{U}$ be a vector space and $\Phi : \mathbb{U} \to (-\infty, \infty]$ be an extended real valued function on $\mathbb{U}$. Its convex conjugate Φ^* on the algebraic dual space $\mathbb{U}^*$ of $\mathbb{U}$ is defined by

$$\Phi^*(\ell) := \sup_{u \in \mathbb{U}} \left\{ \langle \ell, u \rangle_{\mathbb{U}^*, \mathbb{U}} - \Phi(u) \right\} \in [-\infty, \infty], \qquad \ell \in \mathbb{U}^*.$$

We consider a linear map $A : \mathbb{U}^* \to \mathbb{V}^*$ defined on $\mathbb{U}^*$ with values in the algebraic dual space $\mathbb{V}^*$ of some vector space $\mathbb{V}$.

Lemma 4.2 (Abstract dual equality). — *We assume that:*

(a) *Φ is a convex lower $\sigma(\mathbb{U}, \mathbb{U}^*)$-semicontinuous function and there is some $\ell_o \in \mathbb{U}^*$ such that for all $u \in \mathbb{U}$, $\Phi(u) \geqslant \Phi(0) + \langle \ell_o, u \rangle_{\mathbb{U}^*, \mathbb{U}}$;*

(b) *Φ^* has $\sigma(\mathbb{U}^*, \mathbb{U})$-compact level sets: $\{\ell \in \mathbb{U}^* : \Phi^*(\ell) \leqslant a\}$, $a \in \mathbb{R}$;*

(c) *The algebraic adjoint* $A^\dagger$ *of* A *satisfies* $A^\dagger\mathbb{V} \subset \mathbb{U}$.

Then, the dual equality

$$\inf\left\{\Phi^*(\ell); \ell \in \mathbb{U}^*, A\ell = v^*\right\} = \sup_{v\in\mathbb{V}} \left\{\langle v, v^*\rangle_{\mathbb{V},\mathbb{V}^*} - \Phi(A^\dagger v)\right\} \in (-\infty,\infty] \quad (4.1)$$

holds true for any $v^* \in \mathbb{V}^*$.

Proof of Lemma 4.2. — In the special case where $\Phi(0) = 0$ and $\ell_o = 0$, this result is [11, Thm. 2.3]. Considering $\Psi(u) := \Phi(u) - [\Phi(0) + \langle \ell_o, u\rangle]$, $u \in \mathbb{U}$, we see that $\inf \Psi = \Psi(0) = 0$, Ψ is a convex lower $\sigma(\mathbb{U},\mathbb{U}^*)$-semicontinuous function and $\Psi^*(\ell) = \Phi^*(\ell_o+\ell)+\Phi(0)$, $\ell \in \mathbb{U}^*$, has $\sigma(\mathbb{U}^*,\mathbb{U})$-compact level sets. As Ψ^* satisfies the assumptions of [11, Thm. 2.3], we have the dual equality

$$\begin{aligned}\inf\left\{\Psi^*(\ell); \ell \in \mathbb{U}^*, A\ell = v^* - A\ell_o\right\}&\\ = \sup_{v\in\mathbb{V}} &\left\{\langle v, v^* - A\ell_o\rangle_{\mathbb{V},\mathbb{V}^*} - \Psi(A^\dagger v)\right\} \in [0,\infty]\end{aligned}$$

which is (4.1). □

Proof of Theorem 4.1. — Let us denote

$$R_{\mu_0}(\,\cdot\,) := \int_{\mathbb{R}^n} R_x(\,\cdot\,)\,\mu_0(\mathrm{d}x) \in \mathcal{P}(\Omega).$$

R_{μ_0} is a measure on paths, its initial marginal, as a probability measure in $\mathbb{R}^n$, is $R_{\mu_0,0} = \mu_0$. When $\mu_0 = m$, we have $R_m = R$. If U is any bounded functional on paths,

$$\begin{aligned}E_{R_{\mu_0}}(U) &= \int E_R(U \mid X_0 = x)\mu_0(\mathrm{d}x)\\ &= \int E_R(U \mid X_0 = x)\frac{\mathrm{d}\mu_0}{\mathrm{d}R_0}(x)R_0(\mathrm{d}x)\\ &= E_R\big(E_R(U\frac{\mathrm{d}\mu_0}{\mathrm{d}m}(X_0) \mid X_0)\big)\\ &= E_R\big(U\frac{\mathrm{d}\mu_0}{\mathrm{d}m}(X_0)\big).\end{aligned}$$

So $R_{\mu_0}(\,\cdot\,) = \dfrac{\mathrm{d}\mu_0}{\mathrm{d}m}(X_0)\,R(\,\cdot\,)$, we see that for any $P \in \mathcal{P}(\Omega)$ such that $P_0 = \mu_0$,

$$H(P \mid R) = H(\mu_0 \mid m) + H(P \mid R_{\mu_0}). \quad (4.2)$$

Consequently, the minimizer of (3.1) is also the minimizer of

$$H(P \mid R_{\mu_0}) \to \min; \quad P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1 \quad (4.3)$$

and

$$\mathcal{A}(\mu_0,\mu_1) = H(\mu_0 \mid m) + \inf\{H(P \mid R_{\mu_0}),\ P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\}. \quad (4.4)$$

Therefore, all we have to prove is

$$\begin{aligned}&\inf\{H(P\,|\,R_{\mu_0}),\ P\in\mathcal{P}(\Omega):P_0=\mu_0,P_1=\mu_1\}\\&\qquad\qquad=\sup\left\{\int_{\mathbb{R}^n}\psi\,\mathrm{d}\mu_1-\int_{\mathbb{R}^n}\mathcal{Q}^R\psi\,\mathrm{d}\mu_0,\ \psi\in C_b(\mathbb{R}^n)\right\}.\end{aligned}$$

This is an application of Lemma 4.2 with $\mathbb{U}=C_b(\Omega)$, $\mathbb{V}=C_b(\mathbb{R}^n)$ and

$$\begin{aligned}\Phi(u)&=\int_{\mathbb{R}^n}\log\left(\int_\Omega e^u\,\mathrm{d}R^x\right)\mu_0(\mathrm{d}x), &&u\in C_b(\Omega),\\ A^\dagger\psi&=\psi(X_1)\in C_b(\Omega), &&\psi\in C_b(\mathbb{R}^n).\end{aligned}$$

Let $C_b(\Omega)'$ be the topological dual space of $(C_b(\Omega),\|\cdot\|)$ equipped with the uniform norm $\|u\|:=\sup_\Omega|u|$. It is shown in [12, Lem. 4.2] that for any $\ell\in C_b(\Omega)'$,

$$\Phi^*(\ell)=\begin{cases}H(\ell\,|\,R_{\mu_0}), & \text{if } \ell\in\mathcal{P}(\Omega)\text{ and }(X_0)_\#\ell=\mu_0\\ +\infty, & \text{otherwise}\end{cases}\tag{4.5}$$

But according to [10, Lem. 2.1], the effective domain $\{\ell\in C_b(\Omega)^*:\Phi^*(\ell)<\infty\}$ of Φ^* is a subset of $C_b(\Omega)'$. Hence, for any ℓ in the algebraic dual $C_b(\Omega)^*$ of $C_b(\Omega)$, $\Phi^*(\ell)$ is given by (4.5).

The assumption (c) of Lemma 4.2 on $A^\dagger$ is obviously satisfied. Let us show that Φ and Φ^* satisfy the assumptions (a) and (b).

