UNIVERSITÀ CATTOLICA DEL SACRO CUORE
Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di laurea magistrale in Matematica

Variational representations of optimal transport and Schrödinger problem

Relatore
Dr. Luca Tamanini

Correlatore
Dr. Nicolò De Ponti

Candidato
Mattia Garatti
5306287

Anno accademico 2024–2025

There are more things
in heaven and earth, Horatio,
Than are dreamt of
in our philosophy.
William Shakespeare,
Hamlet | Act 1, Scene 5

Abstract

This document, written by Mattia Garatti under the supervision of Dr. Luca Tamanini and the co-supervision of Dr. Nicolò De Ponti, constitutes the final exam paper for the master's degree course in Mathematics offered by *Facoltà di Scienze Matematiche, Fisiche e Naturali* of *Università Cattolica del Sacro Cuore*.

Both optimal transport and the Schrödinger problem are interpolation problems: in the former, one is interested in seeking the optimal deterministic way to send an initial distribution onto a final target, where optimality depends on a given cost to be minimized; in the latter, one aims at finding the most likely evolution for systems of diffusive particles between two different observations.

At first sight, the two problems seem thus quite different, but a more careful insight suggests a strong connection between the two: indeed, the Schrödinger problem can be interpreted as a noised, or regularized, version of the optimal transport problem.

The aim of this thesis is to study some variational representations of these two problems, thus highlighting, at least formally, how optimal transport can be recovered as Γ-limit of the Schrödinger problem. The topics discussed lie at the interface between Mathematical Analysis, Probability Theory and Geometry of the Wasserstein space, thus representing a cutting-edge subject.

Contents

Notations

For sake of clarity, we collect here some standard notations. We emphasize that we adopt the French conventions on ordering.

Basics

χ_E	characteristic function of $E \subseteq X$
$\iota_E = \frac{1-\chi_E}{\lvert\chi_E\rvert}$	indicator function of $E \subseteq X$
$\log x$	natural logarithm of $x > 0$
$\mathbb{N}$	set of natural numbers (so 0 included)
$\overline{\mathbb{R}}$	extended real line
p^i	canonical projection on the i-th factor
p_X	canonical projection on X

Functional analysis

${}^{\perp}E$	left orthogonal of E
$\mathrm{Lip}(f)$	Lipschitz constant of $f \in \mathrm{Lip}(X)$
$\mathrm{Lip}(X)$	space of Lipschitz functions over X
$C(X;Y)$	space of continuous functions from X to Y
$C_b(X;Y)$	space of bounded continuous functions from X to Y
$C_c^\infty(X;Y)$	space of compactly supported smooth functions from X to Y
$D(\Delta) \subseteq H^1(U)$	space of ($\sim_{\mathcal{L}^n}$-equivalence class of) functions of $H^1(U)$ such that $\Delta u \in L^2(U)$

$D(\Delta_{loc}) \subseteq H^1(U)$	space of ($\sim_{\mathcal{L}^n}$-equivalence class of) functions of $H^1(U)$ such that $\Delta u \in L^2_{loc}(U)$
$E^\perp$	right orthogonal of E or (Hilbert) orthogonal of E
$H^1(U)$	Sobolev space $W^{1,2}(U)$
$L^0(X, \mu; Y)$	space of ($\sim_\mu$-equivalence class of) Borel functions from X to Y
$L^p(X, \mu; Y)$	space of ($\sim_\mu$-equivalence class of) functions from X to Y p-summable with respect to μ
$x_n \nearrow x$	increasing convergence of x_n to x
$x_n \searrow x$	decreasing convergence of x_n to x
$x_n \to x$	convergence of x_n to x

Measure theory

δ_x	Dirac delta measure centered in x
$\frac{d\nu}{d\mu}$	Radon–Nikodym derivative of ν with respect to μ
$\mathcal{M}(X)$	set of Borel finite signed measures over X
$\mathcal{M}_+(X)$	set of Borel finite positive measures over X
$\mathcal{P}(X)$	set of Borel finite positive normalized measures over X
$\mathcal{R}(X)$	set of Radon measures over X, i.e. Borel measures inner regular and locally finite
$\mathcal{R}_+(X)$	set of positive Radon measures over X
$\mathscr{B}(X)$	Borel σ-algebra over X
$\nu = \varrho\mu$	$\nu \ll \mu$ and $\frac{d\nu}{d\mu} = \varrho$
$\nu \ll \mu$	ν is absolutely continuous with respect to μ
$\sim_\mu$	equivalence relation induced by the measure μ

Introduction

Let us put ourselves in the following situation: we have a certain quantity of mass distributed in a certain way. We are interested in moving it to a new configuration.

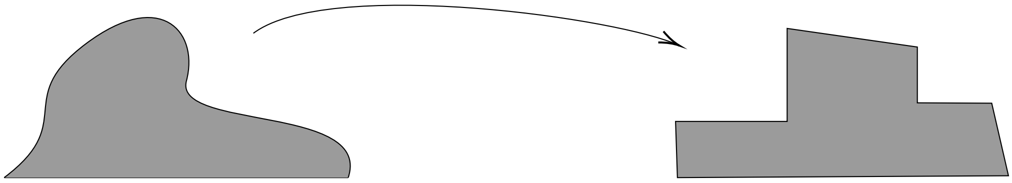

Figure 1: An intuitive representation of the optimal transport problem.

This requires a certain cost, so a question naturally arises, namely

What is the best way to transport the mass from the initial configuration
to the final one in order to minimize the cost of the transport?

The first to tackle this problem, in a mathematically rigorous way, was Gaspar Monge and today it is known as the *optimal transport problem.* The problem has been then reformulated by Leonid Kantorovich.

Let us now shift our attention to a different situation: a system of material points that move with Brownian motion.

Figure 2: Pollen in the air moving of Brownian motion (Unsplash, CC0 Public Domain).

Suppose we know the initial and the final configurations of the system in a certain time interval. What we might ask ourselves is

What is the most likely evolution of the system?

The previous question has been proposed, for the first time, by Erwin Schrödinger and leads to the second variational problem of our interest, the *Schrödinger problem*.

At first sight, the only similarity between the two problems is the fact that both are interpolation problems. We will see that a deeper connection can be established.

In the first three chapters, we focus on the abstract, compared with the sequel, theory regarding the two problems by presenting their fundamental characteristics and the first connections between them.

In **Chapter 1** we present the primal formulations of the two problems: with regard to optimal transport, we start from the historical definition given by Monge and then move on to the less problematic definition provided by Kantorovich for which, under appropriate hypotheses, we show the existence of solutions through the direct method of the Calculus of Variations; in a similar way, we proceed for the Schrödinger problem in order to be able to make a first comparison.

In **Chapter 2** we analyze two dual formulations of the problems induced by convexity: in particular, as regards the optimal transport problem, we present Kantorovich–Rubinstein duality; as for instead the Schrödinger problem, we analyze the representation induced by the variational formulation of the entropy functional.

With **Chapter 3** we close the first abstract part of the thesis presenting the initial properties of the Wasserstein space $(\mathcal{P}_2(X), W_2)$ and some of its geometrical aspects.

In the last three chapters, the setting becomes more tangible: in fact, we place ourselves in the Euclidean environment in order to study the most interesting variational representations of the two problems, focusing on their ties.

In **Chapter 4**, increasing the complexity in the formulation of the problems, we can show a dynamical representation, recovering the intuitive idea that underlies the two problems.

In **Chapter 5** we write both problems as fluid-dynamic problems under appropriate constraints using the Benamou–Brenier formulas.

Chapter 6 is where we take stock by studying the representations of problems through semigroups: as regards optimal transport, we determine explicit representations for Kantorovich potentials through the Hopf–Lax semigroup; regarding the Schrödinger problem, instead, we manage to provide the dual representation induced by the Hopf–Cole semigroup that allows us to connect to the Kantorovich–Rubinstein duality introduced in Chapter 2.

Gaspard Monge (1746–1818)

Gaspard Monge was born on May 9, 1746, in Beaune (France).

Figure 3: Jean Naigeon, *Portrait de Monge*, oil on canvas, 1811, Musée des Beaux-Arts, Beaune, France.

At the age of 16, he went to Lyon to attend the Collège de la Trinité where, only seventeen years old, he would already be assigned to teach a physics course. He completed his education in 1764 and then returned to Beaune where he drew up a plan of the city. For this work, the year later, he was appointed to the École Royale du Génie at Mézières as a draftsman. There, Monge got in touch with Charles Bossut. When the latter was elected to the Académie des Sciences in 1768, the next year Monge succeeded him as professor of mathematics. In 1771, Monge approached Condorcet who recommended to him to present memoirs to the Académie des Sciences about his works. Monge submitted four works on Calculus of Variations, Infinitesimal Geometry, Partial Differential Equations and Combinatorics. In 1780, he was elected as adjoint géomètre at the Académie des Sciences in Paris. On September 21, 1789, given his political alignment in favor of the Revolution, he was offered the post of Minister of the Navy, a position he carried out with little success. After a few months, he immediately returned to the Académie, which was then abolished on August 8, 1793. Nominated by the National Convention on March 11, 1794, to the body for establishing the École Centrale des Travaux Publics, which would soon become the École Polytechnique, he was appointed there as instructor in descriptive geometry on November 9, 1794. His lectures here were the origin of his book "Application de l'analyse à la géométrie", for which he is considered the father of differential geometry. Further information on his life can be found in [51].

He died on July 28, 1818, in Paris (France).

Leonid Kantorovich (1912–1986)

Leonid Vitalyevich Kantorovich was born on January 19, 1912, in St. Petersburg (Russia).

Figure 4: A young Leonid Kantorovich in 1939 (picture from [46]).

Only fourteen years old, he started his mathematical studies at the Leningrad State University, graduating in 1930 at 18 years old, having reached the level equivalent to a doctorate and continuing his research at the Mathematical Department of the Faculty of Physics and Mathematics of Leningrad State University: he would not formally receive a PhD until 1935 because of the abolition of doctoral degrees by the Soviet Union. In 1930, he was appointed as an assistant in the Naval Engineering School and then, the following year, as a research associate in his university. In 1932, he became associate professor in the Department of Numerical Mathematics and from 1934 he was a professor. His young age was, in a certain sense, a problem; for example, his first lecture became famous because many students shouted at him to sit down and wait for the professor to arrive like everyone else: too bad the professor was him. In 1933 he published with Vladimir Ivanovich Krylov the book *Calculus of Variations*, the first of his more than 300 contributions to mathematics, economics and computer science. During his career, he won numerous prizes in both mathematics and economics, including the 1975 Nobel Prize in Economics for his contribution to the optimal transport theory. In the 1980s, Kantorovich himself suggested a way to divide up all his contributions, proposing nine distinct areas: descriptive function theory and set theory, constructive function theory, approximate methods of analysis; functional analysis, functional analysis and applied mathematics, linear programming, hardware and software, optimal planning and optimal prices, the economic problems of a planned economy. Further information on his life can be found in [52].

He died from cancer on April 7, 1986, in Moscow (USSR).

Erwin Schrödinger (1887–1961)

Erwin Rudolf Josef Alexander Schrödinger was born on August 12, 1887, in Erdberg (Austria).

Figure 5: Erwin Schrödinger (picture from [50]).

Up to the age of ten he has been home-schooled by a private tutor and in the autumn of 1898 he entered the Akademisches Gymnasium, later with respect to usual because of a long holiday in England. After the graduation in 1906, he entered the University of Vienna. On May 20, 1910, Schrödinger was awarded his doctorate for the dissertation *On the conduction of electricity on the surface of insulators in moist air* and then undertook voluntary military service. He participated in World War I and, in the spring of 1917, he was sent back to Vienna and he started to teach a meteorology course. In 1926, Schrödinger published his revolutionary work about the general theory of relativity and wave mechanics. With his wife's knowledge, with whom he never had a good relationship, he had many lovers, including the wife of his colleague Arthur March, Hilde. When Alexander Lindemann, head of physics at Oxford University, visited Germany in the spring of 1933, Schrödinger asked him for a position in England for him and Arthur because he decided he could not live in a country in which the persecution of Jews had become national policy. In summer 1933 Hilde became pregnant with Schrödinger's child and on November 4, 1933, Schrödinger, his wife and Hilde arrived in Oxford where, after a while, he discovered he had been awarded the Nobel Prize. From this time, Schrödinger openly had two wives. He went back to Austria and spent the years 1936–1938 in Graz but on August 26, 1938, the Nazis dismissed him for political unreliability: it was the consequence of his decision of 1933. After a year in Gent, he went to Dublin where he remained until he retired in 1956 when he returned to Vienna. Further information on his life can be found in [50].

He died on January 4, 1961, in Vienna (Austria).

Preliminaries

In this initial part, we describe some general topics and results that will be used in the subsequent chapters.

1 inf-convolution of functions

(P.1) Definition *Consider a metric space* (X,d) *and a function* $f : X \to [0,+\infty[$*. We call* inf-convolution of f *the sequence* $(\mathcal{I}_h f)$ *such that*

$$\mathcal{I}_h f(x) = \inf_{y\in X} \left\{\min\left\{f(y), h\right\} + hd(x,y)\right\}.$$

The main properties of the inf-convolution operator are recalled in the following Proposition.

(P.2) Proposition *Consider a metric space* (X,d) *and a function* $f : X \to [0,+\infty[$*. The following facts hold true:*

(a) $(\mathcal{I}_h f)$ *is an increasing sequence,*

(b) *for every* $h \in \mathbb{N}$ *and* $x \in X$

$$\mathcal{I}_h f(x) \geq 0,$$

(c) *for every* $h \in \mathbb{N}$ *and* $x \in X$

$$\mathcal{I}_h f(x) \leq f(x), \qquad\qquad \mathcal{I}_h f(x) \leq h,$$

(d) *for every* $h \in \mathbb{N}$ *the function* $\mathcal{I}_h f \in \mathrm{Lip}_b(X)$*,*

(e) *if* f *is lower semicontinuous, then* $\mathcal{I}_h f \nearrow f$ *pointwise.*

Proof.

(a) Consider $h, k \in \mathbb{N}$ such that $h \leq k$. For every $x \in X$ one has

$$\mathcal{I}_h f(x) = \inf_{y \in X} \{\min\{f(y), h\} + hd(x,y)\} \leq \inf_{y \in X} \{\min\{f(y), k\} + kd(x,y)\} = \mathcal{I}_k f(x),$$

so $(\mathcal{I}_h f)$ is increasing.

(b) It follows directly from the fact that distance is always positive.

(c) It is a straightforward computation: for every $h \in \mathbb{N}$ and $x \in X$

$$\mathcal{I}_h f(x) = \inf_{y \in X} \{\min\{f(y), h\} + hd(x,y)\} \leq \min\{f(x), h\} + hd(x,x) = \min\{f(x), h\}.$$

(d) For every $h \in \mathbb{N}$, fixed $y \in X$, the function $\{x \longmapsto \min\{f(y), h\} + hd(x,y)\}$ is Lipschitz, so $\mathcal{I}_h f$, that is the pointwise infimum with respect to y, is also Lipschitz. Boundedness comes from the second inequality in (c).

(e) First of all, from (a) and the first inequality in (c), for every $x \in X$

$$\lim_h \mathcal{I}_h f(x) = \sup_h \mathcal{I}_h f(x) \leq f(x), \tag{P.3}$$

then $\sup_h \mathcal{I}_h f(x) < +\infty$, since f is finite valued. Now, let us consider a minimizing sequence (y_h) such that

$$\min\{f(y_h), h\} + hd(x, y_h) \leq \mathcal{I}_h f(x) + \frac{1}{h+1}.$$

In particular, since $\min\{f(y_h), h\} \geq 0$,

$$hd(x, y_h) \leq \sup_h \mathcal{I}_h f(x) + \frac{1}{h+1}.$$

Passing to the limit as $h \to +\infty$, the only possibility is $d(x, y_h) \to 0$ or, in other words, $y_h \to x$. Now, since $hd(x, y_h) \geq 0$,

$$\min\{f(y_h), h\} \leq \mathcal{I}_h f(x) + \frac{1}{h+1}$$

and passing to the $\liminf$ as $h \to +\infty$,

$$\liminf_h f(y_h) \leq \sup_h \mathcal{I}_h f(x).$$

Then, by lower semicontinuity of f,

$$f(x) \le \mathcal{I}_h f(x). \tag{P.4}$$

Combining (P.3) and (P.4), we obtain

$$\lim_h \mathcal{I}_h f(x) = \sup_h \mathcal{I}_h f(x) = f(x),$$

as desired. ∎

2 Push forward of measures

(P.5) Definition *Consider two metric spaces $(X, d_X), (Y, d_Y)$ and a Borel function $f : X \to Y$. We call* push forward operator *of f the function $f_\# : \mathcal{R}(X) \to \mathcal{R}(Y)$ such that for all $B \in \mathscr{B}(Y)$*

$$f_\#\mu(B) = \mu(f^{-1}(B)).$$

In particular, $f_\#\mu$ is called push forward measure*.*

Clearly, $(g \circ f)_\# = g_\# \circ f_\#$.

(P.6) Proposition (change of variables formula) *Let $(X, d_X), (Y, d_Y)$ be two metric spaces and $f : X \to Y$, $\varphi : Y \to \mathbb{R}$ Borel functions. If $\mu \in \mathcal{R}(X)$, then*

$$\int_Y \varphi df_\#\mu = \int_X (\varphi \circ f)\, d\mu.$$

In particular, a function $\psi : X \to \overline{\mathbb{R}}$ is $f_\#\mu$-integrable if and only if $\psi \circ f$ is μ-integrable.

Proof. Let us suppose initially $\varphi = \chi_B$ for some $B \in \mathscr{B}(Y)$. Then

$$\begin{aligned}\int_Y \chi_B df_\#\mu &= \int_B df_\#\mu = f_\#\mu(B) = \mu(f^{-1}(B)) = \\ &= \int_{f^{-1}(B)} d\mu = \int_{f^{-1}(B)} 1\, d\mu + \int_{X\setminus f^{-1}(B)} 0\, d\mu = \\ &= \int_X \chi_B(f(x))d\mu(x) = \int_X (\chi_B \circ f)d\mu.\end{aligned}$$

By linearity, the same holds true for every $\mathscr{B}(X)$-simple function $\varphi : Y \to [0, +\infty]$. If $\varphi : Y \to [0, +\infty]$ is Borel, the same holds by the monotone convergence Theorem. The general case and the second part of the statement follow considering the positive and negative parts of the integrands on both sides. ∎

(P.7) Example *Let $(X, d_X), (Y, d_Y)$ be two metric spaces and $\pi \in \mathcal{R}(X \times Y)$. If $p_X : X \times Y \to X$ and $\mu = (p_X)_\# \pi$, then for every Borel function $\varphi : X \to \mathbb{R}$*

$$\int_X \varphi d\mu = \int_X \varphi d(p_X)_\# \pi = \int_{X\times Y} (\varphi \circ p_X) d\pi = \int_{X\times Y} \varphi d\pi.$$

In particular, using the change of variables formula, we can exhibit a quick way to check the condition $f_\# \mu = \nu$ as we can see in the following Proposition.

(P.8) Proposition *Let $(X, d_X), (Y, d_Y)$ be two metric spaces, $f : X \to Y$ a Borel function , $\mu \in \mathcal{R}(X)$ and $\nu \in \mathcal{R}(Y)$. The following facts are equivalent:*

(a) $f_\# \mu = \nu$,

(b) for all $\varphi \in C_b(Y)$, one has

$$\int_Y \varphi d\nu = \int_X (\varphi \circ f)\, d\mu. \tag{P.9}$$

Proof.
$(a) \Longrightarrow (b)$ It is a direct consequence of Proposition (P.6).
$(b) \Longrightarrow (a)$ Let us start by noticing that

$$\left\{B \in \mathscr{B}(Y) : \mu(f^{-1}(B)) = \nu(B)\right\}$$

is a σ-algebra. As a consequence, the condition $f_\# \mu = \nu$ needs only to be checked on a family of generators of $\mathscr{B}(Y)$. In particular, we only need to check $f_\# \mu = \nu$ on the open subsets of Y. Let, therefore, $A \subseteq Y$ be open and for every $n \in \mathbb{N}$ consider

$$A_n = \left\{x \in A : \operatorname{dist}(x, \partial A) \geq \frac{1}{n+1}\right\}$$

and $\varphi_n : Y \to \mathbb{R}$ in $C_b(Y)$ such that

$$\varphi_n(x) = (1 - (n+1)\operatorname{dist}(x, A_n))^+ .$$

Since $\varphi_n \nearrow \chi_A$, using the monotone convergence Theorem, (P.9) is also true for $\varphi = \chi_A$. In particular,

$$f_\# \mu(A) = \mu(f^{-1}(A)) = \int_{f^{-1}(A)} d\mu = \int_X (\chi_A \circ f)\, d\mu = \int_Y \chi_A d\nu = \nu(A). \blacksquare$$

(P.10) Lemma *Let* $(X,\mathcal{U}),(Y,\mathcal{V})$ *be two measurable spaces,* $e : X \to Y$ *a* $(\mathcal{U},\mathcal{V})$*-measurable function,* $\mu \in \mathcal{M}_+(X)$ *and* $v \in L^p(X,\mu;\mathbb{R}^k)$ *with* $p \in]1,+\infty[$*. Then* $e_\#(v\mu) \ll e_\#\mu$ *with density* $w \in L^p(Y,e_\#\mu;\mathbb{R}^k)$ *such that*

$$\|w\|_{L^p(Y,e_\#\mu;\mathbb{R}^k)} \leq \|v\|_{L^p(X,\mu;\mathbb{R}^k)}.$$

Proof. Consider $\varphi : Y \to \mathbb{R}^k$ bounded and measurable. By Hölder's inequality,

$$\left|\int_Y \varphi de_\#(v\mu)\right| = \left|\int_X (\varphi \circ e)(x) d(v\mu)(x)\right| = \left|\int_X (\varphi \circ e) \cdot v d\mu\right| \leq \|v\|_{L^p(X,\mu;\mathbb{R}^k)} \|\varphi\|_{L^{p'}(Y,e_\#\mu;\mathbb{R}^k)}$$

so the linear operator

$$\varphi \mapsto \int_Y \varphi de_\#(v\mu)$$

is continuous with respect to the $L^{p'}(Y,e_\#\mu;\mathbb{R}^k)$ norm. Recalling the density of the set of bounded and measurable functions in $L^{p'}(Y,e_\#\mu;\mathbb{R}^k)$, we can easily extend the previous operator to $L^{p'}(Y,e_\#\mu;\mathbb{R}^k)$, then, by duality, it is representable by some $w \in L^p(Y,e_\#\mu;\mathbb{R}^k)$ such that

$$\|w\|_{L^p(Y,e_\#\mu;\mathbb{R}^k)} \leq \|v\|_{L^p(X,\mu;\mathbb{R}^k)}.$$

The fact that w is the density of $e_\#(v_\mu)$ with respect to $e_\#\mu$ follows by construction. ∎

3 Polish spaces and convergence of measures

(P.11) Definition *Given a measurable space* $(X,\mathcal{U})$ *and a measure* μ *on* $(X,\mathcal{U})$*, we call* $A \in \mathcal{U}$ *an* atom*, if* $\mu(A) > 0$ *and for any* $B \in \mathcal{U}$ *such that* $B \subseteq A$ *one has*

$$0 \in \{\mu(B), \mu(A \setminus B)\}.$$

In particular, we call a measure atomic *if it admits at least an atom and* non-atomic *if it has no atom.*

To better comprehend the previous definition, let us build some concrete examples of atomic and non-atomic measures.

(P.12) Example *Consider* $\mathbb{R}^n$ *endowed with the Euclidean distance. The Lebesgue measure* $\mathcal{L}^n$ *is non-atomic and the Dirac delta measure* δ_0 *is atomic. In particular, the measure* $\mu = \mathcal{L}^n + \delta_0$ *is atomic.*

(P.13) Definition *Given a metric space* (X,d) *and* $\mu \in \mathcal{R}(X)$*, we call* support of μ *the set*

$$\operatorname{supt}(\mu) = \{x \in X : |\mu|(U) > 0 \text{ for all } U \subseteq X \text{ such that } U \text{ is open and } x \in U\}.$$

The previous set is closed and, if (X,d) is separable, one can prove that $\mu \in \mathcal{R}(X)$ is concentrated on $\operatorname{supt}(\mu)$.

Let us now introduce the fundamental abstract setting in which we will work.

(P.14) Definition *Given a topological space* (X,τ)*, we call it* Polish space *if there exists a distance* d *on* X *such that* d *induces* τ *and* (X,d) *is complete and separable.*

In the following, we will always fix a distance for each Polish space, so we will think of a Polish space (X,τ) as a complete and separable metric space (X,d). One can prove that every open subset of a Polish space is also Polish (when equipped with the induced distance): see [1] for the technical details.

(P.15) Lemma (Ulam) *If* (X,d) *is a Polish space, then for every* $\mu \in \mathcal{M}_+(X)$ *and for every* $\varepsilon > 0$ *there exists* $K \subseteq X$ *compact such that* $\mu(X \setminus K) < \varepsilon$*.*

Proof. See [1, Lemma 1.5] ■

For the sake of completeness, let us recall the following fundamental result.

(P.16) Theorem (disintegration) *Let* $(Z,d_Z),(X,d_X)$ *be Polish spaces,* $\mu \in \mathcal{P}(Z)$*,* $\pi : Z \to X$ *a Borel function and* $\nu = \pi_\# \mu \in \mathcal{P}(X)$*. There exists a* ν*-a.e. uniquely determined Borel family* $\{\mu_x\}_{x\in X} \subseteq \mathcal{P}(Z)$ *such that*

$$\mu_x(Z \setminus \pi^{-1}(x)) = 0 \text{ for } \nu\text{-a.e. } x \in X$$

and for every Borel function $f : Z \to [0,+\infty]$

$$\int_Z f(z)d\mu(z) = \int_X \left(\int_{\pi^{-1}(x)} f(z)d\mu_x(z) \right) d\nu(x).$$

Proof. See [19, Chapter III, pp. 78–81]. ■

Despite the tremendous power of the disintegration Theorem, what we will mostly use is the following Corollary.

(P.17) Corollary *Let $(X_1, d_1), (X_2, d_2)$ be Polish spaces, $\mu \in \mathcal{P}(X_1 \times X_2)$ and $\nu = p^1_\# \mu \in \mathcal{P}(X_1)$. There exists a ν-a.e. uniquely determined Borel family $\{\mu_{x_1}\}_{x_1 \in X_1} \subseteq \mathcal{P}(X_2)$ such that for every Borel function $f : X_1 \times X_2 \to [0, +\infty]$*

$$\int_{X_1 \times X_2} f(x_1, x_2) d\mu(x_1, x_2) = \int_{X_1} \left(\int_{X_2} f(x_1, x_2) d\mu_{x_1}(x_2) \right) d\nu(x_1).$$

Proof. It is a direct consequence of the identification of $(p^1)^{-1}(x_1)$ with X_2 and the disintegration Theorem. ■

We can now pass to introduce the notion of weak convergence of measures: it is a key choice for the direct method of the Calculus of Variations and also for the geometric analysis of Wasserstein spaces.

(P.18) Definition *Let (X, d) be a metric space and (μ_n) in $\mathcal{R}(X)$. Given $\mu \in \mathcal{R}(X)$, we say that (μ_n)* weakly converge *to μ, and we denote this fact with $\mu_n \rightharpoonup \mu$, if for every $\varphi \in C_b(X)$*

$$\lim_n \int_X \varphi d\mu_n = \int_X \varphi d\mu.$$

In the following Lemmas, which constitute a proper subset of the Portmanteau Lemma, we explore equivalent definitions for weak convergence of measures in $\mathcal{P}(X)$.

(P.19) Lemma *Let (μ_n) in $\mathcal{P}(X)$ and $\mu \in \mathcal{P}(X)$. The following facts are equivalent:*

(a) $\mu_n \rightharpoonup \mu$ in $\mathcal{P}(X)$,

(b) for all $f \in \mathrm{Lip}_b(X)$ one has

$$\int_X f d\mu_n \to \int_X f d\mu.$$

Proof.
$(a) \implies (b)$ Obvious.
$(b) \implies (a)$ Let $f \in C_b(X)$ and consider $(\mathcal{I}_h f)$ its inf-convolution. We already know that $(\mathcal{I}_h f)$ is a bounded from below increasing sequence of Lipschitz functions and $\mathcal{I}_h f \nearrow f$ pointwise. Consider also $(\mathcal{I}_h(-f))$, a bounded from below and increasing sequence of Lipschitz functions such that $\mathcal{I}_h(-f) \nearrow -f$ pointwise. Up to a change in the sign, we obtain a bounded from above and decreasing sequence $(-\mathcal{I}_h(-f))$ of Lipschitz functions such that $-\mathcal{I}_h(-f) \searrow f$ pointwise.

Now, for every $h \in \mathbb{N}$ one has

$$\liminf_n \int_X f d\mu_n \geq \liminf_n \int_X \mathcal{I}_h f d\mu_n = \int_X \mathcal{I}_h f d\mu$$

so passing to the limit as $h \to +\infty$, by the monotone convergence Theorem, it results

$$\liminf_n \int_X f d\mu_n \geq \int_X f d\mu.$$

Analogously, using $(-\mathcal{I}_h(-f))$ we have

$$\limsup_n \int_X f d\mu_n \leq \int_X f d\mu$$

and the proof is complete. ■

(P.20) Lemma *Let* (μ_n) *in* $\mathcal{P}(X)$ *and* $\mu \in \mathcal{P}(X)$*. The following facts are equivalent:*

(a) $\mu_n \rightharpoonup \mu$ *in* $\mathcal{P}(X)$*,*

(b) *for all* $f : X \to \mathbb{R}$ *lower semicontinuous and lower bounded one has*

$$\liminf_n \int_X f d\mu_n \geq \int_X f d\mu.$$

Proof. It is similar to the proof of Lemma (P.19). ■

(P.21) Lemma *Let* (μ_n) *in* $\mathcal{P}(X)$ *and* $\mu \in \mathcal{P}(X)$*. The following facts are equivalent:*

(a) $\mu_n \rightharpoonup \mu$ *in* $\mathcal{P}(X)$*,*

(b) *for every* $A \subseteq X$ *open, one has*

$$\liminf_n \mu_n(A) \geq \mu(A),$$

(c) *for every* $C \subseteq X$ *closed, one has*

$$\limsup_n \mu_n(C) \leq \mu(C).$$

Proof.

$(a) \implies (b)$ Take χ_A and consider $(\mathcal{I}_h\chi_A)$ its inf-convolution. Using the fact that χ_A is lower semicontinuous because A is open, we can observe that $(\mathcal{I}_h\chi_A)$ is a bounded and increasing sequence of Lipschitz functions such that $\mathcal{I}_h\chi_A \nearrow \chi_A$ pointwise. Now,

$$\liminf_n \mu_n(A) = \liminf_n \int_X \chi_A d\mu_n \geq \liminf_n \int_X \mathcal{I}_h\chi_A d\mu_n,$$

and using Lemma (P.19),

$$\liminf_n \mu_n(A) \geq \int_X \mathcal{I}_h\chi_A d\mu.$$

Passing to the limit as $h \to +\infty$, by the monotone convergence Theorem,

$$\liminf_n \mu_n(A) \geq \int_X \chi_A d\mu = \mu(A).$$

$(b) \iff (c)$ By the fact that μ_n is a finite measure,

$$\mu_n(C) = \mu_n(X) - \mu_n(X \setminus C).$$

Now,

$$\limsup_n \mu_n(C) \leq \limsup_n \mu_n(X) - \liminf_n \mu_n(X \setminus C),$$

but $\mu_n(X) = \mu(X) = 1$ and using (b) we obtain

$$\limsup_n \mu_n(C) \leq \mu(X) - \mu(X \setminus C) = \mu(C).$$

The converse implication is similar.

$(b) \implies (a)$ For every $f \in C_b(X)$ positive, Fubini–Tonelli's Theorem implies

$$\int_X f d\mu_n = \int_0^{\sup f} \mu_n(\{f > t\}) dt.$$

Let us prove that for all $t > 0$ such that $\mu(\{f = t\}) = 0$, one has

$$\lim_n \mu_n(\{f > t\}) = \mu(\{f > t\}).$$

Using (b), we can first of all write

$$\mu(\{f > t\}) \leq \liminf_n \mu_n(\{f > t\}).$$

Now,

$$\mu(\{f > t\}) = \mu(\{f > t\}) + \mu(\{f = t\}) = \mu(\{f \geq t\})$$

and using (c) we can affirm that

$$\mu(\{f > t\}) \leq \liminf_n \mu_n(\{f > t\}) \leq \limsup_n \mu_n(\{f > t\}) \leq$$
$$\leq \limsup_n \mu_n(\{f \geq t\}) \leq \mu(\{f \geq t\}) = \mu(\{f > t\})$$

so

$$\lim_n \mu_n(\{f > t\}) = \mu(\{f > t\}).$$

Using this fact and the dominated convergence Theorem,

$$\lim_n \int_X f d\mu_n = \lim_n \int_0^{\sup f} \mu_n(\{f > t\}) dt = \int_X f d\mu.$$

The general case follows by splitting the positive and negative parts of a function $f \in C_b(X)$. ■

Thanks to the notion of weak convergence, we can define a topology on $\mathcal{R}(X)$: it is sufficient to define closed sets as the set containing the limits of every one of their sequences. We will refer to this construction as *weak topology*. Unless otherwise specified, we will always consider $\mathcal{R}(X)$ endowed with the weak topology. The following is a geometrical analysis result about $\mathcal{P}(X)$: the proof is based on Riesz's Theorem and Banach–Alaoglu–Bourbaki's Theorem, for which we refer to [9].

(P.22) Theorem *If (X, d) is a compact metric space, then $\mathcal{P}(X)$ is weakly compact.*

Proof. By [9, Theorem 4.31], $\mathcal{R}(X)$, endowed with total variation norm, is isometric to its dual, $C(X)$. In this sense, weak convergence induces the weak* topology so, by [9, Theorem 3.16], being $\mathcal{P}(X)$ the closed unit ball of $\mathcal{R}(X)$, the thesis follows. ■

In order to show a compactness criterion in $\mathcal{M}_+(X)$, we introduce the notion of tightness.

(P.23) Definition *Let (X, d) be a Polish space and $\mathcal{F} \subseteq \mathcal{M}_+(X)$ such that $\sup_{\mu \in \mathcal{F}} \mu(X) < +\infty$. We call $\mathcal{F}$* tight *if for every $\varepsilon > 0$ there exists $K \subseteq X$ compact such that for all $\mu \in \mathcal{F}$ one has $\mu(X \setminus K) < \varepsilon$.*

(P.24) Theorem (Prokhorov) *Let (X, d) be a Polish space and $\mathcal{F} \subseteq \mathcal{M}_+(X)$ such that*

$$\sup_{\mu \in \mathcal{F}} \mu(X) < +\infty.$$

Then $\overline{\mathcal{F}}$ is weakly compact if and only if $\mathcal{F}$ is tight.

Proof. Without loss of generality, we can only consider the case $\mathcal{F} \subseteq \mathcal{P}(X)$. Indeed, given

$$\mathcal{F}' = \left\{ \mu' \in \mathcal{P}(X) : \forall\, E \in \mathscr{B}(X)\ \mu'(E) = \frac{\mu(E)}{\mu(X)},\ \mu \in \mathcal{F} \right\},$$

suppose $\overline{\mathcal{F}'}$ is weakly compact. Given (μ_n) in $\mathcal{F}$, consider (μ'_n) in $\mathcal{F}'$. By weak compactness, there exist (μ'_{n_j}) in $\mathcal{F}'$ and $\mu' \in \overline{\mathcal{F}'}$ such that $\mu'_{n_j} \rightharpoonup \mu'$ in $\overline{\mathcal{F}'}$. In particular, $\mu_{n_j} \rightharpoonup \mu$ in $\overline{\mathcal{F}}$ and so $\mathcal{F}$ is weakly compact.

Suppose that $\mathcal{F} \subseteq \mathcal{P}(X)$ is tight. By definition, we can find an increasing sequence (K_h) of compact subsets of X such that

$$\lim_h \sup_{\mu \in \mathcal{F}} \mu(X \setminus K_h) = 0.$$

Let (μ_n) be in $\mathcal{F}$. For every $h \in \mathbb{N}$, if we consider $(\mu_n \lfloor K_h)_n$ in $\mathcal{P}(K_h)$, weakly compact thanks to the compactness of K_h, then there exist $(\mu_{n_j} \lfloor K_h)_j$ in $\mathcal{P}(K_h)$ and $\nu_h \in \mathcal{P}(K_h)$ such that $\mu_{n_j} \lfloor K_h \rightharpoonup \nu_k$ in $\mathcal{P}(K_h)$. Up to defining, for $A \in \mathscr{B}(X)$, $\nu_h(A) = 0$ if $A \cap K_h = \emptyset$, we can view ν_h as a measure in $\mathcal{P}(X)$ with support in K_h and we have that $\mu_{n_j} \lfloor K_h \rightharpoonup \nu_k$ in $\mathcal{P}(X)$. In particular, by the fact that (K_h) is an increasing sequence, $\nu_k \leq \nu_{k+1}$.

Now,

$$1 - \sup_{\mu \in \mathcal{F}} \mu(X \setminus K_h) \leq \mu_{n_j}(Z) - \mu_{n_j}(X \setminus K_h) = \mu_{n_j}(Z \cap K_h) = \mu_{n_j} \lfloor K_h(X) \leq 1,$$

then passing to the limit as $j \to +\infty$ we obtain

$$1 - \sup_{\mu \in \mathcal{F}} \mu(X \setminus K_h) \leq \nu_k(X) \leq 1,$$

therefore, observing that additivity comes from the monotonicity of (ν_h),

$$\nu = \sup_{h \in \mathbb{N}} \nu_h \in \mathcal{P}(X).$$

Now, given $\varphi \in C_b(X)$, one has

$$\left| \int_X \varphi d(\mu_{n_j} - \nu) \right| \leq \left| \int_X \varphi d(\mu_{n_j} - \mu_{n_j} \lfloor K_h) \right| + \left| \int_X \varphi d(\mu_{n_j} \lfloor K_h - \nu_h) \right| + \left| \int_X \varphi d(\nu_h - \nu) \right| \leq$$
$$\leq \sup \varphi \sup_{\mu \in \mathcal{F}} \mu(X \setminus K_h) + + \left| \int_X \varphi d(\mu_{n_j} \lfloor K_h - \nu_h) \right| + \left| \int_X \varphi d(\nu_h - \nu) \right|$$

and taking $j, h \to +\infty$

$$\lim_j \left| \int_X \varphi d(\mu_{n_j} - \nu) \right| = 0,$$

then $\overline{\mathcal{F}}$ is weakly compact.

Suppose now that $\overline{\mathcal{F}}$ is weakly compact. Fixed $\varepsilon > 0$ and (x_i) in X such that

$$\overline{\{x_i : i \in \mathbb{N}\}} = X,$$

we want to prove that for any $j \in \mathbb{N}$ there exists $k_j \in \mathbb{N}$ such that for all $\mu \in \mathcal{F}$,

$$\mu \left(X \setminus \bigcup_{i=0}^{k_j} \mathrm{B} \left(x_i, \frac{1}{j+1} \right) \right) \leq 2^{-j} \varepsilon.$$

Suppose the claim is false, then there exists $j_0 \in \mathbb{N}$ such that for any $k \in \mathbb{N}$ there exists $\mu_k \in \mathcal{F}$ such that

$$\mu_k \left(X \setminus \bigcup_{i=0}^{k} \mathrm{B} \left(x_i, \frac{1}{j_0+1} \right) \right) > 2^{-j_0} \varepsilon.$$

By weak compactness of $\overline{\mathcal{F}}$, there exist (μ_{k_n}) in $\mathcal{P}(X)$ and $\mu \in \mathcal{P}(X)$ such that $\mu_{k_n} \rightharpoonup \mu$. If $n \to +\infty$, one has

$$\mu \left(X \setminus \bigcup_{i=0}^{k} \mathrm{B} \left(x_i, \frac{1}{j_0+1} \right) \right) \geq 2^{-j_0} \varepsilon,$$

but if $k \to +\infty$, we obtain $0 \geq 2^{-j_0}\varepsilon$, in contradiction with $\varepsilon > 0$.

The claim proved implies that $\sup\limits_{\mu \in \mathcal{F}} \mu(X \setminus K) \leq \varepsilon$, where

$$K = \bigcap_{j=0}^{\infty} \overline{\bigcup_{i=0}^{k_j} \mathrm{B} \left(x_i, \frac{1}{j+1} \right)},$$

so $\mathcal{F}$ is tight. ∎

4 Convex Analysis: a toolbox

In the following section, unless otherwise specified, we will assume that X and Y are two sets and $c : X \times Y \to \mathbb{R}$ is a Borel function.

(P.25) Definition *We call* $\mathfrak{G} \subseteq X \times Y$ c-cyclically monotone *if for every* $n \in \mathbb{N} \setminus \{0\}$, $(x_1, y_1), \ldots, (x_n, y_n) \in \mathfrak{G}$ *and every permutation* σ

$$\sum_{i=1}^{n} c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^{n} c(x_i, y_i).$$

(P.26) Definition *Given $\varphi : X \to [-\infty, +\infty[$, we call* c-conjugate of φ, *the function*

$$\varphi^c(y) = \inf_{x \in X} \{c(x, y) - \varphi(x)\} .$$

Analogously, if $\psi : Y \to [-\infty, +\infty[$, we call c-conjugate of ψ, *the function*

$$\psi^c(x) = \inf_{y \in Y} \{c(x, y) - \varphi(y)\} .$$

(P.27) Definition *We call $f : X \times Y \to \mathbb{R}$* c-affine *if it is of the form $c(\cdot, y) + \alpha$ or $c(x, \cdot) + \beta$ for some $x \in X, y \in Y, \alpha, \beta \in \mathbb{R}$. Moreover, we call a function $\varphi : X \to [-\infty, +\infty[$* c-concave *, if it is the infimum of a family of c-affine functions of the form $c(\cdot, y) + \alpha$. Analogously, we call a function $\psi : Y \to [-\infty, +\infty[$* c-concave, *if it is the infimum of a family of c-affine functions of the form $c(x, \cdot) + \beta$.*

One can prove that $\varphi : X \to [-\infty, +\infty[$ is c-concave if and only if it is the c-conjugate of a function ψ.

(P.28) Proposition *Consider (X, d) a metric space and a function $\varphi : X \to \mathbb{R}$. The following facts are equivalent:*

(a) φ is d-concave,

(b) $\varphi \in \mathrm{Lip}(X)$ and $\mathrm{Lip}(\varphi) \leq 1$.

In particular, in this case $\varphi^c = -\varphi$.

Proof.
$(a) \implies (b)$ Let φ be d-concave. Then there exists $\psi : X \to \mathbb{R} \cup \{+\infty\}$ such that for every $x \in X$

$$\varphi(x) = \psi^c(x) = \inf_{y \in X} \{d(x, y) - \psi(y)\} .$$

Fixed $y \in X$, for every $x, z \in X$, by the triangle inequality,

$$d(x, y) - \psi(y) - (d(z, y) - \psi(y)) = d(x, y) - d(z, y) \leq d(x, z)$$

then, the function $\{x \mapsto d(x, y) - \psi(y)\} \in \mathrm{Lip}(X)$ has a Lipschitz constant less than or equal to 1. Being φ the infimum of the family above, it follows that $\varphi \in \mathrm{Lip}(X)$ and $\mathrm{Lip}(\varphi) \leq 1$.
$(b) \implies (a)$ Let $\varphi \in \mathrm{Lip}(X)$ and $\mathrm{Lip}(\varphi) \leq 1$. Let us prove that for every $x \in X$

$$\varphi(x) = \inf_{y \in X} \{d(x, y) + \varphi(y)\} . \tag{P.29}$$

Clearly,

$$\inf_{y\in X} \{d(x,y)+\varphi(y)\} \leq \varphi(x).$$

On the other hand, $|\varphi(x)-\varphi(y)| \leq d(x,y)$ so in particular $\varphi(x)-\varphi(y) \leq d(x,y)$ then

$$\varphi(x) \leq d(x,y)+\varphi(y)$$

and passing to the infimum for $y \in X$ we obtain

$$\varphi(x) \leq \inf_{y\in X} \{d(x,y)+\varphi(y)\}.$$

Observing that (P.29) corresponds to $\varphi = (-\varphi)^c$ and applying this fact to $-\varphi$ we obtain $\varphi^c = -\varphi$. ■

(P.30) Theorem *If $\varphi : X \to [-\infty,+\infty[$ and $\varphi \neq -\infty$, then $(\varphi^c)^c \geq \varphi$.*

In particular, the equality holds if and only if φ is c-concave.

Proof. See [1, Theorem 3.14]. ■

(P.31) Definition *Given $\varphi : X \to [-\infty,+\infty[$ and $x \in X$ such that $\varphi(x) > -\infty$ we call c-*subdifferential of *φ* in *x the set*

$$\partial^c\varphi(x) = \{y \in Y : \varphi(\xi) \leq \varphi(x) - c(x,y) + c(\xi,y), \forall \xi \in X\}.$$

One can prove that $\varphi(x)+\varphi^c(y) = c(x,y)$ if and only if $y \in \partial^c\varphi(x)$, if and only if $x \in \partial^c\varphi(y)$.

(P.32) Theorem *Let $\mathfrak{G} \subseteq X \times Y$ be c-cyclically monotone. Then there exists a c-concave function $\varphi : X \to [-\infty,+\infty[$ such that $\mathfrak{G} \subseteq$* graph*$(\partial^c\varphi)$.*

Proof. Fixed $(x_0,y_0) \in \mathfrak{G}$, the function

$$\varphi(x) = \inf_{\substack{n\in\mathbb{N}\setminus\{0\}\\ (x_1,y_1),\dots,(x_n,y_n)\in \operatorname{supt}(\pi)}} (c(x,y_n)-c(x_n,y_n)+c(x_n,y_{n-1})-c(x_{n-1},y_{n-1})+\cdots+$$
$$+c(x_1,y_0)-c(x_0,y_0))$$

satisfies the required condition and also $\varphi(x_0)=0$. See [1, Theorem 3.18] for further details. ■

5 Absolutely continuous curves and geodesics

In the following section, unless otherwise specified, we will consider a metric space (X, d). We are interested in defining geodesics, so we only consider curves that are not loops, namely functions $\gamma : [a, b] \to X$ such that $\gamma(a) \neq \gamma(b)$. Let us start from the most general container: absolutely continuous curves.

(P.33) Definition *Let $a, b \in \mathbb{R}$. We call $\gamma : [a, b] \to X$ an* absolutely continuous curve*, and we write $\gamma \in AC([a, b]; X)$, if there exists $g \in L^1(a, b)$ such that for all $x, y \in [a, b]$ such that $x \leq y$ one has*

$$d(\gamma(y), \gamma(x)) \leq \int_x^y g(t)dt.$$

If $g \in L^p(a, b)$ for some $p \in]1, +\infty]$, then we write $\gamma \in AC^p([a, b]; X)$.

(P.34) Lemma *Let $a, b \in \mathbb{R}$. If $\gamma : [a, b] \to X$ is an absolutely continuous curve, then it is uniformly continuous.*

Proof. We already know that for all $g \in L^1(a, b)$, for all $\varepsilon > 0$, there exists $\delta > 0$ such that for every $B \in \mathscr{B}([a, b])$ with $\mathcal{L}^1(B) < \delta$ one has

$$\int_B |g| d\mathcal{L}^1 < \varepsilon.$$

Therefore, if $x, y \in [a, b]$, and without loss of generality $x \leq y$, with $\mathcal{L}^1([x, y]) = |x-y| < \delta$ one has

$$d(\gamma(y), \gamma(x)) \leq \int_x^y |g| d\mathcal{L}^1 < \varepsilon$$

as required. ■

For a very general setting, we only define the absolute value of the derivative of an absolutely continuous curve; meanwhile, when we are in the Euclidean case, everything will be much simpler. In any case, existence and uniqueness are given by the following Theorem.

(P.35) Theorem *Let $a, b \in \mathbb{R}$. For any $\gamma \in AC([a, b]; X)$, the limit*

$$\lim_{h \to 0} \frac{d(\gamma(t), \gamma(t+h))}{|h|}$$

exists for a.e. $t \in]a, b[$. In particular, this limit is the minimal g that we can choose in the definition of an absolutely continuous curve, up to $\mathcal{L}^1$-negligible sets.

Proof. By Weierstrass Theorem, up to replacing X with $\gamma([a,b])$, we can assume X to be compact.

Consider (z_i) a countable and dense subset of X and define for every $i \in \mathbb{N}$ a function $f_i : [a,b] \to \mathbb{R}$ such that

$$f_i(t) = d(\gamma(t), z_i).$$

The fact that f_i is well defined comes from the compactness of X. Using the triangle inequality and the definition of an absolutely continuous curve, for every $x, y \in [a,b]$ such that $x \leq y$, we can write for any admissible $g \in L^1(a,b)$

$$|f_i(y) - f_i(x)| \leq d(\gamma(y), \gamma(x)) \leq \int_x^y g d\mathcal{L}^1,$$

So, up to repeating the estimate for $y \leq x$, we can affirm that (f_i) in $AC([a,b]; X)$, then, for every $i \in \mathbb{N}$, f_i' exists a.e. in $]a,b[$ and since

$$|f_i(y) - f_i(x)| = \int_x^y f' d\mathcal{L}^1,$$

we have for every $x, y \in]a,b[$ such that $x \leq y$

$$\int_x^y -g d\mathcal{L}^1 \leq \int_x^y f' d\mathcal{L}^1 \leq \int_x^y g d\mathcal{L}^1$$

so $-g \leq f' \leq g$ a.e. in $]a,b[$. In other words, $|f_i'| \leq g$ a.e. in $]a,b[$. By the fact that

$$\sup_{i \in \mathbb{N}} |f_i'(t)| \leq g \text{ a.e. in }]a,b[\,,$$

then $\sup\limits_{i \in \mathbb{N}} |f_i'(t)| \in L^1(a,b)$. Consider now $t \in]a,b[$ such that $f_i'(t)$ exists for any $i \in \mathbb{N}$, then by triangle inequality

$$\liminf_{h \to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|} \geq \liminf_{h \to 0} \frac{|f_i(t+h) - f_i(t)|}{|h|} = |f_i'(t)|$$

and passing to the supremum

$$\liminf_{h \to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|} \geq \sup_{i \in \mathbb{N}} |f_i'(t)|.$$

Now, if $h > 0$, we have

$$|f_i(t+h) - f_i(t)| \leq \int_t^{t+h} |f_i'| d\mathcal{L}^1 \leq \int_t^{t+h} \sup_{i \in \mathbb{N}} |f_i'| d\mathcal{L}^1.$$

Let us show that

$$\sup_{i\in\mathbb{N}} |d(\gamma(t+h), z_i) - d(\gamma(t), z_i)| = d(\gamma(t+h), \gamma(t)).$$

By triangle inequality, for every $i \in \mathbb{N}$,

$$|d(\gamma(t+h), z_i) - d(\gamma(t), z_i)| \leq d(\gamma(t+h), \gamma(t))$$

so

$$\sup_{i\in\mathbb{N}} |d(\gamma(t+h), z_i) - d(\gamma(t), z_i)| \leq d(\gamma(t+h), \gamma(t)).$$

On the other hand, by density, there exists (z_{i_j}) such that $z_{i_j} \to \gamma(t)$ so, by continuity of the distance,

$$\sup_{i\in\mathbb{N}} |d(\gamma(t+h), z_i) - d(\gamma(t), z_i)| \geq \sup_{j\in\mathbb{N}} |d(\gamma(t+h), z_{i_j}) - d(\gamma(t), z_{i_j})| = d(\gamma(t+h), \gamma(t)).$$

We have therefore arrived at

$$\sup_{i\in\mathbb{N}} |f_i(t+h) - f_i(t)| = \sup_{i\in\mathbb{N}} |d(\gamma(t+h), z_i) - d(\gamma(t), z_i)| = d(\gamma(t+h), \gamma(t))$$

so

$$d(\gamma(t+h), \gamma(t)) \leq \int_t^{t+h} \sup_{i\in\mathbb{N}} |f_i'| d\mathcal{L}^1.$$

Considering t a Lebesgue point of $\sup_{i\in\mathbb{N}} |f_i'|$, we obtain

$$\limsup_{h\to 0^+} \frac{d(\gamma(t+h), \gamma(t))}{h} \leq \lim_{h\to 0^+} \frac{1}{h} \int_t^{t+h} \sup_{i\in\mathbb{N}} |f_i'| d\mathcal{L}^1 = \sup_{i\in\mathbb{N}} |f_i'|(t).$$

Being able to build an analogous estimate for $h < 0$, we obtain for a.e. $t \in \,]a, b[$

$$\lim_{h\to 0} \frac{d(\gamma(t), \gamma(t+h))}{|h|} = \sup_{i\in\mathbb{N}} |f_i'(t)|. \blacksquare$$

(P.36) Definition *Let $a, b \in \mathbb{R}$. For any $\gamma \in AC([a, b]; X)$ we call* metric derivative of *γ the function $|\gamma'| : \,]a, b[\, \to \mathbb{R}$ defined up to negligible sets by*

$$|\gamma'|(t) = \lim_{h\to 0} \frac{d(\gamma(t), \gamma(t+h))}{|h|}.$$

(P.37) Lemma *Let (X, d) be a metric space, $\gamma : [0, 1] \to X$ a curve and $g \in L^2(0, 1)$*

a.e. *positive such that for every* $s, t \in [0,1]$ *with* $s \leq t$

$$d^2(\gamma(s), \gamma(t)) \leq (t-s) \int_s^t g^2 d\mathcal{L}^1.$$

Then $\gamma \in AC^2([0,1]; X)$ *and* $|\gamma'| \leq g$ a.e. *in* $(0,1)$.

