An Introduction to the Mass Transportation Theory and its Applications

Wilfrid Gangbo
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332, USA
gangbo@math.gatech.edu *

June 2004

Contents

*The author of these notes acknowleges support of U.S. N.S.F. grant DMS-00-74037 and DMS-02-00267. Key words: mass transfer, Wasserstein metric. AMS code: 49J40, 82C40 and 47J25.

Abstract

The Monge-Kantorovich mass transportation theory originated in three influential papers. The first one was written by a great geometer, G. Monge [37]. The second and third one were due to Kantorovich [32] [33] , who received a Nobel prize for related work in economics [38]. The Monge-Kantorovich theory is having a growing number of applications in various areas of sciences including economics, optic (e.g. the reflector problem), meteorology, oceanography, kinetic theory, partial differential equations (PDEs) and functional analysis (e.g. geometric inequalities). The purpose of these five hour lectures is to develop basic tools for the Monge-Kantorovich theory. We will briefly mention its impact in partial differential equations and meteorology. These applications are fully developed in the following preprints, [12] [15] [18] [28], which you can download from my webpage at www.math.gatech.edu/ gangbo/publications/.

We have ended this manuscript with a bibliography of a list, far from being exhaustive, of recent contributions to the mass of transportation theory and its applications to geometric inequalities, as well as computer vision and optics (the reflector problem) .

These notes have been written as a set of lectures for the 2004 Summer Institute at Carnegie Mellon University. It is a pleasure to thanks the organizers of that Summer Institute for their support, hospitality and the impressive organization.

1 Introduction

Assume that we are given a pile of sand occupying a region $X \subset \mathbf{R}^d$ and assume that we have another region $Y \subset \mathbf{R}^d$, that consists of holes, of prescribed distribution. Let ρ_o be the distribution of the sand and ρ_1 be the distribution of the holes. We also assume that for each pair of points $(x, y) \in X \times Y$ we have assigned a nonnegative number $c(x, y)$ which represents the cost for transporting a unit mass from x to y. A transport map T is a strategy which tells us that the mass from x will be moved to Tx. It must satisfy a mass conservation condition that is:

$$\int_{T^{-1}(B)]} \rho_o(x)dx = \int_B \rho_1(y)dy, \tag{1}$$

for all Borel sets $B \subset \mathbf{R}^d$. When (1) holds, we say that "T pushes ρ_o forward to ρ_1" or that T is a strategy for transporting ρ_o onto ρ_1 (see definition 2.1) and we write $T_\#\rho_o = \rho_1$.

Pick up a strategy for transporting ρ_o onto ρ_1, or in other words, pick up a map T such that $T_\#\rho_o = \rho_1$. Given a point x, if $\rho_o(x)dx$ is the mass carried by a small neighborhood of x, the cost for transporting $\rho_o(x)dx$ to a small neighborhood of Tx is $c(x, Tx)\rho_o(x)dx$. Thus, the total cost for transporting ρ_o onto ρ_1 is

$$Cost[T] = \int_{\mathbf{R}^d} c(x, Tx)\rho_o(x)dx.$$

The **Monge problem** is to find a minimizer for

$$\inf_T \{ \int_{\mathbf{R}^d} c(x, Tx)\rho_o(x)dx \mid T_\# \rho_o = \rho_1 \}. \tag{2}$$

In 1781, Monge conjectured that when $c(x,y) = |x-y|$ then, there exists an optimal map that transports the pile of sand to the holes. Two hundred years ellapsed before Sudakov claimed in 1976 in [46], to have proven Monge conjecture. It was recently discovered by Ambrosio [4], that Sudakov's proof contains a gap which cannot be fixed in the case $d > 2$. Before that gap was noticed, the proof of Monge conjecture was proven for all dimensional spaces by Evans and the author, in 1999, in [25]. Their proof assumes that the densities ρ_o and ρ_1 are Lipschitz functions of disjoint supports. These results in [25] were recently independently refined by Ambrosio [4], Caffarelli–Feldman–McCann [13] and Trudinger–Wang [47] . In a meanwhile, Caffarelli [11], McCann and the author [28] [29], independently proved Monge conjecture for cost functions that include those of the form $c(x,y) = h(x-y)$, where h is strictly convex. The case $c(x,y) = l(|x-y|)$ where l is strictly concave, which is relevant in economics, was also solved in [28] [29].