Let us start with (a). It is a standard result of the large deviation theory that $u\mapsto\log\int_\Omega e^u\,dR^x$ is convex (a consequence of Hölder's inequality). It follows that Φ is also convex. As Φ is upper bounded on a neighborhood of 0 in $(C_b(\Omega),\|\cdot\|)$:

$$\sup_{u\in C_b(\Omega),\|u\|\leqslant 1}\Phi(u)\leqslant 1<\infty\tag{4.6}$$

(note that Φ is increasing and $\Phi(1)=1$) and its effective domain is the whole space $\mathbb{U}=C_b(\Omega)$, it is $\|\cdot\|$-continuous everywhere. Since Φ is convex, it is also lower $\sigma(C_b(\Omega),C_b(\Omega)')$-semicontinuous and a fortiori lower $\sigma(C_b(\Omega),C_b(\Omega)^*)$-semicontinuous. Finally, a direct calculation shows that $\ell_o=R_{\mu_0}$ is a subgradient of Φ at 0. This completes the verification of (a).

The assumption (b) is also satisfied because the upper bound (4.6) implies that the level sets of Φ^* are $\sigma(C_b(\Omega)^*,C_b(\Omega))$-compact, see [10, Cor. 2.2]. So far, we have shown that the assumptions of Lemma 4.2 are satisfied.

It remains to show that $A\ell=v^*$ corresponds to the final marginal constraint. Since $\{\Phi^*<\infty\}$ consists of probability measures, it is enough to specify the action of A on the vector subspace $\mathcal{M}_b(\Omega)\subset C_b(\Omega)^*$ of all bounded measures on Ω. For any $Q \in \mathcal{M}_b(\Omega)$ and any $\psi \in C_b(\Omega)$, we have

$$\langle \psi, AQ \rangle_{C_b(\mathbb{R}^n),C_b(\mathbb{R}^n)^*} = \left\langle A^\dagger \psi, Q \right\rangle_{C_b(\Omega),C_b(\Omega)^*} = \int_\Omega \psi(X_1)\,\mathrm{d}Q = \int_{\mathbb{R}^n} \psi\,\mathrm{d}Q_1.$$

This means that for any $Q \in \mathcal{M}_b(\Omega)$, $AQ = Q_1 \in \mathcal{M}_b(\mathbb{R}^n)$.

With these considerations, choosing $v^* = \mu_1 \in \mathcal{P}(\mathbb{R}^n)$ in (4.1) leads us to

$$\begin{aligned}\inf &\left\{ H(Q \,|\, R_{\mu_0});\, Q \in \mathcal{P}(\Omega) : Q_0 = \mu_0, Q_1 = \mu_1 \right\} \\ &= \sup_{\psi \in C_b(\mathbb{R}^n)} \left\{ \int_{\mathbb{R}^n} \psi\,\mathrm{d}\mu_1 - \int_{\mathbb{R}^n} \log \left\langle e^{\psi(X_1)}, R^x \right\rangle \mu_0(\mathrm{d}x) \right\}\end{aligned}$$

which is the desired identity. □

Remark 4.3. — Alternatively, considering $R^y := R^{\mu_1}(\,\cdot\,|\, X_1 = y)$, for m-almost all $x \in \mathbb{R}^n$ and

$$R^{\mu_1}(\,\cdot\,) := \int_{\mathbb{R}^n} R^y(\,\cdot\,)\,\mu_1(\mathrm{d}y) \in \mathcal{P}(\Omega)$$

we would obtain a formulation analogous to (1.4).

Remark 4.4. — We didn't use any specific property of the Kolmogorov semigroup. The dual equality can be generalized, without changing its proof, to any reference path measure $R \in \mathcal{P}(\Omega)$ on any Polish state space $\mathcal{X}$.

5. Benamou–Brenier formulation of the entropic cost

We derive some analogue of the Benamou–Brenier formulation (1.5) for the entropic cost.

Theorem 5.1 (Benamou–Brenier formulation of the entropic cost). — *Let V, μ_0 and μ_1 be such that hypothesis (Exi), (Reg1) and (Reg2) stated in Section 3 are satisfied. We have*

$$\mathcal{A}(\mu_0, \mu_1) = H(\mu_0 \,|\, m) + \inf_{(\nu,v)} \int_{\mathbb{R}^n \times [0,1]} \frac{|v_t(z)|^2}{2}\,\nu_t(\mathrm{d}z)\mathrm{d}t, \tag{5.1}$$

where the infimum is taken over all $(\nu_t, v_t)_{0 \leqslant t \leqslant 1}$ such that, $\nu_t(\mathrm{d}z)$ is identified with its density with respect to Lebesgue measure $\nu(t,z) := d\nu_t/\mathrm{d}z$, satisfying $\nu_0 = \mu_0$, $\nu_1 = \mu_1$ and the following continuity equation

$$\partial_t \nu + \nabla \cdot (\nu\,[v - \nabla(V + \log \nu)/2]) = 0, \tag{5.2}$$

is satisfied in a weak sense.

Moreover, these results still hold true when the infimum in (5.1) *is taken among all* (ν, v) *satisfying* (5.2) *and such that* v *is a gradient vector field, that is*

$$v_t(z) = \nabla\psi_t(z), \quad 0 \leqslant t \leqslant 1, z \in \mathbb{R}^n,$$

for some function $\psi \in \mathcal{C}^\infty([0,1) \times \mathbb{R}^n)$.

Remarks 5.2. —

(1) The continuity equation (5.2) is the linear Fokker–Planck equation

$$\partial_t\nu + \nabla \cdot (\nu\,[v - \nabla V/2]) - \Delta\nu/2 = 0.$$

Its solution $(\nu_t)_{0\leqslant t\leqslant 1}$, with v considered as a known parameter, is the time marginal flow $\nu_t = P_t$ of a weak solution $P \in \mathcal{P}(\Omega)$ (if it exists) of the stochastic differential equation

$$\mathrm{d}X_t = [v_t(X_t) - \nabla V(X_t)/2]\,\mathrm{d}t + \mathrm{d}W_t^P, \quad P\text{-a.s.}$$

where W^P is a P-Brownian motion, $P_0 = \mu_0$ and $(X_t)_{1\leqslant t\leqslant 1}$ is the canonical process.

(2) Clearly, one can restrict the infimum in the identity (5.1) to (ν, v) such that

$$\int_{\mathbb{R}^n\times[0,1]} |v_t(z)|^2\,\nu_t(\mathrm{d}z)\mathrm{d}t < \infty. \tag{5.3}$$

Proof. — Because of (3.2) and (4.4), all we have to show is

$$\inf\{H(P\,|\,R_{\mu_0});\ P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\} = \inf_{(\nu,v)} \int_{\mathbb{R}^n\times[0,1]} \frac{|v_t(z)|^2}{2}\,\nu_t(\mathrm{d}z)\mathrm{d}t,$$

where (ν, v) satisfies (5.2), $\nu_0 = \mu_0$ and $\nu_1 = \mu_1$. As R_{μ_0} is Markov, by [15, Prop. 2.10] we can restrict the infimum to the set of all Markov measures $P \in \mathcal{P}(\Omega)$ such that $P_0 = \mu_0, P_1 = \mu_1$ and $H(P\,|\,R_{\mu_0}) < \infty$. For each such Markov measure P, Girsanov's theorem (see for instance [13, Thm. 2.1] for a proof related to the present setting) states that there exists a measurable vector field $\beta_t^P(z)$ such that

$$\mathrm{d}X_t = [\beta_t^P(X_t) - \nabla V(X_t)/2]\,\mathrm{d}t + \mathrm{d}W_t^P, \quad P\text{-a.s.}, \tag{5.4}$$

where W^P is a P-Brownian motion. Moreover, β^P satisfies

$$E_P \int_0^1 |\beta_t^P|^2(X_t)\mathrm{d}t < \infty$$

and

$$H(P\,|\,R_{\mu_0}) = \frac{1}{2}\int_{\mathbb{R}^n\times[0,1]} |\beta_t^P|^2(z)\,P_t(\mathrm{d}z)\mathrm{d}t. \tag{5.5}$$