Proof. First of all, by Young's inequality, for every $s, t \in [0,1]$ with $s \leq t$, using the assumption,

$$d(\gamma(s), \gamma(t)) \leq \sqrt{(t-s) \int_s^t g^2 d\mathcal{L}^1} \leq \frac{1}{2}(t-s) + \frac{1}{2} \int_s^t g^2 d\mathcal{L}^1 = \frac{1}{2} \int_s^t (1+g^2) d\mathcal{L}^1,$$

so $\gamma \in AC([0,1]; X)$. In particular, it is well defined $|\gamma'|(t)$, and by the fact that for a.e. $t \in [0,1]$

$$\frac{d^2(\gamma(t+h), \gamma(t))}{h^2} \leq \frac{1}{h^2} h \int_t^{t+h} g^2 d\mathcal{L}^1,$$

we have for a.e. $t \in [0,1]$, by Lebesgue's points Theorem,

$$|\gamma'|(t) = \lim_{h \to 0} \frac{d(\gamma(t+h), \gamma(t))}{|h|} \leq \lim_{h \to 0} \sqrt{\frac{1}{h^2} h \int_t^{t+h} g^2 d\mathcal{L}^1} = g(t)$$

so $|\gamma'| \in L^2(0,1)$ and the estimate holds. ■

In order to arrive at the definition of geodesics, we have to talk about the length of curves.

(P.38) Definition *Let* $a, b \in \mathbb{R}$. *For any* $\gamma \in AC([a,b]; X)$ *we call* length of γ

$$l(\gamma) = \int_a^b |\gamma'|(t) dt.$$

Arranging what has been said so far, it is clear that $l(\gamma) \geq d(\gamma(a), \gamma(b))$.

(P.39) Definition *Let* $a, b \in \mathbb{R}$. *We say that* $\gamma \in AC([a,b]; X)$ *has* constant speed *if* $|\gamma'|$ *is, up to negligible sets, a constant.*

(P.40) Lemma *Let* $a, b \in \mathbb{R}$. *If* $\gamma \in AC([a,b]; X)$ *has constant speed, then it is Lipschitz.*

Proof. From the definition of absolutely continuous curve, combined with Theorem (P.35), for all $x, y \in [a,b]$ such that $x \leq y$ one has

$$d(\gamma(x),\gamma(y)) \leq \int_x^y |\gamma'|dt \leq |\gamma'||x-y|.$$

If $y \leq x$ the argument is analogous, so the result follows. ∎

Even if constant-speed curves seem like something new, they are actually nothing more than reparameterizations of absolutely continuous curves, as we can see in the next Proposition.

(P.41) Proposition *Let $a, b \in \mathbb{R}$. For any $\gamma \in AC([a,b];X)$, there exists $\tilde{\gamma} \in AC([0,1];X)$ with constant speed equal to $l(\gamma)$ such that $\gamma([a,b]) = \tilde{\gamma}([0,1])$, $\gamma(a) = \tilde{\gamma}(0)$, and $\gamma(b) = \tilde{\gamma}(1)$.*

Proof. We only consider the case when $\left\{t \mapsto l(\gamma|_{[a,t]})\right\}$ is strictly increasing in $[a,b]$. Consider $L : [a,b] \to [0,l(\gamma)]$ such that

$$L(t) = l(\gamma|_{[a,t]}).$$

From the definition of an absolutely continuous curve, for every $u, v \in [0,l(\gamma)]$ such that $u \leq v$,

$$d(\gamma(L^{-1}(u)),\gamma(L^{-1}(v))) \leq \int_{L^{-1}(u)}^{L^{-1}(v)} |\gamma'|d\mathcal{L}^1 = \int_a^{L^{-1}(v)} |\gamma'|d\mathcal{L}^1 - \int_a^{L^{-1}(u)} |\gamma'|d\mathcal{L}^1 = v-u$$

so $\gamma \circ L^{-1} : [0,l(\gamma)] \to X$ is Lipschitz. On the other hand, by the invariance of the length under reparameterization

$$l(\gamma) = l(\gamma \circ L^{-1}) = \int_0^{l(\gamma)} |(\gamma \circ L^{-1})'|d\mathcal{L}^1,$$

so it must be $|(\gamma \circ L^{-1})'| = 1$ a.e. in $]0,l(\gamma)[$. The desired reparameterization is obtained up to a linear rescaling of $\gamma \circ L^{-1}$. ∎

We finally arrive at defining geodesics.

(P.42) Definition *Let $a, b \in \mathbb{R}$. We say that $\gamma \in AC([a,b];X)$ is a* geodesic *if*

$$l(\gamma) = d(\gamma(a),\gamma(b)).$$

(P.43) Lemma *Let $a, b \in \mathbb{R}$. If $\gamma \in AC([a,b];X)$ is a geodesic, then for every $c, d \in [a,b]$ such that $c \leq d$ also $\gamma|_{[c,d]}$ is a geodesic and the metric derivatives coincide* a.e. *in $]c,d[$.*

Proof. We only need to prove that

$$d(\gamma(c),\gamma(d)) \geq \int_c^d |\gamma'| d\mathcal{L}^1.$$

By triangle inequality, Theorem (P.35) and the fact that γ is geodesic, one has

$$\begin{aligned} d(\gamma(a),\gamma(b)) &\leq d(\gamma(a),\gamma(c)) + d(\gamma(c),\gamma(d)) + d(\gamma(d),\gamma(b)) \leq \\ &\leq \int_a^c |\gamma'| d\mathcal{L}^1 + \int_c^d |\gamma'| d\mathcal{L}^1 + \int_d^b |\gamma'| d\mathcal{L}^1 = \\ &= \int_a^b |\gamma'| d\mathcal{L}^1 = d(\gamma(a),\gamma(b)) \end{aligned}$$

so

$$\begin{aligned} d(\gamma(a),\gamma(c)) + d(\gamma(c),\gamma(d)) + d(\gamma(d),\gamma(b)) &= \int_a^c |\gamma'| d\mathcal{L}^1 + \int_c^d |\gamma'| d\mathcal{L}^1 + \int_d^b |\gamma'| d\mathcal{L}^1 \geq \\ &\geq d(\gamma(a),\gamma(c)) + \int_c^d |\gamma'| d\mathcal{L}^1 + d(\gamma(d),\gamma(b)). \end{aligned}$$

In particular,

$$d(\gamma(c),\gamma(d)) \geq \int_c^d |\gamma'| d\mathcal{L}^1$$

and the result follows. ∎

(P.44) Notation We denote with $\mathrm{Geo}(X)$ the set

$$\mathrm{Geo}(X) = \left\{ \gamma \in AC([0,1];X) : |\gamma'| = l(\gamma) = d(\gamma(0),\gamma(1)) \right\}.$$

The set $\mathrm{Geo}(X)$ admits the following characterization.

(P.45) Proposition *Consider a curve $\gamma : [0,1] \to X$. The following facts are equivalent:*

(a) $\gamma \in \mathrm{Geo}(X)$*,*

(b) for every $s, t \in [0,1]$ one has

$$d(\gamma(s),\gamma(t)) = |s-t| d(\gamma(0),\gamma(1)),$$

(c) for every $s, t \in [0,1]$ one has

$$d(\gamma(s), \gamma(t)) \leq |s-t| d(\gamma(0), \gamma(1)).$$

Proof.

$(a) \implies (b)$ If $\gamma \in \mathrm{Geo}(X)$, for every $s, t \in [0,1]$ such that $s \leq t$ also $\gamma|_{[s,t]}$ is a geodesic with the same constant speed $|\gamma'|(t) = l(\gamma) = d(\gamma(0), \gamma(1))$, so

$$d(\gamma(s), \gamma(t)) = l(\gamma|_{[s,t]}) = \int_s^t |\gamma'| d\mathcal{L}^1 = |s-t| d(\gamma(0), \gamma(1)).$$

The case $t \leq s$ is analogous.

$(b) \implies (c)$ Obvious.

$(c) \implies (a)$ Let us start by saying that $\gamma \in AC([0,1]; X)$: indeed, for every $s, t \in [0,1]$ with $s \leq t$ we have

$$d(\gamma(s), \gamma(t)) \leq |s-t| d(\gamma(0), \gamma(1)) = \int_s^t d(\gamma(0), \gamma(1)) d\mathcal{L}^1.$$

In particular, by Theorem (P.35), for a.e. $t \in]0,1[$

$$|\gamma'|(t) \leq d(\gamma(0), \gamma(1)).$$

Integrating with respect to t we have

$$\int_0^1 |\gamma'| d\mathcal{L}^1 \leq d(\gamma(0), \gamma(1))$$

and remembering that

$$\int_0^1 |\gamma'| d\mathcal{L}^1 = l(\gamma) \geq d(\gamma(0), \gamma(1))$$

we obtain

$$d(\gamma(0), \gamma(1)) \leq l(\gamma) \leq d(\gamma(0), \gamma(1)).$$

In other words,

$$l(\gamma) = d(\gamma(0), \gamma(1)),$$

so γ is a geodesic.

It only remains to prove that $|\gamma'| = d(\gamma(0), \gamma(1))$ a.e. in $]0,1[$. Suppose not, then

$$\mathcal{L}^1 \lfloor [0,1](\{|\gamma'| < d(\gamma(0), \gamma(1))\}) > 0.$$

Consider the measurable sets

$$A = \{t \in [0,1] : |\gamma'|(t) < d(\gamma(0), \gamma(1))\}$$

and

$$B = \{t \in [0,1] : |\gamma'|(t) = d(\gamma(0), \gamma(1))\}.$$

Since $|\gamma'| \le d(\gamma(0), \gamma(1))$ a.e. in $]0,1[$, we have

$$1 = \mathcal{L}^1(\{t \in [0,1] : |\gamma'|(t) \le d(\gamma(0), \gamma(1))\}) = \mathcal{L}^1(A) + \mathcal{L}^1(B),$$

so

$$\int_0^1 |\gamma'| d\mathcal{L}^1 = \int_A |\gamma'| d\mathcal{L}^1 + \int_B |\gamma'| d\mathcal{L}^1 < d(\gamma(0), \gamma(1))(\mathcal{L}^1(A) + \mathcal{L}^1(B)) = d(\gamma(0), \gamma(1))$$

that is a contradiction. In conclusion $\gamma \in \mathrm{Geo}(X)$. ∎

(P.46) Lemma $\mathrm{Geo}(X) \subseteq C([0,1]; X)$ *is closed.*

Proof. Consider (γ_h) in $\mathrm{Geo}(X)$ such that $\gamma_h \to \gamma$ in $C([0,1]; X)$. By continuity of distance and the characterization of $\mathrm{Geo}(X)$, we get $\gamma \in \mathrm{Geo}(X)$. ∎

(P.47) Remark *If (X, d) is Polish, then $C([0,1]; X)$ is Polish as well.*

(P.48) Notation Given $t \in [0,1]$, we denote with $e_t : C([0,1]; X) \to X$ the *evaluation map*, namely

$$e_t(\gamma) = \gamma(t).$$

(P.49) Definition *We call* action *the functional $\mathcal{A}_2 : C([0,1]; X) \to [0, +\infty]$ such that*

$$\mathcal{A}_2(\gamma) = \begin{cases} \int_0^1 |\gamma'|^2 d\mathcal{L}^1 & \text{if } \gamma \in AC^2([0,1]; X), \\ +\infty & \text{if } \gamma \in C([0,1]; X) \setminus AC^2([0,1]; X). \end{cases}$$

One can prove that $\mathcal{A}_2$ is lower semicontinuous.

(P.50) Lemma *If $\gamma \in AC^2([0,1]; X)$, then γ is $\frac{1}{2}$-Hölder.*

Proof. If $x, y \in [0,1]$ such that $x \leq y$, one has, by Hölder's inequality,

$$d(\gamma(x), \gamma(y)) \leq \int_x^y |\gamma'| d\mathcal{L}^1 = \int_0^1 \chi_{[x,y]} |\gamma'| d\mathcal{L}^1 \leq \left(\int_0^1 \chi^2_{[x,y]} d\mathcal{L}^1 \right)^{\frac{1}{2}} \left(\int_0^1 |\gamma'|^2 d\mathcal{L}^1 \right)^{\frac{1}{2}}$$

so

$$d(\gamma(x), \gamma(y)) \leq \mathcal{A}_2^{\frac{1}{2}}(\gamma) |x - y|^{\frac{1}{2}}.$$

If $y \leq x$ the argument is analogous and the result follows. ■

(P.51) Lemma *Let $\gamma \in AC([0,1]; X)$. The following facts are equivalent:*

(a) $\gamma \in \mathrm{Geo}(X)$,

(b) $\mathcal{A}_2(\gamma) = d^2(\gamma(0), \gamma(1))$.

Proof.
$(a) \Longrightarrow (b)$ If $\gamma \in \mathrm{Geo}(X)$, $|\gamma'| = l(\gamma) = d(\gamma(0), \gamma(1))$, so

$$\mathcal{A}_2(\gamma) = \int_0^1 d^2(\gamma(0), \gamma(1)) d\mathcal{L}^1 = d^2(\gamma(0), \gamma(1)).$$

$(b) \Longrightarrow (a)$ Using Hölder's inequality, we can write

$$d^2(\gamma(0), \gamma(1)) \leq l^2(\gamma) = \left(\int_0^1 |\gamma'| d\mathcal{L}^1 \right)^2 \leq \int_0^1 |\gamma'|^2 d\mathcal{L}^1 = \mathcal{A}_2^2(\gamma) = d^2(\gamma(0), \gamma(1)),$$

so

$$l(\gamma) = d(\gamma(0), \gamma(1))$$

and by the fact that Hölder's inequality holds as an equality, $|\gamma'|$ has to be a constant up to a negligible set. ■

(P.52) Definition *A metric space (X, d) is called* geodesic *if for every $x, y \in X$ there exists $\gamma \in \mathrm{Geo}(X)$ with $\gamma(0) = x$ and $\gamma(1) = y$.*

(P.53) Notation Unless otherwise specified, we denote a curve $\Upsilon : [0,1] \to \mathcal{P}(X)$ such that $\Upsilon(t) = \mu_t$, where $\{\mu_t\}_{t \in [0,1]}$ in $\mathcal{P}(X)$, with

$$\mu_t : [0,1] \to \mathcal{P}(X).$$

(P.54) Definition *Consider a curve $\mu_t : [0,1] \to \mathcal{P}(X)$. We say that $\eta \in \mathcal{P}(C([0,1];X))$ is a* lifting of μ_t *if for every $t \in [0,1]$ one has*

$$(e_t)_{\#}\eta = \mu_t.$$

(P.55) Proposition *Consider a curve $\mu_t : [0,1] \to \mathcal{P}(X)$. If there exists a lifting $\eta \in \mathcal{P}(C([0,1]);X)$ of μ_t, then μ_t is weakly continuous.*

Proof. Consider $t \in [0,1]$ and (t_h) in $[0,1]$ such that $t_h \to t$. First of all, given $\gamma \in C([0,1];X)$ and (γ_h) in $C([0,1];X)$ such that $\gamma_h \to \gamma$ in $C([0,1];X)$, for every $s \in [0,1]$ we have

$$|e_s(\gamma_h) - e_s(\gamma)| \leq |\gamma_h(s) - \gamma(s)| \leq \|\gamma_h - \gamma\|_\infty$$

so e_s is continuous. In particular, given $f \in C_b(X)$, for every $\gamma \in C([0,1];X)$

$$\lim_h f(e_{t_h}(\gamma)) = f(e_t(\gamma)).$$

Observing also that $|f(e_{t_h}(\gamma))| \leq \|f\|_\infty \in L^1(C([0,1];X),\eta)$, so, applying the dominated convergence Theorem, we obtain

$$\begin{aligned}\lim_h \int_X f d\mu_{t_h} &= \lim_h \int_X f(x) d(e_{t_h})_{\#}\eta(x) = \\ &= \lim_h \int_{C([0,1];X)} f(e_{t_h}(\gamma)) d\eta(\gamma) = \int_{C([0,1];X)} f(e_t(\gamma)) d\eta(\gamma) = \int_X f d\mu_t,\end{aligned}$$

hence μ_t is weakly continuous. ■

The following is a variant of the classical Ascoli–Arzelà's Theorem.

(P.56) Theorem *If $\mathcal{F} \subseteq C([0,1];X)$ is such that $\sup_{\gamma\in\mathcal{F}} \mathcal{A}_2 < +\infty$ and there exists $D \subseteq [0,1]$ such that $\overline{D} = [0,1]$ and for all $t \in D$*

$$\overline{\{\gamma(t) : \gamma \in \mathcal{F}\}}$$

is compact in X, then $\overline{\mathcal{F}}$ is compact in $C([0,1];X)$.

Proof. See [1, Theorem 10.3]. ■

We will need the following result.

(P.57) Theorem (random Ascoli–Arzelà) *If $\mathcal{F} \subseteq \mathcal{P}(C([0,1];X))$ is such that*

$$\sup_{\eta \in \mathcal{F}} \int_{C([0,1];X)} \mathcal{A}_2 d\eta(\gamma) < +\infty$$

and there exists $D \subseteq [0,1]$ such that $\overline{D} = [0,1]$ and, for all $t \in D$, $\{(e_t)_{\#}\eta : \eta \in \mathcal{F}\}$ is tight in $\mathcal{P}(X)$, then $\mathcal{F}$ is tight in $\mathcal{P}(C([0,1];X))$.

Proof. See [1, Theorem 10.4]. ■

Chapter 1

Static representations

1 The Monge problem

Contrary to what usually happens, we want to start with an example to understand, in an intuitive but physical way, the problem we would then like to formulate in an abstract environment.

(1.1) Example *Consider a box filled with* Spectre tiles*.*

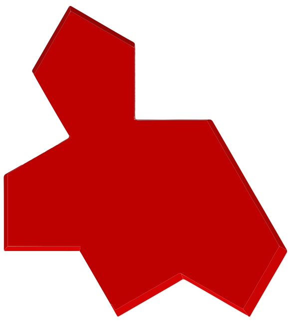

Figure 1.1: A spectre tile (picture from [54]).

Our goal is to make a complete tessellation of the floor using Spectre tiles, spending as little effort as possible. In other words, we have to decide which is the best position for each tile in order to minimize the effort to transport it from the box to the floor.

More details on Spectre tiles can be found in [63]. However, the above example is a modern reformulation of Monge's original idea that can be found in [47]. The essential element is the following: seeking the best function to transport each mass unit from the initial configuration to the final one.

Once we have intuitively understood the problem, we are now ready for the general and abstract definition. We specify that we will only deal with the case of measures with finite and equal mass: in other words, if $(X, d_X), (Y, d_Y)$ are two metric spaces, we will only consider the case of $\mu \in \mathcal{M}_+(X)$, $\nu \in \mathcal{M}_+(Y)$ such that $\mu(X) = \nu(Y) < +\infty$. Without loss of generality, we will also consider only normalized measures, namely $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.

(1.2) Definition *Let $(X, d_X), (Y, d_Y)$ be two metric spaces, $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and $c : X \times Y \to [0, \infty[$ a Borel function called* cost function. *In the so-called* Monge (optimal transport) problem *we look for*

$$\inf \left\{ C_\mu(T) = \int_X c(x, T(x)) d\mu(x) : T : X \to Y \text{ Borel}, T_\# \mu = \nu \right\}.$$

In particular, we call transport map *every $T : X \to Y$ Borel such that $T_\# \mu = \nu$ and* optimal map *every minimizer of the above problem.*

Notice that we only work with finite costs: our goal is, in fact, to apply the general theory to the case in which the cost function is the squared distance, as we will see from Chapter 3. To better comprehend, also quantitatively, the Monge problem, let us analyze a very simple example in which we can solve the problem explicitly. From a physical point of view, the idea is very trivial: the best way to transport the mass of a rod in itself is to not move anything.

(1.3) Example *Consider $X = Y = [0, 1]$, equipped with the Euclidean distance, $\mu = \nu = \mathcal{L}^1 \lfloor [0, 1]$ and $c : [0, 1] \times [0, 1] \to [0, \infty[$ such that*

$$c(x, y) = |x - y|.$$

In this simplified setting, the Monge problem reads as

$$\inf \left\{ C_\mu(T) = \int_0^1 |x - T(x)| dx : T : X \to Y \text{ Borel}, T_\# \mu = \nu \right\}.$$

Given any $B \in \mathscr{B}([0, 1])$, we have

$$T_\# \mu(B) = \mu(T^{-1}(B)) = \mathcal{L}^1(T^{-1}(B))$$

and

$$\nu(B) = \mathcal{L}^1(B),$$

so the constraint can be rewritten as $\mathcal{L}^1(B) = \mathcal{L}^1(T^{-1}(B))$. In other words, every transport map must be a Borel, volume-preserving function. Observing that for every *transport map* $C_\mu(T) \geq 0$*, we directly have*

$$\inf\left\{C_\mu(T) = \int_0^1 |x - T(x)|dx : T : X \to Y \text{ Borel}, T_\# \mu = \nu\right\} \geq 0.$$

Combining the fact that Id *obviously satisfies the constraint and* $C_\mu(\mathrm{Id}) = 0$*, we obtain*

$$\min\left\{C_\mu(T) = \int_0^1 |x - T(x)|dx : T : X \to Y \text{ Borel}, T_\# \mu = \nu\right\} = 0.$$

In particular, by the fact that

$$\int_0^1 |x - T(x)|dx = 0 \Longleftrightarrow T(x) = x \text{ a.e. in } [0,1],$$

we also get that Id *is the unique optimal map in* $L^0([0,1])$*.*

Unfortunately, the Monge problem can be ill posed, as we can see in the following example in which the domain of the problem is the empty set.

(1.4) Example *If* $\mu = \delta_{x_0}$ *for some* $x_0 \in X$*, then for each* $B \in \mathscr{B}(Y)$

$$T_\# \delta_{x_0}(B) = \delta_{x_0}(T^{-1}(B)) = \delta_{T(x_0)}(B).$$

So if $\nu \in \mathcal{P}(Y)$ *is not a Dirac delta, there does not exist any transport map.*

By the innate fallacy of Monge's formulation, the need for a different definition arises.

2 The Kantorovich problem

The aim of this section is to present and analyze the reformulation of the optimal transport problem carried out by Kantorovich.

In the same way as in the previous section, we start our discussion with an intuitive, but physical, example to understand the basic idea under the abstract formulation that we will later see.

(1.5) Example *Suppose that in the province of Brescia there are* $x_1, \ldots, x_n$ *mills and* $y_1, \ldots, y_m$ *bakeries. Every mill produces* μ_i *flour and every bakery needs* ν_j *flour. If we call* $c(x_i, y_j)$ *the unit transport cost from the mill* x_i *to the bakery* y_j *and* π_{ij} *the flour* *transported from the mill* x_i *to the bakery* y_j*, the global transport cost is*

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c(x_i, y_j)\pi_{ij}.$$

Our goal is to minimize the cost by selecting the best way to transport all flour from the mills to the bakeries. Clearly, π_{ij} *has some constraints to respect:*

- *the quantity of flour that came out from each mill must be equal to the sum of the amounts of flour transported from the aforementioned mill to each bakery,*
- *the quantity of flour received by each bakery must be equal to the sum of the amounts of flour transported to the aforementioned bakery from each mill,*
- *the flour transported must be a positive quantity.*

In other words,

$$\sum_{i=1}^{n} \pi_{ij} = \nu_j, \qquad \sum_{j=1}^{m} \pi_{ij} = \mu_i, \qquad \pi_{ij} \geq 0.$$

In the previous example, a new actor enters the scene to replace Monge's transport map. It is the amount of mass transported from the starting space to the target space. In order to relax the problem, following the idea of Kantorovich seen in [34], we use this new way to define the constraint.

(1.6) Definition *Let* $(X, d_X), (Y, d_Y)$ *be two metric spaces. Given* $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$*, we call* $\Gamma(\mu, \nu)$ *the set of* $\pi \in \mathcal{P}(X \times Y)$ *such that for all* $A \in \mathscr{B}(X)$*,* $B \in \mathscr{B}(Y)$ *one has*

$$\pi(A \times Y) = \mu(A), \qquad \pi(X \times B) = \nu(B). \tag{1.7}$$

We call transport plan from μ to ν *every* $\pi \in \Gamma(\mu, \nu)$*.*

Physically speaking, $\pi(A \times B)$ represents the mass, initially in A, sent in B.

One of the main advantages of using a transport plan instead of a transport map is the good shape of $\Gamma(\mu, \nu)$, as shown in the following.

(1.8) Proposition *Let* $(X, d_X), (Y, d_Y)$ *be two metric spaces. If* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$*, then* $\Gamma(\mu, \nu)$ *is not empty and convex.*

Proof. First of all, we have to remember that every measure in $\mathcal{P}(X \times Y)$ is uniquely determined by its value on Cartesian products of Borel sets. Consider now the measure

$\mu \otimes \nu : \mathscr{B}(X \times Y) \to [0,1]$ such that for every $A \in \mathscr{B}(X)$, $B \in \mathscr{B}(Y)$

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B).$$

It is straightforward that $\mu \otimes \nu \in \Gamma(\mu, \nu)$.

Now, given $\pi_1, \pi_2 \in \Gamma(\mu, \nu)$ and $\lambda \in [0,1]$, for every $A \in \mathscr{B}(X)$, $B \in \mathscr{B}(Y)$ one has

$$\begin{aligned}(\lambda\pi_1 + (1-\lambda)\pi_2)(A \times Y) = \lambda\pi_1(A \times Y) + (1-\lambda)\pi_2(A \times Y) = \\ = \lambda\mu(A) + (1-\lambda)\mu(A) = \mu(A)\end{aligned}$$

and

$$\begin{aligned}(\lambda\pi_1 + (1-\lambda)\pi_2)(X \times B) = \lambda\pi_1(X \times B) + (1-\lambda)\pi_2(X \times B) = \\ = \lambda\nu(B) + (1-\lambda)\nu(B) = \nu(B),\end{aligned}$$

so $\lambda\pi_1 + (1-\lambda)\pi_2 \in \Gamma(\mu, \nu)$, namely $\Gamma(\mu, \nu)$ is convex. ∎

It will be useful to highlight equivalent conditions for (1.7).

(1.9) Proposition *Let $(X, d_X), (Y, d_Y)$ be two metric spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Consider $\pi \in \mathcal{P}(X \times Y)$. The following facts are equivalent:*

(a) $\pi \in \Gamma(\mu, \nu)$,

(b) $(p_X)_\# \pi = \mu$ *and* $(p_Y)_\# \pi = \nu$.

Proof. Consider $A \in \mathscr{B}(X)$. By the fact that $p_X^{-1}(A) = A \times X$, the equivalence between the first relations in (1.7) and in (*b*) follows directly. The reasoning for the second ones is similar. ∎

Kantorovich's formulation for the optimal transport problem is the following.

(1.10) Definition *Let $(X, d_X), (Y, d_Y)$ be two metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a Borel function $c : X \times Y \to [0, \infty[$ called* cost function. *In the so-called* Kantorovich (optimal transport) problem *we look for*

$$\inf \left\{ C(\pi) = \int_{X \times Y} c(x,y) d\pi(x,y) : \pi \in \Gamma(\mu, \nu) \right\}.$$

In particular, we call optimal plan *every minimizer of the above problem.*

We highlight that the assumptions are the same as in the Monge problem, but, if we compare Example (1.4) and Proposition (1.8), we understand that the Monge problem and the Kantorovich problem in general are not equivalent: the second is always well posed. When, in the following, we refer to the primal representation of the optimal transport problem, we will always mean in the sense of the Kantorovich problem.

(1.11) Remark *The above problem is symmetric: switching coordinates, we can pass from* $\Gamma(\mu, \nu)$ *to* $\Gamma(\nu, \mu)$*.*

Once given the abstract definition of the problem, we now tackle the problem of the existence of optimal plans for the Kantorovich problem in a Polish setting and under the assumption of lower semicontinuous cost.

Let us start proving the weak lower semicontinuity of C: we use inf-convolution of the cost and we pass to the limit thanks to the monotone convergence Theorem.

(1.12) Proposition *Let* $(X, d_X), (Y, d_Y)$ *be two metric spaces,* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$*. If* $c : X \times Y \to [0, +\infty[$ *is a Borel lower semicontinuous function, then* $\{\pi \mapsto C(\pi)\}$ *is weakly lower semicontinuous in* $\mathcal{P}(X \times Y)$*.*

Proof. Consider $(\mathcal{I}_h c)$, the inf-convolution of the cost. We already know that $(\mathcal{I}_h c)$ is a bounded from below increasing sequence of Lipschitz functions and $\mathcal{I}_h c \nearrow c$ pointwise.

Now, given (π_n) in $\mathcal{P}(X \times Y)$ and $\pi \in \mathcal{P}(X \times Y)$ such that $\pi_n \rightharpoonup \pi$ in $\mathcal{P}(X \times Y)$, by weak convergence

$$\lim_n \int_{X \times Y} \mathcal{I}_h c d\pi_n = \int_{X \times Y} \mathcal{I}_h c d\pi,$$

but, from the fact that $\mathcal{I}_h c \leq c$,

$$\liminf_n C(\pi_n) = \liminf_n \int_{X \times Y} c d\pi_n \geq \liminf_n \int_{X \times Y} \mathcal{I}_h c d\pi_n$$

and the conclusion follows, by the monotone convergence Theorem, passing to the limits as $h \to +\infty$. ∎

Although the proof of the weak lower semicontinuity of C is done in a general metric setting, to prove the compactness of $\Gamma(\mu, \nu)$ we need to strengthen the structure restricting to Polish spaces: we will use Ulam's Lemma and Prokhorov's Theorem to demonstrate the following result.

(1.13) Proposition *Let* $(X, d_X), (Y, d_Y)$ *be Polish spaces,* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$*. The set* $\Gamma(\mu, \nu)$ *is weakly compact.*

Proof. Let us start by saying that the marginal conditions

$$\int_X \varphi d\mu = \int_{X\times Y} \varphi d\pi, \ \forall \varphi \in C_b(X),$$

$$\int_X \psi d\nu = \int_{X\times Y} \psi d\pi, \ \forall \psi \in C_b(Y)$$

imply that $\Gamma(\mu,\nu)$ is weakly closed. By Ulam's Lemma, for every $\varepsilon > 0$ there exist $K \subseteq X, \widetilde{K} \subseteq Y$ compacts such that $\mu(X \setminus K) < \frac{\varepsilon}{2}$ and $\mu(Y \setminus \widetilde{K}) < \frac{\varepsilon}{2}$. Thus,

$$\pi(X \times Y \setminus K \times \widetilde{K}) \leq \pi((X \setminus K) \times Y) + \pi(X \times (Y \setminus \widetilde{K})) < \varepsilon,$$

thanks to the marginal conditions, implies that $\Gamma(\mu,\nu)$ is tight. The conclusion follows from Prokhorov's Theorem. ∎

We are now ready to prove the existence of optimal plans for the Kantorovich problem using the direct method of the Calculus of Variations.

(1.14) Theorem *Let* $(X, d_X), (Y, d_Y)$ *be Polish spaces,* $\mu \in \mathcal{P}(X)$*,* $\nu \in \mathcal{P}(Y)$ *and* $c : X \times Y \to [0, +\infty[$ *lower semicontinuous. There exists an optimal plan.*

Proof. Consider (π_h) in $\Gamma(\mu,\nu)$ such that

$$\lim_h C(\pi_h) = \inf_{\pi \in \Gamma(\mu,\nu)} C(\pi).$$

By Proposition (1.13), we know that $\Gamma(\mu,\nu)$ is weakly compact, so there exist (π_{h_k}) in $\Gamma(\mu,\nu)$ and $\pi_0 \in \Gamma(\mu,\nu)$ such that $\pi_h \rightharpoonup \pi_0$ in $\Gamma(\mu,\nu)$. So, by Proposition (1.12),

$$C(\pi_0) \leq \liminf_k C(\pi_{h_k}) = \inf_{\pi \in \Gamma(\mu,\nu)} C(\pi),$$

then

$$C(\pi_0) = \min_{\pi \in \Gamma(\mu,\nu)} C(\pi)$$

and the result follows. ∎

Before proceeding further, let us analyze in more detail the relationship between Monge's formulation and Kantorovich's one. Given a transport map T, we can define the associated transport plan

$$\pi_T = (\mathrm{Id}, T)_{\#}\mu$$

where $(\mathrm{Id}, T) : X \to X \times Y$ is the function such that $(\mathrm{Id}, T)(x) = (x, T(x))$. By the change of variables formula, it follows

$$\begin{aligned} C(\pi_T) &= \int_{X\times Y} c(x,y)d(\mathrm{Id},T)_\# \mu(x,y) = \\ &= \int_X (c \circ (\mathrm{Id},T))(x)d\mu(x) = \int_X c(x,T(x))d\mu(x) = C_\mu(T). \end{aligned}$$

In particular,

$$\inf_{\pi\in\Gamma(\mu,\nu)} C(\pi) \le C_\mu(T)$$

and passing through the infimum with respect to T,

$$\inf_{\pi\in\Gamma(\mu,\nu)} C(\pi) \le \inf_{\substack{T \\ \text{transport map}}} C_\mu(T).$$

Clearly, the inequality can be strict because the domain of the Monge problem can be the empty set. In an attempt at completeness, we highlight that under some regularity assumptions, the two formulations can be recovered as equivalent.

(1.15) Theorem (Pratelli) *Let $(X, d_X), (Y, d_Y)$ be Polish spaces, $\mu \in \mathcal{P}(X)$ non-atomic, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to [0, +\infty[$ continuous. It holds*

$$\min_{\pi\in\Gamma(\mu,\nu)} C(\pi) = \inf_{\substack{T \\ \text{transport map}}} C_\mu(T).$$

Proof. See [56]. ∎

(1.16) Remark *In the following chapters, we will understand the particular importance of quadratic distance cost, namely*

$$c(x, y) = d^2(x, y),$$

which fully fits into the theory we have presented in this section since it is a continuous, positive and finite function.

3 The entropy functional

In Information Theory, entropy means the amount of information contained in a message and transferred through a communication channel. The first study is due to

Claude Shannon in 1948: the interested reader can find further details in the original article [62].

The most interesting aspect, that we underline just intuitively, is that, under some structural assumptions, there is a link between Shannon's entropy and Boltzmann's one: the phrase "when disorder increases, information is lost" is a good summary. Further information on this topic can be found in [25] or in [31].

Coming back to us, in order to introduce the entropy functional, we first need three technical results.

(1.17) Proposition *Consider a metric space (X, d) and $\mu, \nu \in \mathcal{R}_+(X)$. If $\nu \ll \mu$, then $\frac{d\nu}{d\mu} > 0$ ν-a.e. in X.*

Proof. We already know that $\frac{d\nu}{d\mu} \geq 0$ μ-a.e. in X, so $\frac{d\nu}{d\mu} \geq 0$ ν-a.e. in X. Now, consider

$$A = \left\{x \in X : \frac{d\nu}{d\mu}(x) = 0\right\}.$$

By contradiction, if $\nu(A) > 0$, we obtain

$$0 < \nu(A) = \int_A \frac{d\nu}{d\mu} d\mu = 0. \blacksquare$$

(1.18) Proposition *Consider a Polish space (X, d) and $\mathsf{R} : \mathscr{B}(X) \to [0, +\infty]$ a Borel σ-finite measure. There exists a Borel function $W : X \to [0, +\infty[$ such that*

$$\int_X e^{-W} d\mathsf{R} < +\infty.$$

Proof. By the σ-finiteness of R, there exists (A_n) in $\mathscr{B}(X)$ a sequence of disjoint Borel subsets of X such that

$$X = \bigcup_{n=0}^{\infty} A_n, \qquad \forall\, n \in \mathbb{N} : \mathsf{R}(A_n) < +\infty.$$

If we take $W : X \to [0, +\infty[$ such that

$$W(x) = \begin{cases} -\log \frac{1}{\mathsf{R}(A_0)+1} & \text{if } x \in A_0, \\ \quad\vdots & \\ -\log \frac{1}{2^n(\mathsf{R}(A_n)+1)} & \text{if } x \in A_n, \\ \quad\vdots & \end{cases}$$

then

$$\int_X e^{-W} d\mathsf{R} = \sum_{n=0}^{\infty} \int_{A_n} e^{-W} d\mathsf{R} = \sum_{n=0}^{\infty} \int_{A_n} \frac{1}{2^n(\mathsf{R}(A_n)+1)} d\mathsf{R} \leq \sum_{n=0}^{\infty} \frac{1}{2^n} = 2 < +\infty$$

and the result follows. ∎

(1.19) Proposition *Consider a Polish space (X,d). If $\mathsf{R} \in \mathcal{R}_+(X)$, then R is σ-finite.*

Proof. First of all, there exists $Q \subseteq X$ at most countable such that $\overline{Q} = X$. Now, since $\mathsf{R} \in \mathcal{R}_+(X)$, it is locally finite, so for every $q \in Q$ there exists $U_q \in \mathscr{B}(X)$, a neighborhood of q, such that $\mathsf{R}(U_q) < +\infty$. Up to substituting U_q with $\text{int}(U_q) \subseteq U_q$, we can assume U_q to be open. Clearly,

$$\bigcup_{q \in Q} U_q \subseteq X.$$

Now, if $x \in X$, there exists (q_n) in X such that $q_n \to x$. In particular, there exists $N \in \mathbb{N}$ such that for every $n \geq N$ we have $x \in U_{q_n}$, so

$$X \subseteq \bigcup_{q \in Q} U_q.$$

By the fact that Q is at most countable, the result follows. ∎

We can now proceed to provide the definition of the entropy functional: another name, by which it can be found in literature, is Kullback—Leibler divergence. This last name comes from Statistics, where divergences are particular statistical distances. It is not, actually, a distance.

(1.20) Definition *Consider a Polish space (X,d), $\mathsf{R} \in \mathcal{R}_+(X)$, a Borel function $W : X \to [0,+\infty[$ such that*

$$\int_X e^{-W} d\mathsf{R} < +\infty,$$

$\mathsf{R}_W = \frac{1}{\int_X e^{-W} d\mathsf{R}} e^{-W}\mathsf{R} \in \mathcal{P}(X)$ *and* $\sigma \in \mathcal{P}(X)$ *such that*

$$\int_X W d\sigma < +\infty.$$

We call (Boltzmann–Shannon) entropy of σ relative to R *the extended real number*

$$H(\sigma \mid \mathsf{R}) = H^p(\sigma \mid \mathsf{R}_W) - \int_X W d\sigma - \log\left(\int_X e^{-W} d\mathsf{R}\right),$$

where

$$H^p(\sigma \mid \mathsf{R}_W) = \begin{cases} \int_X \frac{d\sigma}{d\mathsf{R}_W} \log \frac{d\sigma}{d\mathsf{R}_W} d\mathsf{R}_W & \text{if } \sigma \ll \mathsf{R}_W, \\ +\infty & \text{otherwise.} \end{cases}$$

In particular, we define the function H_{R} such that

$$H_{\mathsf{R}}(\sigma) = H(\sigma \mid \mathsf{R}).$$

Even if the previous definition may seem complex, every time entropy is well defined, with some simple computations it turns out

$$H(\sigma \mid \mathsf{R}) = H^p(\sigma \mid \mathsf{R}).$$

Let us underline that if $\sigma \ll \mathsf{R}$,

$$\int_X \frac{d\sigma}{d\mathsf{R}} \log\left(\frac{d\sigma}{d\mathsf{R}}\right) d\mathsf{R} = \int_X \log\left(\frac{d\sigma}{d\mathsf{R}}\right) d\sigma.$$

(1.21) Proposition *The previous definition is well posed.*

Proof. Consider $W' : X \to [0, +\infty[$ another Borel function such that

$$\int_X e^{-W'} d\mathsf{R} < +\infty, \qquad \int_X W' d\sigma < +\infty.$$

First of all, if $H(\sigma \mid \mathsf{R})$ is well defined, then $H(\sigma \mid \mathsf{R}) = H^p(\sigma \mid \mathsf{R})$, so

$$H^p(\sigma \mid \mathsf{R}_W) - \int_X W d\sigma - \log\left(\int_X e^{-W} d\mathsf{R}\right) = H^p(\sigma \mid \mathsf{R}_{W'}) - \int_X W' d\sigma - \log\left(\int_X e^{-W'} d\mathsf{R}\right).$$

Combining this fact with Proposition (1.17), Proposition (1.18) and Proposition (1.19), the result follows. ■

(1.22) Proposition *Consider a Polish space (X, d) and $\mathsf{R}, \sigma \in \mathcal{P}(X)$. Then the entropy is well defined and, in particular,*

$$H(\sigma \mid \mathsf{R}) = \begin{cases} \int_X \frac{d\sigma}{d\mathsf{R}} \log\left(\frac{d\sigma}{d\mathsf{R}}\right) d\mathsf{R} & \text{if } \sigma \ll \mathsf{R}, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Consider $W = 0$. Then

$$\int_X e^{-W'} d\mathsf{R} = 1 < +\infty, \qquad \int_X W' d\sigma = 0 < +\infty.$$

In particular, H_{R} is well defined on $\mathcal{P}(X)$ and $H(\sigma \mid \mathsf{R}) = H^p(\sigma \mid \mathsf{R}_W)$. By the fact that $\mathsf{R}_W = \mathsf{R}$, the result follows. ∎

We will need to consider, in the following, the strictly convex function $h : [0, +\infty] \to [-\frac{1}{e}, +\infty]$ such that

$$h(x) = \begin{cases} x \log x & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases}$$

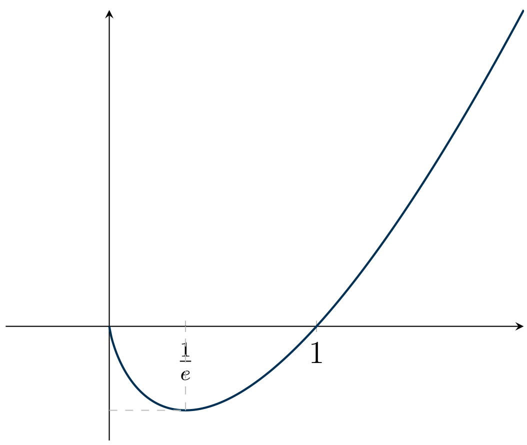

Figure 1.2: The graph of the function h.

We collect in the following Proposition the main properties of the entropy functional.

(1.23) Proposition *Let (X, d) be a Polish space and $\mathsf{R} \in \mathcal{R}_+(X)$. The following facts hold true:*

(a) if $\mathsf{R} \in \mathcal{P}(X)$, then H_{R} is positive,

(b) H_{R} is convex. In particular, where H_{R} is finite, it is strictly convex,

(c) if $\mathsf{R} \in \mathcal{P}(X)$, then $H_{\mathsf{R}}(\sigma) = 0$ if and only if $\sigma = \mathsf{R}$.

Proof.

(a) If σ is not absolutely continuous with respect to R, $H_{\mathsf{R}}(\sigma) = +\infty > 0$. Otherwise, by Jensen's inequality,

$$0 = h(1) = h\left(\frac{1}{\mathsf{R}(X)}\sigma(X)\right) = h\left(\frac{1}{\mathsf{R}(X)}\int_X \frac{d\sigma}{d\mathsf{R}}d\mathsf{R}\right) \leq \frac{1}{\mathsf{R}(X)}\int_X h\left(\frac{d\sigma}{d\mathsf{R}}\right)d\mathsf{R} = H_{\mathsf{R}}(\sigma).$$

(b) Let $\sigma_1, \sigma_2 \in \mathcal{P}(X)$ and $\lambda \in [0,1]$. If at least one, between σ_1 and σ_2, is not absolutely continuous with respect to R, either is $\lambda\sigma_1 + (1-\lambda)\sigma_2$, so

$$H_{\mathsf{R}}(\lambda\sigma_1 + (1-\lambda)\sigma_2) = +\infty = \lambda H_{\mathsf{R}}(\sigma_1) + (1-\lambda)H_{\mathsf{R}}(\sigma_2).$$

Otherwise, if $\sigma_1, \sigma_2 \ll \mathsf{R}$, so is $\lambda\sigma_1 + (1-\lambda)\sigma_2$ and, by the convexity of h and the linearity of Radon–Nikodym derivative,

$$\begin{aligned} H_{\mathsf{R}}(\lambda\sigma_1 + (1-\lambda)\sigma_2) &= \int_X h\left(\frac{d}{d\mathsf{R}}(\lambda\sigma_1 + (1-\lambda)\sigma_2)\right)d\mathsf{R} = \\ &= \int_X h\left(\lambda\frac{d\sigma_1}{d\mathsf{R}} + (1-\lambda)\frac{d\sigma_2}{d\mathsf{R}}\right)d\mathsf{R} \leq \\ &\leq \lambda H_{\mathsf{R}}(\sigma_1) + (1-\lambda)H_{\mathsf{R}}(\sigma_2). \end{aligned}$$

In particular, the strict convexity of h gives us that also H_{R} is strictly convex where it is finite.

(c) If $H_{\mathsf{R}}(\sigma) = 0$, $\sigma \ll \mathsf{R}$. By the fact that

$$0 = H(\sigma \mid \mathsf{R}) = \int_X \left(\frac{d\sigma}{d\mathsf{R}}\log\left(\frac{d\sigma}{d\mathsf{R}}\right) - \frac{d\sigma}{d\mathsf{R}} + 1\right)d\mathsf{R}$$

and the fact that for every $x \geq 0$, $x\log x - x + 1 \geq 0$, where the equality holds if and only if $x = 1$, we obtain $\frac{d\sigma}{d\mathsf{R}} = 1$, so $\sigma = \mathsf{R}$.

Conversely, if $\sigma = \mathsf{R}$, in particular $\sigma \ll \mathsf{R}$ and $\frac{d\sigma}{d\mathsf{R}} = 1$. A straightforward computation then provides

$$H_{\mathsf{R}}(\sigma) = \int_X 0 d\mathsf{R} = 0. \blacksquare$$

The following variational representations will be fundamental.

(1.24) Lemma *Consider a metric space* (X,d), $\mathsf{R} \in \mathsf{R}_+(X)$ *and* $\sigma \in \mathcal{P}(X)$ *such that* $H_{\mathsf{R}}(\sigma)$ *is well defined. The following facts hold true:*

(a) it results

$$H(\sigma \mid \mathsf{R}) = \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi : X \to \mathbb{R} \text{ bounded and Borel}\right\},$$

(b) it results

$$H(\sigma \mid \mathsf{R}) = \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi \in C_b(X)\right\},$$

(c) it results

$$H(\sigma \mid \mathsf{R}) = \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi : X \to \mathbb{R} \text{ Borel}, \int_X e^{\varphi} d\mathsf{R} < +\infty\right\}.$$

Proof.