One can generalize Monge problem to arbitrary measures μ_o and μ_1 when there is no map T such that $T_\# \mu_o = \mu_1$. To do that, one needs to replace the concept of transport map by the concept of *transport schemes* which can be viewed as multivalued maps, coupled with a family of measures. Let us denote by X the support of μ_o and by Y the support of μ_1. As usually done, we denote by 2^Y the set of subsets of Y. We consider maps $T : X \to 2^Y$ and associate to each $x \in X$, a measure γ_x supported by the set Tx, which tells us how to distribute the mass at x through Tx. Therefore, the cost for transporting x to Tx is

$$\int_{Tx} c(x,y)d\gamma_x(y).$$

The total cost for transporting μ_o onto μ_1 is then

$$\bar{I}[T, \{\gamma_x\}_{x\in X}] = \int_X \left[\int_{Tx} c(x,y)d\gamma_x(y) \right] d\mu_o(x).$$

It is more convenient to encode the information in $(T, \{\gamma_x\}_{x\in X})$ in a measure γ defined on $X \times Y$ by

$$\int_{X\times Y} F(x,y)d\gamma(x,y) = \int_X \left[\int_{Tx} F(x,y)d\gamma_x(y) \right] d\mu_o(x).$$

The measure γ is to satisfy the mass conservation condition:

$$\mu_o[A] = \gamma[A \times Y], \qquad \gamma[X \times B] = \mu_1[B]$$

for all Borel sets $A \subset X$ and $B \subset Y$.

In section 2, we introduce Kantorovich problem in terms of γ, as a relaxation of Monge problem. Indeed, we have already extended the set $\mathcal{T}(\mu_o, \mu_1)$ of maps $T : X \to Y$ such that

$T_{\#}\mu_o = \mu_1$, to a bigger set $\Gamma(\mu_o, \mu_1)$. Then, we extend the function $T \to I[T] := \int_X c(x, Tx)\rho_o(x)$ to a function $\bar{I}$ defined on $\Gamma(\mu_o, \mu_1)$ so that if $\mathcal{T}(\mu_o, \mu_1) \neq \emptyset$ then we have that

$$\inf_{\mathcal{T}(\mu_o,\mu_1)} I = \inf_{\Gamma(\mu_o,\mu_1)} \bar{I}.$$

The new problem at the right handside of the previous equality will be called, as usually done in the calculus of variations, a relaxation of the first problem.

In these notes, we first formulate the mass transportation problems and under suitable assumptions, prove existence of solutions for both, the Monge and Kantorovich problems. We incorporate in these notes prerequisites which we don't plan to go over during these five hour lectures. We mention how the mass transportation fits into dynamical systems and fluids mechanic. The Wasserstein distance and its geometry , as a mass transportation problem, which have played an important role in PDEs during the past few years, are studied. We also comment on the applications of the mass transportation theory to PDEs and meteorology. The applications to geometric inequalities will be covered in parallel lectures given by N. Ghoussoub, we omit them here.

2 Formulation of the mass transport problems

2.1 The original Monge-Kantorovich problem

Let us denote by $\mathcal{T}(\mu_o, \mu_1)$ the set of maps that push μ_o forward to μ_1 (see definition 2.1).
Monge problem. Find a minimizer for

$$\inf_T \{ \int_{\mathbf{R}^d} c(x, Tx) d\mu_o(x) \mid T_{\#}\mu_o = \mu_1 \}. \tag{3}$$

Given two measures μ_o and μ_1, proposition 7.18 gives a sufficient condition for the existence of a map T that transports μ_o onto μ_1. Hence, $\mathcal{T}(\mu_o, \mu_1)$ may be empty unless we impose for instance that μ_o is absolutely continuous with respect to Lebesgue measure. In case, $\mathcal{T}(\mu_o, \mu_1)$ is empty, one can replace the transport maps by multivalued maps, coupled with a family of measures as done in the introduction. We go directly to the right concept to use. We refer the reader to the introduction of this manuscript where we have given a more detailed justification of how we introduced the so-called *transport scheme*.