For any P with $P_0 = \mu_0$, $H(\mu_0\,|\,m) < \infty$ and $H(P\,|\,R_{\mu_0}) < \infty$, we have $P_t \ll R_t = m \ll \mathrm{Leb}$ for all t. Taking $\nu = (P_t)_{0\leqslant t\leqslant 1}$ and $v = \beta^P$, the stochastic differential equation (5.4) gives (5.2) and optimizing the left hand side of (5.5) leads us to

$$\inf\{H(P\,|\,R_{\mu_0});\ P\in\mathcal{P}(\Omega): P_0=\mu_0, P_1=\mu_1\} \leqslant \inf_{(\nu,v)}\int_{\mathbb{R}^n\times[0,1]} \frac{|v_t(z)|^2}{2}\,\nu_t(\mathrm{d}z)\mathrm{d}t.$$

On the other hand, it is proved in [23, 15] that the solution P of the Schrödinger problem (4.3) is such that (5.4) is satisfied with $\beta_t^P(z)=\nabla\psi_t(z)$ where ψ is given in (3.6). This completes the proof of the theorem. □

Corollary 5.3. — *Let V, μ_0 and μ_1 be such that the hypotheses stated in Section 3 are satisfied. We have*

$$\mathcal{A}(\mu_0,\mu_1)=\frac{1}{2}[H(\mu_0\,|\,m)+H(\mu_1\,|\,m)] + \inf_{(\rho,v)}\int_{\mathbb{R}^n\times[0,1]}\left(\frac{1}{2}|v_t(z)|^2+\frac{1}{8}|\nabla\log\rho_t(z)|^2\right)\rho_t(z)\,m(\mathrm{d}z)\mathrm{d}t, \quad (5.6)$$

where the infimum is taken over all $(\rho_t,v_t)_{0\leqslant t\leqslant 1}$ such that $\rho_0\,m=\mu_0$, $\rho_1\,m=\mu_1$ and the following continuity equation

$$\partial_t\rho+e^V\nabla\cdot\left(e^{-V}\rho v\right)=0 \quad (5.7)$$

is satisfied in a weak sense.

Moreover, these results still hold true when the infimum in (5.6) *is taken among all (ν,v) satisfying* (5.7) *and such that v is a gradient vector field, that is*

$$v_t(z)=\nabla\theta_t(z),\quad 0\leqslant t\leqslant 1, z\in\mathbb{R}^n,$$

for some function $\theta\in\mathcal{C}^\infty([0,1)\times\mathbb{R}^n)$.

Remark 5.4. — The density ρ in the statement of the corollary must be understood as a density $\rho=\mathrm{d}\nu/\mathrm{d}m$ with respect to the reversing measure m. Indeed, with $\nu(t,z)=\mathrm{d}\nu_t/\mathrm{d}z$, we see that $\nu=e^{-V}\rho$ and the evolution equation (5.7) writes as the current equation $\partial_t\nu+\nabla\cdot(\nu v)=0$.

This result was proved recently by Chen, Georgiou and Pavon in [7] in the case where $V=0$ without any mention to gradient type vector fields. The present proof is essentially the same as in [7]: we take advantage of the time reversal invariance of the relative entropy $H(\cdot\,|\,R)$ with respect to the reversible path measure R.

Proof. — The proof follows almost the same line as Theorem 5.1's one. The additional ingredient is the time-reversal invariance (3.3): $H(P\,|\,R)=H(P^*\,|\,R)$. Let $P\in\mathcal{P}(\Omega)$ be the solution of (3.1). We have already noted that

P^* is the solution of the Schrödinger problem where the marginal constraints μ_0 and μ_1 are inverted. We obtain

$$\begin{aligned} \mathrm{d}X_t &= v_t^P(X_t)\,\mathrm{d}t + \mathrm{d}W_t^P, && P\text{-a.s.} \\ \mathrm{d}X_t &= v_t^{P^*}(X_t)\,\mathrm{d}t + \mathrm{d}W_t^{P^*}, && P^*\text{-a.s.} \end{aligned}$$

where W^P and W^{P^*} are respectively Brownian motions with respect to P and P^* and

$$\begin{aligned} H(P\,|\,R) &= H(\mu_0\,|\,m) + E_P\frac{1}{2}\int_0^1 |\beta_t^P(X_t)|^2\,\mathrm{d}t\,, \\ H(P^*\,|\,R) &= H(\mu_1\,|\,m) + \frac{1}{2}E_{P^*}\int_0^1 |\beta_t^{P^*}(X_t)|^2\,\mathrm{d}t \\ &= H(\mu_1\,|\,m) + \frac{1}{2}E_P\int_0^1 |\beta_{1-t}^{P^*}(X_t)|^2\,\mathrm{d}t \end{aligned}$$

with $v^P = -\nabla V/2 + \beta^P$ and $v^{P^*} = -\nabla V/2 + \beta^{P^*}$. Taking the half sum of the above equations, the identity $H(P\,|\,R) = H(P^*\,|\,R)$ implies that

$$H(P\,|\,R) = \frac{1}{2}[H(\mu_0\,|\,m) + H(\mu_1\,|\,m)] + \frac{1}{4}E_P\int_0^1 (|\beta_t^P|^2 + |\beta_{1-t}^{P^*}|^2)\,\mathrm{d}t.$$

Let us introduce the current velocities of P and P^* defined by

$$\begin{aligned} v_t^{cu,P}(z) &:= v_t^P(z) - \frac{1}{2}\nabla\log\nu_t^P(z) &&= \beta_t^P(z) - \frac{1}{2}\nabla\log\rho_t^P(z), \\ v_t^{cu,P^*}(z) &:= v_t^{P^*}(z) - \frac{1}{2}\nabla\log\nu_t^{P^*}(z) &&= \beta_t^{P^*}(z) - \frac{1}{2}\nabla\log\rho_t^{P^*}(z) \end{aligned}$$

where for any $0 \leqslant t \leqslant 1, z \in \mathbb{R}^n$,

$$\nu_t^P(z) := \frac{\mathrm{d}P_t}{\mathrm{d}z},\ \rho_t^P(z) := \frac{\mathrm{d}P_t}{\mathrm{d}m}(z) \quad \text{and} \quad \nu_t^{P^*}(z) := \frac{\mathrm{d}P_t^*}{\mathrm{d}z},\ \rho_t^{P^*}(z) := \frac{\mathrm{d}P_t^*}{\mathrm{d}m}(z).$$

The naming *current velocity* is justified by the current equations

$$\begin{aligned} \partial_t\nu^P + \nabla\cdot(\nu^P v^{cu,P}) &= 0 \quad \text{and} \quad \partial_t\rho^P + e^V\nabla\cdot(e^{-V}\rho^P v^{cu,P}) &&= 0, \\ \partial_t\nu^{P^*} + \nabla\cdot(\nu^{P^*} v^{cu,P^*}) &= 0 \quad \text{and} \quad \partial_t\rho^{P^*} + e^V\nabla\cdot(e^{-V}\rho^{P^*} v^{cu,P^*}) &&= 0. \end{aligned}$$

To see that the first equation $\partial_t\nu^P + \nabla\cdot(\nu^P v^{cu,P}) = 0$ is valid, remark that ν^P satisfies the Fokker–Planck equation (5.2) with v replaced by β^P. The equation for ρ^P follows immediately and the equations for ν^{P^*} and ρ^{P^*} are derived similarly.