(a) Denote with

$$M = \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi : X \to \mathbb{R} \text{ bounded and Borel}\right\}.$$

Consider initially the case $\sigma \ll \mathsf{R}$. For every $\varphi : X \to \mathbb{R}$ bounded and Borel consider $\psi : X \to]0, +\infty[$ bounded and Borel such that $\psi = e^{\varphi}$. If

$$\int_X \psi d\mathsf{R} = +\infty,$$

then

$$\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) = -\infty$$

and there is nothing to prove. Conversely, in the case

$$\int_X \psi d\mathsf{R} < +\infty,$$

by the fact that $\psi > 0$ R-a.e. in X, we can define the measure $\mathsf{R}' \in \mathcal{P}(X)$ such that for every $B \in \mathcal{B}(X)$

$$\mathsf{R}'(B) = \frac{1}{\int_X \psi d\mathsf{R}} \int_B \psi d\mathsf{R}.$$

By uniqueness of Radon–Nikodym derivative,

$$\frac{d\sigma}{d\mathsf{R}} = \frac{d\sigma}{d\mathsf{R}'}\frac{d\mathsf{R}'}{d\mathsf{R}} = \frac{d\sigma}{d\mathsf{R}'}\frac{\psi}{\int_X \psi d\mathsf{R}},$$

so

$$\begin{aligned} H(\sigma \mid \mathsf{R}) &= \int_X \frac{d\sigma}{d\mathsf{R}} \log\left(\frac{d\sigma}{d\mathsf{R}}\right) d\mathsf{R} = \frac{1}{\int_X \psi d\mathsf{R}} \int_X \frac{d\sigma}{d\mathsf{R}'}\psi \log\left(\frac{d\sigma}{d\mathsf{R}'}\frac{\psi}{\int_X \psi d\mathsf{R}}\right) d\mathsf{R} = \\ &= \frac{1}{\int_X \psi d\mathsf{R}} \int_X \frac{d\sigma}{d\mathsf{R}'} \log\left(\frac{d\sigma}{d\mathsf{R}'}\right) \psi d\mathsf{R} + \frac{1}{\int_X \psi d\mathsf{R}} \int_X \frac{d\sigma}{d\mathsf{R}'}\psi \log \psi d\mathsf{R} + \\ &- \frac{1}{\int_X \psi d\mathsf{R}} \int_X \frac{d\sigma}{d\mathsf{R}'}\psi \log\left(\int_X \psi d\mathsf{R}\right) d\mathsf{R} = \\ &= \int_X \frac{d\sigma}{d\mathsf{R}'} \log\left(\frac{d\sigma}{d\mathsf{R}'}\right) d\mathsf{R}' + \int_X \frac{d\sigma}{d\mathsf{R}'} \log \psi d\mathsf{R}' - \int_X \frac{d\sigma}{d\mathsf{R}'} \log\left(\int_X \psi d\mathsf{R}\right) d\mathsf{R}' = \\ &= H(\sigma \mid \mathsf{R}') + \int_X \log \psi d\sigma - \log\left(\int_X \psi d\mathsf{R}\right) \geq \\ &\geq \int_X \log \psi d\sigma - \log\left(\int_X \psi d\mathsf{R}\right) = \int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right). \end{aligned}$$

In particular,

$$H(\sigma \mid \mathsf{R}) \geq M.$$

If $H(\sigma \mid \mathsf{R}) < +\infty$, consider (φ_n) such that for every $n \in \mathbb{N} \setminus \{0\}$

$$\varphi_n = \log\left(\max\left\{\frac{1}{n}, \min\left\{\frac{d\sigma}{d\mathsf{R}}, n\right\}\right\}\right).$$

First of all,

$$\int_X \frac{d\sigma}{d\mathsf{R}} \log\left(\frac{d\sigma}{d\mathsf{R}}\right) d\mathsf{R} = H(\sigma \mid \mathsf{R}) < +\infty,$$

and, since $x(\log x)^- \leq \frac{1}{e}$,

$$\begin{aligned} \int_X \left(\log\left(\frac{d\sigma}{d\mathsf{R}}\right)\right)^- d\sigma &= \int_X \left(\log\left(\frac{d\sigma}{d\mathsf{R}_W}\right) - W - \log\left(\int_X e^{-W} d\mathsf{R}\right)\right)^- d\sigma \leq \\ &\leq \int_X \frac{d\sigma}{d\mathsf{R}_W} \log\left(\frac{d\sigma}{d\mathsf{R}_W}\right) d\mathsf{R}_W + \|W\|_{L^1(X,\sigma)} + \left|\log\left(\int_X e^{-W} d\mathsf{R}\right)\right| < \\ &< +\infty, \end{aligned}$$

so, $\log\left(\frac{d\sigma}{d\mathsf{R}}\right) \in L^1(X,\sigma)$. In particular, $\frac{d\sigma}{d\mathsf{R}}\log\left(\frac{d\sigma}{d\mathsf{R}}\right) \in L^1(X,\mathsf{R})$. By the fact that,

$$\frac{d\sigma}{d\mathsf{R}}\varphi_n \to \frac{d\sigma}{d\mathsf{R}}\log\left(\frac{d\sigma}{d\mathsf{R}}\right) \ \mathsf{R}\text{-a.e. in } X,$$
$$\left|\frac{d\sigma}{d\mathsf{R}}\left(\varphi_n - \log\left(\frac{d\sigma}{d\mathsf{R}}\right)\right)\right| \le 2\left|\frac{d\sigma}{d\mathsf{R}}\log\left(\frac{d\sigma}{d\mathsf{R}}\right)\right| \in L^1(X,\mathsf{R}),$$

by the dominated convergence Theorem, we have $\frac{d\sigma}{d\mathsf{R}}\varphi_n \to \frac{d\sigma}{d\mathsf{R}}\log\left(\frac{d\sigma}{d\mathsf{R}}\right)$ in $L^1(X,\mathsf{R})$. In the same way, $e^{\varphi_n} \to \frac{d\sigma}{d\mathsf{R}}$ in $L^1(X,\mathsf{R})$. In particular, from

$$M \ge \int_X \varphi_n d\sigma - \log\left(\int_X e^{\varphi_n} d\mathsf{R}\right),$$

using the convergence properties of Lebesgue spaces, the continuity of logarithm and the dominated convergence Theorem, we obtain

$$M \ge \int_X \log\left(\frac{d\sigma}{d\mathsf{R}}\right) d\sigma - \log\sigma(X) = H(\sigma \mid \mathsf{R}).$$

Instead, if $H(\sigma \mid \mathsf{R}) = +\infty$, using the same argument as before, we can find a sequence (φ_n) of positive Borel functions such that $\varphi_n \to \log\left(\frac{d\sigma}{d\mathsf{R}}\right)$ pointwise. As before, applying Fatou's Lemma in place of the dominated convergence Theorem,

$$+\infty = H(\sigma \mid \mathsf{R}) \le M.$$

If σ is not absolutely continuous with respect to R, clearly $H(\sigma \mid \mathsf{R}) = +\infty$ and there exists $S \in \mathscr{B}(X)$ such that $\mathsf{R}(S) = 0$ but $\sigma(S) > 0$. Consider the sequence (φ_n) of Borel functions such that

$$\varphi_n(x) = \begin{cases} n & \text{if } x \in S, \\ 0 & \text{if } x \in X \setminus S, \end{cases}$$

then

$$M \ge \lim_n \left(\int_X \varphi_n d\sigma - \log\left(\int_X e^{\varphi_n} d\mathsf{R}\right)\right) = \lim_n \left(\int_S n d\sigma - \log\left(\int_{X\setminus S} 1 d\mathsf{R}\right)\right) = +\infty.$$

(b) It follows from (a) and the density of $C_b(X)$ in $L^1(X,\mathsf{R})$.

(c) Clearly, using (a),

$$H(\sigma \mid \mathsf{R}) \leq \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi : X \to \mathbb{R} \text{ Borel}, \int_X e^{\varphi} d\mathsf{R} < +\infty\right\}.$$

Conversely, if $H(\sigma \mid \mathsf{R}) = +\infty$ we are done, otherwise we are in the case $\sigma \ll \mathsf{R}$ and, in the same way as is (a),

$$H(\sigma \mid \mathsf{R}) \geq \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi : X \to \mathbb{R} \text{ Borel}, \int_X e^{\varphi} d\mathsf{R} < +\infty\right\}. \blacksquare$$

We want now to prove that the entropy functional is convex and lower semicontinuous: this will be crucial in order to apply the direct method of the Calculus of Variations in the following section.

(1.25) Proposition *Consider a metric space (X, d) and $\mathsf{R} \in \mathcal{R}_+(X)$. The function $\{(\sigma, \mathsf{R}) \mapsto H(\sigma \mid \mathsf{R})\}$, when it is well defined, is convex and weakly lower semicontinuous. In particular, the function H_{R} is weakly lower semicontinuous when it is well defined.*

Proof. First of all, by Lemma (1.24),

$$H(\sigma \mid \mathsf{R}) = \sup\left\{\int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right) : \varphi \in C_b(X)\right\}.$$

Fixed $\varphi \in C_b(X)$, the function

$$\left\{(\sigma, \mathsf{R}) \mapsto \int_X \varphi d\sigma - \log\left(\int_X e^{\varphi} d\mathsf{R}\right)\right\}$$

is linear, hence convex, and weakly continuous, so we also obtain the desired convexity and lower semicontinuity of $\{(\sigma, \mathsf{R}) \mapsto H(\sigma \mid \mathsf{R})\}$. $\blacksquare$

4 Primal Schrödinger problem

In 1931, Erwin Schrödinger proposed in [61] an interpolation problem on Brownian particles that would later be called the Schrödinger problem.

(1.26) Proposition *Consider a Polish space (X, d), $\mathsf{R} \in \mathcal{R}_+(X \times X)$ and $\mu_0, \mu_1 \in \mathcal{P}(X)$. Suppose there exists a Borel function $B : X \to [0, +\infty[$ such that*

$$\int_{X\times X} e^{-B(x)-B(y)} d\mathsf{R}(x, y) < +\infty, \quad \int_X B(x) d\mu_0(x) < +\infty, \quad \int_X B(x) d\mu_1(x) < +\infty.$$

Then H_{R} is well defined on $\Gamma(\mu_0, \mu_1)$ by

$$H_{\mathsf{R}}(\sigma) = H^p(\sigma \mid \mathsf{R}_W) - \int_X B d\mu_0 - \int_X B d\mu_1 - \log\left(\int_X e^{-W} d\mathsf{R}\right).$$

Proof. Consider $W(x,y) = B(x) + B(y)$. By assumption,

$$\int_{X\times X} e^{-W} d\mathsf{R} < +\infty$$

and if $\pi \in \Gamma(\mu_0, \mu_1)$, using the change of variables formula,

$$\int_{X\times X} W d\pi = \int_{X\times X} B(x) d\pi(x,y) + \int_{X\times X} B(y) d\pi(x,y) = \int_X B d\mu_0 + \int_X B d\mu_1 < +\infty,$$

so the result follows. ■

Having shown that entropy is well defined on $\Gamma(\mu_0, \mu_1)$, under a small technicality, we can give the abstract formulation of the problem.

(1.27) Definition *Consider a Polish space (X, d), $\mathsf{R} \in \mathcal{R}_+(X \times X)$ and $\mu_0, \mu_1 \in \mathcal{P}(X)$. Suppose there exists a Borel function $B : X \to [0, +\infty[$ such that*

$$\int_{X\times X} e^{-B(x)-B(y)} d\mathsf{R}(x,y) < +\infty, \quad \int_X B(x) d\mu_0(x) < +\infty, \quad \int_X B(x) d\mu_1(x) < +\infty.$$

In the so-called Schrödinger problem *we look for*

$$\inf \{H_{\mathsf{R}}(\pi) : \pi \in \Gamma(\mu_0, \mu_1)\}.$$

The goal of the section is to prove, under some technical assumptions, existence, uniqueness and also a certain structural formula for the solution of the Schrödinger problem: the proof is an expansion of the one contained in [28]. As regards the first part of the statement, we used the direct method of the Calculus of Variations to prove existence, uniqueness descends directly from the strict convexity of the entropy functional and the structural formula comes from a Corollary of Hahn–Banach's Theorems. The second part gives a realistic case in which the first part is applicable, two functional inequalities tied to the solution of the problem and a theoretical information on the latter: the proof is, somehow, a direct computation and relies, also, on Du Bois-Reymond's Lemma.

(1.28) Theorem *Consider a Polish space* (X,d)*,* $\mathfrak{m} \in \mathcal{R}_+(X)$ *and* $\mathsf{R} \in \mathcal{R}_+(X \times X)$ *such that* $(p^1)_\# \mathsf{R} = (p^2)_\# \mathsf{R} = \mathfrak{m}$ *and* $\mathfrak{m} \otimes \mathfrak{m} \ll \mathsf{R} \ll \mathfrak{m} \otimes \mathfrak{m}$*. Let* $\mu_0 = \varrho_0 \mathfrak{m}, \mu_1 = \varrho_1 \mathfrak{m} \in \mathcal{P}(X)$ *and*

$$P_0 = \{x \in X : \varrho_0(x) > 0\}, \qquad\qquad P_1 = \{x \in X : \varrho_1(x) > 0\},$$

defined up to $\mathfrak{m}$*-negligible sets. If there exists a Borel function* $B : X \to [0, +\infty[$ *such that*

$$\int_{X \times X} e^{-B(x)-B(y)} d\mathsf{R}(x,y) < +\infty, \quad \int_X B(x) d\mu_0(x) < +\infty, \quad \int_X B(x) d\mu_1(x) < +\infty,$$

then the following facts hold true:

(a) *if* $H(\mu_0 \otimes \mu_1 \mid \mathsf{R}) < +\infty$*, then there exists a unique minimizer* γ *of* H_R *in* $\Gamma(\mu_0, \mu_1)$*.*

In particular, there exists two Borel functions $f, g : X \to [0, +\infty[$*, unique* $\mathfrak{m}$*-a.e. in* X *up to a rescaling* $\left\{(f,g) \mapsto (cf, \frac{g}{c})\right\}$ *with some* $c > 0$*, such that*

$$\gamma = (f \otimes g)\mathsf{R},$$

(b) *if* $\varrho_0, \varrho_1 \in L^\infty(X, \mathfrak{m})$ *and there exists* $c > 0$ *such that* $\mathsf{R} \geq c\mathfrak{m} \otimes \mathfrak{m}$ in $P_0 \times P_1$*, then* $H(\mu_0 \otimes \mu_1 \mid \mathsf{R}) < +\infty$*. In particular,* $f, g \in L^1(X, \mathfrak{m}) \cap L^\infty(X, \mathfrak{m})$ *with*

$$\|f\|_{L^\infty(X,\mathfrak{m})} \|g\|_{L^1(X,\mathfrak{m})} \leq \frac{\|\varrho_0\|_{L^\infty(X,\mathfrak{m})}}{c}, \qquad \|f\|_{L^1(X,\mathfrak{m})} \|g\|_{L^\infty(X,\mathfrak{m})} \leq \frac{\|\varrho_1\|_{L^\infty(X,\mathfrak{m})}}{c}$$

and γ *is the only transport plan of the form* $(f' \otimes g')\mathsf{R}$ *for some Borel functions* $f', g' : X \to [0, +\infty[$*.*

Proof.
(a) First of all, by Proposition (1.13), $\Gamma(\mu_0, \mu_1)$ is weakly compact. Furthermore, by Proposition (1.25), H_R is weakly lower semicontinuous, so there exists $\gamma \in \Gamma(\mu_0, \mu_1)$ such that

$$H_\mathsf{R}(\gamma) = \min_{\pi \in \Gamma(\mu_0, \mu_1)} H_\mathsf{R}(\pi).$$

In particular,

$$H_\mathsf{R}(\gamma) \leq H_\mathsf{R}(\mu_0 \otimes \mu_1) < +\infty,$$

so $\gamma \ll \mathsf{R}$ and, using Proposition (1.23), H_R is strictly convex where it is finite, so the uniqueness of γ comes.

For the second part of the statement, let us underline that the uniqueness property of (f, g) comes from the definition of $f \otimes g$. In particular, it is sufficient to prove the existence of an appropriate pair (f, g).

Now, remembering that $\mathfrak{m} \otimes \mathfrak{m} \ll \mathsf{R} \ll \mathfrak{m} \otimes \mathfrak{m}$, let us prove that

$$\frac{d\gamma}{d\mathsf{R}} > 0 \ \mathfrak{m} \otimes \mathfrak{m}\text{-a.e. in } P_0 \times P_1.$$

By contradiction, suppose

$$\mathsf{R}\left((P_0 \times P_1) \cap \left\{x \in X : \frac{d\gamma}{d\mathsf{R}} = 0\right\}\right) > 0.$$

Given $\lambda \in]0, 1[$, using the convexity of h,

$$\frac{h\left(\frac{d\gamma}{d\mathsf{R}} + \lambda\left(\frac{d(\mu_0 \otimes \mu_1)}{d\mathsf{R}} - \frac{d\gamma}{d\mathsf{R}}\right)\right) - h\left(\frac{d\gamma}{d\mathsf{R}}\right)}{\lambda} \leq h\left(\frac{d(\mu_0 \otimes \mu_1)}{d\mathsf{R}}\right) - h\left(\frac{d\gamma}{d\mathsf{R}}\right) \in L^1(X \times X, \mathsf{R})$$

and as $\lambda \to 0^+$

$$\frac{h\left(\frac{d\gamma}{d\mathsf{R}} + \lambda\left(\frac{d(\mu_0 \otimes \mu_1)}{d\mathsf{R}} - \frac{d\gamma}{d\mathsf{R}}\right)\right) - h\left(\frac{d\gamma}{d\mathsf{R}}\right)}{\lambda} \searrow -\infty \ \mathsf{R}\text{-a.e. in } (P_0 \times P_1) \cap \left\{x \in X : \frac{d\gamma}{d\mathsf{R}} = 0\right\},$$

so by the monotone convergence Theorem

$$\lim_{\lambda \to 0^+} \frac{H_{\mathsf{R}}(\gamma + \lambda(\mu_0 \otimes \mu_1 - \gamma)) - H_{\mathsf{R}}(\gamma)}{\lambda} = -\infty \text{ in } (P_0 \times P_1) \cap \left\{x \in X : \frac{d\gamma}{d\mathsf{R}} = 0\right\},$$

but the minimality of γ ensures

$$H_{\mathsf{R}}(\gamma + \lambda(\mu_0 \otimes \mu_1 - \gamma)) - H_{\mathsf{R}}(\gamma) \geq 0 \text{ in } (P_0 \times P_1) \cap \left\{x \in X : \frac{d\gamma}{d\mathsf{R}} = 0\right\},$$

because $\gamma + \lambda(\mu_0 \otimes \mu_1 - \gamma) \in \Gamma(\mu_0, \mu_1)$, so using the sign permanence Theorem, it must be

$$\lim_{\lambda \to 0^+} \frac{H_{\mathsf{R}}(\gamma + \lambda(\mu_0 \otimes \mu_1 - \gamma)) - H_{\mathsf{R}}(\gamma)}{\lambda} \geq 0 \text{ in } (P_0 \times P_1) \cap \left\{x \in X : \frac{d\gamma}{d\mathsf{R}} = 0\right\},$$

that provides the contradiction.

Now, consider the set

$$U = \left\{u \in L^\infty(X \times X, \gamma) : (p^1)_\#(u\gamma) = (p^2)_\#(u\gamma) = 0\right\}.$$

Take $u \in U$ and $\varepsilon \in \left]0, \frac{1}{\|u\|_{L^\infty(X\times X,\gamma)}}\right[$. Observing that $(1+\varepsilon u)\gamma \in \Gamma(\mu_0,\mu_1)$ and $u\frac{d\gamma}{d\mathsf{R}}$ is well defined R-a.e. in $X \times X$, one has

$$\begin{aligned}
\left\| h\left((1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\right)\right\|_{L^1(X\times X,\mathsf{R})} &= \int_{X\times X}\left|(1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\log\left((1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\right)\right| d\mathsf{R} = \\
&= \int_{X\times X}\left|(1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\log(1+\varepsilon u) + (1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\log\left(\frac{d\gamma}{d\mathsf{R}}\right)\right| d\mathsf{R} \leq \\
&\leq \int_{X\times X}\left|(1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\log(1+\varepsilon u)\right| d\mathsf{R} + \\
&+ \int_{X\times X}\left|(1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\log\left(\frac{d\gamma}{d\mathsf{R}}\right)\right| d\mathsf{R}, \\
&\leq \int_{X\times X}(1+\varepsilon u)|\log(1+\varepsilon u)|d\gamma + \int_{X\times X}(1+\varepsilon u)\left|\log\left(\frac{d\gamma}{d\mathsf{R}}\right)\right| d\gamma \\
&\leq \|(1+\varepsilon u)\log(1+\varepsilon u)\|_{L^\infty(X\times X,\gamma)} + \\
&+ \|1+\varepsilon u\|_{L^\infty(X\times X,\gamma)}\int_{X\times X}\left|\log\left(\frac{d\gamma}{d\mathsf{R}}\right)\right| d\gamma \leq \\
&\leq \|(1+\varepsilon u)\log(1+\varepsilon w)\|_{L^\infty(X\times X,\gamma)} + \\
&+ \|1+\varepsilon u\|_{L^\infty(X\times X,\gamma)}\left\|\frac{d\gamma}{d\mathsf{R}}\log\left(\frac{d\gamma}{d\mathsf{R}}\right)\right\|_{L^1(X\times X,\mathsf{R})} < +\infty,
\end{aligned}$$

so $h\left((1+\varepsilon u)\frac{d\gamma}{d\mathsf{R}}\right) \in L^1(X\times X,\mathsf{R})$. In the same way as before, using the monotone convergence Theorem,

$$\lim_{\varepsilon\to 0^+}\frac{H_{\mathsf{R}}((1+\varepsilon u)\gamma) - H_{\mathsf{R}}(\gamma)}{\varepsilon} = \int_{X\times X} u\frac{d\gamma}{d\mathsf{R}}\left(\log\left(\frac{d\gamma}{d\mathsf{R}}\right)+1\right)d\mathsf{R}.$$

Using again minimality of γ and the sign permanence Theorem,

$$\int_{X\times X} u\frac{d\gamma}{d\mathsf{R}}\left(\log\left(\frac{d\gamma}{d\mathsf{R}}\right)+1\right)d\mathsf{R} \geq 0.$$

In particular, since

$$\int_{X\times X} u\frac{d\gamma}{d\mathsf{R}}d\mathsf{R} = \int_{X\times X} u\,d\gamma = 0,$$

we have

$$\int_{X\times X} u\frac{d\gamma}{d\mathsf{R}}\log\left(\frac{d\gamma}{d\mathsf{R}}\right)d\mathsf{R} \geq 0.$$

Up to switching u with $-u$, we obtain

$$\int_{X\times X} u \frac{d\gamma}{d\mathsf{R}} \log\left(\frac{d\gamma}{d\mathsf{R}}\right) d\mathsf{R} \le 0,$$

then

$$\int_{X\times X} u \frac{d\gamma}{d\mathsf{R}} \log\left(\frac{d\gamma}{d\mathsf{R}}\right) d\mathsf{R} = 0, \text{ for every } u \in U,$$

or, in other words, $\log\left(\frac{d\gamma}{d\mathsf{R}}\right) \in {}^{\perp}U$. Consider the subspace of $L^1(X\times X,\gamma)$

$$V = \left\{f \in L^1(X\times X,\gamma) : f = \varphi \oplus \psi, \varphi \in L^0(X,\mathfrak{m}\lfloor P_0), \psi \in L^0(X,\mathfrak{m}\lfloor P_1)\right\}.$$

It is sufficient to prove that ${}^{\perp}U \subseteq V$.

Let us start showing that V is closed in $L^1(X\times X,\gamma)$. To show this, we can initially observe that

$$\begin{aligned} f \in V \Longleftrightarrow \Big(& f \in L^1(X\times X,\gamma) \text{ and } f(x,y) + f(x',y') = f(x,y') + f(x',y) \\ & \text{for } \mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\text{-a.e. } (x,x',y,y') \in P_0\times P_0\times P_1\times P_1\Big). \end{aligned}$$

This comes directly as follows: if $f \in V$, clearly $f \in L^1(X\times X,\gamma)$ and there exist $\varphi \in L^0(X,\mathfrak{m}\lfloor P_0), \psi \in L^0(X,\mathfrak{m}\lfloor P_1)$ such that

$$\begin{aligned} f(x,y) &= \varphi(x) + \psi(y), \ \mathfrak{m}\otimes\mathfrak{m}\text{-a.e. } (x,y) \in P_0\times P_1 \\ f(x',y') &= \varphi(x') + \psi(y'), \ \mathfrak{m}\otimes\mathfrak{m}\text{-a.e. } (x',y') \in P_0\times P_1 \end{aligned}$$

so, summing both sides,

$$f(x,y)+f(x',y') = f(x,y')+f(x',y), \ \mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\text{-a.e. } (x,x',y,y') \in P_0\times P_0\times P_1\times P_1;$$

conversely, if $f \in L^1(X\times X,\gamma)$ and

$$f(x,y)+f(x',y') = f(x,y')+f(x',y), \ \mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\text{-a.e. } (x,x',y,y') \in P_0\times P_0\times P_1\times P_1,$$

by Fubini's Theorem, the function $\{x \mapsto f(x,y)\}$ (resp. $\{y \mapsto f(x,y)\}$) is Borel for $\mathfrak{m}$-a.e. $y \in P_1$ (resp. $\mathfrak{m}$-a.e. $x \in P_0$) so, setting $\varphi(x) = f(x,y')$, $\psi(y) = f(x',y)$ for some admissible $(x',y') \in P_0\times P_1$, it results $f = \varphi + \psi$, then $f \in V$. By the fact that the condition

$$f(x,y)+f(x',y') = f(x,y')+f(x',y), \ \mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\otimes\mathfrak{m}\text{-a.e. } (x,x',y,y') \in P_0\times P_0\times P_1\times P_1$$

is stable with respect to $L^1(X \times X, \gamma)$ convergence, we can affirm that V is closed in $L^1(X \times X, \gamma)$.

Let us now prove that $V^\perp \subseteq U$. If $\tilde{u} \in L^\infty(X \times X, \gamma) \setminus U$, one of the marginals of $\tilde{u}\gamma$ is non-zero. If it is the first marginal to be non-zero, by the fact that $(p_1)_\# \gamma = \mu_0$, for every Borel function $\vartheta : X \to \mathbb{R}$, using Corollary (P.17),

$$\begin{aligned}\int_{X\times X} (\vartheta \circ p^1)(x,y) d(\tilde{u}\gamma)(x,y) &= \int_{X\times X} \vartheta(x) d(\tilde{u}\gamma)(x,y) = \\ &= \int_{X\times X} \vartheta(x)\tilde{u}(x,y) d\gamma(x,y) = \\ &= \int_X \left(\int_X \vartheta(x)\tilde{u}(x,y) d\gamma_x(y) \right) d\mu_0(x) = \\ &= \int_X \vartheta(x) f_0(x) d\mu_0(x) = \int_X \vartheta(x) d(f_0\mu_0)(x),\end{aligned}$$

where

$$f_0(x) = \int_X \vartheta(x)\tilde{u}(x,y) d\gamma_x(y),$$

that is, by Proposition (P.8), $(p^1)_\#(\tilde{u}\gamma) = f_0\mu_0$. In particular, $f_0 \oplus 0 = f_0 \circ p_1 \in V$ and

$$\int_X \tilde{u}(f_0 \oplus 0) d\gamma = \int_X f_0 \circ p^1 d(\tilde{u}\gamma) = \int_X f_0 d(p^1)_\#(\tilde{u}\gamma) = \int_X f_0^2 d\mu_0 > 0,$$

so $\tilde{u} \notin V^\perp$.

We can finally prove that ${}^\perp U \subseteq V$. If $f \in L^1(X \times X, \gamma) \setminus V$, by the fact that V is closed, applying [9, Corollary 1.8], we can find $u \in (L^1(X \times X, \gamma))' \cong L^\infty(X \times X, \gamma)$ such that for every $\hat{f} \in V$

$$\int_X \hat{f} u d\gamma = 0$$

and

$$\int_X f u d\gamma = 1. \tag{1.29}$$

In particular, $u \in V^\perp \subseteq U$ and, by (1.29), $f \notin {}^\perp U$.

(b) First of all, using our assumptions,

$$\begin{aligned}\int_{X\times X} e^{-B(x)-B(y)} d(\mathfrak{m} \otimes \mathfrak{m})\lfloor(P_0 \times P_1) &= \int_{P_0\times P_1} e^{-B(x)-B(y)} d(\mathfrak{m} \otimes \mathfrak{m}) \leq \\ &\leq \frac{1}{c} \int_{P_0\times P_1} e^{-B(x)-B(y)} d\mathsf{R} \leq\end{aligned}$$

$$\leq \frac{1}{c}\int_{X\times X} e^{-B(x)-B(y)} d\mathsf{R} < +\infty.$$

Now, we know that $\mu_0, \mu_1 \ll \mathfrak{m}$, so $\mu_0 \otimes \mu_1 \ll \mathfrak{m} \otimes \mathfrak{m}$. If $A \in \mathscr{B}(X \times X)$ such that

$$\frac{1}{\int_X e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)} \int_A e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m}) = 0,$$

by strict positivity of the integrand, it must be $(\mathfrak{m}\otimes\mathfrak{m})(A) = 0$, so

$$\mathfrak{m}\otimes\mathfrak{m} \ll \frac{e^{-B(x)-B(y)}}{\int_X e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)}(\mathfrak{m}\otimes\mathfrak{m}).$$

In particular,

$$(\mu_0 \otimes \mu_1) \ll \frac{e^{-B(x)-B(y)}}{\int_X e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)$$

and, from this, we have

$$H^p\left(\mu_0 \otimes \mu_1 \mid \frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)\right) =$$
$$= \int_{X\times X} \log\left(\frac{d(\mu_0\otimes\mu_1)}{d\left(\frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)\right)}\right) d(\mu_0\otimes\mu_1),$$

that is

$$\begin{aligned}
&H^p\left(\mu_0 \otimes \mu_1 \mid \frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)\right) = \\
&= \int_{X\times X} \log\left(\frac{d\mu_0}{d\mathfrak{m}\lfloor P_0}\right) d(\mu_0\otimes\mu_1) + \int_{X\times X} \log\left(\frac{d\mu_1}{d\mathfrak{m}\lfloor P_1}\right) d(\mu_0\otimes\mu_1) + \\
&- \int_{X\times X} \log\left(\frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)}}\right) d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1) d(\mu_0\otimes\mu_1) = \\
&= \int_X \log\left(\varrho_0|_{P_0}\right) d\mu_0 + \int_X \log\left(\varrho_1|_{P_1}\right) d\mu_1 + \int_X B d\mu_0 + \int_X B d\mu_1 + \\
&+ \log\left(\int_{X\times X} e^{-B(x)-B(y)} d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0 \times P_1)\right) < +\infty,
\end{aligned}$$

hence, using [2, Lemma 7.2],

$$H(\mu_0 \otimes \mu_1 \mid \mathsf{R}) =$$

$$= H^p\left(\mu_0\otimes\mu_1 \mid \frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)}d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)\right)+$$
$$+\int_{X\times X}\log\frac{d\left(\frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)}d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)\right)}{d\mathsf{R}}d(\mu_0\otimes\mu_1)=$$
$$= H^p\left(\mu_0\otimes\mu_1 \mid \frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)}d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)}(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)\right)+$$
$$+\int_{X\times X}\log\left(\frac{e^{-B(x)-B(y)}}{\int_{X\times X} e^{-B(x)-B(y)}d(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)}\right)d(\mu_0\otimes\mu_1)+$$
$$+\int_{X\times X}\log\left(\frac{(\mathfrak{m}\otimes\mathfrak{m})\lfloor(P_0\times P_1)}{d\mathsf{R}}\right)d(\mu_0\otimes\mu_1)<+\infty.$$

For the second part of the statement, consider $\sigma\in\Gamma(\mu_0,\mu_1)$ such that $\sigma=(f'\otimes g')\mathsf{R}$ for some Borel functions $f,g : X\to[0,+\infty[$. Take $\zeta\in C_c^\infty(X)$. By the fact that $(p^1)_\#((f'\otimes g')\mathsf{R})=\mu_0=\varrho_0\mathfrak{m}$, $(p^1)_\#\mathsf{R}=\mathfrak{m}$ and using Corollary (P.17),

$$\begin{aligned}\int_X\zeta\varrho_0 d\mathfrak{m} = \int_X\zeta d\mu_0 &= \int_X \zeta d(p^1)_\#((f'\otimes g')\mathsf{R}) =\\ &=\int_{X\times X}\zeta(x)d(f'\otimes g')\mathsf{R}=\int_{X\times X}\zeta(x)f'(x)g'(y)d\mathsf{R}=\\ &=\int_X\zeta(x)f'(x)\left(\int_X g'(y)d\mathsf{R}_x(y)\right)d\mathfrak{m}(x),\end{aligned}$$

so

$$\int_X\left(f'(x)\left(\int_X g'(y)d\mathsf{R}_x(y)\right)-\varrho_0(x)\right)\zeta(x)d\mathfrak{m}(x)=0$$

and by Du Bois-Reymond Lemma's,

$$f'(x)\left(\int_X g'(y)d\mathsf{R}_x(y)\right)=\varrho_0(x)<+\infty\ \mathfrak{m}\text{-a.e. } x\in X.$$

In particular, f' vanishes $\mathfrak{m}$-a.e. in $X\setminus P_0$. Since, by assumption, $\mathsf{R}_x\geq c\mathfrak{m}$ in P_1 for $\mathfrak{m}$-a.e. $x\in P_0$, taken a suitable $x\in P_0$,

$$c\|g'\|_{L^1(X,\mathfrak{m})}=\int_X g'd(c\mathfrak{m})\leq\int_X g'd\mathsf{R}_x$$

so $g'\in L^1(X,\mathfrak{m})$ and

$$f'(x)\leq\frac{\|\varrho_0\|_{L^\infty(X,\mathfrak{m})}}{\int_X g'd\mathsf{R}_x}\leq\frac{\|\varrho_0\|_{L^\infty(X,\mathfrak{m})}}{c\|g'\|_{L^1(X,\mathfrak{m})}}.$$

By the fact that f' vanish $\mathfrak{m}$-a.e. in $X \setminus P_0$, we obtain

$$\|f'\|_{L^\infty(X,\mathfrak{m})}\|g'\|_{L^1(X,\mathfrak{m})} \leq \frac{\|\varrho_0\|_{L^\infty(X,\mathfrak{m})}}{c}.$$

Using p^2 in place of p^1, we obtain also

$$\|f'\|_{L^1(X,\mathfrak{m})}\|g'\|_{L^\infty(X,\mathfrak{m})} \leq \frac{\|\varrho_1\|_{L^\infty(X,\mathfrak{m})}}{c}.$$

In particular, $\log f'$ and $\log g'$ are bounded from above.

By the fact that H_{R} is well defined on $\Gamma(\mu_0, \mu_1)$,

$$\int_{X\times X} (\log f' \oplus \log g') d\sigma = H(\sigma \mid \mathsf{R}) > -\infty$$

so $\log f' \circ p^1, \log g' \circ p^2 \in L^1(X \times X, \sigma)$. Now, as already done before, using also the fact that γ and σ have the same marginals,

$$\begin{aligned}\lim_{\lambda\to 0^+} \frac{H(\sigma + \lambda(\gamma-\sigma) \mid \mathsf{R}) - H(\sigma \mid \mathsf{R})}{\lambda} &= \int_{X\times X} (f \otimes g - f' \otimes g') \log f' \otimes g' d\mathsf{R} = \\ &= \int_{X\times X} (\log f' \oplus \log g') d(\gamma - \sigma) = \\ &= \int_X \log f' d(p^1)_\#(\gamma-\sigma) + \\ &+ \int_X \log g' d(p^2)_\#(\gamma - \sigma) = 0.\end{aligned}$$

By convexity of H_{R}, $H(\sigma \mid \mathsf{R}) \leq H(\gamma \mid \mathsf{R})$ and the uniqueness of the minimizer for H_{R} provides $\sigma = \gamma$. ∎

5 Comparison

We want now to analyze the connection between the static representations of the two problems: the goal of the section is to emphasize, at least heuristically, the bond between the two problems using their static representation formulas. What we can, preliminarily, say is that both are interpolation problems defined as minimization problems of convex functionals on the same convex domain. It is, however, possible to highlight a much deeper connection between the two problems.

Consider a Polish space (X, d), $\mathfrak{m} \in \mathcal{R}_+(X)$ and, for every $\varepsilon > 0$, $\mathsf{R}_\varepsilon \in \mathcal{R}_+(X \times X)$ such that $(p_1)_\#\mathsf{R}_\varepsilon = (p_2)_\#\mathsf{R}_\varepsilon = \mathfrak{m}$ and $\mathfrak{m} \otimes \mathfrak{m} \ll \mathsf{R}_\varepsilon \ll \mathfrak{m} \otimes \mathfrak{m}$.

Consider a Borel function $c : X \times X \to [0, +\infty[$ lower semicontinuous such that for every $x \in X$ the function $\{y \longmapsto c(x, y)\}$ is coercive. Suppose that for every $x \in X$, $\{\mathsf{R}_\varepsilon\}_{\varepsilon>0}$ satisfies a large deviation principle with rate function $\{y \longmapsto c(x, y)\}$, namely for every $U, C \subseteq X$, with U open and C closed,

$$\liminf_{\varepsilon\to 0} \varepsilon \log \mathsf{R}_\varepsilon(U) \geq -\inf_{y\in U} c(x, y), \qquad \limsup_{\varepsilon\to 0} \varepsilon \log \mathsf{R}_\varepsilon(C) \leq -\inf_{y\in C} c(x, y).$$

Fix $\mu \in \mathcal{P}(X)$ such that $\mu \ll \mathfrak{m}$. Consider for every $\nu \in \mathcal{P}(X)$ such that $\nu \ll \mathfrak{m}$ and such that there exists a Borel function $B : X \to [0, +\infty[$ such that

$$\int_{X\times X} e^{-B(x)-B(y)} d\mathsf{R}(x, y) < +\infty, \quad \int_X B(x)d\mu_0(x) < +\infty, \quad \int_X B(x)d\mu_1(x) < +\infty,$$

the functionals

$$S_\varepsilon^{(\mu)}(\nu) = \inf \{\varepsilon H_{\mathsf{R}_\varepsilon}(\pi) : \pi \in \Gamma(\mu, \nu)\}$$

and

$$T^{(\mu)}(\nu) = \inf \{C(\pi) : \pi \in \Gamma(\mu, \nu)\}.$$

It results

$$\Gamma - \lim_{\varepsilon\to 0} S_\varepsilon^{(\mu)} = T^{(\mu)},$$

that can be read as "*optimal transport problem is the* Γ*-limit of Schrödinger problem*". In particular, the limit of the sequence of the solutions of (SP) is a solution of (OT). At the moment, it is still quite a complex problem to understand which of the solutions actually is: some works linked to this question are [21] and [4].

Let us analyze, at least intuitively, a specific example.

(1.30) Example *Consider* $\mathbb{R}^n$*, endowed with Euclidean distance. Given* $\varepsilon > 0$*, consider* $\mathsf{R}_{\frac{\varepsilon}{2}} = \mathsf{r}_{\frac{\varepsilon}{2}} \mathcal{L}^n \otimes \mathcal{L}^n$*, where*

$$\mathsf{r}_t(x, y) = \frac{1}{\sqrt{(4\pi t)^n}} e^{-\frac{|x-y|^2}{4t}}.$$

Given $\mu, \nu \in \mathcal{P}(X)$ *such that* $\mu, \nu \ll \mathcal{L}^n$*, the choice* $B = 0$ *guarantees the well-definiteness of* $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$ *on* $\Gamma(\mu, \nu)$*. A simple computation provides*

$$H_{\mathsf{R}_{\frac{\varepsilon}{2}}}(\sigma) = \int \log \left(\frac{d\sigma}{\mathsf{R}_{\frac{\varepsilon}{2}}}\right) d\sigma = \int \log \left(\frac{d\sigma}{d(\mu\otimes\nu)} \frac{d(\mu\otimes\nu)}{d(\mathcal{L}^n\otimes\mathcal{L}^n)} \frac{d(\mathcal{L}^n\otimes\mathcal{L}^n)}{d\mathsf{R}_{\frac{\varepsilon}{2}}}\right) d\sigma =$$
$$= \int \log \left(\frac{d\sigma}{d(\mu\otimes\nu)}\right) d\sigma + \int \log \left(\frac{d(\mu\otimes\nu)}{d(\mathcal{L}^n\otimes\mathcal{L}^n)}\right) d\sigma + \int \log \left(\frac{d(\mathcal{L}^n\otimes\mathcal{L}^n)}{d\mathsf{R}_{\frac{\varepsilon}{2}}}\right) d\sigma.$$

Now, as regards the first term, we recognize

$$\int \log\left(\frac{d\sigma}{d(\mu\otimes\nu)}\right)d\sigma = H(\sigma|\mu\otimes\nu).$$

For the second term,

$$\begin{aligned}\int \log\left(\frac{d(\mu\otimes\nu)}{d(\mathcal{L}^n\otimes\mathcal{L}^n)}\right)d\sigma &= \int \log\left(\frac{d\mu}{d\mathcal{L}^n}(x)\frac{d\nu}{d\mathcal{L}^n}(y)\right)d\sigma(x,y) = \\ &= \int \log\left(\frac{d\mu}{d\mathcal{L}^n}(x)\right)d\sigma(x,y) + \int \log\left(\frac{d\nu}{d\mathcal{L}^n}(y)\right)d\sigma(x,y)\end{aligned}$$

so, using the change of variables formula,

$$\int \log\left(\frac{d(\mu\otimes\nu)}{d(\mathcal{L}^n\otimes\mathcal{L}^n)}\right)d\sigma = \int \log\left(\frac{d\mu}{d\mathcal{L}^n}\right)d\mu + \int \log\left(\frac{d\nu}{d\mathcal{L}^n}\right)d\nu = H(\mu|\mathcal{L}^n) + H(\nu|\mathcal{L}^n).$$

For the third term, observing that

$$\mathcal{L}^n\otimes\mathcal{L}^n = \sqrt{(2\pi\varepsilon)^n}e^{\frac{|x-y|^2}{2\varepsilon}}\mathsf{R}_{\frac{\varepsilon}{2}},$$

we can write

$$\begin{aligned}\int \log\left(\frac{d(\mathcal{L}^n\otimes\mathcal{L}^n)}{d\mathsf{R}_{\frac{\varepsilon}{2}}}\right)d\sigma &= \int \log\left(\sqrt{(2\pi\varepsilon)^n}e^{\frac{|x-y|^2}{2\varepsilon}}\right)d\sigma = \\ &= \frac{n}{2}\log(2\pi\varepsilon)\,\sigma(X\times X) + \frac{1}{2\varepsilon}\int |x-y|^2 d\sigma(x,y) = \\ &= \frac{n}{2}\log(\pi\varepsilon) + \frac{1}{2\varepsilon}C(\sigma).\end{aligned}$$

Resuming,

$$H_{\mathsf{R}_{\frac{\varepsilon}{2}}}(\sigma) = H(\sigma|\mu\otimes\nu) + H(\mu|\mathcal{L}^n) + H(\nu|\mathcal{L}^n) + \frac{n}{2}\log(2\pi\varepsilon) + \frac{1}{2\varepsilon}C(\sigma)$$

and multiplying both sides by ε we get

$$\varepsilon H_{\mathsf{R}_{\frac{\varepsilon}{2}}}(\sigma) = \varepsilon H(\pi|\mu\otimes\nu) + \varepsilon H(\mu|\mathcal{L}^n) + \varepsilon H(\nu|\mathcal{L}^n) + \frac{n}{2}\varepsilon\log(2\pi\varepsilon) + \frac{1}{2}C(\sigma).$$

Passing to the limit as $\varepsilon\to 0$, on the right-hand side we expect that only the last term survives. This provides, at least intuitively, the Γ-limit mentioned above in the abstract setting. As regards the pointwise limit, this latter is clear from the calculations we did.

To formalize the previous discussion, the interested reader has, first of all, to learn how to manage with Γ-convergence. Further details can be found in [39], [15] and [14].

Chapter 2

Dual representations

1 Kantorovich–Rubinstein duality

The Kantorovich problem is a convex constrained optimization problem, so it admits a dual representation. Let us, first of all, understand in what sense with an example.

(2.1) Example *Consider again the situation described in Example* (1.5). *If buying and selling are handled by a broker, who buys at price φ_i from x_i and sells to y_j at ψ_j, to be competitive with respect to the direct selling between mills and bakeries, he must be sure that $\psi_j - \varphi_i \leq c(x_i, y_j)$. In this assumption, the broker is interested in maximizing the profit, that is*

$$\sum_{j=1}^{m} \psi_j \nu_j - \sum_{i=1}^{n} \varphi_i \mu_i .$$

The abstract formulation is then the following.

(2.2) Definition *Let $(X, d_X), (Y, d_Y)$ be two metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a Borel function $c : X \times Y \to [0, +\infty[$ called* cost function. *In the so-called* dual optimal transport problem *we look for*

$$\sup_{(\varphi,\psi)\in I_c} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \right\}$$

where

$$I_c = \{(\varphi, \psi) \in \mathrm{Lip}_b(X) \times \mathrm{Lip}_b(Y) : \varphi + \psi \leq c\} .$$

We call Kantorovich potentials *every maximizing couple of the above problem.*

The aim of this section is to show the equivalence of primal and dual optimal transport problems in a Polish setting with lower semicontinuous cost. The proof will be based essentially on this Proposition.

(2.3) Proposition *Let* $(X, d_X), (Y, d_Y)$ *be Polish spaces,* $\mu \in \mathcal{P}(X)$*,* $\nu \in \mathcal{P}(Y)$*, a continuous function* $c : X \times Y \to [0, +\infty[$ *and* $\pi \in \Gamma(\mu, \nu)$ *an optimal plan such that*

$$\int_{X \times Y} c d\pi < +\infty.$$

Then $\operatorname{supt}(\pi)$ *is c-cyclically monotone.*

Proof. See [1, Theorem 3.17]. ■

(2.4) Theorem (Kantorovich–Rubinstein duality) *Let* $(X, d_X), (Y, d_Y)$ *be Polish spaces,* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$*. If* $c : X \times Y \to [0, +\infty[$ *is lower semicontinuous, then*

$$\min \left\{ C(\pi) = \int_{X \times Y} c(x, y) d\pi(x, y) : \pi \in \Gamma(\mu, \nu) \right\} = \sup_{(\varphi, \psi) \in I_c} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \right\}.$$

If, in addition, $c \in \operatorname{Lip}_b(X \times Y)$*, then the supremum is attained and Kantorovich potentials are of the form* (φ, φ^c) *for some c-concave function* φ*.*

Proof. First of all, by Theorem (1.14), there exists an optimal plan $\pi_0 \in \Gamma(\mu, \nu)$, namely

$$C(\pi_0) = \min_{\pi \in \Gamma(\mu, \nu)} C(\pi).$$

Using the indicator function, we can rewrite the problem as

$$C(\pi_0) = \inf_{\pi \in \mathcal{P}(X \times Y)} \left\{ \int_{X \times Y} c d\pi + \iota_{\Gamma(\mu, \nu)}(\pi) \right\}.$$

Consider the set

$$\mathfrak{B} = \{(\varphi, \psi) : \varphi : X \to \mathbb{R} \text{ Borel}, \psi : Y \to \mathbb{R} \text{ Borel}\}.$$

Let us prove that for every $\pi \in \mathcal{P}(X \times Y)$

$$\iota_{\Gamma(\mu, \nu)}(\pi) = \sup_{(\varphi, \psi) \in \mathfrak{B}} \left\{ \int_X \varphi d\mu - \int_{X \times Y} \varphi d\pi + \int_Y \psi d\nu - \int_{X \times Y} \psi d\pi \right\}.$$

If $\pi \in \Gamma(\mu, \nu)$, marginal conditions provide

$$\sup_{(\varphi, \psi) \in \mathfrak{B}} \left\{ \int_X \varphi d\mu - \int_{X \times Y} \varphi d\pi + \int_Y \psi d\nu - \int_{X \times Y} \psi d\pi \right\} = 0.$$

On the other hand, if $\pi \in \mathcal{P}(X \times Y) \setminus \Gamma(\mu, \nu)$, one of the marginal conditions fails. Without loss of generality, we can think that $(p_X)_{\#}\pi \neq \mu$ so there exists $B \in \mathscr{B}(X)$ such that $(p_X)_{\#}\pi(B) \neq \mu(B)$. If $(p_X)_{\#}\pi(B) < \mu(B)$, let $\varphi_n = n\chi_B$. It holds

$$\int_X \varphi_n d\mu - \int_{X \times Y} \varphi_n d\pi = n\left(\mu(B) - \pi(B \times Y)\right) = n(\mu(B) - (p_X)_{\#}\pi(B))$$

and, letting $n \to +\infty$,

$$\sup_{(\varphi,\psi)\in\mathfrak{B}} \left\{\int_X \varphi d\mu - \int_{X\times Y} \varphi d\pi + \int_Y \psi d\nu - \int_{X\times Y} \psi d\pi\right\} = +\infty.$$

The case $(p_X)_{\#}\pi(B) > \mu(B)$ is analogous: we just have to use $\varphi_n = -n\chi_B$.

Now,

$$\begin{aligned}
\min_{\pi\in\Gamma(\mu,\nu)} C(\pi) = \inf_{\pi\in\mathcal{P}(X\times Y)} & \left\{\int_{X\times Y} c d\pi + \right. \\
& \left. + \sup_{(\varphi,\psi)\in\mathfrak{B}} \left\{\int_X \varphi d\mu - \int_{X\times Y} \varphi d\pi + \int_Y \psi d\nu - \int_{X\times Y} \psi d\pi\right\}\right\} = \\
& = \inf_{\pi\in\mathcal{P}(X\times Y)} \sup_{(\varphi,\psi)\in\mathfrak{B}} \left\{\int_{X\times Y} (c - \varphi - \psi) d\pi + \int_X \varphi d\mu + \int_Y \psi d\nu\right\} \geq \\
& \geq \sup_{(\varphi,\psi)\in\mathfrak{B}} \inf_{\pi\in\mathcal{P}(X\times Y)} \left\{\int_{X\times Y} (c - \varphi - \psi) d\pi + \int_X \varphi d\mu + \int_Y \psi d\nu\right\} \geq \\
& \geq \sup_{(\varphi,\psi)\in I_c} \inf_{\pi\in\mathcal{P}(X\times Y)} \left\{\int_{X\times Y} (c - \varphi - \psi) d\pi + \int_X \varphi d\mu + \int_Y \psi d\nu\right\}
\end{aligned}$$

and, by the fact that $c \geq \varphi + \psi$ in I_c, we obtain

$$\min_{\pi\in\Gamma(\mu,\nu)} C(\pi) \geq \sup_{(\varphi,\psi)\in I_c} \left\{\int_X \varphi d\mu + \int_Y \psi d\nu\right\}.$$

In order to prove the converse inequality, let us consider firstly the case $c \in \mathrm{Lip}_b(X \times Y)$. By Proposition (2.3), $\mathrm{supt}(\pi_0)$ is c-cyclically monotone and, like in Theorem (P.32), fixed $(x_0, y_0) \in \mathrm{supt}(\pi)$, the function φ defined by

$$\begin{aligned}
\varphi(x) = \inf_{\substack{n\in\mathbb{N}\setminus\{0\} \\ (x_1,y_1),\dots,(x_n,y_n)\in\mathrm{supt}(\pi_0)}} & (c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \dots + \\
& + c(x_1, y_0) - c(x_0, y_0))
\end{aligned}$$

is a c-concave Lipschitz and bounded from above function such that $\varphi + \varphi^c = c$ and

$\varphi(x_0) = 0$. By the fact that

$$\varphi^c(y) \le c(x_0, y) - \varphi(x_0) \le \sup c,$$

also φ^c is bounded from above. Now,

$$\varphi(x) = (\varphi^c)^c(x) = \inf_{y \in Y} c(x, y) - \varphi^c(y) \ge \inf c - \sup \varphi^c$$

so φ is also bounded from below. Similarly, φ^c is bounded from below. Consequently, φ and φ^c are bounded Lipschitz functions. Finally,

$$\begin{aligned} \min_{\pi \in \Gamma(\mu,\nu)} C(\pi) &= \int_{X \times Y} c d\pi_0 = \int_{\mathrm{supt}(\pi_0)} c d\pi_0 = \int_{\mathrm{supt}(\pi_0)} (\varphi + \varphi^c) d\pi_0 = \\ &= \int_{\mathrm{supt}(\pi_0)} \varphi d\pi_0 + \int_{\mathrm{supt}(\pi_0)} \varphi^c d\pi_0 \le \sup_{(\varphi,\psi) \in I_c} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \right\}. \end{aligned}$$

In particular, equality holds and (φ, φ^c) are Kantorovich potentials.

In the general case, consider $(\mathcal{I}_n c)$, the inf-convolution of c. By the fact that $\mathcal{I}_n c \in \mathrm{Lip}_b(X \times Y)$ for every $n \in \mathbb{N}$, using the previous case,

$$\min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} \mathcal{I}_n c d\pi = \sup_{(\varphi,\psi) \in I_{\mathcal{I}_n c}} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \right\}.$$

In particular, using the fact that $\mathcal{I}_n c \le c$,

$$\min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} \mathcal{I}_n c d\pi \le \sup_{(\varphi,\psi) \in I_c} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu \right\}. \tag{2.5}$$

Now, let (π_n) in $\Gamma(\mu, \nu)$ a sequence of optimal plans for $(\mathcal{I}_n c)$. By Proposition (1.13), up to a subsequence, $\pi_n \rightharpoonup \overline{\pi}$ for some $\overline{\pi} \in \Gamma(\mu, \nu)$. If $m \in \mathbb{N}$ and $n \ge m$,

$$\min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} \mathcal{I}_n c d\pi = \int_{X \times Y} \mathcal{I}_n c d\pi_n \ge \int_{X \times Y} \mathcal{I}_m c d\pi_n,$$

so passing to the lim inf as $n \to +\infty$, by Lemma (P.19),

$$\liminf_n \min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} \mathcal{I}_n c d\pi \ge \liminf_n \int_{X \times Y} \mathcal{I}_m c d\pi_n = \int_{X \times Y} \mathcal{I}_m c d\overline{\pi}$$

and if $m \to +\infty$ we obtain, using the monotone convergence Theorem,

$$\liminf_n \min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} \mathcal{I}_n c d\pi \ge \lim_m \int_{X \times Y} \mathcal{I}_m c d\overline{\pi} = \int_{X \times Y} c d\overline{\pi} \ge \min_{\pi \in \Gamma(\mu,\nu)} C(\pi).$$

By the fact that the right-hand side in (2.5) does not depend on n,

$$\min_{\pi\in\Gamma(\mu,\nu)} C(\pi) \le \sup_{(\varphi,\psi)\in I_c} \left\{\int_X \varphi d\mu + \int_Y \psi d\nu\right\}$$

and the result follows. ∎

(2.6) Remark *Using Proposition (P.28), arguing in the same way as in the proof of Kantorovich–Rubinstein duality, it is clear that*

$$\sup_{(\varphi,\psi)\in I_d} \left\{\int_X \varphi d\mu + \int_Y \psi d\nu\right\} = \sup_{f\in \mathrm{Lip(X)},\mathrm{Lip}(f)\le 1} \int_X f d(\mu-\nu).$$

2 Dual Schrödinger problem

The Schrödinger problem, like the optimal transport problem, is a convex constraint optimization problem, so we can derive a dual representation.

(2.7) Definition *Consider a Polish space (X,d), $\mathsf{R}\in\mathcal{R}_+(X\times X)$ and $\mu_0,\mu_1\in\mathcal{P}(X)$. Suppose there exists a Borel function $B : X\to [0,+\infty[$ such that*

$$\int_{X\times X} e^{-B(x)-B(y)} d\mathsf{R}(x,y) < +\infty, \quad \int_X B(x)d\mu_0(x) < +\infty, \quad \int_X B(x)d\mu_1(x) < +\infty.$$

In the so-called dual Schrödinger problem *we look for*

$$\sup\left\{\int_X \varphi d\mu_0 + \int_X \psi d\mu_1 - \log\left(\int_{X\times X} e^{\varphi\oplus\psi} d\mathsf{R}\right) : \varphi,\psi : X\to\mathbb{R} \text{ Borel}, \int_{X\times X} e^{\varphi\oplus\psi} d\mathsf{R} < +\infty\right\}.$$

Let us now show that the two formulations are in fact equivalent, as expected: we prove this fact in the case $H(\mu_0\otimes\mu_1 \mid \mathsf{R}) < +\infty$ to have existence, uniqueness, and the structural formula for the minimizer. The argument is completely based on Lemma (1.24). A more general proof can be found in [40].

(2.8) Theorem *Consider a Polish space (X,d), $\mathfrak{m}\in\mathcal{R}_+(X)$ and $\mathsf{R}\in\mathcal{R}_+(X\times X)$ such that $(p_1)_\#\mathsf{R} = (p_2)_\#\mathsf{R} = \mathfrak{m}$ and $\mathfrak{m}\otimes\mathfrak{m} \ll \mathsf{R} \ll \mathfrak{m}\otimes\mathfrak{m}$. Let $\mu_0 = \varrho_0\mathfrak{m}, \mu_1 = \varrho_1\mathfrak{m} \in \mathcal{P}(X)$. If there exists a Borel function $B : X\to[0,+\infty[$ such that*

$$\int_{X\times X} e^{-B(x)-B(y)} d\mathsf{R}(x,y) < +\infty, \quad \int_X B(x)d\mu_0(x) < +\infty, \quad \int_X B(x)d\mu_1(x) < +\infty$$

and $H(\mu_0 \otimes \mu_1 \mid \mathsf{R}) < +\infty$, *then*

$$\min_{\pi \in \Gamma(\mu_0,\mu_1)} H_{\mathsf{R}}(\pi) = \max\left\{\int_X \varphi d\mu_0 + \int_X \psi d\mu_1 - \log\left(\int_{X\times X} e^{\varphi\oplus\psi} d\mathsf{R}\right) : \right.$$
$$\left. \varphi, \psi : X \to \mathbb{R} \text{ Borel}, \int_X e^{\varphi\oplus\psi} d\mathsf{R} < +\infty\right\}.$$

Proof. First of all, by Theorem (1.28), let $\gamma = (f \otimes g)\mathsf{R}$ be the minimizer for H_{R} on $\Gamma(\mu_0, \mu_1)$, for some $f, g : X \to [0, +\infty[$ Borel functions. In particular, $\varphi = \log f$ and $\psi = \log g$ are Borel and

$$\int_{X\times X} e^{\varphi\oplus\psi} d\mathsf{R} = \int_{X\times X} f \otimes g d\mathsf{R} = \gamma(X \times X) = 1 < +\infty.$$

Now, by Lemma (1.24),

$$\begin{aligned}
\min_{\pi \in \Gamma(\mu_0,\mu_1)} H_{\mathsf{R}}(\pi) = H_{\mathsf{R}}(\gamma) = \sup&\left\{\int_X u d\gamma - \log\left(\int_{X\times X} e^u d\mathsf{R}\right) : \right.\\
&\left. u : X \times X \to \mathbb{R} \text{ Borel}, \int_X e^u d\mathsf{R} < +\infty\right\} \geq\\
&\geq \sup\left\{\int_X \varphi d\mu_0 + \int_X \psi d\mu_1 - \log\left(\int_{X\times X} e^{\varphi\oplus\psi} d\mathsf{R}\right) : \right.\\
&\left. \varphi, \psi : X \to \mathbb{R} \text{ Borel}, \int_X e^{\varphi\oplus\psi} d\mathsf{R} < +\infty\right\} \geq\\
&\geq \int_X \log f d\mu_0 + \int_X \log g d\mu_1 - \log\left(\int_{X\times X} e^{\log f \oplus \log g} d\mathsf{R}\right) =\\
&= \int_{X\times X} \log(f \otimes g) d\gamma - \log 1 = \int_{X\times X} \log(f \otimes g) d\gamma =\\
&= H_{\mathsf{R}}(\gamma) = \min_{\pi \in \Gamma(\mu_0,\mu_1)} H_{\mathsf{R}}(\pi),
\end{aligned}$$

so the supremum is a maximum and the result follows. ∎

Chapter 3

The Wasserstein space $(\mathcal{P}_2(X), W_2)$

1 Definition and initial properties

This chapter is mainly dedicated to the study of the metric space $(\mathcal{P}_2(X), W_2)$: the name of this space is a problematic issue in the History of Mathematics because Leonid Nisonovich Vaserstein, the mathematician from whom it takes its name, does not provide an explicit definition of it. Moreover, he was only interested in the case $(\mathcal{P}_1(X), W_1)$, which we do not care about. In the end, Wasserstein distances were introduced, independently, several times throughout the last century: another possible name is, indeed, Kantorovich–Rubinstein distance.

From this point on, we will always consider the square distance as cost in the optimal transport problem. Let us start with the basic definition of the chapter.

(3.1) Definition *Consider a Polish space (X, d) and the set*

$$\mathcal{P}_2(X) = \left\{ \mu \in \mathcal{P}(X) : \int_X d^2(x, x_0) d\mu(x) < +\infty \text{ for some (and thus for all) } x_0 \in X \right\},$$

we call Wasserstein distance on $\mathcal{P}_2(X)$ *the function $W_2 : \mathcal{P}_2(X) \times \mathcal{P}_2(X) \to \mathbb{R}$ such that*

$$W_2^2(\mu, \nu) = \min_{\pi \in \Gamma(\mu,\nu)} C(\pi).$$

Since form and substance are different, we have to prove that W_2 is well defined and, actually, a distance on $\mathcal{P}_2(X)$. We will use the following Lemma whose proof is based on Corollary (P.17).

(3.2) Lemma (Dudley) *Consider Polish spaces $(X_1, d_1), (X_2, d_2), (X_3, d_3)$ and take $\mu_1 \in \mathcal{P}(X_1), \mu_2 \in \mathcal{P}(X_2), \mu_3 \in \mathcal{P}(X_3)$. If $\pi^{12} \in \Gamma(\mu_1, \mu_2)$ and $\pi^{23} \in \Gamma(\mu_2, \mu_3)$, then there* *exists* $\pi \in \mathcal{P}(X_1 \times X_2 \times X_3)$ *such that*

$$(p^{1,2})_\# \pi = \pi^{12}, \qquad (p^{2,3})_\# \pi = \pi^{23}.$$

Proof. By the fact that $\pi^{12} \in \Gamma(\mu_1, \mu_2)$, we have that $\pi^{12} \in \mathcal{P}(X_1 \times X_2)$ and $(p^2)_\# \pi^{12} = \mu_2$. By Corollary (P.17), for every $A \in \mathscr{B}(X_1 \times X_2)$

$$\pi^{12}(A) = \int_{X_2} \pi^{12}_{x_2}(A) d\mu_2(x_2).$$

Analogously, for every $B \in \mathscr{B}(X_2 \times X_3)$

$$\pi^{23}(B) = \int_{X_3} \pi^{23}_{x_2}(B) d\mu_2(x_2).$$

The measure $\pi : \mathscr{B}(X_1 \times X_2 \times X_3) \to [0,1]$ such that

$$\pi(E) = \int_{X_2} (\pi^{12}_{x_2}, \pi^{23}_{x_2})(E) d\mu_2(x_2)$$

has the required properties. ■

(3.3) Theorem *If (X,d) is a Polish space, then $(\mathcal{P}_2(X), W_2)$ is a metric space.*

In particular, the function $E : X \to \mathrm{Im}(E) \subseteq \mathcal{P}_2(X)$ such that

$$E(x) = \delta_x$$

is an isometry.

Proof. First of all, since $\mu \otimes \nu \in \Gamma(\mu, \nu)$, given $x_0 \in X$, by the fact that $\mu, \nu \in \mathcal{P}_2(X)$,

$$W_2^2(\mu,\nu) \leq 2\int_X d^2(x,x_0) d\mu(x) + 2\int_X d^2(y,x_0) d\nu(y) < +\infty.$$

Now, by the symmetry of the Kantorovich problem and d, W_2 is also symmetric and clearly $W_2 \geq 0$. In particular,

$$W_2^2(\mu,\mu) \leq \int_{X\times X} d^2(x,y) d(\mathrm{Id},\mathrm{Id})_\#\mu(x,y) = 0,$$

so $W_2(\mu,\mu)=0$. Conversely, if $W_2(\mu,\nu)=0$, given π an optimal plan, one has

$$0=W_2^2(\mu,\nu)=\int_{X\times X} d^2(x,y)d\pi(x,y),$$

so $d^2(x,y)=0$ π-a.e. in $X\times X$, that is $x=y$ π-a.e. in $X\times X$. What we obtain is that for all bounded and Borel functions f

$$\int_X f(x)d\mu(x)=\int_{X\times X} f(x)d\pi(x,y)=\int_{X\times X} f(y)d\pi(x,y)=\int_X f(y)d\nu(y),$$

so $\mu=\nu$.

Let us now prove the triangle inequality. Let us consider $\mu,\nu,\sigma\in\mathcal{P}_2(X)$ and $X_1=X_2=X_3=X$. Initially, $\mu\in\mathcal{P}(X_1),\nu\in\mathcal{P}(X_2),\sigma\in\mathcal{P}(X_3)$, so for every $\pi^{12}\in\Gamma(\mu,\nu)$ optimal and $\pi^{23}\in\Gamma(\nu,\sigma)$ optimal, by Dudley's Lemma, there exists $\pi\in\mathcal{P}(X_1\times X_2\times X_3)$ such that

$$(p^{1,2})_\#\pi=\pi^{12},\qquad\qquad (p^{2,3})_\#\pi=\pi^{23}.$$

In particular,

$$(p^1)_\#\pi=\mu,\qquad\qquad (p^3)_\#\pi=\sigma,$$

so $(p^{1,2})_\#\pi\in\Gamma(\mu,\sigma)$ and

$$W_2(\mu,\sigma)\le\sqrt{\int_{X\times X} d^2(x_1,x_3)d(p^{1,3})_\#\pi(x_1,x_3)}.$$

Viewing $d^2(x_1,x_3)$ as a function of x_1,x_2 and x_3 not depending on x_2, using the change of variables formula one has

$$W_2(\mu,\sigma)\le\sqrt{\int_{X\times X\times X} d^2(x_1,x_3)d\pi(x_1,x_2,x_3)},$$

so by monotonicity of the function $\{t\mapsto t^2,t\ge 0\}$, the integral and the square root, using the triangle inequality of d, one has

$$W_2(\mu,\sigma)\le\sqrt{\int_{X\times X\times X} (d(x_1,x_2)+d(x_2,x_3))^2 d\pi(x_1,x_2,x_2)}.$$

Finally, viewing $d(x_1,x_2)$ as a function of x_1,x_2 and x_3 not depending on x_3, $d(x_2,x_3)$

as a function of x_1, x_2 and x_3 not depending on x_1 and using the triangle inequality of Lebesgue norms, we arrive at

$$W_2(\mu,\sigma) \le \sqrt{\int_{X\times X\times X} d^2(x_1,x_2)d\pi(x_1,x_2,x_3)} + \sqrt{\int_{X\times X\times X} d^2(x_2,x_3)d\pi(x_1,x_2,x_3)}$$

and using again the change of variables formula

$$W_2(\mu,\sigma) \le \sqrt{\int_{X\times X} d^2(x_1,x_2)d\pi^{12}(x_1,x_2)} + \sqrt{\int_{X\times X} d^2(x_2,x_3)d\pi^{23}(x_2,x_3)}.$$

The desired conclusion follows by the optimality of π^{12} and π^{23}.

Let us now prove that the function $E : X \to \mathrm{Im}(E) \subseteq \mathcal{P}_2(X)$ such that

$$E(x) = \delta_x$$

is an isometry. Clearly it is well defined and bijective. Consider now $x, y \in X$. Let us prove that $\Gamma(\delta_x,\delta_y) = \{\delta_x \times \delta_y\}$. By contradiction, if not, then there exists $\pi \in \Gamma(\delta_x,\delta_y)$ such that $\pi \ne \delta_x \otimes \delta_y = \delta_{(x,y)}$. In particular, there exists $(x',y') \ne (x,y)$ such that $(x',y') \in \mathrm{supt}(\pi)$. Let us consider $r > 0$ such that $x \notin \mathrm{B}(x',r)$ and $\varphi \in C_b(X)$ such that $\varphi \ge 0$, $\varphi(x') > 0$ and $\mathrm{supt}(\varphi) \subseteq \mathrm{B}(x',r)$. One has,

$$0 = \varphi(x) = \int_X \varphi d\delta_x = \int_X \varphi d(p_X)_\#\pi = \int_{\mathrm{B}(x',r)} \varphi d(p_X)_\#\pi > 0$$

that is a contradiction. In conclusion, $W_2(\delta_x,\delta_y) = d(x,y)$, so E is an isometry. ∎

2 Geometrical analysis of $(\mathcal{P}_2(X), W_2)$

We would now like to study some geometric properties of Wasserstein spaces. To do this, we first need a generalization of Dudley's Lemma.

(3.4) Lemma (iterated Dudley) *Given $N \in \mathbb{N}\cup\{+\infty\}$ such that $N \ge 3$, (X_n, d_n) Polish spaces and $\mu_n \in \mathcal{P}(X_n)$ for $1 \le n \le N$. If $\vartheta_n \in \Gamma(\mu_{n-1},\mu_n)$ for $2 \le n \le N$, there exist $\pi_n \in \mathcal{P}(X_1 \times \cdots \times X_n)$, for $1 \le n \le N$ such that the following facts hold true:*

(a) $p^{1,\dots,n-1}_\#\pi_n = \pi_{n-1}$ *for* $2 \le n \le N$,

(b) $p^i_\#\pi_n = \mu_i$ *for* $1 \le i \le n \le N$,

(c) $p^{i-1,i}_\#\pi_n = \vartheta_i$ *for* $2 \le i \le n \le N$.

Proof. The case $N = 3$ follows by Dudley's Lemma. In general, for N finite, we only have to repeatedly apply Dudley's Lemma to

$$X_1 \times \cdots \times X_n = Z_1 \times Z_2 \times Z_3$$

where $Z_1 = X_1 \times \cdots \times X_{n-2}$, $Z_2 = X_{n-1}$, $Z_3 = X_n$, $\pi_{n-1} \in \mathcal{P}(Z_1 \times Z_2)$ and $\vartheta \in \Gamma(\mu_{n-1}, \mu_n) \subseteq \mathcal{P}(Z_2 \times Z_3)$.

The case $N = +\infty$ follows from Kolmogorov's Theorem. ■

Let us start with uniform properties of $(\mathcal{P}_2(X), W_2)$: we are, in particular, interested in completeness. The proof will be based on the fact that, given a measure space $(E, \mathscr{U}, \mu)$, with μ finite, the metric space $L^p(E, \mathscr{U}, \mu; X)$ is complete whenever (X, d) is complete. Since we want to talk about completeness, to be consistent, let us remember that the set (in general is not a vector space)

$$L^p(E, \mathscr{U}, \mu; X) = \Big\{ f \in M(E, \mu; X) : \int_E d^p(f(x), z_0) d\mu(x) < +\infty, \\ \text{for some (and thus all) } z_0 \in X \Big\}$$

is endowed with the distance

$$\left(\int_E d^p(f, g) d\mu \right)^{\frac{1}{p}}.$$

(3.5) Theorem *Consider a Polish space (X, d). $(\mathcal{P}_2(X), W_2)$ is complete.*

Proof. Let (μ_n) be a Cauchy sequence in $(\mathcal{P}_2(X), W_2)$. By the Cauchy property of (μ_n), up to a subsequence, we can assume

$$\sum_{n=0}^{\infty} W_2(\mu_n, \mu_{n+1}) < +\infty.$$

Now, let us consider $\pi_\infty \in \mathcal{P}(\prod_{n=0}^{\infty} X_n)$ in accordance with iterated Dudley's Lemma, applied to (μ_n) and $(X_n) = (X)$.

Observing that

$$\int_{\prod_{n=0}^{\infty} X_n} d^2(p^n, p^{n+1}) d\pi_\infty = \int_{\prod_{n=0}^{\infty} X_n} d^2(x_n, x_{n+1}) d(p^n, p^{n+1})_\# \pi_\infty = W_2^2(\mu_n, \mu_{n+1}),$$

we deduce that the sequence of projections (p^n) is Cauchy in $L^2(\prod_{n=0}^{\infty} X_n, \prod_{n=0}^{\infty} \mathscr{B}_n, \pi_\infty; X)$. Thanks to the completeness of $L^2(\prod_{n=0}^{\infty} X_n, \prod_{n=0}^{\infty} \mathscr{B}_n, \pi_\infty; X)$, (p^n) converges to some $\overline{p}$

in $L^2(\prod_{n=0}^{\infty} X_n, \prod_{n=0}^{\infty} \mathscr{B}_n, \pi_\infty, X)$. Defining

$$\mu_\infty = (\overline{p})_\# \pi_\infty,$$

we have

$$W_2^2(\mu_n, \mu_\infty) \leq \int_{\prod_{n=0}^{\infty} X_n} d^2(p^n, \overline{p}) d\pi_\infty \to 0,$$

so $\mu_n \to \mu_\infty$ and the proof is concluded. ∎

We now want to move on and analyze more topological properties of $(\mathcal{P}_2(X), W_2)$, in particular, compactness and separability. Since W_2 convergence is not particularly easy to handle, let us look at an equivalent condition.

(3.6) Theorem *Consider a Polish space (X, d). Let (μ_n) in $\mathcal{P}_2(X)$ and $\mu \in \mathcal{P}_2(X)$. The following facts are equivalent:*

(a) $\mu_n \to \mu$ in $(\mathcal{P}_2(X), W_2)$,

(b) $\mu_n \rightharpoonup \mu$ and for some (and thus all) $x_0 \in X$ one has

$$\lim_n \int_X d^2(x_0, x) d\mu_n(x) = \int_X d^2(x_0, x) d\mu(x).$$

Proof.
$(a) \implies (b)$ Fix $x_0 \in X$. Let us start proving that if $\nu \in \mathcal{P}_2(X)$, then $\Gamma(\delta_{x_0}, \nu) = \{\delta_{x_0} \otimes \nu\}$. By contradiction, if there exists $\pi \in \Gamma(\delta_{x_0}, \nu)$ such that $\pi \neq \delta_{x_0} \otimes \nu$, then there also exists $x' \in X$ such that for all $y \in X$ one has $(x', y) \in \operatorname{supt}(\pi)$. Let us consider $r > 0$ such that $x_0 \notin \mathrm{B}(x', r)$ and $\varphi \in C_b(X)$ such that $\varphi \geq 0$, $\varphi(x') > 0$ and $\operatorname{supt}(\varphi) \subseteq \mathrm{B}(x', r)$. One has,

$$0 = \varphi(x_0) = \int_X \varphi d\delta_{x_0} = \int_X \varphi d(p^1)_\# \pi = \int_{\mathrm{B}(x', r)} \varphi d(p^1)_\# \pi > 0$$

that is a contradiction.

Using the above result, we can write

$$\left| \sqrt{\int_X d^2(x, x_0) d\mu_n(x)} - \sqrt{\int_X d^2(x, x_0) d\mu(x)} \right| = |W_2(\delta_{x_0}, \mu_n) - W_2(\delta_{x_0}, \mu)|$$

and, using the triangle inequality, one has

$$\left| \sqrt{\int_X d^2(x, x_0) d\mu_n(x)} - \sqrt{\int_X d^2(x, x_0) d\mu(x)} \right| \leq W_2(\mu_n, \mu).$$

We can therefore conclude that

$$\lim_n \int_X d^2(x_0, x) d\mu_n(x) = \int_X d^2(x_0, x) d\mu(x).$$

Let us now consider $f \in \mathrm{Lip}_b(X)$ and $\pi_n \in \Gamma(\mu_n, \mu)$ optimal. One has,

$$\begin{aligned}
&\limsup_n \left| \int_X f(x) d\mu_n(x) - \int_X f(y) d\mu(y) \right| = \\
&\quad = \limsup_n \left| \int_{X\times X} f(x) d\pi_n(x,y) - \int_{X\times X} f(y) d\pi_n(x,y) \right| \leq \\
&\quad \leq \mathrm{Lip}(f) \limsup_n \int_{X\times X} d(x,y) d\pi_n(x,y).
\end{aligned}$$

Using Hölder's inequality and remembering that $\pi_n \in \mathcal{P}(X \times X)$ we obtain

$$\limsup_n \left| \int_X f(x) d\mu_n(x) - \int_X f(y) d\mu(y) \right| \leq \mathrm{Lip}(f) \limsup_n W_2(\mu_n, \mu) = 0.$$

The result follows by Lemma (P.19).

$(b) \implies (a)$ Let us consider firstly the case when (X, d) is compact, so totally bounded, so bounded. Fixed $z \in X$ and consider the closed subset of $C(X)$

$$Z = \{f \in \mathrm{Lip}(X) : \mathrm{Lip}(f) \leq 1, f(z) = 0\}.$$

Let us observe that if $f \in Z$

$$|f(x)| = |f(x) - f(z)| \leq \mathrm{Lip}(f) d(x, z) \leq \mathrm{diam}(X),$$

so $\max|f| \leq \mathrm{diam}(X)$ and $f \in \mathrm{Lip}_b(X)$. If (f_h) in Z, for every $h \in \mathbb{N}$

$$\|f_h\|_\infty = \max f_h \leq \mathrm{diam}(X),$$

so (f_h) is bounded in $(C(X), \| \|_\infty)$. Fixed $\varepsilon > 0$, given any $\delta > 0$ such that $\delta < \varepsilon$, for every $h \in \mathbb{N}$, if $x, y \in X$ such that $d(x, y) < \delta$, then

$$|f_h(x) - f_h(y)| \leq d(x, y) < \delta < \varepsilon,$$

so (f_h) is also equi-uniformly continuous. By Ascoli–Arzelà's Theorem, there exists (f_{h_k}) converging in $(C(X), \| \|_\infty))$. Being Z closed, (f_{h_k}) also converges in Z. Therefore Z is compact.

Let us define, for every $f \in Z$,

$$L_n(f) = \int_X f d\mu_n, \qquad\qquad L(f) = \int_X f d\mu.$$

Fixed $\varepsilon > 0$, by the fact that $\mu_n \rightharpoonup \mu$, using Lemma (P.19), there exists $\bar{h} \in \mathbb{N}$ such that for every $h \geq \bar{h}$

$$\left| \int_X f d\mu_n - \int_X f d\mu \right| < \varepsilon,$$

so, being uniformly bounded, $L_n \to L$ uniformly in Z. By translation invariance, we have

$$\sup_{\mathrm{Lip}(f) \leq 1} \left(\int_X f d(\mu_n - \mu) \right) = \sup_{f \in Z} \left(\int_X f d(\mu_n - \mu) \right) \to 0.$$

Now, using Remark (2.6) and Kantorovich–Rubinstein duality, we deduce the existence of $\pi_n \in \Gamma(\mu_n, \mu)$ optimal for d such that

$$\lim_n \int_{X \times X} d d\pi_n = 0.$$

Being X compact, d is bounded, so there exists $C \in \mathbb{R}$ such that $d \leq C$. Furthermore, observing that

$$\begin{aligned}
\int_{X \times X} d^2 d\pi_n &= \int_{(X \times X) \cap \{d \leq 1\}} d^2 d\pi_n + \int_{(X \times X) \cap \{1 < d \leq C\}} d^2 d\pi_n \leq \\
&\leq \int_{(X \times X) \cap \{d \leq 1\}} d d\pi_n + C \int_{(X \times X) \cap \{1 < d \leq C\}} d d\pi_n \leq (1 + C) \int_{X \times X} d d\pi_n,
\end{aligned}$$

we have

$$W_2^2(\mu_n, \mu) \leq \int_{X \times X} d^2 d\pi_n \leq (1 + C) \int_{X \times X} d d\pi_n$$

and passing to the limit as $n \to +\infty$ we have $\mu_n \to \mu$ in $(\mathcal{P}_2(X), W_2)$.

In the general case, fixed $x_0 \in X$, consider (σ_n) in $\mathcal{P}(X)$ and $\sigma \in \mathcal{P}(X)$ such that for every $B \in \mathscr{B}(X)$

$$\sigma_n(B) = \frac{1}{Z_n} \int_B (1 + d^2(x_0, x)) d\mu_n(x), \qquad \sigma(B) = \frac{1}{Z} \int_B (1 + d^2(x_0, x)) d\mu(x)$$

where

$$Z_n = \int_X (1 + d^2(x_0, x)) d\mu_n(x), \qquad Z = \int_X (1 + d^2(x_0, x)) d\mu(x).$$

Clearly, by (b),

$$\lim_n Z_n = Z$$

and, by Lemma (P.20), for every $A \subseteq X$ open

$$\liminf_n \int_A d^2(x_0, x) d\mu_n(x) = \liminf_n \int_X d^2(x_0, x)\chi_A(x) d\mu_n(x) \geq$$
$$\geq \int_X d^2(x_0, x)\chi_A(x) d\mu(x) = \int_A d^2(x_0, x) d\mu(x).$$

Combining these two facts with Lemma (P.21), we obtain that for every $A \subseteq X$ open

$$\begin{aligned}\liminf_n \sigma_n(A) = \liminf_n \frac{1}{Z_n} \int_A (1 + d^2(x_0, x)) d\mu_n(x) = \\ = \frac{1}{Z} \liminf_n \int_A (1 + d^2(x_0, x)) d\mu_n(x) \geq \\ \geq \frac{1}{Z} \left[\liminf_n \mu_n(A) + \liminf_n \int_A d^2(x_0, x) d\mu_n(x) \right] \geq \sigma(A),\end{aligned}$$

then, applying again Lemma (P.21), $\sigma_n \rightharpoonup \sigma$.

Now, by Prokhorov's Theorem applied to the set $\{\sigma_n : n \in \mathbb{N}\}$, there exists an increasing sequence (K_k) of compact subsets of X such that $x_0 \in K_0$ and

$$\lim_k \sup_n \sigma_n(X \setminus K_k) = 0.$$

Now

$$Z_n \sigma_n(X \setminus K_k) \geq \int_{X \setminus K_k} d^2(x_0, x) d\mu_n(x),$$

so

$$\lim_k \sup_n \int_{X \setminus K_k} d^2(x_0, x) d\mu_n(x) = 0.$$

If we define $\mu_{n,k} = \mu_n \lfloor K_k + (1 - \mu_n(K_k))\delta_{x_0} \in \mathcal{P}(K_k)$, fixing $k \in \mathbb{N}$, we have

$$\mu_{n,k} \rightharpoonup \nu_k = \mu \lfloor K_k + (1 - \mu(K_k))\delta_{x_0}.$$

Using a diagonal argument, being in a compact, we can find $(\mu_{n_j,k})$ such that for every $k \in \mathbb{N}$, $(\mu_{n_j,k})$ converges in $(\mathcal{P}_2(K_k), W_2)$. Up to interpreting $(\mu_{n_j,k})$ as a sequence in $\mathcal{P}_2(X)$ of measures with support in K_k, we obtain, for fixed $k \in \mathbb{N}$, that $(\mu_{n_j,k})$ converges in $(\mathcal{P}_2(X), W_2)$.

Considering now $\pi_{n,k} \in \mathcal{P}(X \times X)$ such that

$$\pi_{n,k} = (\mathrm{Id}, \mathrm{Id})_\# \mu_n \lfloor K_k + (\mathrm{Id}, f)_\# \mu_n,$$

where $f : X \to X$ such that $f(x) = x_0$, we have, for every $B \in \mathscr{B}(X)$,

$$
\begin{aligned}
(p^1)_{\#}\pi_{n,k}(B) &= \pi_{n,k}((p^1)^{-1}(B)) = \pi_{n,k}(B \times X) = \\
&= (\mathrm{Id}, \mathrm{Id})_{\#}\mu_n \lfloor K_k(B \times X) + (\mathrm{Id}, f)_{\#}\mu_n \lfloor (X \setminus K_k)(B \times X) = \\
&= \mu_n \lfloor K_k((\mathrm{Id}, \mathrm{Id})^{-1}(B \times X)) + \mu_n \lfloor (X \setminus K_k)((\mathrm{Id}, f)^{-1}(B \times X)) = \\
&= \mu_n \lfloor K_k(B) + \mu_n \lfloor (X \setminus K_k)(B) = \mu_n(B)
\end{aligned}
$$

and

$$
\begin{aligned}
(p^2)_{\#}\pi_{n,k}(B) &= \pi_{n,k}((p^2)^{-1}(B)) = \pi_{n,k}(X \times B) = \\
&= (\mathrm{Id}, \mathrm{Id})_{\#}\mu_n \lfloor K_k(X \times B) + (\mathrm{Id}, f)_{\#}\mu_n \lfloor (X \setminus K_k)(X \times B) = \\
&= \mu_n \lfloor K_k((\mathrm{Id}, \mathrm{Id})^{-1}(X \times B)) + \mu_n \lfloor (X \setminus K_k)((\mathrm{Id}, f)^{-1}(B \times X)) = \\
&= \mu_n \lfloor K_k(B) + \mu_n \lfloor (X \setminus K_k)(X)\delta_{x_0}(B) = \\
&= \mu_n \lfloor K_k(B) + (1 - \mu_n(X \setminus K_k)\delta_{x_0}(B).
\end{aligned}
$$

Resuming,

$$
(p^1)_{\#}\pi_{n,k} = \mu_n, \qquad (p^2)_{\#}\pi_{n,k} = \mu_{n,k},
$$

so $\pi_{n,k} \in \Gamma(\mu_n, \mu_{k_n})$ and

$$
\begin{aligned}
W_2^2(\mu_n, \mu_{n,k}) &\leq \int_{X \times X} d^2(x, y) d\pi_{n,k}(x, y) = \\
&= \int_{X \times X} d^2(x, y) d(\mathrm{Id}, \mathrm{Id})_{\#}\mu_n \lfloor K_k(x, y) + \int_{X \times X} d^2(x, y) d(\mathrm{Id}, f)_{\#}\mu_n(x, y) = \\
&= \int_X (d^2 \circ (\mathrm{Id}, \mathrm{Id}))(x) d\mu_n \lfloor K_k(x) + \int_X (d^2 \circ (\mathrm{Id}, f))(x) d\mu_n \lfloor K_k(x) = \\
&= \int_X d^2(x, x)\chi_{K_k} d\mu_n(x) + \int_X d^2(x, x_0)\chi_{X \setminus K_k} d\mu_n(x) = \\
&= \int_{X \setminus K_k} d^2(x, x_0) d\mu_n(x).
\end{aligned}
$$

Now, let us prove that the sequence (μ_{n_j}) is Cauchy in $(\mathcal{P}_2(X), W_2)$. Fixed $\varepsilon > 0$, by the fact that $(\mu_{n_j,k})$ converges in $(\mathcal{P}_2(X), W_2)$ for any $k \in \mathbb{N}$, then it is Cauchy. In particular, for every $k \in \mathbb{N}$ there exists $\overline{n}(k) \in \mathbb{N}$ such that for every $n_j, n_l \geq \overline{n}(k)$

$$
W_2^2(\mu_{n_j,k}, \mu_{n_l,k}) < \frac{\varepsilon^2}{8},
$$

then, by the triangle inequality,

$$\begin{aligned}W_2^2(\mu_{n_j}, \mu_{n_l}) &\leq 4W_2^2(\mu_{n_j}, \mu_{n_j,k}) + 4W_2^2(\mu_{n_j,k}, \mu_{n_l,k}) + 4W_2^2(\mu_{n_l,k}, \mu_{n_l}) \leq \\ &\leq 8\sup_n W_2^2(\mu_n, \mu_{n,k}) + 4W_2^2(\mu_{n_j,k}, \mu_{n_l,k}) < \\ &< 8\sup_n \int_{X\setminus K_k} d^2(x, x_0) d\mu_n(x) + \frac{\varepsilon^2}{2}.\end{aligned}$$

If we choose $\overline{k} \in \mathbb{N}$ such that

$$\sup_n \int_{X\setminus K_k} d^2(x, x_0) d\mu_n(x) < \frac{\varepsilon^2}{16},$$

then for every $n_j, n_l \geq \overline{n}(\overline{k})$

$$W_2^2(\mu_{n_j}, \mu_{n_l}) < \varepsilon^2$$

as required. By the completeness of $(\mathcal{P}_2(X), W_2)$, it follows that (μ_{n_j}) converges in $(\mathcal{P}_2(X), W_2)$. We can show that $\mu_{n_j} \to \mu$ in $(\mathcal{P}_2(X), W_2)$. Indeed, let $\nu \in \mathcal{P}_2(X)$ such that $\mu_{n_j} \to \nu$ in $(\mathcal{P}_2(X), W_2)$. Using the first implication, $\mu_{n_j} \rightharpoonup \nu$, but since $\mu_n \rightharpoonup \mu$, then $\mu_{n_j} \rightharpoonup \mu$ and the uniqueness of the weak limit provides $\mu = \nu$.

We have therefore exhibited a subsequence of (μ_n) that converges to μ in $(\mathcal{P}_2(X), W_2)$. Repeating the argument for every subsequence of (μ_n) in place of (μ_n), we obtain $\mu_n \to \mu$ in $(\mathcal{P}_2(X), W_2)$. ∎

As for the compactness, the proof is essentially based on the combination of the weak compactness of $\mathcal{P}(X)$ with the characterization of W_2 convergence that we have proved.

(3.7) Corollary *Consider a Polish space (X, d). If (X, d) is compact, then $(\mathcal{P}_2(X), W_2)$ is compact as well.*

Proof. First of all, by Theorem (P.22), $\mathcal{P}(X)$ is weakly compact. Now, compactness of (X, d) implies that (X, d) is totally bounded, hence bounded. In particular, $\mathcal{P}_2(X) = \mathcal{P}(X)$ because if $\mu \in \mathcal{P}(X)$ for every $x_0 \in X$

$$\int_X d^2(x, x_0) d\mu(x) \leq \operatorname{diam}^2(X) < +\infty.$$

Combining this with Theorem (3.6) the result follows. ∎

With regards to separability, we, more or less directly, trace back to the useful characterization of W_2 convergence that we have given before.

(3.8) Corollary *Consider a Polish space* (X, d). $(\mathcal{P}_2(X), W_2)$ *is separable.*

Proof. First of all, by separability of (X, d), there exists $D \subseteq X$ at most countable such that $\overline{D} = X$. Consider, now, the sets

$$\mathcal{D} = \left\{ \sigma \in \mathcal{P}_2(X) : \sigma = \sum_{i=0}^{n} q_i \delta_{x_i}, x_i \in D, q_i \in \mathbb{Q}^+, n \in \mathbb{N} \right\}$$

and

$$\mathfrak{Z} = \left\{ \sigma \in \mathcal{P}_2(X) : \sigma = \sum_{i=0}^{\infty} t_i \delta_{x_i}, x_i \in X, t_i \in \mathbb{R}^+ \right\}.$$

We have $\mathfrak{Z} \subseteq \overline{\mathcal{D}}$. Let us prove that if $\mu \in \mathcal{P}_2(X)$ has bounded support, then $\mu \in \overline{\mathcal{D}}$. By the fact that μ has bounded support, there exist $x_0 \in X$ and $r > 0$ such that $\mathrm{supt}(\mu) \in \mathrm{B}(x_0, r)$. Now, for any $h \in \mathbb{N} \setminus \{0\}$ let us consider $\left(A_i^{(h)}\right)$ a Borel partition of $B(x_0, r)$ such that $\mathrm{diam}\left(A_i^{(h)}\right) \leq \frac{1}{h}$. Defining

$$\mu_h = \sum_{i=0}^{\infty} \mu\left(A_i^{(h)}\right) \delta_{x_i^{(h)}},$$

where $x_i^{(h)} \in A_i^{(h)}$, if $f \in \mathrm{Lip}_b(X)$, one has

$$\begin{aligned}
\left| \int_{A_i^{(h)}} f d\mu - \mu\left(A_i^{(h)}\right) \int_{A_i^{(h)}} f d\delta_{x_i^{(h)}} \right| &= \left| \int_{A_i^{(h)}} f d\mu - \mu\left(A_i^{(h)}\right) f(x_i^{(h)}) \right| \leq \\
&\leq \int_{A_i^{(h)}} \left| f(x) - f\left(x_i^{(h)}\right) \right| d\mu(x) \leq \\
&\leq \mathrm{Lip}(f) \int_{A_i^{(h)}} d(x, x_i^{(h)}) d\mu(x) \leq \frac{1}{h} \mu\left(A_i^{(h)}\right) \mathrm{Lip}(f),
\end{aligned}$$

so

$$\begin{aligned}
\left| \int_X f d\mu - \int_X f d\mu_h \right| &= \left| \sum_{i=0}^{\infty} \int_{A_i^{(h)}} f d\mu - \sum_{i=0}^{\infty} \mu\left(A_i^{(h)}\right) \int_{A_i^{(h)}} f d\delta_{x_i^{(h)}} \right| \leq \\
&\leq \sum_{i=0}^{\infty} \left| \int_{A_i^{(h)}} f d\mu - \mu\left(A_i^{(h)}\right) \int_{A_i^{(h)}} f d\delta_{x_i^{(h)}} \right| \leq \\
&\leq \frac{1}{h} \mathrm{Lip}(f) \sum_{i=0}^{\infty} \mu\left(A_i^{(h)}\right) = \frac{1}{h} \mathrm{Lip}(f) \mu(\mathrm{B}(x_0, r)).
\end{aligned}$$

Thanks to Lemma (P.19), we obtain $\mu_n \rightharpoonup \mu$. Now, analogously

$$\left| \int_X d^2(x_i^{(h)}, x) d\mu(x) - \int_X d^2(x_i^{(h)}, x) d\mu_h(x) \right| \leq \sum_{i=0}^{\infty} \int_{A_i^{(h)}} \left| d^2(x_i^{(h)}, x) - d^2(x_i^{(h)}, x_i^{(h)}) \right| d\mu(x)$$

therefore

$$\left|\int_X d^2(x_i^{(h)}, x)d\mu(x) - \int_X d^2(x_i^{(h)}, x)d\mu_h(x)\right| \le \frac{1}{h^2}\mu(\mathrm{B}(x_0, r)),$$

so, by Theorem (3.6), we obtain $\mu_h \to \mu$ in $(\mathcal{P}_2(X), W_2)$. Observing that (μ_h) in $\mathfrak{Z}$, it follows that $\mu \in \overline{\mathcal{D}}$.

In the general case of $\mu \in \mathcal{P}_2(X)$, given $\bar{x} \in \mathrm{supt}(\mu)$ and $R > 0$, if we consider the measure

$$\mu_R = \frac{1}{\mu(\mathrm{B}(\bar{x}, R))}\mu\lfloor \mathrm{B}(\bar{x}, R),$$

one has that the support of μ_R is bounded and, if $R \to +\infty$, $\mu_R \to \mu$ in $(\mathcal{P}_2(X), W_2)$, thanks to Theorem (3.6). Combining the case with bounded support and the previous approximation, the result follows. ■

(3.9) Remark *Combining Theorem* (3.5) *and Corollary* (3.8)*, we understand that whenever* (X, d) *is Polish, then* $(\mathcal{P}_2(X), W_2)$ *is Polish as well.*

For further details about Wasserstein spaces, the interested reader may refer to [64] and [3].

We conclude the section with some results due to Brenier and Knott–Smith. We omit the proof, but the interested reader can find further details in [1], [37] or [8]. We just underline that absolute continuity with respect to $\mathcal{L}^n$ assumption is required to avoid atomic behavior that would ill pose Monge's problem.

(3.10) Theorem *Consider* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$*, with* $\mu \ll \mathcal{L}^n$*. The following facts hold true:*

(a) *there exists a unique minimizer* π *for the problem*

$$\min\left\{\int_{\mathbb{R}^{2n}} \frac{1}{2}|x - y|^2 d\pi(x, y) : \pi \in \Gamma(\mu, \nu)\right\},$$

(b) *there exists a unique, up to* μ_0*-negligible sets, minimizer* T *for the problem*

$$\min\left\{\int_{\mathbb{R}^n} \frac{1}{2}|x - T(x)|^2 d\mu(x) : T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel}, T_{\#}\mu = \nu\right\},$$

(c) π *is induced by* T*,*

(d) *there exists a function* $\psi : \mathbb{R}^n \to]-\infty, +\infty]$ *convex, lower semicontinuous and* μ-a.e. *differentiable in* $\mathbb{R}^n$ *such that*

$$T = D\psi.$$

Proof. See [1, Theorem 5.2]. ∎

(3.11) Theorem *Consider* $\mu \in \mathcal{P}_2(\mathbb{R}^n)$, *with* $\mu \ll \mathcal{L}^n$. *If there exists a function* $\psi : \mathbb{R}^n \to]-\infty, +\infty]$ *convex, lower semicontinuous and* μ-a.e. *differentiable in* $\mathbb{R}^n$, *then*

$$T = D\psi,$$

is the unique, up to μ_0*-negligible sets, minimizer for the problem*

$$\min\left\{\int_{\mathbb{R}^n} \frac{1}{2}|x - T(x)|^2 d\mu(x) : T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel, } T_\#\mu = \nu\right\}.$$

Proof. See [1, Theorem 5.2]. ∎

(3.12) Theorem *Consider* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$, *with* $\mu, \nu \ll \mathcal{L}^n$. *If* $T^{\mu\to\nu}$ *is the unique, up to* μ*-negligible sets, minimizer for the problem*

$$\min\left\{\int_{\mathbb{R}^n} \frac{1}{2}|x - T(x)|^2 d\mu(x) : T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel, } T_\#\mu = \nu\right\}$$

and $T^{\nu\to\mu}$ *is the unique, up to* ν*-negligible sets, minimizer for the problem*

$$\min\left\{\int_{\mathbb{R}^n} \frac{1}{2}|x - T(x)|^2 d\nu(x) : T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel, } T_\#\nu = \mu\right\},$$

then

$$T^{\nu\to\mu} \circ T^{\mu\to\nu} = \mathrm{Id}$$

μ-a.e. *in* $\mathbb{R}^n$ *and*

$$T^{\mu\to\nu} \circ T^{\nu\to\mu} = \mathrm{Id}$$

ν-a.e. *in* $\mathbb{R}^n$.

Proof. See [1, Theorem 5.2]. ∎

Chapter 4

Dynamical representations

From now on, we will consider, as an environment, $\mathbb{R}^n$ equipped with the Euclidean distance and we will deal only with the case $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$, to have $W_2(\mu, \nu)$ well defined.

1 Dynamic optimal transport

So far, to talk about the Kantorovich problem, we have always paid attention to the initial and final configurations of mass, or at most on the intermediary that carries out the transport. With the dynamic representation we want to propose in this section, we are instead going to focus our attention on the path that the mass will have to follow to move from the initial configuration to the final one. In this way, by increasing the level of abstraction to formulate the problem, we will instead recover its most intrinsic meaning. Let us analyze a short example.

(4.1) Example *Consider $x, y \in \mathbb{R}^n$ and $\delta_x, \delta_y \in \mathcal{P}_2(\mathbb{R}^n)$. As we have already seen, the only optimal plan is $\delta_x \otimes \delta_y$. As for the path that the mass must travel, we have to consider the geodesic connecting x and y, namely*

$$\gamma(t) = ty + (1-t)x.$$

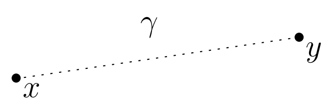

Figure 4.1: The path of a Dirac mass between x and y.

Let us proceed, then, by formalizing what we saw in the previous Example.

(4.2) Definition *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$. In the so-called* dynamic optimal transport problem *we look for*

$$\inf \left\{ \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)), (e_0)_\# \eta = \mu, (e_1)_\# \eta = \nu \right\}.$$

In particular, we will refer to any admissible η as a dynamic transport plan *and any optimal η as a* optimal geodesic plan. *We will denote with* $\mathrm{OptGeo}(\mu,\nu)$ *the set of optimal geodesic plans from μ to ν.*

We want now to prove that the dynamic formulation of the optimal transport problem is equivalent to Kantorovich's one: fundamental will be the fact that fixed $x, y \in \mathbb{R}^n$, there exists a unique $\gamma \in \mathrm{Geo}(\mathbb{R}^n)$ connecting them.

(4.3) Theorem *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$. It results*

$$\min_{\pi \in \Gamma(\mu,\nu)} C(\pi) = \min \left\{ \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)), (e_0)_\# \eta = \mu, (e_1)_\# \eta = \nu \right\}.$$

Proof. First of all, if $\gamma \in AC([0,1];\mathbb{R}^n)$, by Hölder's inequality,

$$\int_0^1 |\gamma'| d\mathcal{L}^1 \leq \left(\int_0^1 |\gamma'|^2 d\mathcal{L}^1 \right)^{\frac{1}{2}}$$

and, by the definition of an absolutely continuous curve and Theorem (P.35),

$$|\gamma(1) - \gamma(0)| \leq \int_0^1 |\gamma'| d\mathcal{L}^1,$$

so

$$|\gamma(1) - \gamma(0)|^2 \leq \int_0^1 |\gamma'|^2 d\mathcal{L}^1.$$

Now, if η is a dynamic transport plan

$$\begin{aligned}\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) &\geq \int_{AC([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) \geq \int_{AC([0,1];\mathbb{R}^n)} \int_0^1 |\gamma'|^2 d\mathcal{L}^1 d\eta(\gamma) \geq \\ &\geq \int_{AC([0,1];\mathbb{R}^n)} |\gamma(1) - \gamma(0)|^2 d\eta(\gamma) = \int_{AC([0,1];\mathbb{R}^n)} |e_1(\gamma) - e_0(\gamma)|^2 d\eta(\gamma) = \\ &= \int |x-y|^2 d(e_0, e_1)_\# \eta(x,y).\end{aligned}$$

Observing that $p^1_\# \circ (e_0, e_1)_\# = (p^1 \circ (e_0, e_1))_\# = (e_0)_\#$ (resp. $p^2_\# \circ (e_0, e_1)_\# = (e_1)_\#$), we can affirm that $(e_0, e_1)_\# \eta \in \Gamma(\mu, \nu)$, so

$$\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) \geq \min_{\pi \in \Gamma(\mu,\nu)} C(\pi).$$

In particular,

$$\min_{\pi \in \Gamma(\mu,\nu)} C(\pi) \leq \min \left\{ \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)), (e_0)_\# \eta = \mu, (e_1)_\# \eta = \nu \right\}.$$

Consider now $\pi \in \Gamma(\mu, \nu)$ optimal. For any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ choose $\gamma_{x,y} : [0, 1] \to \mathbb{R}^n$ such that $\gamma_{x,y}(t) = ty + (1-t)x$. Define the function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathrm{Geo}(\mathbb{R}^n)$ such that

$$\Phi(x, y) = \gamma_{x,y}$$

and the measure $\eta' = \Phi_\# \pi \in \mathcal{P}(C([0, 1]; \mathbb{R}^n))$ which is, by construction, supported in $\mathrm{Geo}(\mathbb{R}^n)$. Using Lemma (P.51), we can write

$$\min_{\pi \in \Gamma(\mu,\nu)} C(\pi) = \int |x-y|^2 d\pi(x, y) = \int_{\mathrm{Geo}(\mathbb{R}^n)} |\gamma(1) - \gamma(0)|^2 d\eta'(\gamma) = \int_{\mathrm{Geo}(\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta'(\gamma)$$

so

$$\min_{\pi \in \Gamma(\mu,\nu)} C(\pi) \geq \min \left\{ \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)), (e_0)_\# \eta = \mu, (e_1)_\# \eta = \nu \right\}.$$

and the proof is concluded. ∎

(4.4) Corollary *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$. The following facts are equivalent:*

(a) $\eta \in \mathrm{OptGeo}(\mu, \nu)$,

(b) $\mathrm{supt}(\eta) \subseteq \mathrm{Geo}(\mathbb{R}^n)$ *and* $(e_0, e_1)_\# \eta \in \Gamma(\mu, \nu)$ *is optimal.*

Proof. It is a direct consequence of Theorem (4.3). ∎

Thanks to this new characterization of the optimal transport problem, we can prove another geometric property of $(\mathcal{P}_2(\mathbb{R}^n), W_2)$.

(4.5) Theorem $(\mathcal{P}_2(\mathbb{R}^n), W_2)$ *is geodesic.*

Proof. Fix $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^n)$. By Theorem (4.3), there exists $\eta \in \mathrm{OptGeo}(\mu, \nu)$ such that

$$W_2^2(\mu, \nu) = \int_{\mathrm{Geo}(\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma).$$

Consider $\mu_t = (e_t)_\# \eta$. Clearly $\mu_0 = \mu$, $\mu_1 = \nu$, and using Lemma (P.51) and the characterization of $\mathrm{Geo}(\mathbb{R}^n)$, for every $s, t \in [0,1]$ we have

$$|s-t|^2 W_2^2(\mu,\nu) = |s-t|^2 \int_{\mathrm{Geo}(\mathbb{R}^n)} |\gamma(1)-\gamma(0)|^2 d\eta(\gamma) = \int_{\mathrm{Geo}(\mathbb{R}^n)} |\gamma(s)-\gamma(t)|^2 d\eta(\gamma) =$$
$$= \int_{C([0,1];\mathbb{R}^n)} |\gamma(s)-\gamma(t)| d\eta(\gamma) = \int |x-y|^2 d(e_s,e_t)_\#\eta(x,y).$$

Since $(e_s,e_t)_\#\eta \in \Gamma(\mu_s,\mu_t)$, we obtain the inequality

$$|s-t|^2 W_2^2(\mu.\nu) \geq W_2^2(\mu_s,\mu_t).$$

By the characterization of $\mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^n))$, $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^n))$, and the result follows. ■

It is interesting to point out that every constant speed geodesic can be lifted to an optimal geodesic plan between its extrema.

(4.6) Theorem *Every* $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^n))$ *has a lifting* $\eta \in \mathrm{OptGeo}(\mu_0,\mu_1)$.