The very definition of P^* implies that $\rho_t^{P^*} = \rho_{1-t}^P$ and the time reversal invariance $R^* = R$ implies that

$$v_t^{cu,P^*}(z) = -v_{1-t}^{cu,P}(z), \quad 0 \leqslant t \leqslant 1, z \in \mathbb{R}^n.$$

Therefore, $\beta^{P^*}_{1-t} = -v^{cu,P}_t + \frac{1}{2}\nabla \log \rho^P_t$ and $\frac{1}{4}(|\beta^P_t|^2 + |\beta^{P^*}_{1-t}|^2) = \frac{1}{2}|v^{cu,P}_t|^2 + \frac{1}{8}|\nabla \log \rho^P_t|^2$. This completes the proof of the first statement of the corollary.

For the second statement about $v = \nabla\theta$, remark that as in Theorem 5.1's proof, the solution P of the Schrödinger problem is such that $\beta^P = \nabla\psi$ for some smooth function ψ. One concludes with $v^{cu,P} = \beta^P - \frac{1}{2}\nabla \log \rho^P$, by taking $\theta = \psi - \log\sqrt{\rho^P}$. □

Remarks 5.5. —

(1) The current velocity $v^{cu,P}$ of a diffusion path measure P has been introduced by Nelson in [17] together with its *osmotic velocity* $v^{os,P} := \frac{1}{2}\nabla \log \rho^P$.

(2) Up to a multiplicative factor, $\int_{\mathbb{R}^n} |\nabla \log \rho_t(z)|^2 \, \rho_t(z)\, m(\mathrm{d}z)$ is the entropy production or Fischer information. The average osmotic action is $A^{os}(P) := \int_{\mathbb{R}^n\times[0,1]} \frac{1}{2}|v^{os,P}|^2 \,\mathrm{d}P_t\mathrm{d}t = \int_{\mathbb{R}^n\times[0,1]} \frac{1}{8}|\nabla \log \rho|^2\rho\,\mathrm{d}m\,\mathrm{d}t$ and it is directly connected to a variation of entropy. It's worth remarking that by considering the dilatation in time of the reference path measure as introduced in Remark 3.6, the osmotic action vanishes in the limit for $\varepsilon \to 0$. Let us define now the osmotic cost

$$I^{\mathrm{os}}(\mu_0,\mu_1) := \inf\{A^{os}(P); P \in \mathcal{P}(\Omega) : P_0 = \mu_0, P_1 = \mu_1\}$$

and the current cost

$$I^{\mathrm{cu}}(\mu_0,\mu_1) := \inf_{(\rho,v)} \int_{\mathbb{R}^n\times[0,1]} \frac{1}{2}|v_t(z)|^2\rho_t(z)\, m(\mathrm{d}z)\mathrm{d}t$$

where the infimum runs through all the (ρ, v) satisfying (5.7). The standard Benamou–Brenier formula precisely states that $I^{\mathrm{cu}}(\mu_0,\mu_1) = W_2^2(\mu_0,\mu_1)/2$. Therefore, Corollary 5.3 implies that

$$\mathcal{A}(\mu_0,\mu_1) \geqslant \frac{1}{2}[H(\mu_0\,|\,m) + H(\mu_1\,|\,m)] + \frac{1}{2}W_2^2(\mu_0,\mu_1) + I^{os}(\mu_0,\mu_1).$$

In particular, by the positivity of the entropic cost I^{os} we obtain the following relation between the entropic and Wasserstein costs:

Corollary 5.6. — *Let V, μ_0 and μ_1 be such that the hypotheses stated in Section 3 are satisfied. We have*

$$\mathcal{A}(\mu_0,\mu_1) \geqslant \frac{1}{2}[H(\mu_0\,|\,m) + H(\mu_1\,|\,m)] + \frac{1}{2}W_2^2(\mu_0,\mu_1).$$

6. Contraction with respect to the entropic cost

The analogy between optimal transport and minimal entropy can also be observed in the context of contractions.

As explained in the introduction, contraction in Wasserstein distance depends on the curvature. Even if there are actually many contraction inequalities in Wasserstein distance, we focus here on two main results. The first one depends on the curvature and the second one includes the dimension. These results can be written for more general semigroups satisfying the curvature-dimension condition as defined in the Bakry-Émery theory.

In the context of the Kolmogorov semigroup of Section 2 with a generator given by (2.1) in $\mathbb{R}^n$, the two main contraction inequalities can be formulated as follows.

- Let us assume that for some real λ, we have $\mathrm{Hess}(V) \geqslant \lambda\,\mathrm{Id}$ in the sense of symmetric matrices. Then for any f, g probability densities with respect to the measure m and any $t \geqslant 0$,
$$W_2(T_t f\, m, T_t g\, m) \leqslant e^{-\frac{\lambda}{2}t} W_2(fm, gm). \tag{6.1}$$
Let us recall that this result was proved in [18] in the general context of Riemannian manifold. Although in the context of Kolmogorov semigroups the proof is easy, its generalization for the entropic cost to a Riemannian setting is not trivial.
- When $L = \Delta/2$, that is $V = 0$, the heat equation in $\mathbb{R}^n$ satifies the following dimension dependent contraction property:
$$W_2^2(T_t f\,\mathrm{Leb}, T_s g\,\mathrm{Leb}) \leqslant W_2^2(f\,\mathrm{Leb}, g\,\mathrm{Leb}) + n(\sqrt{t} - \sqrt{s})^2, \tag{6.2}$$
for any $s, t \geqslant 0$ and any probability densities f, g with respect to the Lebesgue measure Leb. This contraction was proved in a more general context in [4, 9].

The two inequalities (6.1) and (6.2) can be proved in terms of entropic cost. Let us choose the reference path measure R associated with the potential V and take $\varepsilon, u > 0$ and $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^n)$. In order to extend for each $u, \varepsilon > 0$ the dual formulation for the entropic cost of Theorem 4.1, consider the semigroup $(T_{\varepsilon u t})_{t \geqslant 0}$ and the corresponding path measure $R^{\varepsilon u}$: time is dilated by the factor $(\varepsilon u)^{-1}$. Theorem 4.1 implies that

$$\begin{aligned}&\mathcal{A}^{R^{\varepsilon u}}(\mu_0, \mu_1)\\ &\quad = H(\mu_0\,|\,m) + \sup\left\{\int_{\mathbb{R}^n} \psi\, \mathrm{d}\mu_1 - \int_{\mathbb{R}^n} \log T_{\varepsilon u}(e^{\psi})\, \mathrm{d}\mu_0,\ \ \psi \in C_b(\mathbb{R}^n)\right\}.\end{aligned}$$

Now by changing ψ with ψ/ε we see that

$$\begin{aligned}&\varepsilon\mathcal{A}^{R^{\varepsilon u}}(\mu_0, \mu_1)\\ &\quad = \varepsilon H(\mu_0\,|\,m) + \sup\left\{\int_{\mathbb{R}^n} \psi\, \mathrm{d}\mu_1 - \int_{\mathbb{R}^n} \mathcal{Q}_u^{\varepsilon}\psi\, \mathrm{d}\mu_0,\ \ \psi \in C_b(\mathbb{R}^n)\right\}\end{aligned} \tag{6.3}$$

where for any $\psi \in C_b(\mathbb{R}^n)$,

$$\mathcal{Q}_u^\varepsilon \psi = \varepsilon \log T_{\varepsilon u}(e^{\psi/\varepsilon}). \tag{6.4}$$

For simplicity, we denote $\varepsilon \mathcal{A}^{R^{\varepsilon u}} = \mathcal{A}_u^\varepsilon$ and $\mathcal{A}_1^\varepsilon = \mathcal{A}^\varepsilon$.