Proof. Consider $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^n))$. Let $N \in \mathbb{N}$ be such that $N \geq 3$ and take, for every $i = 1,\dots,N$, $\vartheta_i \in \Gamma(\mu_{\frac{i-1}{N}}, \mu_{\frac{i}{N}})$ optimal. By Iterated Dudley's Lemma, there exists $\pi_N \in \mathcal{P}((\mathbb{R}^n)^{N+1})$ such that for every $i = 1,\dots,N$

$$p_\#^{i-1,i}\pi_N = \vartheta_i.$$

For every $i = 1,\dots,N$, consider $\Phi_i : \mathbb{R}^n \times \mathbb{R}^n \to \mathrm{Geo}(\mathbb{R}^n)$ ϑ_i-measurable such that

$$\Phi_i(x,y) = \gamma_{x,y}$$

where $\gamma_{x,y} : [0,1] \to \mathbb{R}^n$ is such that $\gamma_{x,y}(t) = ty + (1-t)x$. Define the function $\Phi^N : (\mathbb{R}^n)^{N+1} \to C([0,1];\mathbb{R}^n)$ such that for every $i = 1,\dots,N$

$$\Phi^N(x_0,\dots,x_N)|_{\left[\frac{i-1}{N},\frac{i}{N}\right]} = x_{i-1} + (x_i - x_{i-1})\left(N\left(t - \frac{i-1}{N}\right)\right).$$

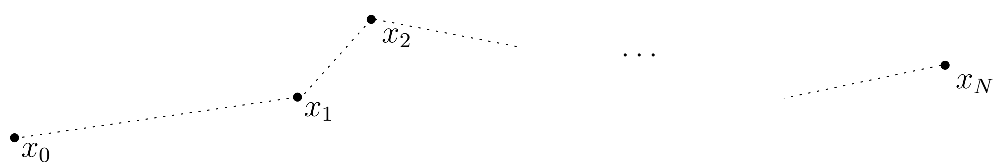

Figure 4.2: A piecewise geodesic connecting $N+1$ points.

Consider now $\eta_N = (\Phi^N)_\# \pi_N \in \mathcal{P}(C([0,1];\mathbb{R}^n))$,

$$\begin{aligned}\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2 d\eta_N &= \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2 d(\Phi^N)_\# \pi_N(\gamma) = \int \mathcal{A}_2 \circ \Phi^N d\pi_N = \\ &= \int \int_0^1 |(\Phi^N)'|^2 d\mathcal{L}^1 d\pi_N = \int \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} |(\Phi^N)'| d\mathcal{L}^1 d\pi_N = \\ &= N^2 \sum_{i=1}^N \int \left(\int_{\frac{i-1}{N}}^{\frac{i}{N}} |x_i - x_{i-1}|^2 d\mathcal{L}^1 \right) d\pi_N(x_0, \ldots, x_N).\end{aligned}$$

Using the change of variables $s = N\left(t - \frac{i-1}{N}\right)$, we obtain

$$\begin{aligned}\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2 d\eta_N &= N \sum_{i=1}^N \int |x_i - x_{i-1}|^2 d\pi_N(x_0, \ldots, x_N) = \\ &= N \sum_{i=1}^N \int |x_i - x_{i-1}|^2 dp_\#^{i-1,1}(\pi_N)(x_{i-1}, x_i) = \\ &= N \sum_{i=1}^N \int |x_i - x_{i-1}|^2 d\vartheta_i(x_{i-1}, x_i) = N \sum_{i=1}^N W_2^2(\mu_{\frac{i-1}{N}}, \mu_{\frac{i}{N}}).\end{aligned}$$

By the fact that μ_t is a geodesic,

$$W_2^2(\mu_{\frac{i-1}{N}}, \mu_{\frac{i}{N}}) = \frac{1}{N^2} W_2^2(\mu_0, \mu_1),$$

so we can write

$$\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2 d\eta_N = W_2^2(\mu_0, \mu_1), \tag{4.7}$$

then

$$\sup_N \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta_N(\gamma) < +\infty.$$

Observing that for every $t \in [0,1]$ and all $\varphi \in \mathrm{Lip}_b(\mathbb{R}^n)$ we have,

$$\begin{aligned}&\left| \int_{C([0,1];\mathbb{R}^n)} \varphi(\gamma(t)) d\eta_N(\gamma) - \int_{C([0,1];\mathbb{R}^n)} \varphi\left(\gamma\left(\frac{\lfloor Nt \rfloor}{N}\right)\right) d\eta_N(\gamma) \right| \leq \\ &\leq \int_{C([0,1];\mathbb{R}^n)} \left| \varphi(\gamma(t)) - \varphi\left(\gamma\left(\frac{\lfloor Nt \rfloor}{N}\right)\right) \right| d\eta_N(\gamma) \leq \\ &\leq \mathrm{Lip}(\varphi) \int_{C([0,1];\mathbb{R}^n)} \left| \gamma(t) - \gamma\left(\frac{\lfloor Nt \rfloor}{N}\right) \right| d\eta_N(\gamma) \leq \\ &\leq \mathrm{Lip}(\varphi) \int_{C([0,1];\mathbb{R}^n)} \int_{\frac{\lfloor Nt \rfloor}{N}}^t |\gamma'| d\mathcal{L}^1 d\eta_N(\gamma),\end{aligned}$$

so, applying Hölder's Inequality and thanks to (4.7),

$$
\begin{aligned}
&\left|\int_{C([0,1];\mathbb{R}^n)} \varphi(\gamma(t))d\eta_N(\gamma) - \int_{C([0,1];\mathbb{R}^n)} \varphi\left(\gamma\left(\frac{\lfloor Nt\rfloor}{N}\right)\right) d\eta_N(\gamma)\right| \leq \\
&= \mathrm{Lip}(\varphi)\sqrt{t-\frac{\lfloor Nt\rfloor}{N}}\int_{C([0,1];\mathbb{R}^n)} \sqrt{\int_{\frac{\lfloor Nt\rfloor}{N}}^{t} |\gamma'|^2 d\mathcal{L}^1} d\eta_N(\gamma) \leq \\
&\leq \mathrm{Lip}(\varphi)\sqrt{t-\frac{\lfloor Nt\rfloor}{N}}\int_{C([0,1];\mathbb{R}^n)} \sqrt{\int_0^1 |\gamma'|^2 d\mathcal{L}^1} d\eta_N(\gamma) = \\
&= \mathrm{Lip}(\varphi)\sqrt{t-\frac{\lfloor Nt\rfloor}{N}}\int_{C([0,1];\mathbb{R}^n)} \sqrt{\mathcal{A}_2(\gamma)} d\eta_N(\gamma) \leq \\
&\leq \mathrm{Lip}(\varphi)\sqrt{t-\frac{\lfloor Nt\rfloor}{N}}\sqrt{\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta_N(\gamma)} = \mathrm{Lip}(\varphi)\sqrt{t-\frac{\lfloor Nt\rfloor}{N}} W_2(\mu_0,\mu_1).
\end{aligned}
$$

We can therefore write

$$
\lim_N \int \varphi d(e_t)_{\#}\eta_N = \lim_N \int \varphi d(e_{\frac{\lfloor Nt\rfloor}{N}})_{\#}\eta_N = \lim_N \int \varphi d\mu_{\frac{\lfloor Nt\rfloor}{N}} = \int_X \varphi d\mu_t.
$$

In other words, $(e_t)_{\#}\eta_N \rightharpoonup \mu_t$ and, thanks to Prokhorov's Theorem, $((e_t)_{\#}\eta_N)$ is tight in $\mathcal{P}(\mathbb{R}^n)$. Then, random Ascoli–Arzelà's Theorem gives us the existence of (η_{N_j}) such that $\eta_{N_j} \rightharpoonup \eta$ in $\mathcal{P}(C([0,1];\mathbb{R}^n))$ for some $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$. Writing (4.7) for (η_{N_j}) and passing to the limit as $j \to +\infty$, remembering that $\mathcal{A}_2$ is lower semicontinuous, we obtain, by Lemma (P.20),

$$
\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) \leq W_2^2(\mu_0,\mu_1).
$$

By the fact that

$$
\int_X \varphi d(e_t)_{\#}\eta = \lim_N \int_X \varphi d(e_t)_{\#}\eta_N = \int_X \varphi d\mu_t,
$$

we also obtain $(e_t)_{\#}\eta = \mu_t$ for every $t \in [0,1]$. ∎

Since in $\mathbb{R}^n$ we have the uniqueness of geodesics, it is interesting to understand if the same result holds in $\mathcal{P}_2(\mathbb{R}^n)$. What we can say, in general, is the following result.

(4.8) Proposition *Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. If there exists a unique $\pi \in \Gamma(\mu_0,\mu_1)$ optimal, then there exists a unique $\eta \in \mathrm{OptGeo}(\mu_0,\mu_1)$ and a unique $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^n))$ connecting μ_0 and μ_1. In particular, $\mu_t = (e_t)_{\#}\eta$.*

Proof. As seen in Theorem (4.3), $\eta = \Phi_{\#}\pi \in \mathrm{OptGeo}(\mu_0, \mu_1)$. If $\overline{\eta}$ is another optimal geodesic plan between μ_0 and μ_1, we have $\mathrm{supt}(\overline{\eta}) \subseteq \mathrm{Geo}(\mathbb{R}^n)$ and $(e_0, e_1)_{\#}\overline{\eta} = \pi$. By the disintegration Theorem, $\overline{\eta} = \eta$, so the uniqueness of the optimal geodesic plan follows. The result therefore follows from Theorem (4.6). ∎

More specifically, if $\mu_0 \ll \mathcal{L}^n$, we can combine Theorem (3.10) and the previous Proposition to get the uniqueness of geodesics.

(4.9) Corollary *Consider $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. If $\mu_0 \ll \mathcal{L}^n$, then there exists a unique $\mu_t \in \mathrm{Geo}(\mathcal{P}_2(\mathbb{R}^n))$ joining μ_0, μ_1 and a unique $\eta \in \mathrm{OptGeo}(\mu_0, \mu_1)$.*

In particular,

$$\mu_t = (T_t)_{\#}\mu_0,$$

where T is the optimal transport map from μ_0 to μ_1 and

$$T_t = (1-t)\mathrm{Id} + tT.$$

Proof. Since, by Theorem (3.10), $\pi = (\mathrm{Id}, T)_{\#}\mu_0$ is the unique optimal plan from μ_0 to μ_1, the result follows from Proposition (4.8). ∎

We conclude the section with a semi-explicit (we omit computations) example of geodesics between two non-atomic measures: further technical details can be found in [23], [29], [37], [45], [53] and [43], but we want to focus only on the results.

(4.10) Example *Fix K_0 and K_1 two symmetric and positive definite matrices and $m_0, m_1 \in \mathbb{R}^n$. Consider*

$$\mu_0 = \frac{1}{\sqrt{(2\pi)^n \det(K_0)}} e^{-\frac{1}{2}(x-m_0)\cdot K_0^{-1}(x-m_0)}\mathcal{L}^n,$$
$$\mu_1 = \frac{1}{\sqrt{(2\pi)^n \det(K_1)}} e^{-\frac{1}{2}(x-m_1)\cdot K_1^{-1}(x-m_1)}\mathcal{L}^n.$$

It can be shown

$$\min_{\pi \in \Gamma(\mu_0,\mu_1)} C(\pi) = W_2^2(\mu_0, \mu_1) = |m_0 - m_1|^2 + \mathrm{Tr}(K_0) + \mathrm{Tr}(K_1) - 2\mathrm{Tr}(K_1^{\frac{1}{2}} K_0 K_1^{\frac{1}{2}})^{\frac{1}{2}},$$

where Tr *is the trace (of a matrix) operator, and the geodesic between μ_0 and μ_1 is*

$$\mu_t = \frac{1}{\sqrt{(2\pi)^n \det(K_t)}} e^{-\frac{1}{2}(x-m_t)\cdot K_t^{-1}(x-m_t)}\mathcal{L}^n,$$

where $m_t = tm_1 + (1-t)m_0$ *and* $K_t = t^2K_1 + (1-t)^2K_0 + t(1-t)\left((K_0K_1)^{\frac{1}{2}} + (K_1K_0)^{\frac{1}{2}}\right)$.

2 Dynamic Schrödinger problem

Even if the formulations given so far for the Schrödinger problem are the simplest, since what we are studying is an interpolation problem, the most natural thing is to study it from the dynamic point of view. In other words, we would like to recover its most intuitive meaning and formalize it: We will see in this section how to translate the statement "*finding the most likely evolution for systems of diffusive particles between two different observations*" into abstract mathematical language and how to manage this formulation. Even though we will treat the theory with a fairly high level of generality, let us keep in mind that the basic idea is given by the following example.

(4.11) Example *Consider a system of i.i.d. Brownian particles. Suppose we know, with a certain accuracy, the positions at two different times. We look for the Brownian bridge that connects the observations: in other words, we look for the trajectories of the Brownian particles.*

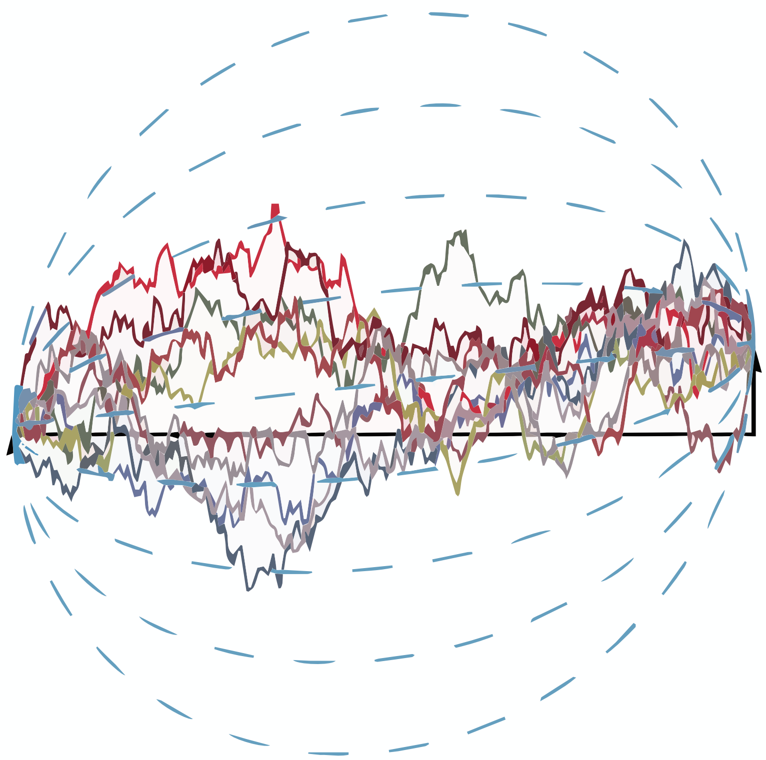

Figure 4.3: An example of Brownian bridge (picture from [38]).

Coming back to us, before going on, we show the following technical result that is necessary for us to have the consistency of the definition of the dynamic Schrödinger problem we want to propose.

(4.12) Proposition *Consider* $\mathsf{R} \in \mathcal{R}_+(C([0,1];\mathbb{R}^n))$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. *If for some (and thus all)* $z \in \mathbb{R}^n$

$$\int e^{-|x-z|^2-|y-z|^2} d(e_0, e_1)_\# \mathsf{R}(x,y) < +\infty,$$

then, for every $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ *such that* $(e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1$, $H_{\mathsf{R}}(\eta)$ *is well defined.*

Proof. Consider $W = |e_0 - z|^2 + |e_1 - z|^2$. Using the change of variables formula,

$$\int_{C([0,1];\mathbb{R}^n))} e^{-W} d\mathsf{R} = \int e^{-|x-z|-|y-z|} d(e_0, e_1)_\# \mathsf{R}(x,y) < +\infty$$

and if $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ such that $(e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1$, using the fact that $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$,

$$\begin{aligned} \int_{C([0,1];\mathbb{R}^n))} W d\eta = \int_{C([0,1];\mathbb{R}^n))} |e_0 - z|^2 d\eta + \int_{C([0,1];\mathbb{R}^n))} |e_1 - z|^2 d\eta = \\ = \int |x-z|^2 d\mu_0(x) + \int |y-z|^2 d\mu_1(y) < +\infty. \blacksquare \end{aligned}$$

(4.13) Definition *Consider* $\mathsf{R} \in \mathcal{R}_+(C([0,1];\mathbb{R}^n))$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. *Suppose for some (and thus all)* $z \in \mathbb{R}^n$

$$\int e^{-|x-z|^2-|y-z|^2} d(e_0, e_1)_\# \mathsf{R}(x,y) < +\infty.$$

In the so-called dynamic Schrödinger problem *we look for*

$$\inf \left\{ H_{\mathsf{R}}(\eta) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)) : (e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1 \right\}.$$

(4.14) Proposition *Consider* $\mathsf{R} \in \mathcal{R}_+(C([0,1];\mathbb{R}^n))$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. *Suppose* $H(\mu_0 \otimes \mu_1 | (e_0.e_1)_\# \mathsf{R}) < +\infty$ *and for some (and thus all)* $z \in \mathbb{R}^n$

$$\int e^{-|x-z|^2-|y-z|^2} d(e_0, e_1)_\# \mathsf{R}(x,y) < +\infty.$$

If the dynamic Schrödinger problem admits a solution, the latter is unique.

Proof. Using Proposition (1.23), H_{R} is strictly convex where it is finite, so the uniqueness follows. ■

(4.15) Lemma *Consider* $\mathsf{R} \in \mathcal{R}_+(C([0,1];\mathbb{R}^n))$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. *Suppose for some (and thus all)* $z \in \mathbb{R}^n$

$$\int e^{-|x-z|^2-|y-z|^2} d(e_0,e_1)_\#\mathsf{R}(x,y) < +\infty.$$

For every $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ *such that* $(e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1$

$$H_{\mathsf{R}}(\eta) = H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\eta) + \int H_{\mathsf{R}_{xy}}(\eta_{xy}) d(e_0,e_1)_\#\eta(x,y).$$

Proof. It is sufficient to write [41, Theorem (1.6), (c)] for η and R, applying both sides log, integrating with respect to η and then applying Proposition (P.6). ■

Let us now analyze the connection between the primal formulation and the dynamic formulation of the Schrödinger problem.

(4.16) Theorem (Föllmer) *Consider* $\mathsf{R} \in \mathcal{R}_+(C([0,1];\mathbb{R}^n))$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. *Suppose* $H(\mu_0 \otimes \mu_1 | (e_0.e_1)_\#\mathsf{R}) < +\infty$ *and for some (and thus all)* $z \in \mathbb{R}^n$

$$\int e^{-|x-z|^2-|y-z|^2} d(e_0,e_1)_\#\mathsf{R}(x,y) < +\infty.$$

The following facts hold true:

(a) *if* $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ *is the solution to the dynamic Schrödinger problem, then for every* $B \in \mathscr{B}(C([0,1];\mathbb{R}^n))$

$$\eta(B) = \int \mathsf{R}_{xy}(B) d(e_0,e_1)_\#\eta(x,y)$$

and $(e_0,e_1)_\#\eta$ *is the solution to the Schrödinger problem,*

(b) *if* γ *is the solution to the Schrödinger problem, then* $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ *such that, for every* $B \in \mathscr{B}(C([0,1];\mathbb{R}^n))$,

$$\eta(B) = \int \mathsf{R}_{xy}(B) d\gamma(x,y),$$

is the solution to the dynamic Schrödinger problem.

Proof. First of all, by Lemma (4.15), if $\eta \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ such that $(e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1$, then

$$H_{\mathsf{R}}(\eta) = H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\eta) + \int H_{\mathsf{R}_{xy}}(\eta_{xy})d(e_0,e_1)_\#\eta(x,y).$$

In particular, since $\mathsf{R}_{xy}, \eta_{xy} \in \mathcal{P}(C([0,1],\mathbb{R}^n))$, by Proposition (1.23), $H_{\mathsf{R}_{xy}}(\eta_{xy}) \geq 0$, so

$$H_{\mathsf{R}}(\eta) \geq H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\eta)$$

with equality if and only if $\eta_{xy} = \mathsf{R}_{xy}$ for $(e_0,e_1)_\#\eta$-a.e. $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$.

(a) Let us start proving that for every for every $B \in \mathscr{B}(C([0,1];\mathbb{R}^n))$

$$\eta(B) = \int \mathsf{R}_{xy}(B)d(e_0,e_1)_\#\eta(x,y).$$

Consider, for every $B \in \mathscr{B}(C([0,1];\mathbb{R}^n))$

$$\overline{\eta}(B) = \int \mathsf{R}_{xy}(B)d(e_0,e_1)_\#\eta(x,y).$$

By construction, we have $\overline{\eta}_{xy} = \mathsf{R}_{xy}$ for $(e_0,e_1)_\#\eta$-a.e. $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ so

$$H_{\mathsf{R}}(\overline{\eta}) = H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\overline{\eta}) \leq H_{\mathsf{R}}(\eta).$$

By minimality of η, the previous inequality must be an equality, and using the uniqueness of the minimizer, it follows that $\overline{\eta} = \eta$.

Now, consider any $\pi \in \Gamma(\mu_0,\mu_1)$ and define for every $B \in \mathscr{B}(C([0,1];\mathbb{R}^n))$

$$\overline{\pi}(B) = \int \mathsf{R}_{xy}(B)d\pi(x,y).$$

By minimality of η, it must be $H_{\mathsf{R}}(\eta) \leq H_{\mathsf{R}}(\overline{\pi})$. By the way,

$$H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\eta) = H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\eta) + \int H_{\mathsf{R}_{xy}}(\eta_{xy})d(e_0,e_1)_\#\eta(x,y) = H_{\mathsf{R}}(\eta) \leq$$
$$\leq H_{\mathsf{R}}(\overline{\pi}) = H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\overline{\pi}) = H_{(e_0,e_1)_\#\mathsf{R}}(\pi),$$

so $(e_0,e_1)_\#\eta$ is the solution to the Schrödinger problem.

(b) Consider any $\overline{\eta} \in \mathcal{P}(C([0,1];\mathbb{R}^n))$ such that $(e_0)_\#\overline{\eta} = \mu_0$ and $(e_1)_\#\overline{\eta} = \mu_1$. Then, by minimality of γ,

$$H_{\mathsf{R}}(\overline{\eta}) \geq H_{(e_0,e_1)_\#\mathsf{R}}((e_0,e_1)_\#\overline{\eta}) \geq H_{(e_0,e_1)_\#\mathsf{R}}(\gamma) = H_{\mathsf{R}}(\eta),$$

and the result follows. ∎

(4.17) Corollary *Consider* $\mathsf{R} \in \mathcal{R}_+(C([0,1];\mathbb{R}^n))$ *and* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$. *Suppose* $H(\mu_0 \otimes \mu_1 | (e_0.e_1)_\#\mathsf{R}) < +\infty$ *and for some (and thus all)* $z \in \mathbb{R}^n$

$$\int e^{-|x-z|^2-|y-z|^2} d(e_0,e_1)_\#\mathsf{R}(x,y) < +\infty.$$

It results

$$\inf_{\Gamma(\mu_0,\mu_1)} H_{(e_0,e_1)_\#\mathsf{R}} = \inf \left\{ H_{\mathsf{R}}(\eta) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)) : (e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1 \right\}.$$

Proof. It is a matter of reasoning exactly as in Föllmer's Theorem. ∎

3 Comparison

With the introduction of the Wiener measure $\overline{\mathsf{R}}_{\frac{\varepsilon}{2}}$, namely that measure having $\mathsf{r}_{\frac{\varepsilon}{2}}\delta_0 \otimes \mathcal{L}^n$ as joint law at the times 0 and 1 and whose disintegration with respect to (e_0, e_1) is a family (indexed by $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$) of Brownian bridges with variance $\frac{\varepsilon}{2}$, it is possible to dynamically visualize the example (1.30).

By the fact that $\left\{\overline{\mathsf{R}}_{\frac{\varepsilon}{2}}\right\}_{\varepsilon>0}$ satisfies a large deviation principle with rate function $I = \frac{1}{2}\mathcal{A}_2$, using [44, Theorem 3.4] and [44, Theorem 3.5], we get

$$\Gamma - \lim_{\varepsilon\to 0} \varepsilon H_{\overline{\mathsf{R}}_{\frac{\varepsilon}{2}}} = I.$$

In particular, using the coercivity of I,

$$\begin{aligned} &\lim_{\varepsilon\to 0} \min \left\{ H_{\overline{\mathsf{R}}_{\frac{\varepsilon}{2}}}(\eta) : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)) : (e_0)_\#\eta = \mu_0, (e_1)_\#\eta = \mu_1 \right\} = \\ &= \min \left\{ \int_{C([0,1];\mathbb{R}^n)} I d\eta : \eta \in \mathcal{P}(C([0,1];\mathbb{R}^n)), (e_0)_\#\eta = \mu, (e_1)_\#\eta = \nu \right\} \end{aligned}$$

and the (unique) minimizer of $H_{\overline{\mathsf{R}}_{\frac{\varepsilon}{2}}}$ converges to the (unique) minimizer of the dynamic optimal transport problem.

Chapter 5

Benamou–Brenier formulas

Also in this chapter, we will consider, as an environment, $\mathbb{R}^n$ equipped with the Euclidean distance.

1 Continuity equation

In the Theory of Continuous Bodies, the Law of Conservation of Mass in Eulerian form, namely written in a reference system that fixes space rather than the material points moving within it, is expressed by the so-called continuity equation, namely

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(v\rho) = 0,$$

where ρ is the mass density function of the body and v its velocity field, both in Eulerian coordinates. Further information on continuous bodies can be found in [5] or, for non-Italian readers, in [32].

Since we describe mass using the language of Measure Theory, and to allow a more in-depth theoretical study on the existence of solutions, we need a distributional form of the equation.

(5.1) Definition *Let $T \in]0, +\infty]$, $c > 0$, $\mu_t :]0, T[\to \mathcal{M}(\mathbb{R}^n)$ be weakly continuous and a Borel function $v :]0, T[\times \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$(t, x) \longmapsto v_t(x),$$

where the function $\{x \longmapsto |v_t|\}$ belongs to $L^1(\mu_t)$ and the function $\left\{t \longmapsto \|v_t\|_{L^1(\mu_t)}\right\}$ belongs to $L^1_{loc}(]0, T[)$, and $\overline{\mu} \in \mathcal{M}(\mathbb{R}^n)$. We say that μ_t is a distributional solution *of* *the* continuity problem*, namely*

$$\begin{cases} \frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = 0, \\ \mu_0 = \overline{\mu}, \end{cases}$$

if for every $\varphi \in C_c^\infty([0,T[\times \mathbb{R}^n)$ *it holds*

$$\int_0^T \int \left(\frac{\partial \varphi}{\partial t}(t,x) + v_t(x) \cdot D\varphi(t,x) \right) d\mu_t(x)dt + \int \varphi(0,x) d\overline{\mu} = 0.$$

In particular, we say that $\mu_t :]0,T[\to \mathcal{M}(\mathbb{R}^n)$ *weakly continuous is a* distributional solution *of the* continuity equation*, namely*

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = 0,$$

if for every $\varphi \in C_c^\infty(]0,T[\times \mathbb{R}^n)$ *it holds*

$$\int_0^T \int \left(\frac{\partial \varphi}{\partial t}(t,x) + v_t(x) \cdot D\varphi(t,x) \right) d\mu_t(x)dt = 0.$$

The pair (μ_t, v_t)*, where* μ_t *is a distributional solution of the continuity equation defined by* v_t*, is called* continuity pair*.*

The technical assumptions are required to have the integrals well defined. By the way, the basic idea of the definition is the interpretation of a measure as a distribution. The reader interested in further information about Distribution Theory can read [59, Part II].

Proving that a curve of measures is a distributional solution of the continuity equation can be tricky using the definition. Let us look at a useful equivalent formulation that speeds up the process. We prove first a technical Lemma.

(5.2) Lemma *The family*

$$D = \left\{ \sum_{i=1}^N \alpha_i \varphi_i : \alpha_i \in C_c^\infty(]0,+\infty[), \varphi_i \in C_c^\infty(\mathbb{R}^n), N \in \mathbb{N} \right\}$$

is dense in $C_c^\infty(]0,+\infty[\times \mathbb{R}^n)$*.*

Proof. It is sufficient to apply a variant of the Stone–Weierstrass Theorem. ∎

(5.3) Proposition *Let* $\mu_t :]0,+\infty[\to \mathcal{M}(\mathbb{R}^n)$ *be weakly continuous and a Borel* *function* $v :]0, +\infty[\times \mathbb{R}^n \to \mathbb{R}^n$ *such that*

$$(t, x) \longmapsto v_t(x),$$

with the function $\{x \longmapsto |v_t|\}$ *belonging to* $L^1(\mu_t)$ *and the function* $\left\{t \longmapsto \|v_t\|_{L^1(\mu_t)}\right\}$ *belonging to* $L^1_{loc}(]0, +\infty[)$. *The following facts are equivalent:*

(*a*) μ_t *is a distributional solution of the continuity equation,*

(*b*) *for every* $\varphi \in C_c^\infty(\mathbb{R}^n)$ *the function* $\left\{t \mapsto \int \varphi(x) d\mu_t(x)\right\} \in AC_{loc}(]0, +\infty[)$ *and its weak derivative is*

$$\frac{d}{dt} \int \varphi d\mu_t(x) = \int D\varphi(x) \cdot v_t(x) d\mu_t(x).$$

Proof. Let us start by saying that, by Lemma (5.2), linearity and the dominated convergence Theorem, it is sufficient to consider tests of the form $\alpha\varphi$, where $\alpha \in C_c^\infty(]0, +\infty[), \varphi \in C_c^\infty(\mathbb{R}^n)$. Then

$$\int_0^\infty \int (\alpha'(t)\varphi(x) + \alpha(t) v_t(x) \cdot D\varphi(x)) \, d\mu_t(x) dt =$$
$$= \int_0^\infty \left(\alpha'(t) \int \varphi(x) d\mu_t(x) + \alpha(t) \int v_t(x) \cdot D\varphi(x) d\mu_t(x) \right) dt.$$

The previous equality proves both implications. ■

Let us immediately see a first application to the optimal transport theory of the distributional interpretations of the continuity equation: geodesics in $\mathcal{P}_2(\mathbb{R}^n)$ are solutions of a continuity equation.

(5.4) Proposition *Consider* $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ *with* $\mu_0 \ll \mathcal{L}^n$. *Then* μ_t, *the geodesic connecting* μ_0 *and* μ_1, *is a distributional solution of a continuity equation.*

Proof. By Theorem (3.10), there exists a unique, up to μ_0-negligible sets, minimizer T for the problem

$$\min \left\{ \int_{\mathbb{R}^n} \frac{1}{2} |x - T(x)|^2 d\mu_0(x) : T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel, } T_\# \mu_0 = \mu_1 \right\}.$$

Furthermore, by Corollary (4.9), we also know that the constant speed geodesic from μ_0 to μ_1 is

$$\mu_t = (T_t)_\# \mu_0$$

where

$$T_t = (1-t)\mathrm{Id} + tT.$$

Now, for every $\varphi \in C_c^\infty(\mathbb{R}^n)$ and for a.e. $t \in]0,+\infty[$ we have, by differentiation under the integral sign Theorem,

$$\frac{d}{dt}\int_{\mathbb{R}^n} \varphi d\mu_t = \frac{d}{dt}\int_{\mathbb{R}^n} \varphi(T_t(x))d\mu_0(x) = \int D\varphi(T_t(x))\cdot (T-\mathrm{Id})(x)d\mu_0(x) =$$
$$= \int D\varphi(x)\cdot \left((T-\mathrm{Id})\circ T_t^{-1}\right)(x)d\mu_t(x).$$

Naming

$$v_t = (T-\mathrm{Id})\circ T_t^{-1} = \frac{\mathrm{Id} - T_t^{-1}}{t}$$

and using Proposition (5.3), we obtain a continuity equation for μ_t, namely

$$\frac{\partial \mu_t}{\partial t} + \mathrm{div}\,(v_t\mu_t) = 0,$$

interpreted in the distributional sense. ■

We conclude the section by reporting the following Lemma, without proof, which will be useful later.

(5.5) Lemma *Let (μ_t, v_t) be a continuity pair and $f \in H^1(\mathbb{R}^n)$. Then the function $\left\{t \longmapsto \int f d\mu_t\right\}$ is absolutely continuous and for* a.e. $t \in [0,1]$

$$\left|\frac{d}{dt}\int f d\mu_t\right| \leq \int |Df||v_t|d\mu_t,$$

where the exceptional set can be chosen to be independent of f.

Moreover, if the function $\{t \longmapsto f_t\}$ belongs to

$$AC([0,1];L^2(\mathbb{R}^n)) \cap L^\infty([0,1];H^1(\mathbb{R}^n)),$$

then the function $\left\{t \longmapsto \int f_t d\mu_t\right\}$ is absolutely continuous and for a.e. $t \in [0,1]$

$$\frac{d}{ds}\Big(\int f_s d\mu_s\Big)|_{s=t} = \int \frac{d}{ds} f_s(t)d\mu_t + \frac{d}{ds}\Big(\int f_t d\mu_s\Big)|_{s=t}.$$

Proof. See [27]. ■

2 Optimal transport

We are now interested in rewriting the optimal transport problem as a distributional fluid-dynamic problem: in particular, a problem of curves of measures subject to the continuity constraint.

First of all, let us introduce the functional we want to minimize: the quadratic action. From a physical viewpoint, it can be interpreted as a rescaling of the kinetic energy.

(5.6) Definition *Consider $\mu \in \mathcal{P}_2(\mathbb{R}^n)$ and $v : \mathbb{R}^n \to \mathbb{R}^n$ a Borel vector field. We define their* quadratic action *as*

$$\mathcal{A}(v, \mu) = \int |v|^2 d\mu.$$

In the following, we will need to use the standard Gaussian several times, so we establish the following Notation.

(5.7) Notation We denote with ϱ the function in $C^\infty(\mathbb{R}^n)$ such that

$$\varrho(x) = \frac{e^{-|x|^2}}{\int e^{-|x|^2} d\mathcal{L}^n(x)}$$

and for $\varepsilon > 0$,

$$\varrho_\varepsilon(x) = \frac{1}{\varepsilon^n} \varrho\left(\frac{x}{\varepsilon}\right).$$

(5.8) Remark *Switching from Cartesian coordinates to generalized spherical ones, we can state that*

$$\int |x|^2 \varrho(x) d\mathcal{L}^n(x) \sim \int_0^\infty r^{n+1} e^{-r^2} dr < +\infty.$$

(5.9) Lemma *Consider $\mu \in \mathcal{P}_2(\mathbb{R}^n)$. For every $\varepsilon > 0$, $\mu * \varrho_\varepsilon \in C^\infty(\mathbb{R}^n)$ and $\mu * \varrho_\varepsilon > 0$. Furthermore, as $\varepsilon \to 0^+$, $(\mu * \varrho_\varepsilon)\mathcal{L}^n \to \mu$ in $(\mathcal{P}_2(\mathbb{R}^n), W_2)$.*

Proof. First of all,

$$\mu * \varrho_\varepsilon(x) = \int \varrho_\varepsilon(x - y) d\mu(y).$$

In particular, $\mu * \varrho_\varepsilon \in C^\infty(\mathbb{R}^n)$ and $\mu * \varrho_\varepsilon > 0$.

Now, consider for every $B \in \mathscr{B}(\mathbb{R}^n \times \mathbb{R}^n)$

$$\Sigma_\varepsilon(B) = \int \int_{B_x} \varrho_\varepsilon(x - y) d\mathcal{L}^n(y) d\mu(x),$$

where $B_x = B \cap \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : y \in \mathbb{R}\}$, namely a fiber of B. Then, using the change of variables $z = \frac{x-y}{\varepsilon}$,

$$\begin{aligned} W_2^2(\mu, \mu * \varrho_\varepsilon \mathcal{L}^n) \leq \int |x-y|^2 d\Sigma_\varepsilon(x,y) &= \int \int |x-y|^2 \varrho_\varepsilon(x-y) d\mathcal{L}^n(y) d\mu(x) = \\ &= \frac{1}{\varepsilon^n} \int \int |x-y|^2 \varrho\left(\frac{x-y}{\varepsilon}\right) d\mathcal{L}^n(y) d\mu(x) = \\ &= \frac{\varepsilon^2}{\varepsilon^n} \varepsilon^n \int \int |z|^2 \varrho(z) d\mathcal{L}^n(z) d\mu = \varepsilon^2 \mu(\mathbb{R}^n) \int |z|^2 \varrho(z) d\mathcal{L}^n(z) = \\ &= \varepsilon^2 \int |z|^2 \varrho(z) d\mathcal{L}^n(z), \end{aligned}$$

so $W_2^2(\mu, \mu * \varrho_\varepsilon \mathcal{L}^n) \to 0$ as $\varepsilon \to 0^+$. ∎

The next is the main result of the section, where we show that the optimal transport problem is equivalently reformulated through the minimization of the quadratic action functional subject to the continuity constraint. As regards the $\leq$ inequality, arguing by density, using a convolution of measures, we have proved that it is sufficient to deal with the case of sub-linear growth of the vector field v_t: for this specific case, the proof is a direct computation that requires the flow of the vector field. For the $\geq$ inequality and the existence of the minimizers, we explicitly build the curve of measures and the vector field using geodesics in $\mathbb{R}^n$ and Lemma (P.10). All the ideas are expansions of the contents taken from [1].

(5.10) Theorem (Benamou–Brenier formula) *For every $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ it holds*

$$W_2^2(\mu_0, \mu_1) = \min \left\{ \int_0^1 \mathcal{A}(v_t, \mu_t) dt : (\mu_t, v_t) \text{ continuity pair, } \mu_t : [0,1] \to \mathcal{P}_2(\mathbb{R}^n) \right\}.$$

Proof. First of all, let us start proving

$$W_2^2(\mu_0, \mu_1) \leq \inf \left\{ \int_0^1 \mathcal{A}(v_t, \mu_t) dt : (\mu_t, v_t) \text{ continuity pair, } \mu_t : [0,1] \to \mathcal{P}_2(\mathbb{R}^n) \right\}.$$

Take μ_t a distributional solution of

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = 0$$

with v_t smooth and with no more than linear growth. By [1, Proposition 16.4],

$$\mu_t = (X_t)_\# \mu_0,$$

where X_t is the flow of v_t. Considering the admissible plan $\Sigma = (\mathrm{Id}, X_1)_\# \mu_0$,

$$W_2^2(\mu_0, \mu_1) \le \int |x-y|^2 d\Sigma(x,y) = \int |X_0(x) - X_1(x)|^2 d\mu_0(x) =$$
$$= \int \left| \int_0^1 \frac{d}{dt} X_t(x) dt \right|^2 d\mu_0(x) = \int \left| \int_0^1 v_t(X_t(x)) dt \right|^2 d\mu_0(x)$$

so, by Hölder's inequality and Fubini–Tonelli's Theorem,

$$W_2^2(\mu_0, \mu_1) \le \int \int_0^1 |v_t(X_t(x))|^2 dt d\mu_0(x) = \int_0^1 \int |v_t(X_t(x))|^2 d\mu_0(x) dt =$$
$$= \int_0^1 \int |v_t|^2 d\mu_0 dt = \int_0^1 \mathcal{A}(v_t, \mu_t) dt.$$

In the general case, consider for every $t \in [0,1]$

$$\mu_t^\varepsilon = \mu_t * \varrho_\varepsilon \mathcal{L}^n, \qquad v_t^\varepsilon = \frac{v_t \mu_t * \varrho_\varepsilon \mathcal{L}^n}{\mu_t^\varepsilon}.$$

Observing that μ_t^ε is a distributional solution of

$$\frac{\partial \mu_t}{\partial t} + \mathrm{div}(v_t^\varepsilon \mu_t) = 0,$$

by the previous case (see [3, Proposition 8.1.8] for details on the growth),

$$W_2^2(\mu_0^\varepsilon, \mu_1^\varepsilon) \le \int_0^1 \mathcal{A}(v_t^\varepsilon, \mu_t^\varepsilon) dt.$$

Now, denoting for every $B \in \mathscr{B}(\mathbb{R}^n)$ and every $x \in \mathbb{R}^n$

$$\nu_x(B) = \frac{1}{\int \varrho_\varepsilon(x-y) d\mu_t(y)} \int_B \varrho_\varepsilon(x-y) d\mu_t(y),$$

by Hölder's inequality,

$$\left| \int v_t(y) d\nu_x(y) \right| \le \int |v_t(y)|^2 d\nu_x(y)$$

so

$$\left| \int \frac{v_t(y) \varrho_\varepsilon(x-y)}{\int \varrho_\varepsilon(x-y) d\mu_t(y)} d\mu_t(y) \right|^2 \le \int \frac{|v_t(y)|^2 \varrho_\varepsilon(x-y)}{\int \varrho_\varepsilon(x-y) d\mu_t(y)} d\mu_t(y)$$

or, in other words,

$$\frac{\left|\int v_t(y)\varrho_\varepsilon(x-y)d\mu_t(y)\right|^2}{|\int \varrho_\varepsilon(x-y)d\mu_t(y)|^2} \leq \frac{\int |v_t(y)|^2\varrho_\varepsilon(x-y)d\mu_t(y)}{\int \varrho_\varepsilon(x-y)d\mu_t(y)}.$$

Multiplying both sides by

$$\int \varrho_\varepsilon(x-y)d\mu_t(y)$$

and integrating in x we get

$$\begin{aligned}\int \frac{\left|\int v_t(y)\varrho_\varepsilon(x-y)d\mu_t(y)\right|^2}{|\int \varrho_\varepsilon(x-y)d\mu_t(y)|^2}\int \varrho_\varepsilon(x-y)d\mu_t(y)d\mathcal{L}^n(x) \leq \\ \leq \int \frac{\int |v_t(y)|^2\varrho_\varepsilon(x-y)d\mu_t(y)}{\int \varrho_\varepsilon(x-y)d\mu_t(y)}\int \varrho_\varepsilon(x-y)d\mu_t(y)d\mathcal{L}^n(x),\end{aligned}$$

so for every $t \in [0,1]$

$$\mathcal{A}(v_t^\varepsilon, \mu_t^\varepsilon) = \int |v_t^\varepsilon|^2 d\mu_t^\varepsilon \leq \int |v_t(y)|^2 \left(\int \varrho_\varepsilon(x-y)d\mathcal{L}^n(x)\right) d\mu_t(y) = \mathcal{A}(v_t, \mu_t).$$

Reassessing,

$$W_2^2(\mu_0^\varepsilon, \mu_1^\varepsilon) \leq \int_0^1 \mathcal{A}(v_t, \mu_t)dt,$$

therefore

$$\begin{aligned}W_2(\mu_0, \mu_1) \leq W_2(\mu_0, \mu_0^\varepsilon) + W_2(\mu_0^\varepsilon, \mu_1^\varepsilon) + W_2(\mu_1^\varepsilon, \mu_1) \leq \\ \leq W_2(\mu_0, \mu_0^\varepsilon) + \left(\int_0^1 \mathcal{A}(v_t, \mu_t)dt\right)^{\frac{1}{2}} + W_2(\mu_1^\varepsilon, \mu_1)\end{aligned}$$

and, by Lemma (5.9), passing to the limit as $\varepsilon \to 0^+$, we obtain

$$W_2^2(\mu_0, \mu_1) \leq \int_0^1 \mathcal{A}(v_t, \mu_t)dt$$

so, passing to the infimum for μ_t, the desired inequality comes.

Let us now prove the converse inequality and the existence of a minimizer. Fix $\Sigma \in \Gamma(\mu_0, \mu_1)$ and for every $t \in]0,1[$ consider the function $e_t : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that

$$e_t(x, y) = (1-t)x + ty = x + t(y-x).$$

Take the geodesic

$$\mu_t = (e_t)_\# \Sigma.$$

Clearly,

$$\int |x-y|^2 d\Sigma(x,y) < +\infty,$$

so, by Lemma (P.10), for every $t \in]0,1[$ there exists $v_t \in L^2(\mathbb{R}^n, \mu_t; \mathbb{R}^n)$ such that $(e_t)_\#(y-x)\Sigma = v_t \mu_t$ and

$$\|v_t\|_{L^2(\mathbb{R}^n,\mu_t;\mathbb{R}^n)} \le \|y-x\|_{L^2(\mathbb{R}^n\times\mathbb{R}^n,\Sigma;\mathbb{R}^n)} = W_2(\mu_0,\mu_1).$$

In particular,

$$W_2^2(\mu_0,\mu_1) \ge \|v_t\|^2_{L^2(\mathbb{R}^n,\mu_t;\mathbb{R}^n)} = \int_{\mathbb{R}^n} |v_t|^2 d\mu_t = \mathcal{A}(v_t,\mu_t)$$

so, integrating both sides in t,

$$W_2^2(\mu_0,\mu_1) \ge \int_0^1 \mathcal{A}(v_t,\mu_t)dt. \tag{5.11}$$

Now, given $\varphi \in C_c^\infty(\mathbb{R}^n)$

$$\frac{d}{dt}\int \varphi d\mu_t = \frac{d}{dt}\int \varphi(e_t(x,y))d\Sigma(x,y)$$

and, by the differentiation under the integral sign Theorem,

$$\frac{d}{dt}\int \varphi d\mu_t = \int D\varphi(e_t(x,y))\cdot(y-x)d\Sigma(x,y) = \int D\varphi\cdot v_t d\mu_t,$$

so, by Proposition (5.3), μ_t is a distributional solution of

$$\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t\mu_t) = 0,$$

then μ_t is a minimizer and the proof is complete. ■

(5.12) Remark *Combining Benamou–Brenier Formula with Theorem (4.3), we obtain*

$$\min\left\{\int_{C([0,1];X)} \mathcal{A}_2(\gamma)d\eta(\gamma) : \eta \in \mathcal{P}(C([0,1];X)), (e_0)_\#\eta = \mu, (e_1)_\#\eta = \nu\right\} =$$
$$= \min\left\{\int_0^1 \mathcal{A}(v_t,\mu_t)dt : (\mu_t, v_t) \text{ continuity pair, } \mu_t : [0,1] \to \mathcal{P}_2(\mathbb{R}^n)\right\}.$$

Let us see an application of the Benamou–Brenier formula to functional inequalities taken from [1].

(5.13) Proposition *Let (μ_t, v_t) be a continuity pair such that $\{t \longmapsto \mathcal{A}(v_t, \mu_t)\}$ belongs to $L^1(0,1)$. Then $\mu_t \in AC^2([0,1]; \mathcal{P}_2(\mathbb{R}^n))$ and $|\mu_t'|^2 \leq \mathcal{A}(v_t, \mu_t)$ for* a.e. $t \in]0,1[$.

Proof. Up to a rescaling in the Benamou–Brenier Formula, for every $s, t \in [0,1]$ with $s \leq t$ it holds

$$W_2^2(\mu_s, \mu_t) \leq (t-s) \int_s^t \mathcal{A}(v_\tau, \mu_t au) d\mathcal{L}^1(\tau),$$

so the results follow by Lemma (P.37). ■

Up to a generalization of the technique used to prove the $\geq$ inequality in the Benamou–Brenier formula, we obtain the following result.

(5.14) Theorem *If $\mu_t \in AC^2([0,1]; \mathcal{P}_2(\mathbb{R}^n))$, then there exists a time-dependent vector field v_t such that (μ_t, v_t) is a continuity pair and for* a.e. $t \in]0,1[$ *it holds*

$$\mathcal{A}(v_t, \mu_t) = |\mu_t'|^2.$$

Proof. First of all, by [1, Remark 10.8], there exists $\eta \in P(C([0,1]); \mathbb{R}^n)$ such that $\mathrm{supt}(\eta) \subseteq AC^2([0,1]; \mathbb{R}^n)$ and

(5.15)
$$\int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma) d\eta(\gamma) \leq \int_0^1 |\mu_t'|^2 d\mathcal{L}^1(t) < +\infty.$$

Using Fubini–Tonelli's Theorem,

$$\int_0^1 \left(\int_{C([0,1];\mathbb{R}^n)} |\gamma'|^2(t) d\eta(\gamma) \right) d\mathcal{L}^1(t) < +\infty,$$

so for a.e. $t \in]0,1[$ it holds

$$\int_{C([0,1];\mathbb{R}^n)} |\gamma'|^2(t) d\eta(\gamma) < +\infty.$$

In particular, for a.e. $t \in]0,1[$ is well defined $\{\gamma \mapsto \gamma'(t)\} \in L^2(C([0,1];\mathbb{R}^n), \eta, \mathbb{R}^n)$, therefore, by Lemma (P.10), for a.e. $t \in]0,1[$ there exists $v_t \in L^2(\mathbb{R}^n, \mu_t; \mathbb{R}^n)$ such that $(e_t)_\#(\gamma'(t)\eta) = v_t \mu_t$ and

(5.16)
$$\|v_t\|_{L^2(\mathbb{R}^n, \mu_t; \mathbb{R}^n)} \leq \|\gamma'(t)\|_{L^2(C([0,1];\mathbb{R}^n), \eta; \mathbb{R}^n)}.$$

Given $\varphi \in C_c^\infty(\mathbb{R}^n)$, by differentiation under the integral sign Theorem,

$$\frac{d}{dt}\int \varphi d\mu_t = \frac{d}{dt}\int \varphi(\gamma(t))d\eta(\gamma) = \int_{C([0,1];\mathbb{R}^n)} D\varphi(\gamma(t))\cdot\gamma'(t)d\eta(\gamma) =$$
$$= \int D\varphi\cdot d(e_t)_\#(\gamma'(t)\eta) = \int D\varphi\cdot d(v_t\mu_t) = \int D\varphi\cdot v_t d\mu_t,$$

so, by Proposition (5.3), (μ_t, v_t) is a continuity pair.

Now, by (5.16), Fubini–Tonelli's Theorem and (5.15),

$$\int_0^1 \mathcal{A}(v_t,\mu_t)d\mathcal{L}^1(t) = \int_0^1\left(\int |v_t|^2 d\mu_t\right)d\mathcal{L}^1(t) \le \int_0^1\left(\int_{C([0,1];\mathbb{R}^n)} |\gamma'|^2(t)d\eta(\gamma)\right)d\mathcal{L}^1(t) =$$
$$= \int_{C([0,1];\mathbb{R}^n)}\left(\int_0^1 |\gamma'|^2(t)d\mathcal{L}^1(t)\right)d\eta(\gamma) = \int_{C([0,1];\mathbb{R}^n)} \mathcal{A}_2(\gamma)d\eta(\gamma) \le$$
$$\le \int_0^1 |\mu_t'|^2 d\mathcal{L}^1(t)$$

and, by Proposition (5.13), for a.e. $t \in]0,1[$

$$|\mu_t'|^2 \le \mathcal{A}(v_t,\mu_t). \tag{5.17}$$

Let us prove that for a.e. $t\in]0,1[$ it holds $\mathcal{A}(v_t,\mu_t) = |\mu_t'|^2$. By contradiction, suppose there exists $B \in \mathscr{B}(]0,1[)$ such that $\mathcal{L}^1(B) > 0$ and for every $t \in B$

$$\mathcal{A}(v_t,\mu_t) \ne |\mu_t'|^2,$$

that is, by (5.17), for a.e. $t \in B$

$$|\mu_t'|^2 < \mathcal{A}(v_t,\mu_t).$$

Observing that

$$\int_0^1 \mathcal{A}(v_t,\mu_t)d\mathcal{L}^1(t) = \int_B \mathcal{A}(v_t,\mu_t)d\mathcal{L}^1(t) + \int_{]0,1[\setminus B} \mathcal{A}(v_t,\mu_t)d\mathcal{L}^1(t) >$$
$$> \int_B |\mu_t'|^2 d\mathcal{L}^1(t) + \int_{]0,1[\setminus B}\mathcal{A}(v_t,\mu_t)d\mathcal{L}^1(t) =$$
$$= \int_0^1 |\mu_t'|^2 d\mathcal{L}^1(t)$$

we arrive at the desired contradiction, so the result follows. ■

Finally, thanks to the previous Theorem, we can exhibit this useful estimate.

(5.18) Corollary *Consider $\mu_t \in AC^2([0,1];\mathcal{P}_2(\mathbb{R}^n))$. For every $f \in \mathrm{Lip}_b(\mathbb{R}^n) \cap C^1(\mathbb{R}^n)$ it holds*

$$\left|\int f d\mu_1 - \int f d\mu_0\right| \leq \int_0^1 \left(\int |Df|^2 d\mu_t\right)^{\frac{1}{2}} |\mu_t'| d\mathcal{L}^1(t).$$

Proof. By Theorem (5.14), there exists a time-dependent vector field v_t such that (μ_t, v_t) is a continuity pair and for a.e. $t \in]0,1[$ it holds

$$\|v_t\|^2_{L^2(\mathbb{R}^n,\mu_t;\mathbb{R}^n)} = \int |v_t|^2 d\mu_t = \mathcal{A}(v_t,\mu_t) = |\mu_t'|^2,$$

then

$$\|v_t\|_{L^2(\mathbb{R}^n,\mu_t;\mathbb{R}^n)} = |\mu_t'|.$$

For every $h \in \mathbb{N}$ consider a cut-off function $\zeta \in C_c^\infty(\mathbb{R}^n)$ such that $0 \leq \zeta \leq 1$ in $\mathbb{R}^n$ and $\zeta = 1$ on $\overline{\mathrm{B}(0,h)}$. Take

$$f_h = \zeta \mathcal{R}_h f \in C_c^\infty(\mathbb{R}^n).$$

By Proposition (5.3),

$$\left|\int f_h d\mu_1 - \int f_h d\mu_0\right| \leq \left|\int_0^1 \frac{d}{dt}\left(\int f_h d\mu_t\right) d\mathcal{L}^1\right| = \left|\int_0^1 \left(\int Df_h \cdot v_t d\mu_t\right) d\mathcal{L}^1(t)\right|,$$

so, by Hölder's inequality,

$$\begin{aligned}\left|\int f_h d\mu_1 - \int f_h d\mu_0\right| &\leq \int_0^1 \left(\int |Df_h|^2 d\mu_t\right)^{\frac{1}{2}} \|v_t\|_{L^2(\mathbb{R}^n,\mu_t;\mathbb{R}^n)} d\mathcal{L}^1(t) = \\ &= \int_0^1 \left(\int |Df_h|^2 d\mu_t\right)^{\frac{1}{2}} |\mu_t'| d\mathcal{L}^1(t).\end{aligned}$$

The estimate follows by the dominated convergence Theorem. ■

3 Fokker–Planck equations

The simplest equations of (continuous) motion for a system of floating particles are the Fokker–Planck equation backward or forward, namely

$$-\frac{\partial \rho}{\partial t} + \mathrm{div}(v\rho) = c\Delta\rho, \qquad \frac{\partial \rho}{\partial t} + \mathrm{div}(v\rho) = c\Delta\rho,$$

where ρ is the probability distribution of the system and v its velocity field, both in Eulerian coordinates. They are mainly used to describe small Brownian systems, a current in an electrical circuit and the electric field in a laser. Further information can be found in [58]. As in the case of the continuity equation, we need a distributional form.

(5.19) Definition *Let* $T \in]0,+\infty]$, $c > 0$, $\mu_t :]0,T[\to \mathcal{M}(\mathbb{R}^n)$ *be weakly continuous and a Borel function* $v :]0,T[\times \mathbb{R}^n \to \mathbb{R}^n$ *such that*

$$(t,x) \longmapsto v_t(x),$$

where the function $\{x \longmapsto |v_t|\}$ *belongs to* $L^1(\mu_t)$ *and the function* $\left\{t \longmapsto \|v_t\|_{L^1(\mu_t)}\right\}$ *belongs to* $L^1_{loc}(]0,T[)$, *and* $\overline{\mu} \in \mathcal{M}(\mathbb{R}^n)$. *We say that* μ_t *is a* distributional solution *of the* backward Fokker–Planck problem, *namely*

$$\begin{cases} -\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = c\Delta \mu_t, \\ \mu_0 = \overline{\mu}, \end{cases}$$

if for every $\varphi \in C_c^\infty([0,T[\times \mathbb{R}^n)$ *it holds*

$$\int_0^T \int \left(-\frac{\partial \varphi}{\partial t}(t,x) + v_t(x) \cdot D\varphi(t,x) - c\Delta\varphi(t,x) \right) d\mu_t(x)dt + \int \varphi(0,x) d\overline{\mu} = 0.$$

In particular, we say that $\mu_t :]0,T[\to \mathcal{M}(\mathbb{R}^n)$ *weakly continuous is a* distributional solution *of the* backward Fokker–Planck equation, *namely*

$$-\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = c\Delta \mu_t,$$

if for every $\varphi \in C_c^\infty(]0,T[\times \mathbb{R}^n)$ *it holds*

$$\int_0^T \int \left(-\frac{\partial \varphi}{\partial t}(t,x) + v_t(x) \cdot D\varphi(t,x) - c\Delta\varphi(t,x) \right) d\mu_t(x)dt = 0.$$

The pair (μ_t, v_t), *where* μ_t *is a distributional solution of the backward Fokker–Planck equation defined by* v_t, *is called* backward Fokker–Planck pair.

(5.20) Definition *Let* $T \in]0,+\infty]$, $c > 0$, $\mu_t :]0,T[\to \mathcal{M}(\mathbb{R}^n)$ *be weakly continuous and a Borel function* $v :]0,T[\times \mathbb{R}^n \to \mathbb{R}^n$ *such that*

$$(t,x) \longmapsto v_t(x),$$

where the function $\{x \longmapsto |v_t|\}$ *belongs to* $L^1(\mu_t)$ *and the function* $\left\{t \longmapsto \|v_t\|_{L^1(\mu_t)}\right\}$

belongs to $L^1_{loc}(]0,T[)$*, and* $\overline{\mu} \in \mathcal{M}(\mathbb{R}^n)$*. We say that* μ_t *is a* distributional solution *of the* forward Fokker–Planck problem*, namely*

$$\begin{cases} \frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = c\Delta \mu_t, \\ \mu_0 = \overline{\mu}, \end{cases}$$

if for every $\varphi \in C_c^\infty([0,T[\times \mathbb{R}^n)$ *it holds*

$$\int_0^T \int \left(\frac{\partial \varphi}{\partial t}(t,x) + v_t(x) \cdot D\varphi(t,x) - c\Delta\varphi(t,x) \right) d\mu_t(x)dt + \int \varphi(0,x) d\overline{\mu} = 0.$$

In particular, we say that $\mu_t :]0,T[\to \mathcal{M}(\mathbb{R}^n)$ *weakly continuous is a* distributional solution *of the* forward Fokker–Planck equation*, namely*

$$-\frac{\partial \mu_t}{\partial t} + \operatorname{div}(v_t \mu_t) = c\Delta \mu_t,$$

if for every $\varphi \in C_c^\infty(]0,T[\times \mathbb{R}^n)$ *it holds*

$$\int_0^T \int \left(\frac{\partial \varphi}{\partial t}(t,x) + v_t(x) \cdot D\varphi(t,x) - c\Delta\varphi(t,x) \right) d\mu_t(x)dt = 0.$$

The pair (μ_t, v_t)*, where* μ_t *is a distributional solution of the forward Fokker–Planck equation defined by* v_t*, is called* forward Fokker–Planck pair*.*

Also in this case we can exhibit a useful equivalent formulation that speeds up the process of proving that a curve of measures is a solution of the Fokker–Planck equations.

(5.21) Proposition *Let* $c > 0$ *and* $\mu_t :]0,+\infty[\to \mathcal{M}(\mathbb{R}^n)$ *be weakly continuous and a Borel function* $v :]0,+\infty[\times \mathbb{R}^n \to \mathbb{R}^n$ *such that*

$$(t,x) \longmapsto v_t(x),$$

with the function $\{x \longmapsto |v_t|\}$ *belonging to* $L^1(\mu_t)$ *and the function* $\left\{t \longmapsto \|v_t\|_{L^1(\mu_t)}\right\}$ *belonging to* $L^1_{loc}(]0,+\infty[)$*. The following facts are equivalent:*

(a) μ_t *is a distributional solution of the backward Fokker–Planck equation,*

(b) for every $\varphi \in C_c^\infty(\mathbb{R}^n)$ *the function* $\left\{t \mapsto \int \varphi(x) d\mu_t(x)\right\} \in AC_{loc}(]0,+\infty[)$ *and its weak derivative is*

$$\frac{d}{dt} \int \varphi d\mu_t(x) = -\int \left(D\varphi(x) \cdot v_t(x) + c\Delta\varphi(x) \right) d\mu_t(x).$$

Proof. It is similar to the proof of Proposition (5.3). ∎

(5.22) Proposition *Let $c > 0$ and $\mu_t :]0, +\infty[\to \mathcal{M}(\mathbb{R}^n)$ be weakly continuous and a Borel function $v :]0, +\infty[\times \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$(t, x) \longmapsto v_t(x),$$

with the function $\{x \longmapsto |v_t|\}$ belonging to $L^1(\mu_t)$ and the function $\left\{t \longmapsto \|v_t\|_{L^1(\mu_t)}\right\}$ belonging to $L^1_{loc}(]0, +\infty[)$. The following facts are equivalent:

(a) *μ_t is a distributional solution of the forward Fokker–Planck equation,*

(b) *for every $\varphi \in C_c^\infty(\mathbb{R}^n)$ the function $\left\{t \mapsto \int \varphi(x) d\mu_t(x)\right\} \in AC_{loc}(]0, +\infty[)$ and its weak derivative is*

$$\frac{d}{dt} \int \varphi d\mu_t(x) = \int \left(D\varphi(x) \cdot v_t(x) + c\Delta\varphi(x)\right) d\mu_t(x).$$

Proof. It is similar to the proof of Proposition (5.3). ∎

We conclude the section by reporting the following Lemma, without proof, which will be useful later.

(5.23) Lemma *Let $c > 0$ and (μ_t, v_t) be a backward (resp. forward) Fokker–Planck pair and $f \in D(\Delta) \subseteq H^1(\mathbb{R}^n)$. Then the function $\left\{t \longmapsto \int f d\mu_t\right\}$ is absolutely continuous and for* a.e. $t \in [0, 1]$

$$\left|\frac{d}{dt} \int f d\mu_t\right| \leq \int \left(|Df||v_t| + c|\Delta f|\right) d\mu_t,$$

where the exceptional set can be chosen to be independent of f.

Moreover, if the function $\{t \longmapsto f_t\}$ belongs to

$$AC([0, 1], L^2(\mathbb{R}^n)) \cap L^\infty([0, 1], H^1(\mathbb{R}^n))$$

and the function $\{t \longmapsto \Delta f_t\}$ belongs to

$$L^\infty([0, 1], L^2(\mathbb{R}^n)),$$

then the function $\left\{t \longmapsto \int f_t d\mu_t\right\}$ *is absolutely continuous and for* a.e. $t \in [0,1]$

$$\frac{d}{ds}\Big(\int f_s d\mu_s\Big)|_{s=t} = \int \left(\frac{d}{ds} f_s\right)|_{s=t} d\mu_t + \frac{d}{ds}\Big(\int f_t d\mu_s\Big)|_{s=t}.$$

Proof. See [27]. ■

4 Schrödinger problem

We have already underlined in Example (1.30) that given $\varepsilon > 0$, considering $\mathsf{R}_{\frac{\varepsilon}{2}} = \mathsf{r}_{\frac{\varepsilon}{2}} \mathcal{L}^n \otimes \mathcal{L}^n$, where

$$\mathsf{r}_{\frac{\varepsilon}{2}}(x,y) = \frac{1}{\sqrt{(2\pi\varepsilon)^n}} e^{-\frac{|x-y|^2}{2\varepsilon}},$$

and given $\mu, \nu \in \mathcal{P}(X)$ such that $\mu, \nu \ll \mathcal{L}^n$, $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$ is well defined on $\Gamma(\mu,\nu)$. In this section, we would like to focus on this situation and show that the Schrödinger problem can also be reformulated, producing a Benamou–Brenier formula. In particular, we will consider only the case of $\mu_0 = \varrho_0 \mathcal{L}^n, \mu_1 = \varrho_1 \mathcal{L}^n \in \mathcal{P}(\mathbb{R}^n)$ with bounded densities and supports following, mainly, [27] in which can be found the generalization in the curved and possibly non-smooth setting. With these assumptions, it is not restrictive to assume that every continuity (resp. backward Fokker–Planck or forward Fokker–Planck) pair (μ_t, v_t) verifies

$$\exists\, C > 0, \forall\, t \in [0,1] : \mu_t \le C\mathcal{L}^n. \tag{5.24}$$

Once the framework of hypotheses has been defined, the first thing we can say is that we continue to have existence, uniqueness and the structural formula for the solution of the Schrödinger problem stated in Theorem (1.28).

(5.25) Corollary *For every* $\varepsilon > 0$*, there exists a unique minimizer* γ *of* $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$ *in* $\Gamma(\mu_0, \mu_1)$.

In particular, there exist $f^\varepsilon, g^\varepsilon \in L^1(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$*, unique a.e. in* $\mathbb{R}^n$ *up to a rescaling* $\left\{(f^\varepsilon, g^\varepsilon) \mapsto (cf^\varepsilon, \frac{g^\varepsilon}{c})\right\}$ *with some* $c > 0$*, with supports included in* $\mathrm{supt}(\mu_0)$ *and* $\mathrm{supt}(\mu_1)$ *respectively, such that*

$$\gamma = (f^\varepsilon \otimes g^\varepsilon)\mathsf{R}_{\frac{\varepsilon}{2}}.$$

Proof. It is a direct consequence of the Theorem (1.28). ■

(5.26) Notation For every $\varepsilon > 0$, for every $x \in \mathbb{R}^n$, let us denote

$$f_t^\varepsilon(x) = \begin{cases} f^\varepsilon(x) & \text{if } t = 0, \\ \displaystyle\int f^\varepsilon(y)\mathsf{r}_{\frac{t\varepsilon}{2}}(x,y)d\mathcal{L}^n(y) & \text{if } 0 < t \leq 1, \end{cases}$$

$$g_t^\varepsilon(x) = \begin{cases} \displaystyle\int g^\varepsilon(y)\mathsf{r}_{\frac{(1-t)\varepsilon}{2}}(x,y)d\mathcal{L}^n(y) & \text{if } 0 \leq t < 1, \\ g_t^\varepsilon(x) & \text{if } t = 1. \end{cases}$$

Furthermore, let us denote

$$\varrho_t^\varepsilon = \begin{cases} \varrho_0 & \text{if } t = 0, \\ f_t^\varepsilon g_t^\varepsilon & \text{if } 0 < t < 1, \\ \varrho_1 & \text{if } t = 1, \end{cases} \qquad \mu_t^\varepsilon = \begin{cases} \mu_0 & \text{if } t = 0, \\ \varrho_t^\varepsilon \mathcal{L}^n & \text{if } 0 < t < 1, \\ \mu_1 & \text{if } t = 1. \end{cases}$$

Finally, let us denote for every $t \in [0,1]$

$$\varphi_t^\varepsilon = \varepsilon \log f_t^\varepsilon, \qquad \psi_t^\varepsilon = \varepsilon \log g_t^\varepsilon$$

and $\vartheta_t^\varepsilon = \frac{1}{2}(\psi_t^\varepsilon - \varphi_t^\varepsilon)$.

We collect in the following Lemmas the main technical results necessary for the purposes of the section.

(5.27) Lemma *For every $\varepsilon > 0$, the following facts hold true:*

(a) *for every $t \in [0,1]$ the functions $f_t^\varepsilon, g_t^\varepsilon, \varrho_t^\varepsilon, \varphi_t^\varepsilon, \psi_t^\varepsilon$ and ϑ_t^ε are well defined,*

(b) *for every $t \in [0,1]$ $\mu_t^\varepsilon \in \mathcal{P}_2(\mathbb{R}^n)$,*

(c) *the function $\{t \longmapsto H(\mu_t^\varepsilon|\mathcal{L}^n)\}$ is continuous on $[0,1]$,*

(d) *for every $t \in [0,1]$ it holds $f_t^\varepsilon, g_t^\varepsilon \in D(\Delta) \subseteq H^1(\mathbb{R}^n)$,*

(e) *for every $t \in]0,1[$ it holds $\varrho_t^\varepsilon \in D(\Delta) \subseteq H^1(\mathbb{R}^n)$,*

(f) *for every $t \in [0,1]$ it holds $\varphi_t^\varepsilon, \psi_t^\varepsilon, \vartheta_t^\varepsilon \in D(\Delta_{loc}) \subseteq H^1(\mathbb{R}^n)$,*

(g) *the functions $\{t \longmapsto f_t^\varepsilon\}, \{t \longmapsto g_t^\varepsilon\}$ belong to*

$$C([0,1], L^2(\mathbb{R}^n)) \cap AC_{loc}([0,1], H^1(\mathbb{R}^n)) \cap L^\infty([0,1], L^\infty(\mathbb{R}^n))$$

and for a.e. $t \in [0,1]$

$$\frac{\partial f_t^\varepsilon}{\partial t} = \frac{\varepsilon}{2}\Delta f_t^\varepsilon, \qquad \frac{\partial g_t^\varepsilon}{\partial t} = -\frac{\varepsilon}{2}\Delta g_t^\varepsilon,$$

(h) *the function* $\{t \longmapsto \varrho_t^\varepsilon\}$ *belong to*

$$C([0,1], L^2(\mathbb{R}^n)) \cap AC([0,1], L^2(\mathbb{R}^n)) \cap AC_{loc}(]0,1[\,, H^1(\mathbb{R}^n)) \cap L^\infty([0,1], L^\infty(\mathbb{R}^n))$$

and for a.e. $t \in [0,1]$

$$\frac{\partial \varrho_t^\varepsilon}{\partial t} + \operatorname{div}(D\vartheta_t^\varepsilon \varrho_t^\varepsilon) = 0,$$

(i) *for every* $\overline{x} \in \mathbb{R}^n$*, for every compact* K *in* $[0,1]$*, there exists* $M > 0$*, depending only on* $\varrho_0, \varrho_1, K, \overline{x}$*, such that the functions* $\{t \longmapsto \varphi_t^\varepsilon\}$*,*$\{t \longmapsto \psi_t^\varepsilon\}$*,*$\{t \longmapsto \vartheta_t^\varepsilon\}$ *belong to*

$$AC(K, H^1(\mathbb{R}^n, e^{-V}\mathcal{L}^n)),$$

where $V(x) = M|x - \overline{x}|^2$*, and for* a.e. $t \in [0,1]$

$$\frac{\partial \varepsilon_t^\varepsilon}{\partial t} = \frac{1}{2}|D\varphi_t^\varepsilon|^2 + \frac{\varepsilon}{2}\Delta\varphi_t^\varepsilon, \qquad -\frac{\partial \psi_t^\varepsilon}{\partial t} = \frac{1}{2}|D\psi_t^\varepsilon|^2 + \frac{\varepsilon}{2}\Delta\psi_t^\varepsilon,$$

$$\frac{\partial \vartheta_t^\varepsilon}{\partial t} + \frac{1}{2}|D\vartheta_t^\varepsilon|^2 = -\frac{\varepsilon^2}{8}\left(2\Delta \log \varrho_t^\varepsilon + |D \log \varrho_t^\varepsilon|^2\right),$$

(j) *for every* $\delta \in \,]0,1[$*, for every* $\overline{x} \in \mathbb{R}^n$ *there exist* $C, C' > 0$*, depending only on* $\varrho_0, \varrho_1, \overline{x}$*, and* $C'' > 0$*, depending only on* $\varrho_0, \varrho_1, \overline{x}, \delta$*, such that*

$$\forall\, t \in [0,1], \text{ a.e. } x \in \mathbb{R}^n : \varrho_t^\varepsilon(x) \leq Ce^{-C'|x-\overline{x}|^2}$$

$$\forall\, t \in [\delta,1], \text{ a.e. in } \mathbb{R}^n : \operatorname{Lip}(\varphi_t^\varepsilon) + \operatorname{Lip}(\psi_{1-t}^\varepsilon) \leq C''(1 + |\cdot - \overline{x}|),$$

(k) *it holds*

$$\int\int_0^1 |D\varphi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^2 < +\infty, \quad \int\int_0^1 |D\psi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^2 < +\infty, \quad |D\vartheta_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^2 < +\infty.$$

Proof. See [28]. ■

(5.28) Lemma *For every* $\varepsilon, t \in \,]0,1[$ *and any* $p \in \mathbb{N}$*, let us denote with* h_t^ε *any of*

$\varphi_t^\varepsilon, \psi_t^\varepsilon, \vartheta_t^\varepsilon, \log\varrho_t^\varepsilon$ and with H_t^ε any of

$$\varrho_t^\varepsilon |Dh_t^\varepsilon|^p, \qquad \varrho_t^\varepsilon \log\varrho_t^\varepsilon |Dh_t^\varepsilon|^p, \qquad |D\varrho_t^\varepsilon||Dh_t^\varepsilon|^p, \qquad \Delta\varrho_t^\varepsilon |Dh_t^\varepsilon|^p, \qquad \varrho_t^\varepsilon Dh_t^\varepsilon \cdot D(\Delta h_t^\varepsilon).$$

Then, $H_t^\varepsilon \in L^1(\mathbb{R}^n)$. Moreover, for every $\delta \in \left]0, \frac{1}{2}\right[$ and every $\overline{x} \in \mathbb{R}^n$ we have

$$\lim_{R\to+\infty} \sup_{t\in[\delta,1-\delta]} \int_{X\setminus \mathrm{B}(\overline{x},R)} |H_t^\varepsilon| d\mathcal{L}^n = 0.$$

Finally, the function $\left\{t \longmapsto \int H_t^\varepsilon d\mathcal{L}^n\right\}$ defined on $]0,1[$ is continuous.

Proof. See [28]. ■

(5.29) Lemma *Consider $u \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ positive* a.e. *in $\mathbb{R}^n$, $\delta > 0$ and, for every $t \geq 0$,*

$$\mathfrak{C}_t^\delta(x) = \log\left(\int \mathsf{r}_t(x,y)u(y)d\mathcal{L}^n(y) + \delta\right).$$

The following facts hold true:

(a) there exists $C > 0$ such that for every $t \geq 0$

$$\|\mathfrak{C}_t^\delta\|_{L^\infty(\mathbb{R}^n)} \leq C,$$

(b) for all $x \in \mathbb{R}^n$ and $M > 0$, the function $\left\{t \longmapsto \mathfrak{C}_t^\delta\right\}$ belongs to

$$C([0,+\infty[\,, L^2(\mathbb{R}^n, e^{-V}\mathcal{L}^n)) \cap AC_{loc}(]0,+\infty[\,, L^2(\mathbb{R}^n, e^{-V}\mathcal{L}^n)),$$

where $V(y) = M|x-y|^2$, and for a.e. *$t > 0$*

$$\frac{\partial \mathfrak{C}_t^\delta}{\partial t} = |D\mathfrak{C}_t^\delta|^2 + \Delta\mathfrak{C}_t^\delta,$$

(c) the functions $\left\{t \longmapsto |D\mathfrak{C}_t^\delta|\right\}$, $\left\{t \longmapsto \Delta\mathfrak{C}_t^\delta\right\}$ belong to $L^\infty_{loc}(]0,+\infty[\,, L^2(\mathbb{R}^n))$,

(d) let μ_t in $\mathcal{P}(\mathbb{R}^n)$, where $t \geq 0$, be weakly continuous with $\mu_t \leq C\mathcal{L}^n$ for some $C > 0$, independent of t. Set $\eta_t = \frac{d\mu_t}{d\mathcal{L}^n}$ and denote with H_t^δ any of

$$\mathfrak{C}_t^\delta \eta_t, \qquad |\mathfrak{C}_t^\delta|^2\eta_t, \qquad |D\mathfrak{C}_t^\delta|\eta_t, \qquad |\mathfrak{C}_t^\delta|^2\eta_t.$$

Then, $H_t^\delta \in L^1(\mathbb{R}^n)$ for every $t, \delta > 0$ and, for any compact K in $]0,+\infty[$, the function $\left\{(t,x) \longmapsto H_t^\delta(x)\right\}$ belongs to $L^1(K \times \mathbb{R}^n, \mathcal{L}^1 \otimes \mathcal{L}^n)$.

Proof. See [27]. ■

Now that we have summarized the entire technical system, we can proceed to show some dynamical representation formulas. The whole proof essentially relies on the use of the Gauss–Green formula and on the explicit representation of the minimizer of the Schrödinger problem.

(5.30) Proposition *For every $\varepsilon > 0$, one has*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \frac{\varepsilon}{2}\left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n)\right) + \\ + \int\int_0^1 \left(\frac{1}{2}|D\vartheta_t^\varepsilon|^2 + \frac{\varepsilon^2}{8}|D\log\varrho_t^\varepsilon|^2\right)\varrho_t^\varepsilon dt d\mathcal{L}^n.$$

Proof. Consider $R > 0$ and a cut-off function $\zeta \in C_c^\infty(\mathrm{B}(0, R+1))$ such that $0 \le \zeta \le 1$ and $\zeta = 1$ on $\mathrm{B}(0, R)$. By Lemma (5.27), $\varrho_t^\varepsilon \in AC([0,1], L^2(\mathbb{R}^n))$ and for every compact K in $[0,1]$ there exists $M > 0$ such that $\vartheta_t^\varepsilon \in AC(K, H_0^1(\mathbb{R}^n, e^{-V}\mathcal{L}^n))$, with $V(x) = M|x|^2$. In particular, the function $\left\{t \longmapsto \int \zeta\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n\right\}$ belongs to $AC_{loc}(]0,1[)$ and for a.e. $t \in]0,1[$, using properties of the Bochner integral (see, for example, [26]) and Lemma (5.27),

$$\frac{d}{dt}\int \zeta\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n = \int \zeta\frac{d}{dt}\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n + \int \zeta\vartheta_t^\varepsilon\frac{d}{dt}\varrho_t^\varepsilon d\mathcal{L}^n$$

so, by Lemma (5.27),

$$\begin{aligned}\frac{d}{dt}\int \zeta\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n &= \int \zeta\left(-\frac{1}{2}|D\vartheta_t^\varepsilon|^2 - \frac{\varepsilon^2}{8}\left(2\Delta\log\varrho_t^\varepsilon + |D\log\varrho_t^\varepsilon|^2\right)\right)\varrho_t^\varepsilon d\mathcal{L}^n + \\ &\quad - \int \zeta\vartheta_t^\varepsilon \mathrm{div}(D\vartheta_t^\varepsilon\varrho_t^\varepsilon)d\mathcal{L}^n = \\ &= -\frac{1}{2}\int \zeta|D\vartheta_t^\varepsilon|^2\varrho_t^\varepsilon d\mathcal{L}^n - \frac{\varepsilon^2}{4}\int \zeta\Delta\log\varrho_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n + \\ &\quad - \frac{\varepsilon^2}{8}\int \zeta|D\log\varrho_t^\varepsilon|^2\varrho_t^\varepsilon d\mathcal{L}^n - \int \zeta\vartheta_t^\varepsilon\mathrm{div}(D\vartheta_t^\varepsilon\varrho_t^\varepsilon)d\mathcal{L}^n.\end{aligned}$$

Using the Gauss–Green formula, and the fact that $\mathbb{R}^n$ has no boundary,

$$\begin{aligned}\int \zeta\Delta\log\varrho_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n &= -\int D(\zeta\varrho_t^\varepsilon)\cdot D\log\varrho_t^\varepsilon d\mathcal{L}^n = -\int D(\zeta\varrho_t^\varepsilon)\cdot\frac{1}{\varrho_t^\varepsilon}D\varrho_t^\varepsilon d\mathcal{L}^n = \\ &= -\int D\zeta\cdot D\varrho_t^\varepsilon d\mathcal{L}^n - \int \zeta\frac{1}{\varrho_t^\varepsilon}D\varrho_t^\varepsilon\cdot\frac{1}{\varrho_t^\varepsilon}D\varrho_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n = \\ &= -\int D\zeta\cdot D\varrho_t^\varepsilon d\mathcal{L}^n - \int \zeta|D\log\varrho_t^\varepsilon|^2\varrho_t^\varepsilon d\mathcal{L}^n\end{aligned}$$

and

$$\int \zeta\vartheta_t^\varepsilon \operatorname{div}(D\vartheta_t^\varepsilon\varrho_t^\varepsilon)d\mathcal{L}^n = -\int D(\zeta\vartheta_t^\varepsilon)\cdot D\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n = \\ = -\int \zeta|D\vartheta_t^\varepsilon|^2\varrho_t^\varepsilon d\mathcal{L}^n - \int \vartheta_t^\varepsilon D\zeta\cdot D\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n,$$

so

$$\frac{d}{dt}\int \zeta\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n = \frac{1}{2}\int \zeta|D\vartheta_t^\varepsilon|^2\varrho_t^\varepsilon d\mathcal{L}^n + \frac{\varepsilon^2}{8}\int \zeta|D\log\varrho_t^\varepsilon|^2\varrho_t^\varepsilon d\mathcal{L}^n + \\ + \int D\zeta\cdot D\varrho_t^\varepsilon d\mathcal{L}^n + \int \vartheta_t^\varepsilon D\zeta\cdot D\vartheta_t^\varepsilon\varrho_t^\varepsilon d\mathcal{L}^n.$$

Take $\delta \in \left]0,\frac{1}{2}\right[$ and integrate both sides with respect to t between δ and $1-\delta$ to obtain, again using the properties of the Bochner integral,

$$\int \zeta\vartheta_{1-\delta}^\varepsilon\varrho_{1-\delta}^\varepsilon d\mathcal{L}^n - \int \zeta\vartheta_\delta^\varepsilon\varrho_\delta^\varepsilon d\mathcal{L}^n = \frac{1}{2}\int\int_\delta^{1-\delta} \zeta|D\vartheta_t^\varepsilon|^2\varrho_t^\varepsilon dtd\mathcal{L}^n + \\ + \frac{\varepsilon^2}{8}\int\int_\delta^{1-\delta} \zeta|D\log\varrho_t^\varepsilon|^2\varrho_t^\varepsilon dtd\mathcal{L}^n + \\ + \int\int_\delta^{1-\delta} D\zeta\cdot D\varrho_t^\varepsilon dtd\mathcal{L}^n + \\ + \int\int_\delta^{1-\delta} \vartheta_t^\varepsilon D\zeta\cdot D\vartheta_t^\varepsilon\varrho_t^\varepsilon dtd\mathcal{L}^n.$$

Now, we want to pass to the limit as $R\to\infty$. As regards the terms

$$\frac{1}{2}\int\int_\delta^{1-\delta} \zeta|D\vartheta_t^\varepsilon|^2\varrho_t^\varepsilon dtd\mathcal{L}^n \qquad \frac{\varepsilon^2}{8}\int\int_\delta^{1-\delta} \zeta|D\log\varrho_t^\varepsilon|^2\varrho_t^\varepsilon dtd\mathcal{L}^n,$$

using the monotone convergence Theorem, we can pass the limit under the integral sign. For the other terms, we can apply the dominated convergence Theorem: indeed, Lemma (5.28) ensures the validity of the a.e. convergence and the $L^1(\mathbb{R}^n)$ bounds. We then get

$$\int \vartheta_{1-\delta}^\varepsilon\varrho_{1-\delta}^\varepsilon d\mathcal{L}^n - \int \vartheta_\delta^\varepsilon\varrho_\delta^\varepsilon d\mathcal{L}^n = \int\int_\delta^{1-\delta}\left(\frac{1}{2}|D\vartheta_t^\varepsilon|^2 + \frac{\varepsilon^2}{8}|D\log\varrho_t^\varepsilon|^2\right)\varrho_t^\varepsilon dtd\mathcal{L}^n. \tag{5.31}$$

Now, by the fact that $\delta\in\left]0,\frac{1}{2}\right[\subseteq\,]0,1[$,

$$\varepsilon\log\varrho_\delta^\varepsilon = \varepsilon\log f_\delta^\varepsilon g_\delta^\varepsilon = \varepsilon\log f_\delta^\varepsilon + \varepsilon\log g_\delta^\varepsilon = \varphi_\delta^\varepsilon + \psi_\delta^\varepsilon,$$

so $\varphi_\delta^\varepsilon = \varepsilon\log\varrho_\delta^\varepsilon - \psi_\delta^\varepsilon$, then $\vartheta_\delta^\varepsilon = \psi_\delta^\varepsilon - \frac{\varepsilon}{2}\log\varrho_\delta^\varepsilon$. Analogously, we can prove the identity

$\vartheta^\varepsilon_{1-\delta} = -\varphi^\varepsilon_{1-\delta} + \frac{\varepsilon}{2}\log\varrho^\varepsilon_{1-\delta}$, so we can rewrite (5.31) as

$$\frac{\varepsilon}{2}\left(\int \log\varrho^\varepsilon_\delta\varrho^\varepsilon_\delta d\mathcal{L}^n + \int \log\varrho^\varepsilon_{1-\delta}\varrho^\varepsilon_{1-\delta} d\mathcal{L}^n\right) + \\ -\left(\int \psi^\varepsilon_\delta\varrho^\varepsilon_\delta d\mathcal{L}^n + \int \varphi^\varepsilon_{1-\delta}\varrho^\varepsilon_{1-\delta} d\mathcal{L}^n\right) = \int\int_\delta^{1-\delta}\left(\frac{1}{2}|D\vartheta^\varepsilon_t|^2 + \frac{\varepsilon^2}{8}|D\log\varrho^\varepsilon_t|^2\right)\varrho^\varepsilon_t dt d\mathcal{L}^n.$$

Passing to the limit as $\delta\to 0$, in the right-hand side we can use the monotone convergence Theorem and in the left-hand side the dominated convergence Theorem combined with Lemma (5.27) and Lemma (5.28), we obtain

$$\frac{\varepsilon}{2}\left(\int \log\varrho_0\varrho_0 d\mathcal{L}^n + \int \log\varrho_1\varrho_1 d\mathcal{L}^n\right) + \\ -\left(\int \psi^\varepsilon_0\varrho_0 d\mathcal{L}^n + \int \varphi^\varepsilon_1\varrho_1 d\mathcal{L}^n\right) = \int\int_0^1\left(\frac{1}{2}|D\vartheta^\varepsilon_t|^2 + \frac{\varepsilon^2}{8}|D\log\varrho^\varepsilon_t|^2\right)\varrho^\varepsilon_t dt d\mathcal{L}^n,$$

that can be rewritten, using the facts that $\varphi^\varepsilon_0 + \psi^\varepsilon_0 = \varepsilon\log\varrho_0$ and $\varphi^\varepsilon_1 + \psi^\varepsilon_1 = \varepsilon\log\varrho_1$, as

$$\int \varphi^\varepsilon_0\varrho_0 d\mathcal{L}^n + \int \psi^\varepsilon_1\varrho_1 d\mathcal{L}^n + \\ -\frac{\varepsilon}{2}\left(\int \log\varrho_0\varrho_0 d\mathcal{L}^n + \int \log\varrho_1\varrho_1 d\mathcal{L}^n\right) = \int\int_0^1\left(\frac{1}{2}|D\vartheta^\varepsilon_t|^2 + \frac{\varepsilon^2}{8}|D\log\varrho^\varepsilon_t|^2\right)\varrho^\varepsilon_t dt d\mathcal{L}^n,$$

or, in other words,

$$\int \varphi^\varepsilon_0\varrho_0 d\mathcal{L}^n + \int \psi^\varepsilon_1\varrho_1 d\mathcal{L}^n = \frac{\varepsilon}{2}\left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n)\right) + \\ + \int\int_0^1\left(\frac{1}{2}|D\vartheta^\varepsilon_t|^2 + \frac{\varepsilon^2}{8}|D\log\varrho^\varepsilon_t|^2\right)\varrho^\varepsilon_t dt d\mathcal{L}^n.$$

By the explicit representation formula of the minimizer of $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$,

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = H_{\mathsf{R}_{\frac{\varepsilon}{2}}}(f^\varepsilon\otimes g^\varepsilon\mathsf{R}_{\frac{\varepsilon}{2}}) = \int \varphi^\varepsilon_0\varrho_0 d\mathcal{L}^n + \int \psi^\varepsilon_1\varrho_1 d\mathcal{L}^n,$$

the result follows. ■

(5.32) Proposition *For every $\varepsilon > 0$, one has*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \frac{1}{2}\int\int_0^1 |D\psi^\varepsilon_t|^2\varrho^\varepsilon_t dt d\mathcal{L}^n.$$

Proof. Let us observe that, using the very definition of ϑ_t^ε in the continuity equation solved by ϱ_t^ε,

$$\frac{\partial \varrho_t^\varepsilon}{\partial t} + \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\psi_t^\varepsilon) = \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\varphi_t^\varepsilon).$$

Now, using the fact that $\varphi_t^\varepsilon + \psi_t^\varepsilon = \varepsilon \log \varrho_t^\varepsilon$,

$$\frac{\partial \varrho_t^\varepsilon}{\partial t} + \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\psi_t^\varepsilon) = \frac{\varepsilon}{2}\mathrm{div}(\varrho_t^\varepsilon D \log \varrho_t^\varepsilon) - \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\psi_t^\varepsilon),$$

then we obtain the forward Fokker–Planck equation

$$\frac{\partial \varrho_t^\varepsilon}{\partial t} + \mathrm{div}(\varrho_t^\varepsilon D\psi_t^\varepsilon) = \frac{\varepsilon}{2}\Delta \varrho_t^\varepsilon.$$

Arguing in the same way as for Proposition (5.30), we can then find

$$\int \varphi_1^\varepsilon d\mu_1 - \int \varphi_1^\varepsilon d\mu_0 = -\frac{1}{2}\int\int_0^1 |D\varphi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^n,$$

that can be rewritten, using the identity $\varphi_0^\varepsilon + \psi_0^\varepsilon = \varrho \log \varrho_0$ and the explicit representation formula of the minimizer of $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$, as

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \frac{1}{2}\int\int_0^1 |D\psi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^n. \blacksquare$$

(5.33) Proposition *For every $\varepsilon > 0$, one has*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_1 \mid \mathcal{L}^n) + \frac{1}{2}\int\int_0^1 |D\varphi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^n.$$

Proof. Let us observe that, using the very definition of ϑ_t^ε in the continuity equation solved by ϱ_t^ε,

$$\frac{\partial \varrho_t^\varepsilon}{\partial t} - \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\varphi_t^\varepsilon) = -\frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\psi_t^\varepsilon).$$

Now, using the fact that $\varphi_t^\varepsilon + \psi_t^\varepsilon = \varepsilon \log \varrho_t^\varepsilon$,

$$\frac{\partial \varrho_t^\varepsilon}{\partial t} - \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\psi_t^\varepsilon) = -\frac{\varepsilon}{2}\mathrm{div}(\varrho_t^\varepsilon D \log \varrho_t^\varepsilon) + \frac{1}{2}\mathrm{div}(\varrho_t^\varepsilon D\varphi_t^\varepsilon),$$

then we obtain the forward Fokker–Planck equation

$$-\frac{\partial}{\partial t}\varrho_t^\varepsilon + \mathrm{div}(\varrho_t^\varepsilon D\varphi_t^\varepsilon) = \frac{\varepsilon}{2}\Delta \varrho_t^\varepsilon.$$

Arguing in the same way as for Proposition (5.30), we can then find

$$\int \varphi_1^\varepsilon d\mu_1 - \int \varphi_1^\varepsilon d\mu_0 = -\frac{1}{2}\int\int_0^1 |D\varphi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^n$$

that can be rewritten, using the identity $\varphi_0^\varepsilon + \psi_0^\varepsilon = \varrho \log \varrho_0$ and the explicit representation formula of the minimizer of $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$, as

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_1 \mid \mathcal{L}^n) + \frac{1}{2}\int\int_0^1 |D\varphi_t^\varepsilon|^2 \varrho_t^\varepsilon dt d\mathcal{L}^n. \blacksquare$$

We are now ready for the main result of the section: the Benamou–Brenier formula for the Schrödinger problem. By the fact that we can dynamically represent the problem in three different ways using the continuity equation or the two Fokker–Planck equations, we obtain also three different Benamou–Brenier formulas. As regards the first form, inequality $\geq$ follows directly from Proposition (5.30) and the fact that ϱ_t^ε solves the continuity equation with velocity $D\vartheta_t^\varepsilon$. The converse inequality can instead be proved using a technique similar to the one seen for Proposition (5.30).

(5.34) Theorem (Benamou–Brenier formula, I form) *For every $\varepsilon > 0$, one has*

$$\begin{aligned}\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} &= \frac{\varepsilon}{2}\left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n)\right) + \\ &+ \min_{\substack{(\eta_t\mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0=\varrho_0, \eta_1=\varrho_1}} \left\{\int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n\right\}.\end{aligned}$$

In particular, on the right-hand side, the (unique) minimizer is $(\varrho_t^\varepsilon, D\vartheta_t^\varepsilon)$.

Proof. Let us start proving that

$$\begin{aligned}\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} &\geq \frac{\varepsilon}{2}\left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n)\right) + \\ &+ \inf_{\substack{(\eta_t\mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0=\varrho_0, \eta_1=\varrho_1}} \left\{\int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n\right\}.\end{aligned}$$

By Proposition (5.30), we know that

$$\begin{aligned}\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} &= \frac{\varepsilon}{2}\left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n)\right) + \\ &+ \int\int_0^1 \left(\frac{1}{2}|D\vartheta_t^\varepsilon|^2 + \frac{\varepsilon^2}{8}|D\log\varrho_t^\varepsilon|^2\right)\varrho_t^\varepsilon dt d\mathcal{L}^n,\end{aligned}$$

but, using Lemma (5.27), we can also say that $(\varrho_t^\varepsilon \mathcal{L}^n, D\vartheta_t^\varepsilon)$ is a continuity pair. In particular, we can pass to the infimum on the right-hand side, obtaining the desired inequality.

For the converse inequality, namely

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} \leq \frac{\varepsilon}{2}\left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n)\right) + \\ + \inf_{\substack{(\eta_t\mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0=\varrho_0, \eta_1=\varrho_1}} \left\{ \int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n \right\},$$

we can, first of all, observe that the assumptions on μ_0,μ_1 guarantee the finiteness of $\min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$. In particular, if

$$\inf_{\substack{(\eta_t\mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0=\varrho_0, \eta_1=\varrho_1}} \left\{ \int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n \right\} = +\infty,$$

the inequality is trivially satisfied. Let us, then, focus on the case

$$\inf_{\substack{(\eta_t\mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0=\varrho_0, \eta_1=\varrho_1}} \left\{ \int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n \right\} < +\infty.$$

We can then restrict to the case where $(\eta_t\mathcal{L}^n, v_t)$ is a continuity pair with $\eta_0 = \varrho_0, \eta_1 = \varrho_1$ and

$$\int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n < +\infty. \tag{5.35}$$

Consider $R > 0$ and a cut-off function $\zeta \in C_c^\infty(\mathrm{B}(0, R+1))$ such that $0 \leq \zeta \leq 1$ and $\zeta = 1$ on $\mathrm{B}(0, R)$. Given $\delta > 0$, define for every $t \in [0, 1]$

$$\varphi_t^{\varepsilon,\delta} = \varepsilon\log(f_t^\varepsilon + \delta), \qquad \psi_t^{\varepsilon,\delta} = \varepsilon\log(g_t^\varepsilon + \delta), \qquad \vartheta_t^{\varepsilon,\delta} = \frac{1}{2}\left(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta}\right).$$

By Lemma (5.29), $\zeta\vartheta_t^{\varepsilon,\delta} \in AC_{loc}(]0,1[\,, L^2(\mathbb{R}^n)) \cap L^\infty_{loc}(]0,1[\,, H^1(\mathbb{R}^n))$, so, given $t_0, t_1 \in]0,1[$, with $t_0 < t_1$, Lemma (5.5), applied to $(\eta_t\mathcal{L}^n, v_t)$ and the function $\left\{t \longmapsto \zeta\vartheta_t^{\varepsilon,\delta}\right\}$ on $[t_0, t_1]$, gave us for a.e. $t \in [t_0, t_1]$

$$\frac{d}{ds}\Big(\int \zeta\vartheta_s^{\varepsilon,\delta}\eta_s d\mathcal{L}^n\Big)|_{s=t} = \int \zeta\frac{d}{dt}\vartheta_t^{\varepsilon,\delta}\eta_t d\mathcal{L}^n + \frac{d}{ds}\left(\int \zeta\vartheta_t^{\varepsilon,\delta}\eta_s d\mathcal{L}^n\right)|_{s=t}.$$

As regards the first term in the right-hand side, by the very definition of $\vartheta_t^{\varepsilon,\delta}$ and a linear transformation in Lemma (5.27),

$$\begin{aligned}\int \zeta \frac{d}{dt}\vartheta_t^{\varepsilon,\delta}\eta_t d\mathcal{L}^n &= \frac{1}{2}\int \zeta \frac{d}{dt}\psi_t^{\varepsilon,\delta}\eta_t d\mathcal{L}^n - \frac{1}{2}\int \zeta \frac{d}{dt}\varphi_s^{\varepsilon,\delta}\eta_t d\mathcal{L}^n = \\ &= -\frac{1}{2}\int \zeta \left(\frac{1}{2}|D\psi_t^{\varepsilon,\delta}|^2 + \frac{\varepsilon}{2}\Delta\psi_t^{\varepsilon,\delta}\right)\eta_t d\mathcal{L}^n + \\ &\quad -\frac{1}{2}\int \zeta \left(\frac{1}{2}|D\varphi_t^{\varepsilon,\delta}|^2 + \frac{\varepsilon}{2}\Delta\varphi_t^{\varepsilon,\delta}\right)\eta_t d\mathcal{L}^n,\end{aligned}$$

and, using the Gauss–Green formula, with the fact that $\mathbb{R}^n$ has no boundary, combined with Young's inequality and the triangle inequality,

$$\begin{aligned}\int \zeta \left(\frac{d}{ds}\vartheta_s^{\varepsilon,\delta}\right)|_{s=t}\eta_t d\mathcal{L}^n &= -\frac{1}{4}\int \zeta\left(|D\psi_t^{\varepsilon,\delta}|^2 + |D\varphi_t^{\varepsilon,\delta}|^2\right)\eta_t d\mathcal{L}^n + \\ &\quad + \frac{\varepsilon}{4}\int \zeta D(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})\cdot D\log\eta_t \eta_t d\mathcal{L}^n + \\ &\quad + \frac{\varepsilon}{4}\int D(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})\cdot D\zeta\eta_t d\mathcal{L}^n \leq \\ &\leq -\frac{1}{4}\int \zeta\left(|D\psi_t^{\varepsilon,\delta}|^2 + |D\varphi_t^{\varepsilon,\delta}|^2\right)\eta_t d\mathcal{L}^n + \\ &\quad + \frac{1}{8}\int \zeta|D(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})|^2\eta_t d\mathcal{L}^n + \frac{\varepsilon^2}{8}\int \zeta|D\log\eta_t|^2\eta_t d\mathcal{L}^n + \\ &\quad + \frac{\varepsilon}{4}\int D(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})\cdot D\zeta\eta_t d\mathcal{L}^n \leq \\ &\leq -\frac{1}{8}\int \zeta\left(|D\psi_t^{\varepsilon,\delta}|^2 + |D\varphi_t^{\varepsilon,\delta}|^2\right)\eta_t d\mathcal{L}^n + \\ &\quad + \frac{\varepsilon^2}{8}\int \zeta|D\log\eta_t|^2\eta_t d\mathcal{L}^n + \\ &\quad + \frac{\varepsilon}{4}\int D(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})\cdot D\zeta\eta_t d\mathcal{L}^n.\end{aligned}$$

For the second term, instead, using Proposition (5.3), combined with the fact that $\zeta\vartheta_t^{\varepsilon,\delta} \in H^1(\mathbb{R}^n)$, the density of $C_c^\infty(\mathrm{B}(0,R+1))$ in $H^1(\mathrm{B}(0,R+1))$ (see [24], for example, for technical details), weighted Young's inequality and the triangle inequality,

$$\begin{aligned}\frac{d}{ds}\left(\int \zeta\vartheta_t^{\varepsilon,\delta}\eta_s d\mathcal{L}^n\right)|_{s=t} &= \frac{1}{2}\int\left(\zeta D(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})\cdot v_t + (\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})D\zeta\cdot v_t\right)\eta_t d\mathcal{L}^n \leq \\ &\leq \frac{1}{8}\int \zeta|D(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})|^2\eta_t d\mathcal{L}^n + \frac{1}{2}\int \zeta|v_t|^2\eta_t d\mathcal{L}^n + \\ &\quad + \frac{1}{2}\int \zeta(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})D\zeta\cdot v_t\eta_t d\mathcal{L}^n \leq \\ &\leq \frac{1}{8}\int \zeta\left(|D\psi_t^{\varepsilon,\delta}|^2 + |D\varphi_t^{\varepsilon,\delta}|^2\right)\eta_t d\mathcal{L}^n + \frac{1}{2}\int \zeta|v_t|^2\eta_t d\mathcal{L}^n +\end{aligned}$$

$$+\frac{1}{2}\int \zeta(\psi_t^{\varepsilon,\delta}-\varphi_t^{\varepsilon,\delta})D\zeta\cdot v_t\eta_t d\mathcal{L}^n.$$

Resuming,

$$\begin{aligned}\frac{d}{ds}\Big(\int \zeta\vartheta_s^{\varepsilon,\delta}\eta_s d\mathcal{L}^n\Big)|_{s=t} \leq &\int \zeta\left(\frac{1}{2}|v_t|^2+\frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t d\mathcal{L}^n+\\ &+\frac{\varepsilon}{4}\int D(\psi_t^{\varepsilon,\delta}+\varphi_t^{\varepsilon,\delta})\cdot D\zeta\eta_t d\mathcal{L}^n+\\ &+\frac{1}{2}\int \zeta(\psi_t^{\varepsilon,\delta}-\varphi_t^{\varepsilon,\delta})D\zeta\cdot v_t\eta_t d\mathcal{L}^n.\end{aligned}$$

Integrating both sides over $[t_0,t_1]$, we get

$$\begin{aligned}\int \zeta\vartheta_{t_1}^{\varepsilon,\delta}\eta_{t_1}d\mathcal{L}^n-\int \zeta\vartheta_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n \leq &\int\int_{t_0}^{t_1} \zeta\left(\frac{1}{2}|v_t|^2+\frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dtd\mathcal{L}^n+\\ &+\frac{\varepsilon}{4}\int\int_{t_0}^{t_1} D(\psi_t^{\varepsilon,\delta}+\varphi_t^{\varepsilon,\delta})\cdot D\zeta\eta_t dtd\mathcal{L}^n+\\ &+\frac{1}{2}\int\int_{t_0}^{t_1} \zeta(\psi_t^{\varepsilon,\delta}-\varphi_t^{\varepsilon,\delta})D\zeta\cdot v_t\eta_t dtd\mathcal{L}^n.\end{aligned}$$

Now, we want to pass to the limit as $R\to+\infty$. As regards the first term in the right-hand sides, we can use, by monotonicity, the monotone convergence Theorem and for the others we have to combine Lemma (5.29), (5.35) and the dominated convergence Theorem. We get

$$\int \vartheta_{t_1}^{\varepsilon,\delta}\eta_{t_1}d\mathcal{L}^n-\int \vartheta_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n \leq \int\int_{t_0}^{t_1}\left(\frac{1}{2}|v_t|^2+\frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dtd\mathcal{L}^n.$$

Let us now pass to the limit as $t_0\to 0^+$. As regards the right-hand side, we can use, by monotonicity, the monotone convergence Theorem. For the second term in the left-hand side, we can, firstly, observe that

$$\int \vartheta_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n=\frac{1}{2}\left(\int \psi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int \varphi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n\right).$$

Let us show that

$$\lim_{t_0\to 0^+}\int \psi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n=\int \psi_0^{\varepsilon,\delta}d\mu_0.$$

Using the triangle inequality,

$$\left|\int \psi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int \psi_0^{\varepsilon,\delta}d\mu_0\right| \leq \int\left|\psi_{t_0}^{\varepsilon,\delta}\eta_{t_0}-\psi_0^{\varepsilon,\delta}\right|\eta_{t_0}d\mathcal{L}^n+\left|\int \psi_0^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int \psi_0^{\varepsilon,\delta}d\mu_0\right|.$$

By the fact that

$$\sup_{x\in]\delta,+\infty[}\left|\frac{1}{x}\right|\leq\frac{1}{\delta},$$

we have that $\log\in\mathrm{Lip}(]\delta,+\infty[)$, with Lipschitz constant $\mathrm{Lip}(\log|_{]\delta,+\infty[})\leq\frac{1}{\delta}$. By (5.24), $\eta_t\leq C$ for every $t\in[0,1]$ so, combined with the fact that $\eta_{t_0}\mathcal{L}^n\in\mathcal{P}(\mathbb{R}^n)$ and Hölder's inequality,

$$\begin{aligned}\left|\int\psi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\psi_0^{\varepsilon,\delta}d\mu_0\right|&\leq\varepsilon\int\left|\log(g_{t_0}^\varepsilon+\delta)-\log(g_0^\varepsilon+\delta)\right|\eta_{t_0}d\mathcal{L}^n+\\&+\left|\int\psi_0^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\psi_0^{\varepsilon,\delta}d\mu_0\right|\leq\\&\leq\frac{\varepsilon}{\delta}\int|g_{t_0}^\varepsilon-g_0^\varepsilon|\eta_{t_0}d\mathcal{L}^n+\left|\int\psi_0^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\psi_0^{\varepsilon,\delta}d\mu_0\right|\leq\\&\leq|g_{t_0}^\varepsilon-g_0^\varepsilon|^2\eta_{t_0}d\mathcal{L}^n+\\&+\left|\int\psi_0^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\psi_0^{\varepsilon,\delta}d\mu_0\right|\leq\\&\leq\frac{\varepsilon}{\delta}C\|g_{t_0}^\varepsilon-g_0^\varepsilon\|_{L^2(\mathbb{R}^n)}+\left|\int\psi_0^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\psi_0^{\varepsilon,\delta}d\mu_0\right|\end{aligned}$$

so, using the fact that $\{t\longmapsto g_t^\varepsilon\}$ is continuous with respect to L^2 norm, the first term in the right-hand side disappears in the limit as $t_0\to0^+$. As regards the second, using the maximum principle and the regularity of the solution of the heat equation (see, for example, [24]), by the fact that $g^\varepsilon\in L^\infty(\mathbb{R}^n)$ and it has compact support, we can state that $g_0^\varepsilon\in C_b(\mathbb{R}^n)$. In particular, $\psi_0^{\varepsilon,\delta}\in C_b(\mathbb{R}^n)$, so, by weak continuity of $\eta_t\mathcal{L}^n$, also this term vanishes. Let us also show that

$$\lim_{t_0\to0^+}\int\varphi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n=\int\varphi_0^{\varepsilon,\delta}d\mu_0.$$

Using again the triangle inequality,

$$\left|\int\varphi_{t_0}^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\varphi_0^{\varepsilon,\delta}d\mu_0\right|\leq\int\left|\varphi_{t_0}^{\varepsilon,\delta}\eta_{t_0}-\varphi_0^{\varepsilon,\delta}\right|\eta_{t_0}d\mathcal{L}^n+\left|\int\varphi_0^{\varepsilon,\delta}\eta_{t_0}d\mathcal{L}^n-\int\varphi_0^{\varepsilon,\delta}d\mu_0\right|.$$

For the first term in the right-hand side, we can argue analogously to the previous estimate. As regards the second, by the fact that f^ε has compact support, $\varphi_0^{\varepsilon,\delta}$ is constant outside a bounded set: in particular, for every $\alpha>0$, there exists $q\in C_b(\mathbb{R}^n)$ such that

$$\|\varphi_0^{\varepsilon,\delta}-q\|_{L^1(\mathbb{R}^n)}<\alpha.$$

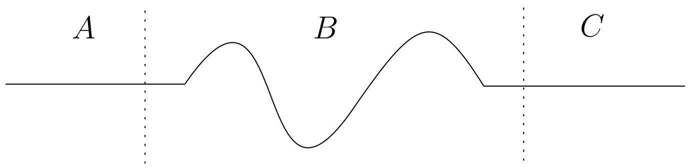

Figure 5.1: A representation of the density argument in dimension 1. We only have to consider a piecewise function q defined as follows: unless we take B large enough, there we can use the density of $C_c^\infty(B)$ in $L^1(B)$; in zones A and C we can take the (constant) value that $\varphi_0^{\varepsilon,\delta}$ have, so $\varepsilon \log \delta$.