As explained in Remark 3.6, we have

$$\lim_{\varepsilon \to 0} \mathcal{A}_u^\varepsilon(\mu_0, \mu_1) = W_2^2(\mu_0, \mu_1)/2u. \tag{6.5}$$

The entropic cost associated to the Kolmogorov semigroup has the following properties.

Theorem 6.1 (Contraction in entropic cost). — *Let $\varepsilon > 0$ be fixed.*

(a) *If V satisfies* $\mathrm{Hess}(V) \geqslant \lambda\,\mathrm{Id}$ *for some $\lambda \in \mathbb{R}$, then for any $t \geqslant 0$,*

$$\begin{aligned} &\mathcal{A}_b^\varepsilon(T_{u_t(b)} fm, T_t gm) \\ &\qquad \leqslant \mathcal{A}_{v_t(b)}^\varepsilon(fm, gm) + \varepsilon[H(T_{u_t(b)} fm | m) - H(fm \,|\, m)], \end{aligned} \tag{6.6}$$

where f, g are probability densities with respect to m, and

$$\begin{aligned} u_t(b) &= t + \frac{1}{\lambda} \log\left(\frac{e^{-\varepsilon\lambda b}}{1 + e^{\lambda t}(e^{-\varepsilon\lambda b} - 1)}\right), \\ v_t(b) &= -\frac{1}{\lambda\varepsilon} \log(1 + e^{\lambda t}(e^{-\varepsilon\lambda b} - 1)) \end{aligned} \tag{6.7}$$

where: if $\lambda \leqslant 0$, $b \in (0, \infty)$ and if $\lambda > 0$, $b \in (0, -\frac{1}{\lambda\varepsilon}\log(1 - e^{-\lambda t}))$.

(b) *If $V = 0$ then for any $t \geqslant 0$,*

$$\begin{aligned} &\mathcal{A}^\varepsilon(T_t fm, T_s gm) \\ &\qquad \leqslant \mathcal{A}^\varepsilon(fm, gm) + \frac{n}{2}(\sqrt{t} - \sqrt{s})^2 + \varepsilon[H(T_t fm \,|\, m) - H(fm \,|\, m)]. \end{aligned}$$

The proof of this theorem relies on the following commutation property between the Markov semigroup T_t and the semigroup $\mathcal{Q}_t^\varepsilon$ defined in (6.4). Let us notice that the second statement of next lemma was proved in [4, Section 5].

Lemma 6.2 (Commutation property). — *Let $s, t \geqslant 0$, $\varepsilon > 0$ and $f : \mathbb{R}^n \to \mathbb{R}$ be any bounded measurable function.*

(a) *If* $\mathrm{Hess}(V) \geqslant \lambda\,\mathrm{Id}$ *for some real λ, then*

$$\mathcal{Q}_{v_t(b)}^\varepsilon(T_t f) \leqslant T_{u_t(b)}(\mathcal{Q}_b^\varepsilon f) \tag{6.8}$$

where for each $t \geqslant 0$, the numbers $u_t(b)$, $v_t(b)$ and b are given in (6.7). Moreover, for ε small enough and $t > 0$ fixed, (6.8) is valid for all b positive.

(b) *If* $V=0$ *then*

$$\mathcal{Q}_1^\varepsilon(T_t f) \leqslant T_s(\mathcal{Q}_1^\varepsilon f) + \frac{n}{2}(\sqrt{t}-\sqrt{s})^2.$$

Proof. — We only have to prove the first statement (a). Let us define for each $s \leqslant t$ the function

$$\Lambda(s) = T_\alpha \mathcal{Q}_\beta^\varepsilon(T_{t-s} f)$$

with $\alpha : [0,t] \to [0,\infty)$ an increasing function such that $\alpha(0)=0$, and $\beta : [0,t] \to [0,\infty)$ and we call $\beta(t)=b$. Setting $g=\exp(T_{t-s}f/\varepsilon)$, using the chain rule for the diffusion operator L we obtain

$$\begin{aligned}
&\Lambda'(s)\\
&= \varepsilon T_\alpha\left[\alpha' L \log T_{\varepsilon\beta} g + \frac{1}{T_{\varepsilon\beta} g} T_{\varepsilon\beta}\left(\varepsilon\beta' Lg - gL\log g\right)\right]\\
&= \varepsilon T_\alpha\left[\alpha'\left(\frac{LT_{\varepsilon\beta}g}{T_{\varepsilon\beta}g} - \frac{|\nabla T_{\varepsilon\beta} g|^2}{2(T_{\varepsilon\beta}g)^2}\right) + \frac{1}{T_{\varepsilon\beta}g} T_{\varepsilon\beta}\left(\varepsilon\beta' Lg - Lg + \frac{|\nabla g|^2}{2g}\right)\right] \qquad (6.9)\\
&= \varepsilon T_\alpha\left[\frac{1}{T_{\varepsilon\beta}g}\left(LT_{\varepsilon\beta}g(\alpha'+\varepsilon\beta'-1) + T_{\varepsilon\beta}\left(\frac{|\nabla g|^2}{2g}\right) - \alpha'\frac{|\nabla T_{\varepsilon\beta}g|^2}{2T_{\varepsilon\beta}g}\right)\right]\\
&\geqslant \varepsilon T_\alpha\left[\frac{1}{T_{\varepsilon\beta}g}\left(LT_{\varepsilon\beta}g(\alpha'+\varepsilon\beta'-1) + \frac{1}{2}T_{\varepsilon\beta}\left(\frac{|\nabla g|^2}{g}\right)(1-e^{-\lambda\varepsilon\beta}\alpha')\right)\right]
\end{aligned}$$

where the last inequality is given by the commutation,

$$\frac{|\nabla T_t g|^2}{T_t g} \leqslant e^{-\lambda t} T_t\left(\frac{|\nabla g|^2}{g}\right)$$

which is implied by the condition $\mathrm{Hess}(V) \geqslant \lambda \,\mathrm{Id}$ (see for instance [3, Section 3.2]). If the following conditions on α and β hold

$$\begin{cases} \alpha' + \varepsilon\beta' - 1 = 0 \\ 1 - e^{-\lambda\varepsilon\beta}\alpha' = 0, \end{cases} \qquad (6.10)$$

we have $\Lambda'(s) \geqslant 0$ for each $0 \leqslant s \leqslant t$. In particular $\Lambda(0) \leqslant \Lambda(t)$ for each $t \geqslant 0$, that is

$$\mathcal{Q}_{v_t(b)}^\varepsilon(T_t f) \leqslant T_{u_t(b)}(\mathcal{Q}_b^\varepsilon f)$$

where $v_t(b)=\beta(0)$ and $u_t(b)=\alpha(t)$. Finally solving system (6.10) together with the conditions $\alpha(0)=0$, $\beta(t)=b$, we can compute the explicit formulas for v and u as in statement (a). In particular, substituting α' in the second equation of the system and integrating from 0 to t we obtain the following relation

$$e^{-\varepsilon\lambda\beta(0)} = 1 + e^{\lambda t}(e^{-\varepsilon\lambda b} - 1). \qquad (6.11)$$

If we assume for a while that the term on the right hand side is positive, we obtain

$$\beta(0) = -\frac{1}{\lambda\varepsilon}\log(1 + e^{\lambda t}(e^{-\varepsilon\lambda b} - 1))$$

and

$$\alpha(t) = t + \frac{1}{\lambda}\log\left(\frac{e^{-\varepsilon\lambda b}}{1+e^{\lambda t}(e^{-\varepsilon\lambda b}-1)}\right).$$

Let us study now the sign of the right hand side in (6.11).