These facts, combined with $\eta_t \leq C$ for all $t \in [0,1]$, that comes again from (5.24), provide

$$\begin{aligned}\left|\int \varphi_{t_0}^{\varepsilon,\delta}\eta_{t_0} d\mathcal{L}^n - \int \varphi_0^{\varepsilon,\delta} d\mu_0\right| &\leq \frac{\varepsilon}{\delta} C \|f_{t_0}^\varepsilon - f_0^\varepsilon\|_{L^2(\mathbb{R}^n)} + \\ &+ \int |\varphi_0^{\varepsilon,\delta} - q|\eta_{t_0} d\mathcal{L}^n + \left|\int q\eta_{t_0} d\mathcal{L}^n - \int q d\mu_0\right| + \\ &+ \int |\varphi_0^{\varepsilon,\delta} - q| d\mu_0 \leq \\ &\leq \frac{\varepsilon}{\delta} C \|f_{t_0}^\varepsilon - f_0^\varepsilon\|_{L^2(\mathbb{R}^n)} + 2C\alpha + \left|\int q\eta_{t_0} d\mathcal{L}^n - \int q d\mu_0\right|.\end{aligned}$$

Using the continuity of $\{t \longmapsto f_t^\varepsilon\}$ with respect to L^2 norm, weak continuity of $\eta_t \mathcal{L}^n$ and letting $\alpha \to 0^+$, we arrive at the conclusion. Resuming, we have

$$\lim_{t_0 \to 0^+} \int \vartheta_{t_0}^{\varepsilon,\delta} \eta_{t_0} d\mathcal{L}^n = \int \vartheta_0^{\varepsilon,\delta} d\mu_0,$$

so

$$\int \vartheta_{t_1}^{\varepsilon,\delta}\eta_{t_1} d\mathcal{L}^n - \int \vartheta_0^{\varepsilon,\delta} d\mu_0 \leq \int\int_0^{t_1} \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n.$$

The limit $t_1 \to 1^-$ is completely analogous and provides

$$\int \vartheta_1^{\varepsilon,\delta} d\mu_1 - \int \vartheta_0^{\varepsilon,\delta} d\mu_0 \leq \int\int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8}|D\log\eta_t|^2\right)\eta_t dt d\mathcal{L}^n.$$

Now,

$$\int \vartheta_1^{\varepsilon,\delta} d\mu_1 = \int \psi_1^{\varepsilon,\delta} d\mu_1 - \frac{\varepsilon}{2}\int \log\left((f_1^\varepsilon + \delta)(g_1^\varepsilon + \delta)\right) d\mu_1,$$

so, passing to the limit as $\delta \to 0^+$, by monotonicity, we get

$$\lim_{\delta \to 0^+} \int \vartheta_1^{\varepsilon,\delta} d\mu_1 = \int \psi_1^\varepsilon d\mu_1 - \frac{\varepsilon}{2} H(\mu_1|\mathcal{L}^n),$$

where both terms in the right-hand side are finite: the second by the assumption on ϱ_1 and the first by the fact that

$$\int \psi_1^\varepsilon d\mu_1 = \varepsilon H(\mu_1|\mathcal{L}^n) - \int \varphi_1^\varepsilon d\mu_1$$

and $\varphi_1^\varepsilon \varrho_1 \in L^1(\mathbb{R}^n)$ by Lemma (5.28). Analogous considerations can be made for

$$\int \vartheta_0^{\varepsilon,\delta} d\mu_0,$$

so

$$\begin{aligned}\int \psi_1^\varepsilon d\mu_1 + \int \varphi_0^\varepsilon d\mu_0 \leq \frac{\varepsilon}{2} \left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n) \right) + \\ + \int \int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8} |D \log \eta_t|^2 \right) \eta_t dt d\mathcal{L}^n,\end{aligned}$$

By the explicit representation formula of the minimizer of $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$ and passing to the infimum over $(\eta_t \mathcal{L}^n, v_t)$ we get the desired inequality.

Combining the two inequalities proved with Proposition (5.30), we get the first part of the statement.

Let us now tackle the second part of the statement, namely that the problem

$$\min_{\substack{(\eta_t \mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0 = \varrho_0, \eta_1 = \varrho_1}} \left\{ \int \int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8} |D \log \eta_t|^2 \right) \eta_t dt d\mathcal{L}^n \right\}$$

has a unique minimizer $(\varrho_t^\varepsilon, D\vartheta_t^\varepsilon)$. Consider the convex set

$$\Gamma = \{ (\eta_t \mathcal{L}^n, m_t) = (\eta_t \mathcal{L}^n, \eta_t v_t) : (\eta_t \mathcal{L}^n, v_t) \text{is a continuity pair, } \eta_0 = \varrho_0, \eta_1 = \varrho_1 \}$$

and the convex and lower semicontinuous function $\Phi : [0, +\infty] \times \mathbb{R} \to [0, +\infty]$ such that

$$\Phi(x, y) = \begin{cases} \frac{y^2}{x} & \text{if } x > 0, \\ 0 & \text{if } (x, y) = (0, 0), \\ +\infty & \text{otherwise.} \end{cases}$$

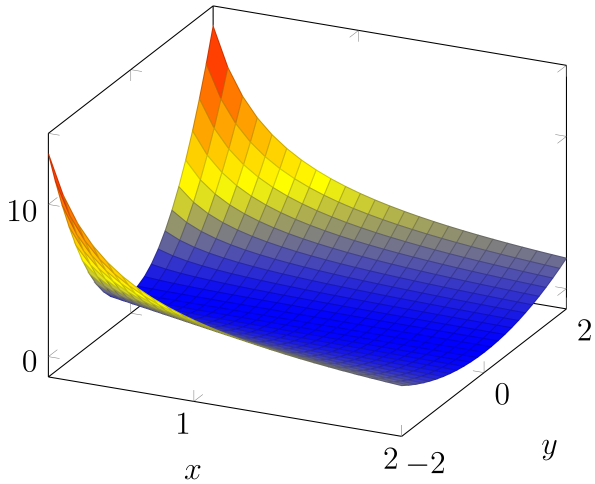

Figure 5.2: The graph of the function $f(x,y) = \frac{y^2}{x}$ for $x > 0$.

In particular, the functionals $\mathcal{H}, \mathcal{F}, \mathcal{A} : \Gamma \to [0, +\infty]$ such that

$$\mathcal{H}(\eta_t, m_t) = \frac{1}{2} \int \int_0^1 \frac{|m_t|^2}{\eta_t} dt d\mathcal{L}^n,$$

$$\mathcal{F}(\eta_t, m_t) = \begin{cases} \frac{\varepsilon^2}{8} \int \int_0^1 \frac{|D\eta_t|^2}{\eta_t} dt d\mathcal{L}^n & \text{if } \eta_t \in H^1_{loc}(\mathbb{R}^n), \\ +\infty & \text{otherwise,} \end{cases}$$

$$\mathcal{A} = \mathcal{H} + \mathcal{F},$$

are convex too. Thus, if $(\overline{\eta}_t, \overline{v}_t)$ is a minimizer of

$$\min_{\substack{(\eta_t \mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0 = \varrho_0, \eta_1 = \varrho_1}} \left\{ \int \int_0^1 \left(\frac{1}{2}|v_t|^2 + \frac{\varepsilon^2}{8} |D \log \eta_t|^2 \right) \eta_t dt d\mathcal{L}^n \right\},$$

setting $\overline{m}_t = \overline{\eta}_t \overline{v}_t$, by Proposition (5.30), $\mathcal{A}(\overline{\eta}_t, \overline{m}_t) = \mathcal{A}(\varrho^\varepsilon_t, \varrho^\varepsilon_t D\vartheta^\varepsilon_t)$ so, using the convexity of $\mathcal{A}$, given $\lambda \in]0,1[$ and setting

$$\eta^\lambda_t = (1-\lambda)\eta_t + \lambda \varrho^\varepsilon_t, \qquad m^\lambda_t = (1-\lambda)m_t + \lambda \varrho^\varepsilon_t D\vartheta^\varepsilon_t,$$

we get

$$\mathcal{A}(\eta^\lambda_t, m^\lambda_t) = (1-\lambda)\mathcal{A}(\overline{\eta}_t, \overline{m}_t) + \lambda \mathcal{A}(\varrho^\varepsilon_t, \varrho^\varepsilon_t D\vartheta^\varepsilon_t).$$

The same identity holds for $\mathcal{F}$, because of the convexity of $\mathcal{H}$ and $\mathcal{F}$, then

$$\Phi(\eta^\lambda_t, m^\lambda_t) = (1-\lambda)\Phi(\overline{\eta}_t, \overline{m}_t) + \lambda \Phi(\varrho^\varepsilon_t, \varrho^\varepsilon_t D\vartheta^\varepsilon_t).$$

By the fact that Φ is linear only on the lines passing through the origin, there must exist, for all $t \in]0,1[$ and for a.e. $x \in X$, $\alpha_t(x)$ such that

$$(\overline{\eta}_t, |D\overline{\eta}_t|) = \alpha_t(\varrho_t^\varepsilon, |D\varrho_t^\varepsilon|).$$

Since $\varrho_t^\varepsilon > 0$, it must be $\alpha_t(x) \geq 0$. Observing that

$$\overline{\eta}_t = \frac{\overline{\eta}_t}{\varrho_t^\varepsilon}\varrho_t^\varepsilon,$$

it must be

$$\alpha_t = \frac{\overline{\eta}_t}{\varrho_t^\varepsilon}.$$

In particular, by the fact that $\varrho_t^\varepsilon \in H^1(\mathbb{R}^n)$, and it is locally bounded away from 0, and $\overline{\eta}_t \in H^1_{loc}(\mathbb{R}^n)$ for a.e. $t \in]0,1]$, then $\alpha_t \in H^1_{loc}(\mathbb{R}^n)$ for a.e. $t \in]0,1]$. In a similar way, we also obtain

$$(\eta_t^\lambda, |D\eta_t^\lambda|) = ((1-\lambda)\alpha_t + \lambda)(\varrho_t^\varepsilon, |D\varrho_t^\varepsilon|).$$

In particular,

$$|D\eta_t^\lambda|^2 = ((1-\lambda)\alpha_t+\lambda)^2|D\varrho_t^\varepsilon|^2+(1-\lambda)^2(\varrho_t^\varepsilon)^2|D\alpha_t|^2+2(1-\lambda)((1-\lambda)\alpha_t+\lambda)\varrho_t^\varepsilon D\varrho_t^\varepsilon\cdot D\alpha_t.$$

Resuming, for every $\lambda \in]0,1[$

$$|D\alpha_t|^2 = -2\left(\alpha_t + \frac{\lambda}{1-\lambda}\right) D\log\varrho_t^\varepsilon \cdot D\alpha_t.$$

By the fact that α_t does not depend on λ, it must be $|D\alpha_t|^2 = 0$, then we can say that α_t is constant. Since $\overline{\eta}_t\mathcal{L}^n, \varrho_t^\varepsilon\mathcal{L}^n \in \mathcal{P}(\mathbb{R}^n)$, it must be $\alpha_t = 1$, namely for every $t \in [0,1]$ we have $\overline{\eta}_t = \varrho_t^\varepsilon$. The conclusion follows now from the strict convexity of the function $\left\{v_t \longmapsto \int\int_0^1|v_t|^2\varrho_t^\varepsilon dtd\mathcal{L}^n\right\}$. ■

(5.36) Theorem (Benamou–Brenier formula, II form) *For every $\varepsilon > 0$, one has*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \\ + \frac{1}{2}\min\left\{\int_0^1\int |v_t|^2 d\nu_t dt : \frac{\partial \nu_t}{\partial t} + \mathrm{div}(v_t\nu_t) = \frac{\varepsilon}{2}\Delta\nu_t, \nu_0 = \mu_0, \nu_1 = \mu_1\right\}.$$

In particular, on the right-hand side, the (unique) minimizer is $(\varrho_t^\varepsilon\mathcal{L}^n, D\psi_t^\varepsilon)$.

Proof. It is a matter of replacing Proposition (5.30) with Proposition (5.32) in the proof of the first form. ■

(5.37) Theorem (Benamou–Brenier formula, III form) *For every $\varepsilon > 0$, one has*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_1 \mid \mathcal{L}^n) + \\ + \frac{1}{2} \min \left\{ \int_0^1 \int |v_t|^2 d\nu_t dt : -\frac{\partial \nu_t}{\partial t} + \operatorname{div}(v_t \nu_t) = \frac{\varepsilon}{2} \Delta \nu_t, \nu_0 = \mu_0, \nu_1 = \mu_1 \right\}.$$

In particular, on the right-hand side, the (unique) minimizer is $(\varrho_t^\varepsilon \mathcal{L}^n, D\varphi_t^\varepsilon)$.

Proof. It is sufficient to swap μ_0 and μ_1 in the second form of the Benamou–Brenier formula. ■

5 Comparison

Using the Benamou–Brenier formulas, we can provide a different point of view to show the connection between the optimal transport problem and the Schrödinger problem. If we compare, for example, the statement of the Benamou–Brenier formula for optimal transport and the first form of the analogous formula for the Schrödinger problem, namely

$$\min_{\Gamma(\mu,\nu)} C = \min \left\{ \int_0^1 \int |v_t|^2 d\mu_t dt : (\mu_t, v_t) \text{ continuity pair, } \mu_t : [0,1] \to \mathcal{P}_2(\mathbb{R}^n) \right\}$$

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \frac{\varepsilon}{2} \left(H(\mu_0 \mid \mathcal{L}^n) + H(\mu_1 \mid \mathcal{L}^n) \right) + \\ + \min_{\substack{(\eta_t \mathcal{L}^n, v_t) \text{ continuity pair,} \\ \eta_0 = \varrho_0, \eta_1 = \varrho_1}} \left\{ \int \int_0^1 \left(\frac{1}{2} |v_t|^2 + \frac{\varepsilon^2}{8} |D \log \eta_t|^2 \right) \eta_t dt d\mathcal{L}^n \right\}.$$

we again can state, formally, that in the limit as $\varepsilon \to 0$, from the Schrödinger problem we pass to the optimal transport problem. It is another piece of evidence of the fact that the Schrödinger problem is a regularized version of the optimal transport one. Analogous considerations can be made with the other two forms of the Benamou–Brenier formula for the Schrödinger problem leading to the same evidence.

It is interesting to note that using the Benamou–Brenier formulas, the comparison can be made without further calculations, while, instead, using the primal formulations of the two problems, it was required to expand the entropy functional.

Chapter 6

Semigroup representations

Also in this chapter, we will consider, as an environment, $\mathbb{R}^n$ equipped with the Euclidean distance.

1 Hamilton–Jacobi equation and viscosity solution

The formulation of classical mechanics closest to quantum mechanics is written using the Hamilton–Jacobi equation, in which the motion of a particle is represented as a wave. For our purposes, the problem can be defined as follows.

(6.1) Definition *Consider $g \in \mathrm{Lip}(\mathbb{R}^n)$. We say that u is a* classical solution *of the initial-value problem for the* Hamilton–Jacobi equation*, namely*

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times \,]0, +\infty[\,, \\ u = g & \text{on } \mathbb{R}^n \times \{0\}\,, \end{cases}$$

if $u \in C^1(\mathbb{R}^n \times \,]0, +\infty[) \cap C(\mathbb{R}^n \times [0, +\infty[)$, $\frac{\partial u}{\partial t}(x,t) - \frac{1}{2}|Du|^2(x,t) = 0$ for every $(x,t) \in \mathbb{R}^n \times \,]0, +\infty[$ and $u(x,0) = g(x)$ for every $x \in \mathbb{R}^n$.

Further mechanical information can be found, for example, in [30] or in [60]. From a mathematical point of view, however, which is what interests us most, the Hamilton–Jacobi equation appears when we are interested in describing generalizations of the problems of the Calculus of Variations: see, for example, [33].

We are interested in particular "weak" solutions of the previous initial-value problem, namely viscosity solutions.

(6.2) Definition *Consider $g \in \mathrm{Lip}(\mathbb{R}^n)$. We say that u is a* viscosity solution *of the* *initial-value problem for the* Hamilton–Jacobi equation*, namely*

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times]0,+\infty[\,, \\ u = g & \text{on } \mathbb{R}^n \times \{0\}\,, \end{cases}$$

if $u(x,0) = g(x)$ *for every* $x \in \mathbb{R}^n$ *and for every* $v \in C^\infty(\mathbb{R}^n \times]0,+\infty[)$ *the following facts hold true:*

(*i*) *if* $(x_0,t_0) \in \mathbb{R}^n \times]0,+\infty[$ *is a local maximum point for* $u - v$*, then*

$$\frac{\partial u}{\partial t}(x_0,t_0) - \frac{1}{2}|Du|^2(x_0,t_0) \le 0,$$

(*ii*) *if* $(x_0,t_0) \in \mathbb{R}^n \times]0,+\infty[$ *is a local minimum point for* $u - v$*, then*

$$\frac{\partial u}{\partial t}(x_0,t_0) - \frac{1}{2}|Du|^2(x_0,t_0) \ge 0.$$

The construction of this last definition and its consistency can be found in [24].

(6.3) Theorem *Consider* $g \in \mathrm{Lip}(\mathbb{R}^n)$*. There exists at most one viscosity solution of the initial-value problem for the Hamilton–Jacobi equation, namely*

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times]0,+\infty[\,, \\ u = g & \text{on } \mathbb{R}^n \times \{0\}\,. \end{cases}$$

Proof. See [24]. ∎

Further information on the Hamilton–Jacobi equation and viscosity solution can be found, for example, in [7] or [22].

2 Hopf–Lax semigroup and optimal transport

The aim of this section is to determine a pair of Kantorovich potentials for the optimal transport problem along geodesics. The first, and most important, ingredient we need is the Hopf–Lax semigroup.

(6.4) Definition *We call* Hopf–Lax semigroup *the family of operators* $Q_t : C_b(\mathbb{R}^n) \to \mathbb{R}$, $t \in [0,+\infty]$*, such that for every* $f \in C_b(\mathbb{R}^n)$ *and* $x \in \mathbb{R}^n$

$$Q_0 f(x) = f(x)$$

and, for $t > 0$,

$$Q_t f(x) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2t}|x-y|^2 \right\}.$$

Through the Hopf–Lax semigroup, it is possible to find a representation formula for the viscosity solution of the initial-value problem for the Hamilton–Jacobi equation.

(6.5) Theorem *Consider $g \in \mathrm{Lip}(\mathbb{R}^n)$ bounded. The unique viscosity solution of the initial-value problem for the Hamilton–Jacobi equation, namely*

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times \left]0, +\infty\right[\,, \\ u = g & \text{on } \mathbb{R}^n \times \{0\}\,, \end{cases}$$

is given by

$$u(x,t) = Q_t g(x).$$

Proof. See [24]. ∎

Other properties of the Hopf–Lax semigroup can be found in the Appendix.

We have seen in Proposition (5.4) that if $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, with $\mu_0 \ll \mathcal{L}^n$, and μ_t is the geodesic connecting μ_0 and μ_1, then (μ_t, v_t), with

$$v_t = (T - \mathrm{Id}) \circ T_t^{-1} = \frac{\mathrm{Id} - T_t^{-1}}{t},$$

where T is the unique, up to μ_0-negligible sets, minimizer for the problem

$$\min \left\{ \int_{\mathbb{R}^n} \frac{1}{2}|x - T(x)|^2 d\mu_0(x) : T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel, } T_{\#}\mu_0 = \mu_1 \right\},$$

and $T_t = (1-t)\mathrm{Id} + tT$, is a continuity pair. In particular, $\mu_t = (T_t)_{\#}\mu_0$. Fix, now $t \in [0, +\infty[$ and consider (φ_t, ψ_t), Kantorovich potential for μ_0, μ_t, whose existence follows from Kantorovich–Rubinstein duality. In accordance with Theorem (3.10) and Theorem (3.11),

$$T_t(x) = D\psi_t(x) = -D(-\psi_t(x) + \frac{1}{2}|x|^2) + \frac{1}{2}D|x|^2$$

but $\varphi_t(x) = -\psi_t(x) + \frac{1}{2}|x|^2$ so

$$T_t = \mathrm{Id} - D\varphi_t.$$

In particular,

$$\mathrm{Id} - D\varphi_t = T_t = (1-t)\mathrm{Id} + t(\mathrm{Id} - D\varphi)$$

so

$$D\varphi_t = tD\varphi,$$

then a good choice might be $\varphi_t(x) = t\varphi(x)$ and

$$\psi_t(x) = \inf_{y\in\mathbb{R}^n} \left\{-t\varphi(x) + \frac{1}{2}|x-y|^2\right\}.$$

We will show, as reported in [1], that this informal reasoning is correct. To do this, we need, first of all, an elementary inequality.

(6.6) Proposition *Let $a, b \in \mathbb{R}$ and $s, t > 0$. It holds*

$$\frac{a^2}{s} + \frac{b^2}{t} \geq \frac{(a+b)^2}{s+t}.$$

In particular, the equality holds if and only if $at = bs$.

Proof. By the Cauchy–Schwarz inequality

$$\forall (\xi,\eta),(\nu,\sigma) \in \mathbb{R}\times\mathbb{R} : [(\xi,\eta)\cdot(\nu,\sigma)]^2 \leq (\xi^2+\eta^2)(\nu^2+\sigma^2).$$

Choose $\xi = \frac{a}{\sqrt{s}}, \eta = \frac{b}{\sqrt{t}}, \nu = \sqrt{s}$ and $\sigma = \sqrt{t}$ to obtain

$$(a+b)^2 \leq \left(\frac{a^2}{s} + \frac{b^2}{t}\right)(s+t)$$

or in other words

$$\frac{a^2}{s} + \frac{b^2}{t} \geq \frac{(a+b)^2}{s+t}.$$

Remembering that the Cauchy–Schwarz inequality holds as an equality if and only if the vectors are linearly dependent, we get the second part of the statement. ∎

We are therefore ready to show that our informal reasoning is indeed correct. Let us just point out that we already know, from Kantorovich–Rubinstein duality and the very definition of c–conjugate, that $(\varphi, Q_1(-\varphi))$ is a pair of Kantorovich potentials from μ_0 to μ_1. The goal of the proof is, indeed, to prove that the representation is true for every time along the geodesic connecting μ_0 to μ_1. Following the idea in [1], we will use the dynamical representation of the optimal transport problem to be able to apply the previous elementary inequality.

(6.7) Theorem *Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^n)$, with $\mu_0 \ll \mathcal{L}^n$, and $(\varphi, Q_1(-\varphi))$ be a pair of Kantorovich potentials from μ_0 to μ_1. Then, given μ_t the geodesic connecting μ_0 and μ_1,* *for all* $t \in [0,1]$

$$(t\varphi, tQ_t(-\varphi))$$

is a pair of Kantorovich potentials from μ_0 *to* μ_t.

Proof. The case $t = 0$ is trivial. We then focus on the case $t \in]0,1]$. By definition of the Hopf–Lax semigroup, for every $x, y \in \mathbb{R}^n$

$$t\varphi(x) + Q_1(-t\varphi)(y) \leq \frac{1}{2}|x-y|^2,$$

so $(t\varphi, Q_1(-t\varphi))$ is admissible. By Theorem (4.3), there exists $\eta \in \mathrm{OptGeo}(\mu_0, \mu_t)$ such that $\Sigma = (e_0, e_t)_\# \eta \in \Gamma(\mu_0, \mu_t)$ is optimal. We only need to prove that

$$\Sigma\text{-a.e. } (x,y) \in \mathbb{R}^n \times \mathbb{R}^n : t\varphi(x) + Q_1(-t\varphi)(y) \geq \frac{1}{2}|x-y|^2. \tag{6.8}$$

First of all, (6.8) is equivalent to

$$\eta\text{-a.e. } \gamma \in \mathrm{Geo}(\mathbb{R}^n) : t\varphi(\gamma(0)) + Q_1(-t\varphi)(\gamma(t)) \geq \frac{1}{2}|\gamma(0)-\gamma(1)|^2. \tag{6.9}$$

Now, by the optimality of $(\varphi, Q_1(-\varphi))$,

$$\eta\text{-a.e. } \gamma \in \mathrm{Geo}(\mathbb{R}^n) : t\varphi(\gamma(0)) + Q_1(-t\varphi)(\gamma(1)) = \frac{1}{2}|\gamma(0)-\gamma(1)|^2, \tag{6.10}$$

so we only need to prove (6.9) for intermediate points, namely if $x = \gamma(0)$ and $y = \gamma(1)$, we need to prove

$$\eta\text{-a.e. } \gamma \in \mathrm{Geo}(\mathbb{R}^n) : \varphi(x) + Q_t(-\varphi)(z) \geq \frac{1}{2t}|x-z|^2 \tag{6.11}$$

for $z = \gamma(t) \in \mathbb{R}^n$ such that $|x-z| = td|x-y|$ and $|z-y| = (1-t)|x-y|$. Observing that

$$\begin{aligned} \varphi(x) + Q_t(-\varphi)(z) &= \varphi(x) + \inf_{w \in \mathbb{R}^n} \left\{ -\varphi(w) + \frac{1}{2t}|w-z|^2 \right\} = \\ &= \inf_{w \in \mathbb{R}^n} \left\{ \varphi(x) - \varphi(w) + \frac{1}{2t}|w-z|^2 \right\}, \end{aligned}$$

the definition of Hopf–Lax semigroup, combined with (6.9), gives us

$$\begin{aligned} \varphi(x) + Q_t(-\varphi)(z) = \inf_{w \in \mathbb{R}^n} &\left\{ \frac{1}{2}|x-y|^2 - Q_1(-\varphi)(y) - \varphi(w) + \frac{1}{2t}|w-z|^2 \right\} \geq \\ &\geq \frac{1}{2}|x-y|^2 + \inf_{w \in \mathbb{R}^n} \left\{ -\frac{1}{2}|w-y|^2 + \frac{1}{2t}|w-z|^2 \right\}. \end{aligned}$$

In other words, to prove (6.11) it is sufficient to show

$$\frac{1}{2}|x-y|^2 + \inf_{w\in\mathbb{R}^n}\left\{-\frac{1}{2}|w-y|^2 + \frac{1}{2t}|w-z|^2\right\} \geq \frac{1}{2t}|x-z|^2,$$

that can be rewritten, using the properties of z, in the form

$$\frac{1}{2(1-t)^2}|z-y|^2 + \inf_{w\in\mathbb{R}^n}\left\{-\frac{1}{2}|w-y|^2 + \frac{1}{2t}|w-z|^2\right\} \geq \frac{1}{2(t-1)^2}|z-y|^2$$

that is

$$\inf_{w\in\mathbb{R}^n}\left\{-\frac{1}{2}|w-y|^2 + \frac{1}{2t}|w-z|^2\right\} \geq -\frac{1}{2(1-t)}|y-z|^2. \tag{6.12}$$

By Proposition (6.6) and the triangle inequality, for every $w \in \mathbb{R}^n$

$$\frac{|y-z|^2}{2(1-t)} + \frac{|w-z|^2}{2t} \geq \frac{|w-y|^2}{2},$$

then (6.12) follows and the proof is complete. ∎

3 Hopf–Cole semigroup and Schrödinger problem

Replacing the Hopf–Lax semigroup with an analogous semigroup built through the heat kernel, namely the Hopf–Cole semigroup, it is possible to provide a similar semigroup representation also for the Schrödinger problem.

(6.13) Definition *We call* Hopf–Cole semigroup *the family of operators* $\overline{Q}_t : C_b(\mathbb{R}^n) \to \mathbb{R}$, $t \in [0, +\infty]$, *such that for every* $f \in C_b(\mathbb{R}^n)$ *and* $x \in \mathbb{R}^n$

$$\overline{Q}_0 f(x) = f(x)$$

and, for $t > 0$,

$$\overline{Q}_t f(x) = \log\left(\int e^{f(y)}\mathsf{r}_t(x,y)d\mathcal{L}^n(y)\right).$$

In other words, a Kantorovich–Rubinstein duality holds also for the Schrödinger problem. Following the approach reported in [27], we first need to consider the forward (resp. backward) Hamilton–Jacobi–Bellman equation, namely given $c > 0$

$$\frac{\partial u}{\partial t} - \frac{1}{2}|Du|^2 = c\Delta u \qquad \left(\text{resp. } -\frac{\partial u}{\partial t} - \frac{1}{2}|Du|^2 = c\Delta u\right),$$

for which we need a notion of supersolution.

(6.14) Definition *Let $T \in]0, +\infty[$ and $c > 0$ and a function $u : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ such that, for every $t \in [0, T]$, the function $u_t : \mathbb{R}^n \to \mathbb{R}$ such that*

$$u_t(x) = u(t, x)$$

is Borel. We say that u is a (strong) supersolution of the forward (resp. backward) Hamilton–Jacobi–Bellman equation *if the following facts hold true:*

(i) *there exists $C > 0$ such that for every $t \in [0, T]$*

$$\|u_t\|_{L^\infty(\mathbb{R}^n)} \leq C,$$

(ii) *for every $t \in [0, T]$ $u_t \in D(\Delta_{loc}) \subseteq H^1(\mathbb{R}^n)$,*

(iii) *the functions $\{t \longmapsto |Du_t|\}$ and $\{t \longmapsto \Delta u_t\}$ belong to $L^\infty(]0, T[\,, L^2(\mathbb{R}^n))$,*

(iv) *there exists $x \in \mathbb{R}^n$ and $M > 0$ such that the function $\{t \longmapsto u_t\}$ belongs to $AC([0, T], L^2(\mathbb{R}^n, e^{-V}\mathcal{L}^n))$, where $V(y) = |x - y|^2$, and for* a.e. $t \in [0, T]$

$$\frac{\partial}{\partial t}u - \frac{1}{2}|Du|^2 \geq c\Delta u \qquad \left(\text{resp. } -\frac{\partial}{\partial t}u - \frac{1}{2}|Du|^2 \geq c\Delta u\right).$$

Further information on the Hamilton–Jacobi–Bellman equation can be found, for example, in [35] or [36].

As a preliminary result, already interesting by itself, we have the following Theorems, whose proof relies on the technique used to prove the Benamou–Brenier formula for the Schrödinger problem.

(6.15) Theorem *Let $c > 0$. The following facts hold true:*

(a) *if $\varphi : [0, 1] \times \mathbb{R}^n \to \mathbb{R}$ is a supersolution of the backward Hamilton–Jacobi–Bellman equation and (ν_t, v_t) is a forward Fokker–Planck pair, then*

$$\int \varphi_1 d\nu_1 - \int \varphi_0 d\nu_0 \leq \frac{1}{2}\int_0^1 \int |v_t|^2 d\nu_t dt,$$

(b) *for any $\varepsilon > 0$*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \sup\left\{\int \varphi_1 d\mu_1 - \int \varphi_0 d\mu_0 : \varphi : [0, 1] \times \mathbb{R}^n \to \mathbb{R},\right.$$

$$\text{supersolution of the backward Hamilton–Jacobi–Belmann equation}\Big\}.$$

Proof.

(a) Fix $\varphi : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ a supersolution of the backward Hamilton–Jacobi–Bellman equation and (ν_t, v_t) a forward Fokker–Planck pair. Consider $R > 0$ and a cut-off function $\zeta \in C_c^\infty(\mathrm{B}(0, R+1))$ such that $0 \le \zeta \le 1$ and $\zeta = 1$ on $\mathrm{B}(0,R)$. By the very definition of a supersolution of the backward Hamilton–Jacobi–Bellman equation, the function $\{t \longmapsto \zeta\varphi_t\}$ belongs to

$$AC([0,1], L^2(\mathbb{R}^n)) \cap L^\infty([0,1], H^1(\mathbb{R}^n))$$

and the function $\{t \longmapsto \Delta(\zeta\varphi_t)\}$ belongs to $L^\infty([0,1], L^2(\mathbb{R}^n))$. In particular, by Lemma (5.23), applied to (ν_t, v_t) and to the function $\{t \longmapsto \zeta\varphi_t\}$, for a.e. $t \in [0,1]$

$$\frac{d}{ds}\Big(\int \zeta\varphi_s d\nu_s\Big)|_{s=t} = \int \zeta \left(\frac{d}{ds}\varphi_s\right)|_{s=t} d\nu_t + \frac{d}{ds}\Big(\int \zeta\varphi_t d\nu_s\Big)|_{s=t}.$$

For the first term in the right-hand side, using again the fact that φ is a supersolution of the backward Hamilton–Jacobi–Bellman equation,

$$\frac{d}{ds}\Big(\int \zeta\varphi_s d\nu_s\Big)|_{s=t} \le -\int \zeta\left(\frac{1}{2}|D\varphi_t|^2 + c\Delta\varphi_t\right) d\nu_t$$

and as regards the second, using the fact that (ν_t, v_t) is a forward Fokker–Planck pair and $\zeta\varphi_t \in D(\Delta) \subseteq H^1(\mathbb{R}^n)$, arguing by density and using Young's inequality, we can write

$$\begin{aligned}\int \zeta\left(\frac{d}{ds}\varphi_s\right)|_{s=t} d\nu_t &= \int \left(\zeta D\varphi_t \cdot v_t + \varphi_t D\zeta \cdot v_t\right) d\nu_t + \\ &+ c\int \left(\zeta\Delta\varphi_t + 2D\zeta\cdot D\varphi_t + \varphi_t\Delta\zeta\right) d\nu_t \le \\ &\le \int \left(\zeta\left(\frac{1}{2}|D\varphi_t|^2 + \frac{1}{2}|v_t|^2\right) + \varphi_t D\zeta\cdot v_t\right) d\nu_t + \\ &+ c\int \left(\zeta\Delta\varphi_t + 2D\zeta\cdot D\varphi_t + \varphi_t\Delta\zeta\right) d\nu_t.\end{aligned}$$

Resuming,

$$\frac{d}{ds}\Big(\int \zeta\varphi_s d\nu_s\Big)|_{s=t} \le \frac{1}{2}\int \zeta|v_t|^2 d\nu_t + \int \varphi_t D\zeta\cdot v_t d\nu_t + c\int \left(2D\zeta\cdot D\varphi_t + \varphi_t\Delta\zeta\right) d\nu_t.$$

Integrating both sides in t over $[0,1]$, we get

$$\int \zeta\varphi_1 d\nu_1 - \int \zeta\varphi_0 d\nu_0 \leq \frac{1}{2}\int_0^1 \int \zeta|v_t|^2 d\nu_t dt + \\ + \int_0^1 \int \varphi_t D\zeta \cdot v_t d\nu_t dt + c\int_0^1 \int (2D\zeta \cdot D\varphi_t + \varphi_t \Delta\zeta)\, d\nu_t dt.$$

Passing through the limit as $R \to +\infty$, in the same way as in the proof of the first form of the Benamou–Brenier formula for the Schrödinger problem, the inequality follows.

(b) Let us start proving the inequality

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} \geq \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \sup\left\{\int \varphi_1 d\mu_1 - \int \varphi_0 d\mu_0 : \varphi : [0,1]\times\mathbb{R}^n \to \mathbb{R},\right. \\ \left.\text{supersolution of the backward Hamilton–Jacobi–Belmann equation}\right\}.$$

By the second form of the Benamou–Brenier formula for the Schrödinger problem,

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \\ + \frac{1}{2}\min\left\{\int_0^1 \int |v_t|^2 d\nu_t dt : -\frac{\partial \nu_t}{\partial t} + \operatorname{div}(v_t\nu_t) = \frac{\varepsilon}{2}\Delta\nu_t, \nu_0 = \mu_0, \nu_1 = \mu_1\right\},$$

so, by (a),

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} \geq \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \int \varphi_1 d\nu_1 - \int \varphi_0 d\nu_0,$$

and passing to the supremum the inequality follows.

As concerns the $\leq$ inequality, given $\delta, s > 0$, consider

$$\varphi_t^{\delta,s}(x) = \varepsilon \log\left(\int g^\varepsilon(y) \mathsf{r}_{\frac{(1-t)\varepsilon}{2}+s}(x,y) d\mathcal{L}^n(y) + \delta\right).$$

By Lemma (5.29), $\varphi^{\delta,s}$ is a supersolution to the backward Hamilton–Jacobi–Bellman equation on $[0,1]$, so

$$\int \varphi_1^{\delta,s} d\mu_1 - \int \varphi_0^{\delta,s} d\mu_0 \leq \sup\left\{\int \varphi_1 d\mu_1 - \int \varphi_0 d\mu_0 : \varphi : [0,1]\times\mathbb{R}^n \to \mathbb{R},\ \text{supersolution}\right. \\ \left.\text{of the backward Hamilton–Jacobi–Belmann equation}\right\}.$$

Now, we want to pass to the limit as $s \to 0^+$: using the continuity of the function

$\left\{s \longmapsto \varphi_t^{\delta,s}\right\}$ with respect to the $L^2(\mathbb{R}^n, e^{-V}\mathcal{L}^n)$ norm, where $V(y) = M|x-y|^2$, and the compactness of the support of μ_0, μ_1, we get

$$\lim_{s\to 0^+} \int \varphi_0^{\delta,s} d\mu_0 = \int \psi_0^{\varepsilon,\delta} d\mu_0, \qquad \lim_{s\to 0^+} \int \varphi_1^{\delta,s} d\mu_1 = \int \psi_1^{\varepsilon,\delta} d\mu_1.$$

Arguing now in the same way as in the proof of the first form of the Benamou–Brenier formula for the Schrödinger problem, passing through the limit as $\delta \to 0^+$ we get

$$\int \psi_1^{\varepsilon} d\mu_1 - \int \psi_0^{\varepsilon} d\mu_0 \leq \sup\Bigg\{\int \varphi_1 d\mu_1 - \int \varphi_0 d\mu_0 : \varphi : [0,1]\times\mathbb{R}^n \to \mathbb{R}, \text{ supersolution}$$
$$\text{of the backward Hamilton–Jacobi–Belmann equation}\Bigg\}.$$

Combining the previous inequality, the identity $\psi_0^{\varepsilon} = -\varphi_0^{\varepsilon} + \varepsilon \log \varrho_0$ and the explicit representation formula of the minimizer of $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$, the desired inequality follows. ■

(6.16) Theorem *Let $c > 0$. The following facts hold true:*

(*a*) *if $\varphi : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ is a supersolution of the forward Hamilton–Jacobi–Bellman equation and (ν_t, v_t) is a backward Fokker–Planck pair, then*

$$\int \varphi_0 d\nu_0 - \int \varphi_1 d\nu_1 \leq \frac{1}{2}\int_0^1 \int |v_t|^2 d\nu_t dt,$$

(*b*) *for any $\varepsilon > 0$*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_1 \mid \mathcal{L}^n) + \sup\Bigg\{\int \varphi_0 d\mu_0 - \int \varphi_1 d\mu_1 : \varphi : [0,1]\times\mathbb{R}^n \to \mathbb{R},$$
$$\text{supersolution of the forward Hamilton–Jacobi–Belmann equation}\Bigg\}.$$

Proof. It is a matter of reversing time and then using the same strategy as in the previous Theorem. ■

The following are the main results of the section. Using the Hopf–Cole semigroup, as said at the beginning of the section, we can provide a dual variational formula for the Schrödinger problem. As regards the first results, following what is reported in [27], the $\leq$ inequality follows directly from the explicit representation formula for the minimizer of the entropy. The converse inequality, instead, is a consequence of Theorem (6.15).

(6.17) Theorem *For any $\varepsilon > 0$,*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \sup\left\{ \int u d\mu_1 - \int \varepsilon \overline{Q}_{\frac{\varepsilon}{2}} \left(\frac{u}{\varepsilon}\right) d\mu_0 : u : \mathbb{R} \to \overline{\mathbb{R}}, \right.$$
$$\left. e^{\frac{u}{\varepsilon}} \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n) \right\}.$$

Proof. Observing that $e^{\varphi_0^\varepsilon} \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$ and $\varphi_1^\varepsilon = \varepsilon \overline{Q}_{\frac{\varepsilon}{2}} \left(\frac{\varphi_0^\varepsilon}{\varepsilon}\right)$, the inequality

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} \leq \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \sup\left\{ \int u d\mu_1 - \int \varepsilon \overline{Q}_{\frac{\varepsilon}{2}} \left(\frac{u}{\varepsilon}\right) d\mu_0 : u : \mathbb{R} \to \overline{\mathbb{R}}, \right.$$
$$\left. e^{\frac{u}{\varepsilon}} \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n) \right\}$$

comes directly from the explicit representation formula for the minimizer of $H_{\mathsf{R}_{\frac{\varepsilon}{2}}}$ and the identity $\varphi_0^\varepsilon + \psi_0^\varepsilon = \varepsilon \log \varrho_0$.

As regards the converse inequality, consider $\delta, s > 0$ and define for every $t \in [0,1]$ and every $u : \mathbb{R}^n \to \overline{\mathbb{R}}$ such that $e^{\frac{u}{\varepsilon}} \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$

$$\overline{Q}^{\delta,s}_{\frac{\varepsilon t}{2}} \left(\frac{u}{\varepsilon}\right)(x) = \log\left(\int e^{\frac{u(y)}{\varepsilon}} \mathsf{r}_{\frac{\varepsilon t}{2}+s}(x,y) d\mathcal{L}^n(y) + \delta \right).$$

By Lemma (5.29), the function $\left\{(t,x) \longmapsto \overline{Q}^{\delta,s}_t u(x)\right\}$ is a supersolution of the backward Hamilton–Jacobi–Bellman equation, so, by Theorem (6.15),

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} \geq \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \int \varepsilon \overline{Q}^{\delta,s}_0 \left(\frac{u}{\varepsilon}\right) d\mu_1 - \int \varepsilon \overline{Q}^{\delta,s}_{\frac{\varepsilon}{2}} \left(\frac{u}{\varepsilon}\right) d\mu_0.$$

Passing through the limit as $s \to 0^+$ and $\delta \to 0^+$, as in the proof of the first form of the Benamou–Brenier formula for the Schrödinger problem, we get

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} \geq \varepsilon H(\mu_0 \mid \mathcal{L}^n) + \int u d\mu_1 - \int \varepsilon \overline{Q}_{\frac{\varepsilon}{2}} \left(\frac{u}{\varepsilon}\right) d\mu_0$$

and the conclusion comes passing to the supremum on u. ∎

(6.18) Theorem *For any $\varepsilon > 0$,*

$$\varepsilon \min_{\Gamma(\mu_0,\mu_1)} H_{\mathsf{R}_{\frac{\varepsilon}{2}}} = \varepsilon H(\mu_1 \mid \mathcal{L}^n) + \sup\left\{ \int u d\mu_0 - \int \varepsilon \overline{Q}_{\frac{\varepsilon}{2}} \left(\frac{u}{\varepsilon}\right) d\mu_1 : u : \mathbb{R} \to \overline{\mathbb{R}}, \right.$$
$$\left. e^{\frac{u}{\varepsilon}} \in L^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n) \right\}.$$

Proof. It is similar to the proof of the previous Theorem. ■

4 Comparison

Following what is reported in [44], we understand that the connection between the optimal transport problem and the Schrödinger problem can also be read using the Hopf–Lax semigroup and the Hopf–Cole semigroup. In other words, we can close the circle open in Chapter 2 by establishing that the dual Schrödinger problem representation via Hopf–Cole converges to the dual representation of the optimal transport problem induced by the Kantorovich–Rubinstein duality.

By exploiting the connection between the optimal transport problem and the Schrödinger problem, we can also establish duality relations among the differential equations that allow their representation. First of all, being able to represent, by the Benamou–Brenier formula, the optimal transport problem using the continuity equation, we have understood that we can operate dually, representing the problem using a differential equation, but replacing the previous one with the Hamilton–Jacobi equation. Analogously, the same duality can also be established between the backward (resp. forward) Fokker–Planck equation and the forward (resp. backward) Hamilton–Jacobi–Bellman equation.

Further developments

On the Benamou-Brenier formula for the Schrödinger problem

The assumptions contained in the Benamou-Brenier formula for the Schrödinger problem do not all appear equally necessary. The boundedness of densities, required to have finite entropy and to apply the regularization property of the heat kernel, is important for the uniqueness of the solution and the regularity of the interpolating potentials. On the other hand, the boundedness of the supports is required to have a compact support property, so a question arises, namely

Is it possible to remove this last assumption?

The issue is not only theoretical: at the moment, indeed, the Benamou-Brenier formula for the Schrödinger problem is not applicable to the case μ_0, μ_1 Gaussian. This is also important in the application since, as we have already seen in Chapter 4, this is a class of measure for which we can do explicit computations. The first next goal is then to try to relax the hypotheses for the Benamou-Brenier formula for the Schrödinger problem starting from the question posed. The hope of a positive result comes from the works of Marcel Nutz and collaborators, for example [49], in which they study the Schrödinger problem without any compactness assumption.

Functional inequalities

It is possible to apply the Schrödinger problem theory to the study of functional inequalities. Some results have been obtained regarding the recovery of already known functional inequalities or, at least in positive curvature, the improvement of them. As an example, we can cite [17]: starting from the convexity inequality of the entropy written along an entropic interpolation and deriving, the Author has managed to generalize the

Talagrand inequality, namely

$$W_2^2(\nu, \gamma) \leq 2H(\nu|\gamma),$$

where $\nu \in \mathcal{P}(M)$, with M an appropriate manifold, and γ is a standard Gaussian measure, to the entropic case. He actually obtained a stochastic version of the Talagrand inequality. This is useful because, for example, in a manifold with Ricci curvature strictly positive, the stochastic Talagrand inequality is stronger than the classical one. A possible way forward is then to apply these ideas to other notable functional inequalities, such as the log-Sobolev inequality or the concentration inequality. Concerning this topic, it is important to understand, also, if it is possible to obtain a dimensional improvement for those functional inequalities that are dimension-free.

Unbalanced optimal transport

As we have specified in Chapter 1, we have focused on transport between measures with the same mass. A possible generalization is then the unbalanced case, also called unbalanced optimal transport. In particular, in [42], the Hellinger–Kantorovich distance has been introduced as a generalization of the Wasserstein distance. Although, in [16], this latter has been generalized to the case $p \neq 2$, the theory, in its entirety, is not actually developed like its balanced counterpart. The importance of deepening this side of transport theory would open up a deeper understanding of unbalanced optimal transport. Once the problem is sufficiently understood, it would be interesting to study its entropic regularization: a static version (see [11]) and a dynamic version (see [6]) are known, but in-depth studies regarding their connection are still lacking.