- If $\lambda \leqslant 0$, it is positive for each $b \in \mathbb{R}$;
- If $\lambda > 0$ in order to be positive, we need the condition for b,

$$b < -\frac{1}{\varepsilon\lambda}\log(1-e^{-\lambda t}) := b_0.$$

Finally let us consider the case when $\varepsilon > 0$ is small. From (6.11) we obtain the relation

$$\beta(0) = be^{\lambda t} + o(\varepsilon)$$

for each $\lambda \in \mathbb{R}$ and b positive.

This completes the proof of the lemma. □

Proof of Theorem 6.1. — The proof is based on the dual formulation stated in Theorem 4.1. Let $\psi \in C_b(\mathbb{R}^n)$, by Lemma 6.2 under the condition $\mathrm{Hess}(V) \geqslant \lambda\,\mathrm{Id}$ and by time reversibility,

$$\begin{aligned}\int_{\mathbb{R}^n}\psi\, T_t g\,\mathrm{d}m - \int_{\mathbb{R}^n}\mathcal{Q}^\varepsilon_b\psi\, T_{u_t(b)}f\,\mathrm{d}m &= \int_{\mathbb{R}^n}T_t\psi\, g\,\mathrm{d}m - \int_{\mathbb{R}^n}T_{u_t(b)}\mathcal{Q}^\varepsilon_b\psi\, f\,\mathrm{d}m\\ &\leqslant \int_{\mathbb{R}^n}T_t\psi\, g\,\mathrm{d}m - \int_{\mathbb{R}^n}\mathcal{Q}^\varepsilon_{v_t(b)}T_t\psi\, f\,\mathrm{d}m\\ &\leqslant \mathcal{A}^\varepsilon_{v_t(b)}(fm, gm) - \varepsilon H(fm\,|\,m).\end{aligned}$$

Finally taking the supremum over $\psi \in C_b(\mathbb{R}^n)$ we obtain the desired inequality in (a). The same argument can be used to prove the contraction property in (b), applying the second commutation inequality in Lemma 6.2. □

Remark 6.3. — Let us observe that if $\lambda < 0$, the function $\beta(s)$, for $s \in [0, t]$, is decreasing, while for $\lambda > 0$ it is increasing and if $\lambda = 0$ it is the constant function $\beta(t) = b$. In particular by choosing $b = 1$ (6.8) writes as follows:

$$\mathcal{Q}^\varepsilon_1(T_t f) \leqslant T_t(\mathcal{Q}^\varepsilon_1 f).$$

Remark 6.4. — Lemma 6.2 can be proved in the general context of a Markov diffusion operator under the Bakry-Émery curvature-dimension condition. Its application to more general problems is actually a working paper of the third author.

Remarks 6.5. — Let us point out two converse assertions.

(1) The contraction in entropic cost in Theorem 6.1 implies back the contraction in Wasserstein cost. Indeed, under the assumptions of Section 3, it can be easily checked that when $\varepsilon \to 0$, we have $u(t) \to t$ and $v(t) \to be^{\lambda t}$. Therefore, with (6.5) and (6.6), one recovers (6.1). Analogous arguments can be applied to recover the contraction of the Wasserstein cost (6.2) when $V = 0$.

(2) The commutation property in Lemma 6.2 implies back the convexity of the potential V. This can be seen by differentiating (6.8) with respect to b around 0. We believe also that for $\varepsilon > 0$ fixed, inequality (6.6) implies back the convexity of the potential.

7. Examples

In this section we will compute explicitly the results discussed in the previous sections, between two given measures. We first compute the Wasserstein cost, its dual and Benamou–Brenier formulations and the displacement interpolation, as exposed in the introduction. Then, we'll do the same for the entropic cost, taking in consideration two different reference path measures R. In particular, we'll compute (6.3), for $u = 1$ and $\varepsilon > 0$ and look at the behavior in the limit $\varepsilon \to 0$ to recover the classical results of the optimal transport. For abuse of notation we will denote with μ_t both the interpolation and its density with respect to the Lebesgue measure $\mathrm{d}x$. We introduce for Gaussian measures the following notation: for any $m \in \mathbb{R}^n$ and $v \in \mathbb{R}$, the density with respect to the Lebesgue measure of $\mathcal{N}(m, v^2)$ is given by

$$(2\pi v^2)^{-n/2} \exp\Big(- \frac{|x-m|^2}{2v^2} \Big).$$

As marginal measures we consider for $x_0, x_1 \in \mathbb{R}^n$

$$\mu_0(x) := \mathcal{N}(x_0, 1), \quad \mu_1(x) := \mathcal{N}(x_1, 1). \tag{7.1}$$

Note that the entropic interpolation between two Dirac measures δ_x and δ_y should be the Bridge R^{xy} between x and y with respect to the reference measure R. But unfortunately $H(\delta_x \,|\, m), H(\delta_y \,|\, m) = \infty$, hence we consider only marginal measures with a density with respect to m.

7.1. Wasserstein cost

The Wasserstein cost between μ_0, μ_1 as in (7.1), is

$$W_2^2(\mu_0, \mu_1) = \mathrm{d}(x_0, x_1)^2.$$

In its dual formulation, the supremum is reached by the function

$$\psi(x) = (x_1 - x_0)x$$

and in the Benamou–Brenier formulation the minimizer vector field is

$$v^{MC} = x_1 - x_0$$

The displacement or McCann interpolation is given by

$$\mu_t^{MC} = \mathcal{N}(x_t, 1) \tag{7.2}$$

where $x_t = (1-t)x_0 + tx_1$. In other words using the push-forward notation (1.6),

$$\mu_t^{MC} = (\hat{x}_t^{MC})_{\#}\mu_0$$

with $\hat{x}_t^{MC}(x) := (1-t)x + t(x + x_1 - x_0)$ a trajectory whose associated velocity field is $v^{MC} = x_1 - x_0$.

7.2. Schrödinger cost

Heat semigroup

As a first example we consider on the state space $\mathbb{R}^n$ the heat (or Brownian) semigroup, that corresponds to the case $V = 0$ in our main example in Section 2, whose infinitesimal generator is the Laplacian $L = \Delta/2$ and the invariant reference measure is the Lebesgue measure $\mathrm{d}x$. Since we are interested in the ε-entropic interpolation, with $\varepsilon > 0$, we take in consideration the heat semigroup with a dilatation in time, whose density kernel is given by

$$p_t^{\varepsilon}(x,y) = (2\pi\varepsilon t)^{-n/2} \exp\left(-\frac{|x-y|^2}{2\varepsilon t}\right)$$

i.e. $p_t^{\varepsilon}(x,y) = \mathcal{N}(y, \varepsilon t)$ for $t > 0, (x,y) \in \mathbb{R}^n \times \mathbb{R}^n$.

- The entropic interpolation (3.5) is

$$\mu_t^{\varepsilon} = \mathcal{N}(x_t, D_t^{\varepsilon}) \tag{7.3}$$

  where x_t is like in (7.2) and $D_t^{\varepsilon} : [0,1] \to \mathbb{R}^+$ is a polynomial function given by

$$D_t^{\varepsilon} = \alpha^{\varepsilon} t(1-t) + 1$$

  with $\alpha^{\varepsilon} = \delta^2/(1+\delta)$ where $\delta = (\varepsilon - 2 + \sqrt{4+\varepsilon^2})/2$. We observe that D_t^{ε} is such that $D_0 = D_1 = 1$ with a maximum in $t = 1/2$ for each $\varepsilon > 0$, (see Figure 7.1).