Multi-marginal and grand canonical problems

If in unbalanced optimal transport we move from marginals with the same mass to unbalanced marginals, another possible way forward is to move beyond the idea of having only two marginals. This idea leads to the so-called multi-marginal optimal transport problem, ubiquitous in different disciplines such as economics (see, for example, [13]), statistics (see, for example, [12]), image processing (see, for example, [57]) and quantum physics and chemistry, in the framework of density functional theory (see, for example, [10] or [18]). A precise definition of the problem can be given as follows.

(6.19) Definition *Consider* $\mu_1, \dots, \mu_\ell \in \mathcal{P}(\mathbb{R}^n)$ *and a Borel function* $c : \mathbb{R}^{n\ell} \to [0, +\infty]$.

In the so-called multi-marginal optimal transport problem *we look for*

$$\inf\left\{\int c d\pi : \pi \in \Gamma(\mu_1, \dots, \mu_\ell)\right\},$$

where $\Gamma(\mu_1, \dots, \mu_\ell) = \left\{\pi \in \mathcal{P}(\mathbb{R}^{n\ell}) : (p^i)_\# \pi = \mu_i, i = 1, \dots, \ell\right\}$.

Of particular interest is the application to quantum chemistry: the problem models the electron-electron repulsion, when the cost function is the Coulomb potential, namely

$$c(x_1, \dots, x_\ell) = \sum_{i<j} \frac{1}{|x_i - x_j|},$$

and the marginals are all the electron density ρ. In this case, the problem leads to considering the Lieb functional, namely

$$\mathcal{F}_L[\rho] = \inf\left\{\mathrm{tr}(H_\ell \Gamma) : \Gamma \in \mathfrak{S}(\mathcal{H}) \text{ self-adjoint}, \rho_\Gamma = \rho\right\},$$

where, $\mathfrak{S}(\mathcal{H})$ is the space of bounded operators, on a well-chosen Hilbert space $\mathcal{H}$, with finite trace and $H_\ell = -\Delta + \sum_{i<j} \frac{1}{|x_i - x_j|}$.

From the multi-marginal case, another direction for a generalization is then to remove also the idea that the number of marginals must be deterministic. In this case we obtain the so-called *grand canonical optimal transport problem.*

A first direction of study is to generalize the convergence rates of the entropic cost established in [48] by substituting the signature conditions introduced in [55] with a less restrictive assumption on the cost function, and by pushing Minty's trick further to obtain quadratic detachments as done in [14] and [48]. A second direction, which would represent a major improvement also for the classical two marginals case, is to tackle the question of the convergence of entropic plans in the entropic regularization of the problem to the case in which there are several, possibly non-deterministic, unregularized optimal plans.

Appendix A

Some properties of the Hopf–Lax semigroup

In this Appendix we just want to collect the main properties of the Hopf–Lax semigroup. The proof is an expansion of the ideas reported in [1]. We only consider, as in Chapter 6, the Euclidean case. Further information, also in a more general setting, can be found, for example, in [1], [2] or [20].

(1.1) Theorem *Consider $f \in C_b(\mathbb{R}^n)$ and the function $\{(t,x) \longmapsto Q_t f(x)\}$. The following facts hold true:*

(a) *for every $t \in [0, +\infty]$*

$$\inf f \leq Q_t f \leq f \leq \sup f,$$

(b) *if $t \to 0$, then $Q_t f(x) \nearrow f(x)$ for every $x \in \mathbb{R}^n$. In particular, if f is uniformly continuous, then $Q_t f \to f$ uniformly on $\mathbb{R}^n$,*

(c) *for every $\varepsilon > 0$, $\{(t,x) \longmapsto Q_t f(x)\}$ is Lipschitz on $[\varepsilon, +\infty[\times \mathbb{R}^n$,*

(d) *for* a.e. $x \in \mathbb{R}^n$ *and* a.e. $t \in]0, +\infty[$ *one has*

$$\frac{d}{dt} Q_t f(x) + \frac{1}{2} |D(Q_t f)(x)|^2 = 0,$$

(e) *if f is also Lipschitz, then $\{(t,x) \longmapsto Q_t f(x)\}$ is Lipschitz on $[0, +\infty[\times \mathbb{R}^n$,*

(f) *for every $(t,x) \in [0, +\infty[\times \mathbb{R}^n$ it holds*

$$\frac{d^+}{dt} Q_t f(x) + \frac{1}{2} |D^*(Q_t f)|^2(x) \leq 0,$$

where

$$\frac{d^+}{dt} Q_t f(x) = \limsup_{h \to 0^+} \frac{Q_{t+h} f(x) - Q_t f(x)}{h}$$

is the upper Dini derivative *of* $Q_t f$ *and*

$$|D^* (Q_t f)|(x) = \lim_{r \to 0} \sup_{\substack{y,z \in \mathrm{B}(x,r) \\ y \neq z}} \frac{|Q_t f(z) - Q_t f(y)|}{|z - y|}$$

is the asymptotic Lipschitz constant *of* $Q_t f$.

(g) $\{Q_t\}_{t \geq 0}$ *is a semigroup.*

Proof.

(a) It follows directly from the definition of Hopf–Lax semigroup and the properties of infimum and supremum.

(b) Clearly, for every $x \in \mathbb{R}^n$, if $0 \leq t_1 \leq t_2$, then $Q_{t_2} f(x) \leq Q_{t_1} f(x)$ and, from (a),

$$\sup_{t \geq 0} Q_t f(x) \leq f(x).$$

Let us prove the converse inequality, namely $\sup_{t \geq 0} Q_t f(x) \geq f(x)$. For every $x \in \mathbb{R}^n$ and $t > 0$, given $R_t = \sqrt{2t(\sup f - \inf f)}$, consider $\overline{\mathrm{B}(x, R_t)}$. We have

$$\inf_{y \in \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x - y|^2 \right\} \leq f(x) \leq \sup f$$

and

$$\begin{aligned} \inf_{y \in \mathbb{R}^n \setminus \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x - y|^2 \right\} &\geq \inf_{\mathbb{R}^n \setminus \overline{\mathrm{B}(x,R_t)}} f + \inf_{y \in \mathbb{R}^n \setminus \overline{\mathrm{B}(x,R_t)}} \frac{1}{2t}|x - y|^2 \geq \\ &\geq \inf f + \frac{R_t^2}{2t} = \sup f, \end{aligned}$$

so

$$\inf_{y \in \mathbb{R}^n \setminus \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x - y|^2 \right\} \geq \inf_{y \in \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x - y|^2 \right\}.$$

Observing that

$$Q_t f(x) = \min \left\{ \inf_{y \in \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x - y|^2 \right\}, \inf_{y \in \mathbb{R}^n \setminus \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x - y|^2 \right\} \right\},$$

we get

$$Q_t f(x) = \inf_{y \in \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x-y|^2 \right\}.$$

At this point, using the positivity of distance,

$$Q_t f(x) = \inf_{y \in \overline{\mathrm{B}(x,R_t)}} \left\{ f(y) + \frac{1}{2t}|x-y|^2 \right\} \geq \inf_{y \in \overline{\mathrm{B}(x,R_t)}} f(y),$$

so

$$\sup_{t \geq 0} Q_t f(x) \geq \sup_{t \geq 0} \inf_{y \in \overline{\mathrm{B}(x,R_t)}} f(y).$$

Now, if $0 < t_1 \leq t_2$,

$$\inf_{y \in \overline{\mathrm{B}(x,R_{t_2})}} f(y) \leq \inf_{y \in \overline{\mathrm{B}(x,R_{t_1})}} f(y)$$

and taken for every $t > 0$, $x_t \in \overline{\mathrm{B}(x,R_t)}$ such that

$$f(x_t) \leq \inf_{\overline{\mathrm{B}(x,R_t)}} f + t,$$

by the fact that for $t \to 0$, $R_t \to 0$ so $x_t \to x$, then, using the fact that $f \in C_b(\mathbb{R}^n)$, $f(x_t) \to f(x)$ so

$$\sup_{t \geq 0} \inf_{\overline{\mathrm{B}(x,R_t)}} f \geq f(x).$$

By the fact that the converse inequality is obvious,

$$\sup_{t \geq 0} \inf_{\overline{\mathrm{B}(x,R_t)}} f = f(x).$$

In particular,

$$\sup_{t \geq 0} Q_t f(x) = f(x).$$

Consider now f also uniformly continuous. Fixed $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\forall\, x, y \in \mathbb{R}^n : |x-y| < \delta \Longrightarrow |f(x) - f(y)| < \frac{\varepsilon}{2} \Longrightarrow f(y) > f(x) - \frac{\varepsilon}{2}.$$

Consider $\bar{t} \geq 0$ such that $R_{\bar{t}} < \delta$. For every $t \leq \bar{t}$ we have $R_t < R_{\bar{t}} < \delta$ and

$$\forall\, x, y \in \mathbb{R}^n : |x-y| \leq R_t \Longrightarrow f(y) > f(x) - \frac{\varepsilon}{2}.$$

In particular, for every $x \in \mathbb{R}^n$

$$f(x) + \frac{\varepsilon}{2} \geq f(x) \geq Q_t f(x) \geq \inf_{y \in \overline{B(x,R_t)}} f(y) \geq f(x) - \frac{\varepsilon}{2}$$

or in other words

$$|Q_t f(x) - f(x)| \leq \frac{\varepsilon}{2} < \varepsilon$$

so $Q_t f \to f$ uniformly on $\mathbb{R}^n$.

(c) Take $t_1, t_2 \in [\varepsilon, +\infty[$ and $x_1, x_2 \in \mathbb{R}^n$. Recalling what we proved in (b),

$$Q_{t_1} f(x_1) = \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ f(y) + \frac{1}{2t_1} |x_1 - y|^2 \right\}$$

and, from the definition of Hopf–Lax semigroup,

$$Q_{t_2} f(x_2) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2t_2} |x_2 - y|^2 \right\} \leq \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ f(y) + \frac{1}{2t_2} |x_2 - y|^2 \right\},$$

so, observing that

$$\inf f = \inf \{g + (f - g)\} \geq \inf g + \inf \{f - g\},$$

then

$$\inf f - \inf g \geq \inf \{f - g\},$$

from

$$\begin{aligned} Q_{t_1} f(x_1) - Q_{t_2} f(x_2) \geq & \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ f(y) + \frac{1}{2t_1} |x_1 - y|^2 \right\} + \\ & - \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ f(y) + \frac{1}{2t_2} |x_2 - y|^2 \right\}, \end{aligned}$$

we have

$$\begin{aligned} Q_{t_1} f(x_1) - Q_{t_2} f(x_2) \geq & \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ f(y) + \frac{1}{2t_1} |x_1 - y|^2 - f(y) - \frac{1}{2t_2} |x_2 - y|^2 \right\} = \\ = & \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ \frac{1}{2t_1} |x_1 - y|^2 - \frac{1}{2t_2} |x_2 - y|^2 \right\}. \end{aligned}$$

Resuming,

$$Q_{t_1} f(x_1) - Q_{t_2} f(x_2) \geq \inf_{y \in \mathrm{B}(x_1, R_{t_1})} \left\{ \frac{1}{2t_1} |x_1 - y|^2 - \frac{1}{2t_2} |x_2 - y|^2 \right\}. \tag{1.2}$$

Now, if we write (1.2) with $x_1 = x_2 = x$,

$$Q_{t_1} f(x) - Q_{t_2} f(x) \geq \inf_{y \in \mathrm{B}(x, R_{t_1})} \left\{ \left(\frac{1}{2t_1} - \frac{1}{2t_2} \right) |x - y|^2 \right\}$$

so

$$Q_{t_1}f(x) - Q_{t_2}f(x) \geq \inf_{y\in \mathrm{B}(x,R_{t_1})} \left\{ -\left|\frac{1}{2t_1} - \frac{1}{2t_2}\right| |x-y|^2 \right\},$$

then

$$\begin{aligned} Q_{t_1}f(x) - Q_{t_2}f(x) &\geq \left|\frac{1}{2t_1} - \frac{1}{2t_2}\right| \inf_{y\in \mathrm{B}(x,R_{t_1})} \left\{ -|x-y|^2 \right\} = \\ &= -\left|\frac{1}{2t_1} - \frac{1}{2t_2}\right| \sup_{y\in \mathrm{B}(x,R_{t_1})} |x-y|^2, \end{aligned}$$

therefore

$$\begin{aligned} Q_{t_1}f(x) - Q_{t_2}f(x) &\geq -\left|\frac{1}{2t_1} - \frac{1}{2t_2}\right| R_{t_1}^2 = -\left|\frac{1}{2t_1} - \frac{1}{2t_2}\right| 2t_1(\sup f - \inf f) = \\ &= -\left|\frac{1}{t_1} - \frac{1}{t_2}\right| t_1(\sup f - \inf f). \end{aligned}$$

Analogously,

$$Q_{t_2}f(x) - Q_{t_1}f(x) \geq -\left|\frac{1}{t_1} - \frac{1}{t_2}\right| t_2(\sup f - \inf f)$$

or, in other words,

$$Q_{t_1}f(x) - Q_{t_2}f(x) \leq \left|\frac{1}{t_1} - \frac{1}{t_2}\right| t_2(\sup f - \inf f),$$

so

$$-\left|\frac{1}{t_1} - \frac{1}{t_2}\right| t_1(\sup f - \inf f) \leq Q_{t_1}f(x) - Q_{t_2}f(x) \leq \left|\frac{1}{t_1} - \frac{1}{t_2}\right| t_2(\sup f - \inf f),$$

then

$$\begin{aligned} |Q_{t_1}f(x) - Q_{t_2}f(x)| &\leq \left|\frac{1}{t_1} - \frac{1}{t_2}\right| \max\{t_1, t_2\} (\sup f - \inf f) = \\ &= \frac{\sup f - \inf f}{\min\{t_1, t_2\}} |t_1 - t_2| \leq \frac{\sup f - \inf f}{\varepsilon} |t_1 - t_2|. \end{aligned}$$

Writing instead (1.2) with $t_1 = t_2 = t$, using the triangle inequality,

$$\begin{aligned} Q_t f(x_1) - Q_t f(x_2) &\geq \inf_{y\in \mathrm{B}(x_1,R_t)} \left\{ \frac{1}{2t} \left(|x_1-y|^2 - |x_2-y|^2 \right) \right\} = \\ &= \frac{1}{2t} \inf_{y\in \mathrm{B}(x_1,R_t)} \left\{ \left(|x_1-y| - |x_2-y|\right)\left(|x_1-y| + |x_2-y|\right) \right\} \\ &\geq \frac{1}{2t} \inf_{y\in \mathrm{B}(x_1,R_t)} \left\{ -|x_1-x_2| \left(2|x_1-y| + |x_1-x_2|\right) \right\} \geq \end{aligned}$$

$$\geq -\frac{1}{2t} \sup_{y \in \overline{\mathrm{B}(x_1,R_t)}} \left\{ 2|x_1 - y||x_1 - x_2| + |x_1 - x_2|^2 \right\}$$

so

$$\begin{aligned} Q_t f(x_1) - Q_t f(x_2) &\geq -\frac{R_t}{t}|x_1 - x_2| - \frac{1}{2t}|x_1 - x_2|^2 = \\ &= -\frac{\sqrt{2(\sup f - \inf f)}}{\sqrt{t}}|x_1 - x_2| - \frac{1}{2t}|x_1 - x_2|^2. \end{aligned}$$

Analogously,

$$Q_t f(x_2) - Q_t f(x_1) \geq -\frac{\sqrt{2(\sup f - \inf f)}}{\sqrt{t}}|x_1 - x_2| - \frac{1}{2t}|x_1 - x_2|^2.$$

or, in other words,

$$Q_t f(x_1) - Q_t f(x_2) \leq \frac{\sqrt{2(\sup f - \inf f)}}{\sqrt{t}}|x_1 - x_2| + \frac{1}{2t}|x_1 - x_2|^2,$$

so

$$\begin{aligned} |Q_t f(x_1) - Q_t f(x_2)| &\leq \frac{\sqrt{2(\sup f - \inf f)}}{\sqrt{\varepsilon}}|x_1 - x_2| + \frac{1}{2\varepsilon}|x_1 - x_2|^2 \leq \\ &\leq \frac{\sqrt{2(\sup f - \inf f)}}{\sqrt{\varepsilon}}|x_1 - x_2| + \frac{1}{\varepsilon}|x_1 - x_2|^2. \end{aligned}$$

Now, for every $\xi \in \mathbb{R}^n$ we have

$$\begin{aligned} &\sup_{\substack{x_1,x_2 \in \mathbb{R}^n \\ x_1 \neq x_2}} \frac{|Q_t f(x_1) - Q_t f(x_2)|}{|x_1 - x_2|} = \\ &= \max \left\{ \sup_{\substack{x_1,x_2 \in \overline{\mathrm{B}(\xi,R_\varepsilon)} \\ x_1 \neq x_2}} \frac{|Q_t f(x_1) - Q_t f(x_2)|}{|x_1 - x_2|}, \sup_{\substack{x_1,x_2 \in \mathbb{R}^n \setminus \overline{\mathrm{B}(\xi,R_\varepsilon)} \\ x_1 \neq x_2}} \frac{|Q_t f(x_1) - Q_t f(x_2)|}{|x_1 - x_2|} \right\}, \end{aligned}$$

so

$$\begin{aligned} \sup_{\substack{x_1,x_2 \in \mathbb{R}^n \\ x_1 \neq x_2}} \frac{|Q_t f(x_1) - Q_t f(x_2)|}{|x_1 - x_2|} &\leq \max \left\{ \frac{2\sqrt{2(\sup f - \inf f)}}{\sqrt{\varepsilon}}, \frac{\sqrt{2}\|Q_t f\|_\infty}{\sqrt{\varepsilon(\sup f - \inf f)}} \right\} \leq \\ &\leq \max \left\{ \frac{2\sqrt{2(\sup f - \inf f)}}{\sqrt{\varepsilon}}, \frac{\sqrt{2}\|f\|_\infty}{\sqrt{\varepsilon(\sup f - \inf f)}} \right\} < +\infty. \end{aligned}$$

Finally,

$$|Q_{t_1}f(x_1) - Q_{t_2}f(x_2)| \leq |Q_{t_1}f(x_1) - Q_{t_2}f(x_1)| + |Q_{t_2}f(x_1) - Q_{t_2}f(x_2)| \leq$$
$$\leq \frac{\sup f - \inf f}{\varepsilon}|t_1 - t_2| + \max\left\{\frac{2\sqrt{2(\sup f - \inf f)}}{\sqrt{\varepsilon}}, \frac{\sqrt{2}\|f\|_\infty}{\sqrt{\varepsilon(\sup f - \inf f)}}\right\}|x_1 - x_2|,$$

then, set

$$C_\varepsilon(f) = \frac{\sup f - \inf f}{\varepsilon} + \max\left\{\frac{2\sqrt{2(\sup f - \inf f)}}{\sqrt{\varepsilon}}, \frac{\sqrt{2}\|f\|_\infty}{\sqrt{\varepsilon(\sup f - \inf f)}}\right\},$$

we obtain

$$|Q_{t_1}f(x_1) - Q_{t_2}f(x_2)| \leq C_\varepsilon(f)\left(|t_1 - t_2| + |x_1 - x_2|\right) \leq \sqrt{2}C_\varepsilon(f)\sqrt{(t_1 - t_2)^2 + (x_1 - x_2)^2},$$

therefore $Q_t f$ is Lipschitz on $[\varepsilon, +\infty[\times \mathbb{R}^n$.

(d) Using (c) and Rademacher's Theorem (see, for example, [24]), for a.e. $x \in \mathbb{R}^n$, for a.e. $t \in]0, +\infty[$, the function $\{y \longmapsto Q_t f(y)\}$ is differentiable at x. Let us also observe that $\mathbb{R}^n$ is a reflexive Banach space and the function $\left\{y \longmapsto f(y) + \frac{1}{2t}|x - y|^2\right\}$ is proper and weakly lower semicontinuous, being continuous and bounded. Furthermore, it is also coercive: indeed, using the triangle inequality of the norm,

$$f(y) + \frac{1}{2t}|x - y|^2 \geq \frac{1}{2t}|y|^2 - \frac{|x|}{t}|y| + (\inf f + |x|) \geq A|y| + B,$$

for some $A > 0$, $B \in \mathbb{R}$. In particular, the direct method of the Calculus of Variations gives us the existence of a minimizer $J_t(x)$. By [1, Lecture 14, Section 3], we know that for every $x \in \mathbb{R}^n$, for a.e. $t \in]0, +\infty[$

$$\frac{d}{dt}Q_t f(x) = -\frac{|x - J_t(x)|^2}{2t^2}.$$

Let us prove that for every differentiability point x of the function $\{y \longmapsto Q_t f(y)\}$ it holds

$$D(Q_t f)(x) = \frac{x - J_t(x)}{t}.$$

Let us start observing that

$$-Q_t f(x) + \frac{1}{2t}|x|^2 = -\inf_{y \in \mathbb{R}^n}\left\{f(y) + \frac{1}{2t}|x - y|^2\right\} + \frac{1}{2t}|x|^2 =$$

$$= \sup_{y \in \mathbb{R}^n} \left\{ -f(y) - \frac{1}{2t}|x-y|^2 \right\} + \frac{1}{2t}|x|^2 =$$
$$= \sup_{y \in \mathbb{R}^n} \left\{ -f(y) - \frac{1}{2t}|x-y|^2 + \frac{1}{2t}|x|^2 \right\} =$$
$$= \sup_{y \in \mathbb{R}^n} \left\{ \frac{1}{t} y \cdot x - \left(f(y) + \frac{1}{2t}|y|^2 \right) \right\}$$

so, being the supremum of a family of affine functions, $\left\{ x \longmapsto -Q_t f(x) + \frac{1}{2t}|x|^2 \right\}$ is a convex function. In other words, $-Q_t f$ is $-\frac{1}{t}$-convex. By the fact that

$$\begin{aligned} Q_t f(y) - Q_t f(x) &\leq f(J_t(x)) + \frac{1}{2t}|y - J_t(x)|^2 - \left(f(J_t(x)) + \frac{1}{2t}|x - J_t(x)|^2 \right) = \\ &= \frac{1}{2t}\left((y - J_t(x)) \cdot (y - J_t(x)) + |x - J_t(x)|^2 \right) = \\ &= \frac{1}{2t}\Big[((y-x) + (x - J_t(x))) \cdot ((y-x) + (x - J_t(x))) + \\ &+ |x - J_t(x)|^2 \Big] = \\ &= \frac{1}{2t}\left(|y-x|^2 + 2(x - J_t(x)) \cdot (y-x) \right) = \\ &= \frac{x - J_t(x)}{t} \cdot (y-x) + \frac{1}{2t}|y-x|^2, \end{aligned}$$

we have that

$$-Q_t f(y) - (-Q_t f(x)) \geq -\frac{x - J_t(x)}{t} \cdot (y-x) - \frac{1}{2t}|y-x|^2,$$

so

$$-\frac{x - J_t(x)}{t} \in \partial_{-\frac{1}{t}}(-Q_t f)(x).$$

Being $Q_t f$ differentiable at x, we obtain

$$D(-Q_t f)(x) = -\frac{x - J_t(x)}{t}$$

then, by linearity,

$$D(Q_t f)(x) = \frac{x - J_t(x)}{t}.$$

In particular,

$$\frac{d}{dt} Q_t f(x) = -\frac{|x - J_t(x)|^2}{2t^2} = -\frac{1}{2}|D(Q_t f)(x)|^2,$$

so the result follows.

(e) First of all, for every $x \in \mathbb{R}^n$ and $t > 0$, given $R_t = 2t\mathrm{Lip}(f)$, consider $\overline{\mathrm{B}(x, R_t)}$. We have

$$\inf_{y\in \mathrm{B}(x,R_t)} \left\{f(y)+\frac{1}{2t}|x-y|^2\right\} \le f(x)$$

and, by the fact that f is Lipschitz, for every $y \in \mathbb{R}^n \setminus \overline{\mathrm{B}(x,R_t)}$ we have

$$\begin{aligned} f(y)+\frac{1}{2t}|x-y|^2 = f(y)+\frac{1}{2t}|x-y||x-y| > \\ > f(y)+\frac{1}{2t}R_t|x-y| = f(y)+\mathrm{Lip}(f)|x-y| \ge \\ \ge f(y)+f(x)-f(y)=f(x) \end{aligned}$$

so

$$\inf_{y\in \mathbb{R}^n\setminus \overline{\mathrm{B}(x,R_t)}} \left\{f(y)+\frac{1}{2t}|x-y|^2\right\} \ge f(x) \ge \inf_{y\in \mathrm{B}(x,R_t)} \left\{f(y)+\frac{1}{2t}|x-y|^2\right\}.$$

Observing that

$$Q_tf(x)=\min\left\{\inf_{y\in \mathrm{B}(x,R_t)} \left\{f(y)+\frac{1}{2t}|x-y|^2\right\}, \inf_{y\in \mathbb{R}^n\setminus \overline{\mathrm{B}(x,R_t)}} \left\{f(y)+\frac{1}{2t}|x-y|^2\right\}\right\},$$

we get

$$Q_tf(x)=\inf_{y\in \mathrm{B}(x,R_t)} \left\{f(y)+\frac{1}{2t}|x-y|^2\right\}. \tag{1.3}$$

Consider $t_1,t_2 \in [0,+\infty[$ and $x_1,x_2\in\mathbb{R}^n$. If $t_1,t_2>0$, in a similar way to what we have done in (c), we can prove that

$$Q_{t_1}f(x_1)-Q_{t_2}f(x_2) \ge \inf_{y\in \mathrm{B}(x_1,R_{t_1})} \left\{\frac{1}{2t_1}|x_1-y|^2-\frac{1}{2t_2}|x_2-y|^2\right\}. \tag{1.4}$$

Writing (1.4) with $t_1=t_2=t$ and using the triangle inequality,

$$\begin{aligned} Q_tf(x_1)-Q_tf(x_2) &\ge \inf_{y\in \mathrm{B}(x_1,R_t)} \left\{\frac{1}{2t}\left(|x_1-y|^2-|x_2-y|^2\right)\right\} = \\ &= \frac{1}{2t}\inf_{y\in \mathrm{B}(x_1,R_t)} \left\{(|x_1-y|-|x_2-y|)\,(|x_1-y|+|x_2-y|)\right\} \\ &\ge \frac{1}{2t}\inf_{y\in \mathrm{B}(x_1,R_t)} \left\{-|x_1-x_2|\,(2|x_1-y|+|x_1-x_2|)\right\} \ge \\ &\ge -\frac{1}{2t}\sup_{y\in \mathrm{B}(x_1,R_t)} \left\{2|x_1-y||x_1-x_2|+|x_1-x_2|^2\right\} \end{aligned}$$

so

$$Q_tf(x_1) - Q_tf(x_2) \geq -\frac{R_t}{t}|x_1 - x_2| - \frac{1}{2t}|x_1 - x_2|^2 =$$
$$= -2\mathrm{Lip}(f)|x_1 - x_2| - \frac{1}{2t}|x_1 - x_2|^2.$$

Analogously,

$$Q_tf(x_2) - Q_tf(x_1) \geq -2\mathrm{Lip}(f)|x_1 - x_2| - \frac{1}{2t}|x_1 - x_2|^2$$

or, in other words,

$$Q_tf(x_1) - Q_tf(x_2) \leq 2\mathrm{Lip}(f)|x_1 - x_2| + \frac{1}{2t}|x_1 - x_2|^2,$$

so

$$|Q_tf(x_1) - Q_tf(x_2)| \leq 2\mathrm{Lip}(f)|x_1 - x_2| + \frac{1}{2t}|x_1 - x_2|^2.$$

In particular, if $|x_1 - x_2| \leq t$,

$$|Q_tf(x_1) - Q_tf(x_2)| \leq \left(2\mathrm{Lip}(f) + \frac{1}{2}\right)|x_1 - x_2|.$$

Conversely, if we are in the case $|x_1 - x_2| \geq t$, consider (y_h) in $\mathbb{R}^n$ such that

$$Q_tf(x_1) \geq f(y_h) + \frac{1}{2t}|x_1 - y_h|^2 - \frac{1}{h+1}.$$

Using the Lipschitzianity of f, $f(y_h) - f(x_1) \geq -\mathrm{Lip}(f)|x_1 - y_h|$, so

$$Q_tf(x_1) \geq f(x_1) - \mathrm{Lip}(f)|x_1 - y_h| + \frac{1}{2t}|x_1 - y_h|^2 - \frac{1}{h+1}.$$

Expanding the trivial inequality $\frac{1}{2t}\left(\mathrm{Lip}(f) - |x_1 - y_h|\right)^2 \geq 0$, we obtain

$$-L|x_1 - y_h| + \frac{1}{2t}|x_1 - y_h|^2 + \frac{\mathrm{Lip}^2(f)t}{2} \geq 0,$$

so

$$Q_tf(x_1) \geq f(x_1) - \frac{\mathrm{Lip}^2(f)t}{2} - \frac{1}{h+1}$$

and passing to the limit as $h \to +\infty$

$$Q_tf(x_1) \geq f(x_1) - \frac{\mathrm{Lip}^2(f)t}{2},$$

then, by the fact that $Q_t f(x_2) \leq f(x_2)$,

$$Q_t f(x_1) - Q_t f(x_2) \geq f(x_1) - \frac{\mathrm{Lip}^2(f)t}{2} - f(x_2) \geq -\mathrm{Lip}(f)|x_1 - x_2| - \frac{\mathrm{Lip}^2(f)t}{2}.$$

Analogously,

$$Q_t f(x_2) - Q_t f(x_1) \geq -\mathrm{Lip}(f)|x_1 - x_2| - \frac{\mathrm{Lip}^2(f)t}{2}$$

or, in other words,

$$Q_t f(x_1) - Q_t f(x_2) \leq \mathrm{Lip}(f)|x_1 - x_2| + \frac{\mathrm{Lip}^2(f)t}{2},$$

so

$$|Q_t f(x_1) - Q_t f(x_2)| \leq \mathrm{Lip}(f)|x_1 - x_2| + \frac{\mathrm{Lip}^2(f)t}{2} \leq \frac{1}{2}\mathrm{Lip}(f)\left(1 + \mathrm{Lip}(f)\right)|x_1 - x_2|.$$

Resuming,

$$|Q_t f(x_1) - Q_t f(x_2)| \leq \max\left\{\frac{1}{2}\mathrm{Lip}(f)\left(1 + \mathrm{Lip}(f)\right), 2\mathrm{Lip}(f) + \frac{1}{2}\right\}|x_1 - x_2|, \tag{1.5}$$

so, fixed $t > 0$, the function $\{x \longmapsto Q_t f(x)\}$ is Lipschitz. Now, if $t_1 \leq t_2$, by (c), the function $Q_t f$ is Lipschitz on $[t_1, +\infty[\times \mathbb{R}^n$ so, in particular, on $[t_1, t_2] \times \mathbb{R}^n$. Therefore, for every $x \in \mathbb{R}^n$

$$|Q_{t_1} f(x) - Q_{t_2} f(x)| \leq \int_{t_1}^{t_2} \left|\frac{d}{dt} Q_t(x)\right| dt.$$

By (d), there exists (x_h) in $\mathbb{R}^n$ such that $x_h \to x$ and for a.e. $t > 0$ and every $h \in \mathbb{N}$

$$\frac{d}{dt} Q_t f(x_h) + \frac{1}{2}|D\left(Q_t f\right)(x_h)|^2 = 0,$$

so

$$|Q_{t_1} f(x_h) - Q_{t_2} f(x_h)| \leq \frac{1}{2}\int_{t_1}^{t_2} |D\left(Q_t f\right)(x_h)|^2 dt$$

and using (1.5),

$$|Q_{t_1} f(x_h) - Q_{t_2} f(x_h)| \leq \frac{1}{2}\left(\max\left\{\frac{1}{2}\mathrm{Lip}(f)\left(1 + \mathrm{Lip}(f)\right), 2\mathrm{Lip}(f) + \frac{1}{2}\right\}\right)^2 |t_1 - t_2|.$$

Using again the fact that $\{x \longmapsto Q_t f(x)\}$ is Lipschitz, we can pass to the limit and obtain

$$|Q_{t_1}f(x) - Q_{t_2}f(x)| \le \frac{1}{2}\left(\max\left\{\frac{1}{2}\mathrm{Lip}(f)\left(1+\mathrm{Lip}(f)\right), 2\mathrm{Lip}(f)+\frac{1}{2}\right\}\right)^2 |t_1 - t_2|. \tag{1.6}$$

The case $t_2 \le t_1$ is analogous. Using now the triangle inequality, (1.5) and (1.6), we obtain

$$\begin{aligned} |Q_{t_1}f(x_1) - Q_{t_2}f(x_2)| \le |Q_{t_1}f(x_1) - Q_{t_2}f(x_1)| + |Q_{t_2}f(x_1) - Q_{t_2}f(x_2)| \le \\ \sqrt{2}\max\left\{\frac{1}{2}\mathrm{Lip}(f)\left(1+\mathrm{Lip}(f)\right), 2\mathrm{Lip}(f)+\frac{1}{2}\right\} \\ \left(1+\max\left\{\frac{1}{2}\mathrm{Lip}(f)\left(1+\mathrm{Lip}(f)\right), 2\mathrm{Lip}(f)+\frac{1}{2}\right\}\right) \\ \sqrt{|x_1-x_2|^2+|t_1-t_2|^2}, \end{aligned}$$

so $\{(t,x) \longmapsto Q_tf(x)\}$ is Lipschitz on $]0,+\infty[\times\mathbb{R}^n$. Furthermore, there exists a unique Lipschitz extension to $]0,+\infty[\times\mathbb{R}^n$ and, by (b) it coincides with $Q_0f(x)$ so the global Lipschitzianity follows.

(f) Suppose, initially, that $f \in \mathrm{Lip}_b(\mathbb{R}^n)$. For every $(t,x) \in\,]0,+\infty[\times\mathbb{R}^n$, consider $x_t \in \mathbb{R}^n$ such that

$$Q_tf(x) = f(x_t) + \frac{1}{2t}|x-x_t|^2$$

and such that for any other minimizer ξ of $Q_tf(x)$,

$$|x-\xi|^2 \le |x-x_t|.$$

For every $t, h > 0$

$$\begin{aligned} Q_{t+h}f(x) - Q_tf(x) \le f(x_t) + \frac{1}{2(t+h)}|x-x_t|^2 - \left(f(x_t)+\frac{1}{2t}|x-x_t|^2\right) \le \\ \le -\frac{h}{2t(t+h)}|x-x_t|^2, \end{aligned}$$

so

$$\frac{Q_{t+h}f(x)-Q_tf(x)}{h} \le -\frac{|x-x_t|^2}{2t(t+h)}$$

and for $h \to 0^+$

$$\frac{d^+}{dt}Q_tf(x) \le -\frac{|x-x_t|^2}{2t^2}.$$

In particular, by [2, Proposition 3.2], we get that the function $\{(t,x) \longmapsto |x-x_t|\}$ is upper semicontinuous. Then, considering y_t such that

$$Q_t f(y) = f(y_t) + \frac{1}{2t}|y - y_t|,$$

we have

$$\begin{aligned}\lim_{r\to 0} \sup_{y,z\in \mathrm{B}(x,r)} \frac{Q_t f(z) - Q_t f(y)}{|y-z|} &\leq \lim_{r\to 0} \sup_{y,z\in \mathrm{B}(x,r)} \frac{1}{2t}\frac{|z-y_t|^2 - |y-y_t|^2}{|y-z|} = \\ &= \lim_{r\to 0} \sup_{y,z\in \mathrm{B}(x,r)} \frac{1}{2t}\frac{(|z-y_t| - |y-y_t|)\,(|z-y_t| + |y-y_t|)}{|y-z|}\end{aligned}$$

so

$$\begin{aligned}\lim_{r\to 0} \sup_{y,z\in \mathrm{B}(x,r)} \frac{Q_t f(z) - Q_t f(y)}{|y-z|} &\leq \lim_{r\to 0} \sup_{y,z\in \mathrm{B}(x,r)} \frac{1}{2t}(|z-y_t| + |y-y_t|) \leq \\ &\leq \lim_{r\to 0} \sup_{y,z\in \mathrm{B}(x,r)} \frac{1}{2t}(|z-y| + 2|y-y_t|) \leq \\ &\leq \lim_{r\to 0} \sup_{y\in \mathrm{B}(x,r)} \frac{1}{2t}(2r + 2|y-y_t|) \leq \\ &\leq \frac{1}{t} \lim_{r\to 0} \sup_{y\in \mathrm{B}(x,r)} |y-y_t| \leq \frac{|x-x_t|}{t}.\end{aligned}$$

Resuming,

$$|D^*(Q_t f)|(x) \leq \frac{|x-x_t|}{t}$$

and, in particular,

$$\frac{d^+}{dt}Q_t f(x) + \frac{1}{2}|D^*\,(Q_t f)|^2(x) \leq -\frac{|x-x_t|^2}{2t^2} + \frac{|x-x_t|^2}{2t^2} = 0.$$

The general case is similar up to change minimizers with minimizing sequences.

(g) If at least one between s, t is equal to 0, it is trivial. We then consider only the case $s, t > 0$. Let us start proving that for every $s, t \geq 0$,

$$Q_{s+t} f \leq (Q_s \circ Q_t) f.$$

For every $x, y, z \in \mathbb{R}^n$, by Proposition (6.6) and the triangle inequality,

$$f(y) + \frac{1}{2(s+t)}|x-y|^2 \leq f(y) + \frac{1}{2t}|y-z|^2 + \frac{1}{2s}|x-z|^2.$$

Passing to the infimum, with respect to y, both sides we obtain

$$Q_{s+t}f(x) \leq \inf_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2t}|y-z|^2 + \frac{1}{2s}|x-z|^2 \right\} \leq$$
$$\leq \inf_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2t}|y-z|^2 \right\} + \frac{1}{2s}|x-z|^2$$

and if now we also pass to the infimum, with respect to z, we get

$$Q_{s+t}f(x) \leq \inf_{z\in\mathbb{R}^n} \left\{ \inf_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2t}|y-z|^2 \right\} + \frac{1}{2s}|x-z|^2 \right\} = (Q_s \circ Q_t)f(x).$$

Now, let us prove that for every $s, t > 0$ and for every $x, z \in \mathbb{R}^n$, it holds

$$\inf_{y\in\mathbb{R}^n} \left\{ \frac{1}{2s}|x-y|^2 + \frac{1}{2t}|y-z|^2 \right\} \leq \frac{1}{2(t+s)}|x-z|^2.$$

On the segment connecting x and z, consider $\overline{y}$ such that

$$|x-\overline{y}| = \frac{t}{s}|z-\overline{y}|.$$

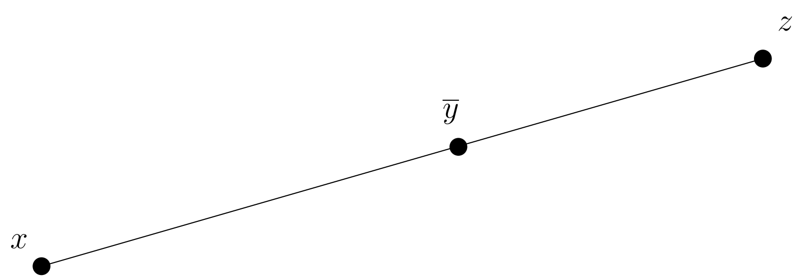

Figure A.1: A point $\overline{y}$ on the segment connecting x and z.

In particular, Proposition (6.6) holds an equality, then

$$\frac{|x-\overline{y}|}{2s} + \frac{|z-\overline{y}|}{2t} = \frac{(|x-\overline{y}| + |z-\overline{y}|)^2}{2(t+s)} = \frac{|x-z|^2}{2(t+s)}.$$

Passing to the infimum, with respect to $y \in \mathbb{R}^n$, both sides we obtain

$$\inf_{y\in\mathbb{R}^n} \left\{ \frac{1}{2s}|x-y|^2 + \frac{1}{2t}|y-z|^2 \right\} \leq \frac{|x-z|^2}{2(t+s)}.$$

Consider now (y_h) in $\mathbb{R}^n$ such that

$$\frac{1}{2s}|x-y_h|^2+\frac{1}{2t}|y_h-z|^2<\inf_{y\in\mathbb{R}^n}\left\{\frac{1}{2s}|x-y|^2+\frac{1}{2t}|y-z|^2\right\}+\frac{1}{h+1}\leq$$
$$\leq\frac{1}{2(t+s)}|x-z|^2+\frac{1}{h+1}.$$

Adding up both sides $f(z)$ and passing to the infimum, with respect to z, we get

$$\inf_{z\in\mathbb{R}^n}\left\{f(z)+\frac{1}{2t}|y_h-z|^2\right\}+\frac{1}{2s}|x-y_h|^2\leq Q_{s+t}f(x)+\frac{1}{h+1}$$

so, passing to the infimum, with respect to y, we arrive at

$$(Q_s\circ Q_t)f(x)\leq Q_{s+t}f(x)+\frac{1}{h+1}.$$

Letting $h\to+\infty$,

$$(Q_s\circ Q_t)f(x)\leq Q_{s+t}f(x),$$

then

$$Q_{s+t}f=(Q_s\circ Q_t)f$$

and the proof is concluded. ∎

Acknowledgements

The first person I would like to thank is my supervisor, *Dr. Luca Tamanini*: I am truly grateful to him for his great expertise, his support in moments of indecision, the patience with which he explained to me, more than once, what was not clear to me and the time he dedicated to me, also for our interminable meetings.

Secondly, I would like to thank my co-supervisor, *Dr. Nicolò De Ponti*, for his availability for discussion and collaboration, even in the future, and, above all, for having made himself open to reviewing the thesis.

In addition, I would like to thank all those faculty members in the *Facoltà di Scienze Matematiche, Fisiche e Naturali* who have always been available to clarify doubts and correct errors, or even just provide insights to make this work as complete as possible.

I cannot then escape from thanking my *family members* for their support: I am grateful to those who did not stand in my way but rather facilitated the completion of this work. In particular, I would like to thank my brother, *Lorenzo*, for his help, not always voluntary, regarding my doubts about the English language.

Finally, I would like to thank all my *friends*. Whether near or far, at the time when I needed advice, or even just to complain, you were an invaluable support. In particular, I would like to thank you for your patience in listening to my doubts about stylistic choices and grammatical conventions, even when they were not always shared.

GRAZIE!

July 15^{th}, 2025

Mattia Garatti

Bibliography

[1] Luigi Ambrosio, Elia Brué, and Daniele Semola. *Lectures on Optimal Transport*. Cham: Springer, 2021.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below". In: *Inventiones mathematicae* 195.2 (2014), pp. 289–391.

[3] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[4] Shrey Aryan and Promit Ghosal. "Entropic Selection Principle for Monge's Optimal Transport". In: *arXiv preprint arXiv:2502.16370* (2025).

[5] Carlo Banfi. *Introduzione alla meccanica dei continui*. CEDAM, Padova, 1990.

[6] Aymeric Baradat and Hugo Lavenant. "Regularized unbalanced optimal transport as entropy minimization with respect to branching brownian motion". In: *Astérisque* 458 (2025).

[7] Martino Bardi and Italo Capuzzo Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Vol. 12. Springer, 1997.

[8] Yann Brenier. "Décomposition polaire et réarrangement monotone des champs de vecteurs". In: *CR Acad. Sci. Paris Sér. I Math.* 305 (1987), pp. 805–808.

[9] Haïm Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. 2011.

[10] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. "Optimal-transport formulation of electronic density-functional theory". In: *Physical Review A—Atomic, Molecular, and Optical Physics* 85.6 (2012), p. 062502.

[11] Maciej Buze and Manh Hong Duong. "Entropic regularisation of unbalanced optimal transportation problems". In: *arXiv preprint:2305.02410* (2023).

[12] Guillaume Carlier, Victor Chernozhukov, and Alfred Galichon. “Vector quantile regression: an optimal transport approach”. In: *The Annals of Statistics* 44.3 (2016).

[13] Guillaume Carlier and Ivar Ekeland. “Matching for teams”. In: *Economic theory* 42 (2010), pp. 397–418.

[14] Guillaume Carlier, Paul Pegon, and Luca Tamanini. “Convergence rate of general entropic optimal transport costs”. In: *Calculus of Variations and Partial Differential Equations* 62.4 (2023), p. 116.

[15] Guillaume Carlier et al. “Convergence of entropic schemes for optimal transport and gradient flows”. In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418.

[16] Lenaic Chizat. “Unbalanced optimal transport: Models, numerical methods, applications”. PhD thesis. Université Paris sciences et lettres, 2017.

[17] Giovanni Conforti. “A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost”. In: *Probability Theory and Related Fields* 174.1 (2019), pp. 1–47.

[18] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. “Density functional theory and optimal transportation with Coulomb cost”. In: *Communications on Pure and Applied Mathematics* 66.4 (2013), pp. 548–599.

[19] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential, vol. 29 of North-Holland Mathematics Studies*. 1978.

[20] Daniela Di Donato. “The intrinsic Hopf-Lax semigroup vs. the intrinsic slope”. In: *Journal of Mathematical Analysis and Applications* 523.2 (2023), p. 127051.

[21] Simone Di Marino and Jean Louet. “The entropic regularization of the Monge problem on the real line”. In: *SIAM Journal on Mathematical Analysis* 50.4 (2018), pp. 3451–3477.

[22] Italo Capuzzo Dolcetta and Federica Dragoni. “Equazioni di Hamilton-Jacobi”. In: *Lecture Notes* (2004).

[23] D.C. Dowson and B.V. Landau. “The Fréchet distance between multivariate normal distributions”. In: *Journal of multivariate analysis* 12.3 (1982), pp. 450–455.

[24] Lawrence Craig Evans. *Partial differential equations*. Vol. 19. American Mathematical Society, 2022.

[25] Xiang Gao, Emilio Gallicchio, and Adrian Roitberg. “The generalized Boltzmann distribution is the only distribution in which the Gibbs-Shannon entropy equals the thermodynamic entropy”. In: *The Journal of chemical physics* 151.3 (2019).

[26] Nicola Gigli, Enrico Pasqualetto, et al. *Lectures on nonsmooth differential geometry.* Vol. 2. Springer, 2020.

[27] Nicola Gigli and Luca Tamanini. “Benamou–Brenier and duality formulas for the entropic cost on $RCD^*(K,N)$ spaces”. In: *Probability Theory and Related Fields* 176.1 (2020), pp. 1–34.

[28] Nicola Gigli and Luca Tamanini. “Second order differentiation formula on $RCD^*(K,N)$ spaces”. In: *Journal of the European Mathematical Society* 23.5 (2021), pp. 1727–1795.

[29] Clark Givens and Rae Michael Shortt. “A class of Wasserstein metrics for probability distributions.” In: *Michigan Mathematical Journal* 31.2 (1984), pp. 231–240.

[30] Herbert Goldstein et al. *Classical mechanics.* 2002.

[31] Andreas Greven, Gerhard Keller, and Gerald Warnecke. *Entropy.* Princeton university press, 2014.

[32] Morton Edward Gurtin. *An introduction to continuum mechanics.* Mathematics in Science and Engineering, Academic press, 1982.

[33] Rudolf Emil Kálmán. “The theory of optimal control and the calculus of variations”. In: *Mathematical optimization techniques* 309 (1963), p. 329.

[34] Leonid Kantorovitch. “On the Translocation of Masses”. In: *Management Science* 5.1 (1958), pp. 1–4.

[35] Isabelle Kemajou-Brown. “Brief history of optimal control theory and some recent developments”. In: *Probability on Algebraic And Geometric Structures.* Vol. 668. Contemporary Mathematics, 2016, pp. 119–130.

[36] Donald Evan Kirk. *Optimal control theory: an introduction.* Courier Corporation, 2004.

[37] Martin Knott and Cyril Smith. “On the optimal mapping of distributions”. In: *Journal of Optimization Theory and Applications* 43 (1984), pp. 39–49.

[38] Andrzej Kozlowski. *"Brownian Bridge" Wolfram Demonstrations Project.* 2011. URL: https://demonstrations.wolfram.com/BrownianBridge/.

[39] Christian Léonard. "From the Schrödinger problem to the Monge–Kantorovich problem". In: *Journal of Functional Analysis* 262.4 (2012), pp. 1879–1920.

[40] Christian Léonard. "Minimization of energy functionals applied to some inverse problems". In: *Applied mathematics and optimization* 44 (2001), pp. 273–297.

[41] Christian Léonard. "Some properties of path measures". In: *Séminaire de Probabilités XLVI* (2014), pp. 207–230.

[42] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures". In: *Inventiones mathematicae* 211.3 (2018), pp. 969–1117.

[43] Anton Mallasto, Augusto Gerolin, and Hà Quang Minh. "Entropy-regularized 2-Wasserstein distance between Gaussian measures". In: *Information Geometry* 5.1 (2022), pp. 289–323.

[44] Mauro Mariani. "A Γ-convergence approach to large deviations". In: *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 18.3 (2018), pp. 951–976.

[45] Robert John McCann. "A convexity principle for interacting gases". In: *Advances in mathematics* 128.1 (1997), pp. 153–179.

[46] Centro PRISTEM - Progetto Ricerche Storiche E Metodologiche. *Leonid Kantorovich*. URL: https://matematica.unibocconi.eu/matematici/leonid-kantorovich.

[47] Gaspard Monge. *Mémoire sur la théorie des déblais et des remblais*. Imprimerie royale, 1781.

[48] Luca Nenna and Paul Pegon. "Convergence rate of entropy-regularized multi-marginal optimal transport costs". In: *Canadian Journal of Mathematics* 77.3 (2025), pp. 1072–1092.

[49] Marcel Nutz and Johannes Wiesel. "Entropic optimal transport: Convergence of potentials". In: *Probability Theory and Related Fields* 184.1 (2022), pp. 401–424.

[50] John Joseph O'Connor and Edmund Frederick Robertson. *Erwin Rudolf Josef Alexander Schrödinger*. 2003. URL: https://mathshistory.st-andrews.ac.uk/Biographies/Schrodinger/.

[51] John Joseph O'Connor and Edmund Frederick Robertson. *Gaspard Monge*. 1999. URL: https://mathshistory.st-andrews.ac.uk/Biographies/Monge/.

[52] John Joseph O'Connor and Edmund Frederick Robertson. *Leonid Vitalyevich Kantorovich*. 2014. URL: https://mathshistory.st-andrews.ac.uk/Biographies/Kantorovich/.

[53] Ingram Olkin and Friedrich Pukelsheim. "The distance between two random vectors with given dispersion matrices". In: *Linear Algebra and its Applications* 48 (1982), pp. 257–263.

[54] Maurizio Paolini. *Attività con i 300 megatasselli "spettro"*. 2023. URL: https://dmf.unicatt.it/paolini/einstein/activities/.

[55] Brendan Pass. "Uniqueness and Monge solutions in the multimarginal optimal transportation problem". In: *SIAM Journal on Mathematical Analysis* 43.6 (2011), pp. 2758–2775.

[56] Aldo Pratelli. "On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation". In: *Annales de l'Institut Henri Poincare (B) Probability and Statistics*. Vol. 43. 1. Elsevier. 2007, pp. 1–13.

[57] Julien Rabin et al. "Wasserstein barycenter and its application to texture mixing". In: *International conference on scale space and variational methods in computer vision*. Springer. 2011, pp. 435–446.

[58] Hannes Risken and Hannes Risken. *Fokker-planck equation*. Springer, 1996.

[59] Walter Rudin. "Functional analysis". In: *International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York* (1991).

[60] Jun John Sakurai. *Modern Quantum Mechanics*. 1994.

[61] Erwin Schrödinger. *Über die umkehrung der naturgesetze*. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u. Company, 1931.

[62] Claude Elwood Shannon. "A mathematical theory of communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423.

[63] David Smith et al. "An aperiodic monotile". In: *Combinatorial Theory* 4.1 (2024).

[64] Cédric Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2008.

List of Figures

List of symbols

Index