It's worth to point out how we managed to derive an explicit formula for the entropic interpolation. The key point is the resolution of the Schrödinger system (3.7) that in our example writes as

$$\begin{cases} (2\pi)^{-n/2}\exp\left(-\frac{|x-x_0|^2}{2}\right) = f(x)\int g(y)(2\pi\varepsilon t)^{-n/2}\exp\left(-\frac{|x-y|^2}{2\varepsilon t}\right)\mathrm{d}y \\ (2\pi)^{-n/2}\exp\left(-\frac{|x-x_1|^2}{2}\right) = g(x)\int f(y)(2\pi\varepsilon t)^{-n/2}\exp\left(-\frac{|x-y|^2}{2\varepsilon t}\right)\mathrm{d}y. \end{cases}$$

By taking f and g exponential functions of the type

$$e^{a_2x^2+a_1x+a_0} \quad \text{with} \quad a_0,a_1,a_2\in\mathbb{R}$$

we can solve the system explicitely by determining the coefficients of f and g.

One can also express the entropic interpolation through the *push-forward* notation as introduced in (3.9), $\mu_t^\varepsilon = (\hat{x}_t^\varepsilon)_\#\mu_0$ where

$$\hat{x}_t^\varepsilon(x) = \sqrt{D_t^\varepsilon}(x-x_0)+x_t.$$

Furthermore $\hat{x}_t^\varepsilon$ satisfies the differential equation

$$\dot{x}_t^\varepsilon = v^{cu,\varepsilon}(x_t^\varepsilon) \tag{7.4}$$

where $v^{cu,\varepsilon}$ is the current velocity given by

$$v^{cu,\varepsilon} = \frac{\dot{D}_t^\varepsilon}{2D_t^\varepsilon}(x-x_t)+x_1-x_0.$$

It can be finally verified that the entropic interpolation (7.3) satisfies the *PDE*

$$\dot{\mu}_t^\varepsilon + \nabla\cdot(\mu_t^\varepsilon v^{cu,\varepsilon}) = 0. \tag{7.5}$$

Remark 7.1. — Let us observe that if $x_0 = x_1$, μ_t^ε is not constant in time, unlike the McCann interpolation.

- Denoting $P\in\mathcal{P}(\Omega)$ the path measure whose flow is given by (7.3) and that minimizes $H(\cdot\,|\,R)$, the entropic cost between μ_0,μ_1 as in (7.2) is
$$\mathcal{A}_u^\varepsilon(\mu_0,\mu_1) = H(P\,|\,R).$$
The easiest way to compute this quantity is to use the Benamou–Brenier formulation in Section 5. The resulting formula has not a nice and interesting form, therefore we don't report it explicitly.
- In the dual formulation proved in Section 3, the supremum is reached by the function $\psi\in C_b(\mathbb{R}^n)$ given, up to a constant term, by
$$\psi_t(x) = -\frac{1}{2}\frac{\delta}{1+\delta(1-t)}x^2 - \frac{1}{2}\frac{\gamma}{1+\delta(1-t)}x \tag{7.6}$$
where δ as in (7.3) and $\gamma = 2[x_0(1+\delta)-x_1]$.

- In the Benamou–Brenier formulation in Theorem 5.1 the minimizer vector field is
$$v^H = \nabla\psi_t.$$
where ψ_t is given by (7.6) and $\nabla\psi_t$ represents the forward velocity. It can be easily verified that the equation
$$\partial_t\mu_t + \nabla\cdot\left(\mu_t\left[\nabla\psi_t - \frac{\nabla\mu_t}{2\mu_t}\right]\right) = 0 \tag{7.7}$$
is satisfied.

Ornstein–Uhlenbeck semigroup

As a second example, we consider on the state space $\mathbb{R}^n$ the Ornstein–Uhlenbeck semigroup, that corresponds to the case $V(x) = |x|^2/2$ for the Kolmogorov semigroup in Section 2, whose infinitesimal generator is given by $L = (\Delta - x\cdot\nabla)/2$ and the invariant measure is the standard Gaussian in $\mathbb{R}^n$. Here again we consider the kernel representation with a dilatation in time; in other words, for $\varepsilon > 0$, the kernel with respect to the Lebesgue measure is given by
$$p_t(x,y) = (2\pi(1-e^{-\varepsilon t}))^{-n/2}\exp\left(-\frac{|y-xe^{-\varepsilon t/2}|^2}{2(1-e^{-\varepsilon t})}\right)$$
i.e. $p_t(x,y) = \mathcal{N}(xe^{-\varepsilon t/2}, 1-e^{-\varepsilon t})(y)$.

- The entropic interpolation (3.5) is given by
$$\mu_t^\varepsilon = \mathcal{N}(m_t, D_t^\varepsilon) \tag{7.8}$$
where $m_t = a_t[(e^{-\varepsilon t/2} - e^{-\varepsilon(1-t/2)})x_0 + (e^{-\varepsilon(1-t)/2} - e^{-\varepsilon(1+t)/2})x_1]$ with
$$a_t := \frac{1+\delta-\delta e^{-\varepsilon}}{(1-e^{-\varepsilon})[\delta(1+\delta)(e^{-\varepsilon t}+e^{-\varepsilon(1-t)}) - 2\delta^2 e^{-\varepsilon}]}$$
with δ as in (7.10), and $D_t^\varepsilon : [0,1] \to \mathbb{R}^+$ defined as
$$D_t^\varepsilon := -1 + 2(1-e^{-\varepsilon})a_t$$
satisfying, as in the case of the Heat semigroup, $D_0^\varepsilon = D_1^\varepsilon = 1$.

  Furthermore, we have $\mu_t^\varepsilon = (\hat{x}_t^\varepsilon)_\#\mu_0$ where
$$\hat{x}_t^\varepsilon := \sqrt{D_t^\varepsilon}(x-x_0) + m_t.$$

  It can be verified that equations (7.5) and (7.4) hold true also in the Ornstein Uhlenbeck case, with the current velocity given by
$$v_{cu}^\varepsilon(x) = \frac{\dot{D}_t^\varepsilon}{2D_t}(x-x_0) + \dot{m}_t \tag{7.9}$$

- The entropic cost between μ_0, μ_1 can be computed as in the Heat semigroup case by
$$\mathcal{A}_u^\varepsilon(\mu_0, \mu_1) = H(P \,|\, R)$$
where P is the path measure associated to the flow (7.8) which minimizes $H(\,\cdot\,|\,R)$.
- In the dual formulation in Section 3, the supremum is reached, up to a constant term, by the function
$$\psi_t(x) = -\frac{1}{2}\frac{\varepsilon\delta e^{-\varepsilon(1-t)}}{1+\delta(1-e^{-\varepsilon(1-t)})}x^2 + \frac{\varepsilon\gamma e^{-\varepsilon(1-t)/2}}{1+\delta(1-e^{-\varepsilon(1-t)})}x \tag{7.10}$$
where $\delta = (e^{-\varepsilon} - \sqrt{e^{-2\varepsilon} - e^{-\varepsilon} + 1})/(e^{-\varepsilon} - 1)$ and $\gamma = (x_0 e^{-\varepsilon/2} - x_1(1+\delta-\delta e^{-\varepsilon}))/(1-e^{-\varepsilon})$.
- In the Benamou–Brenier formulation (Theorem 5.1) the minimizer vector field is
$$v^{OU} = \nabla\psi_t.$$

Remark 7.2. — Let us observe that both in the heat and Ornstein–Uhlenbeck cases, if we take the limit $\varepsilon \to 0$ of the entropic interpolation, of the velocities v^H, v^{OU} and of the function ψ_t, we recover the respective results for the Wasserstein cost stated in Subsection 7.1.

In the following figures we refer to the McCann interpolation with a dotted line, the Heat semigroup with a dashed line and the Ornstein–Uhlenbeck semigroup with a continuous line. We fix $\varepsilon = 1$ and consider marginal measures in one dimension. Figure 7.1 represents the variance of the three interpolations, independent from the initial and final means x_0, x_1. Figures 7.2 and 7.3 correspond to the mean in the three cases respectively with the initial and final means symmetric w.r.t the origin, and for any means. It's worth to remark from these images that the McCann interpolation and the entropic interpolation in the case of the heat semigroup, have the same mean. Finally Figures 7.4 and 7.5 represent the three interpolations at time $t = 0, 1/2, 1$ respectively with different marginal data, as before.

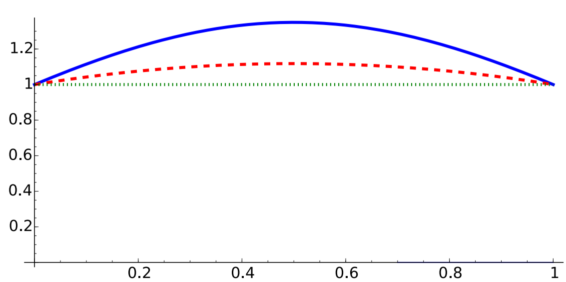

Figure 7.1. Variance, $\varepsilon = 1$

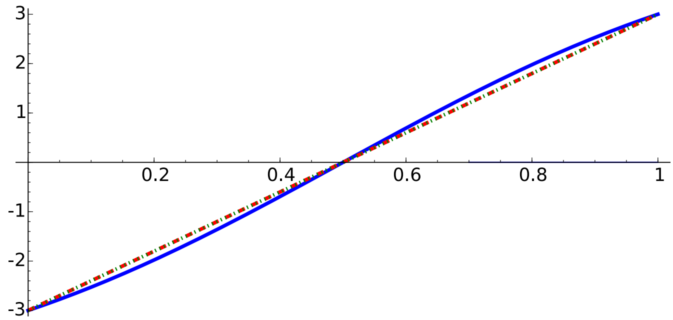

Figure 7.2. Mean with $x_0 = -3, x_1 = 3, \varepsilon = 1$

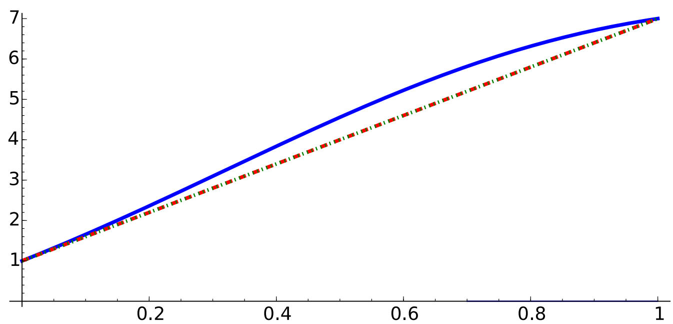

Figure 7.3. Mean with $x_0 = 1, x_1 = 7, \varepsilon = 1$

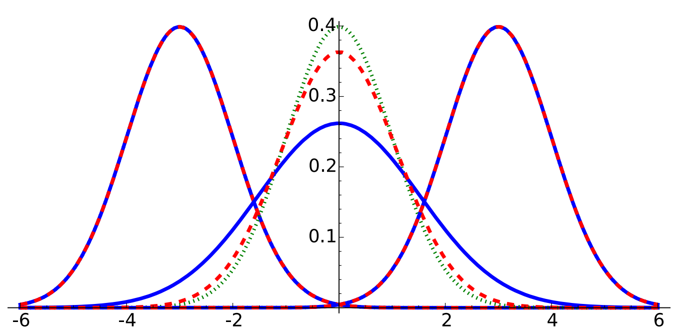

Figure 7.4. Interpolations at time $t = 0, 1/2, 1,$ $x_0 = -3, x_1 = 3, \varepsilon = 1$

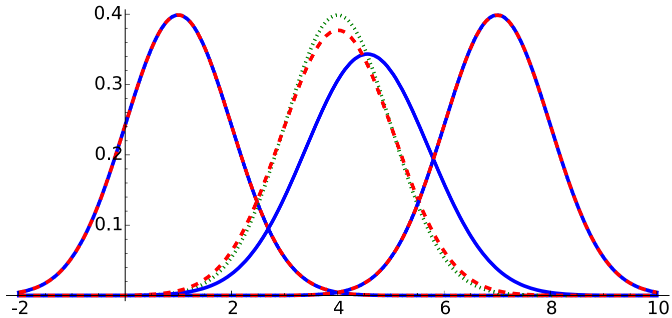

Figure 7.5. Interpolations at time $t = 0, 1/2, 1$, $x_0 = 1, x_1 = 7, \varepsilon = 1$

Bibliography

[1] L. Ambrosio, N. Gigli & G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Lectures in Mathematics, Birkhäuser, 2008, vii+334 pages.

[2] D. Bakry, F. Bolley, I. Gentil & P. Maheux, "Weighted Nash inequalities", *Rev. Mat. Iberoam.* **28** (2012), no. 3, p. 879-906.

[3] D. Bakry, I. Gentil & M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften, vol. 348, Springer, 2014, xx+552 pages.

[4] ———, "On Harnack inequalities and optimal transportation", *Ann. Sc. Norm. Super. Pisa* **14** (2015), no. 3, p. 705-727.

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

[6] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna & G. Peyré, "Iterative Bregman projections for regularized transportation problems", *SIAM J. Sci. Comput.* **37** (2015), no. 2, p. A1111-A1138.

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

[8] L. Kantorovich, "On the translocation of masses", *C. R. (Dokl.) Acad. Sci. URSS* **37** (1942), p. 199-201.

[9] K. Kuwada, "Space-time Wasserstein controls and Bakry-Ledoux type gradient estimates.", *Calc. Var. Partial Differ. Equ.* **54** (2015), no. 1, p. 127-161.

[10] C. Léonard, "Convex conjugates of integral functionals", *Acta Math. Hung.* **93** (2001), no. 4, p. 253-280.

[11] ———, "Minimization of energy functionals applied to some inverse problems", *Appl. Math. Optim.* **44** (2001), no. 3, p. 273-297.

[12] ———, "From the Schrödinger problem to the Monge-Kantorovich problem", *J. Funct. Anal.* **262** (2012), no. 4, p. 1879-1920.

[13] ———, "Girsanov theory under a finite entropy condition", in *Séminaire de Probabilités XLIV*, Lecture Notes in Math., vol. 2046, Springer, 2012, p. 429-465.

[14] ———, "Some properties of path measures", in *Séminaire de probabilités de Strasbourg, vol. 46.*, Lecture Notes in Mathematics, vol. 2123, Springer, 2014, p. 207-230.

[15] ———, "A survey of the Schrödinger problem and some of its connections with optimal transport", *Discrete Contin. Dyn. Syst.* **34** (2014), no. 4, p. 1533-1574.

[16] T. Mikami & M. Thieullen, "Duality theorem for the stochastic optimal control problem", *Stoch. Proc. Appl.* **116** (2006), no. 12, p. 1815-1835.

[17] E. Nelson, *Dynamical theories of Brownian motion*, Princeton University Press, 1967, 142 pages.

[18] M.-K. von Renesse & K.-T. Sturm, "Transport inequalities, gradient estimates, entropy, and Ricci curvature", *Commun. Pure Appl. Math.* **58** (2005), no. 7, p. 923-940.

[19] G. Royer, *Une initiation aux inégalités de Sobolev logarithmiques*, Cours spécialisés (Paris), vol. 5, Société Mathématique de France, 1999, 114 pages.

[20] D. W. Stroock, *Partial differential equations for probabilists*, Cambridge Studies in Advanced Mathematics, vol. 112, Cambridge University Press, 2008, xv+215 pages.

[21] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, 2003, xvi+370 pages.

[22] ———, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, 2009, xxii+973 pages.

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