Definition 2.1 (Transport maps and schemes). *Assume that μ is a measure on X and that ν is a measure on Y. (i) We say that $T : X \to Y$ transports μ onto ν and we write $T_{\#}\mu = \nu$ if*

$$\nu[B] = \mu[T^{-1}(B)] \tag{4}$$

for all Borel set $B \subset Y$. We sometimes say that T is a measure-preserving map with respect to (μ, ν) or T pushes μ forward to ν. We denote by $\mathcal{T}(\mu, \nu)$ the set of T such that $T_{\#}\mu = \nu$.

(ii) If γ is a measure on $X \times Y$ then its projection $proj_X\gamma$ is a measure on X and its projection $proj_Y\gamma$ is a measure on Y defined by $proj_X\gamma[A] = \gamma[A \times Y]$, $proj_Y\gamma[B] = \gamma[X \times \mathbf{R}^d]$.

(iii) A measure γ on $X \times Y$ has μ and ν as its marginals if $\mu = proj_X\gamma$ and $\nu = proj_Y\gamma$. We write that $\gamma \in \Gamma(\mu, \nu)$ and call γ a transport scheme for μ and ν.

Exercise 2.2. *Assume that μ_o and μ_1 are two probability measures on $\mathbf{R}^d$. Assume that $T : \mathbf{R}^d \to \mathbf{R}^d$ is a Borel map and that γ is a Borel measure on $\mathbf{R}^d \times \mathbf{R}^d$. (i) Show that $T_{\#}\mu_o = \mu_1$ if and only if*

$$\int_{\mathbf{R}^d} G(Tx) d\mu_o(x) = \int_{\mathbf{R}^d} G(y) dy$$

for all $G \in L^1(\mathbf{R}^d, \mu_1)$.

(ii) Show that $\gamma \in \Gamma(\mu_o, \mu_1)$ if and only if

$$\int_{\mathbf{R}^d} F(x) d\mu_o(x) = \int_{\mathbf{R}^d \times \mathbf{R}^d} F(x) d\gamma(x,y), \quad \int_{\mathbf{R}^d \times \mathbf{R}^d} G(y) d\gamma(x,y) = \int_{\mathbf{R}^d} G(y) d\mu_1(y),$$

for all $F \in L^1(\mathbf{R}^d, \mu_o)$ and $G \in L^1(\mathbf{R}^d, \mu_1)$.

Remark 2.3. *(i) Note that (4) expresses a mass conservation condition between the two measures.*

(ii) While the set $\Gamma(\mu_o, \mu_1)$ always contains the product measure $\mu_o \times \mu_1$ when $\mu_o[\mathbf{R}^d] = \mu_1[\mathbf{R}^d] = 1$, the set $\mathcal{T}(\mu_o, \mu_1)$ maybe empty. For instance, assume that x, y, z are three distinct elements of $\mathbf{R}^d$, set $\mu_o = 1/2(\delta_x + \delta_y)$ and $\mu_1 = 1/3(\delta_x + \delta_y + \delta_z)$. Then there is no map T that transports μ_o onto μ_1.

(iii) If $\gamma \in \Gamma(\mu_o, \mu_1)$, (x, y) being in the support of γ, expresses the fact that the mass $d\gamma(x, y)$ is transported from x to y. Here, the support of γ is the smallest closed set $K \subset \mathbf{R}^d \times \mathbf{R}^d$ such that $\gamma[K] = \gamma[\mathbf{R}^d \times \mathbf{R}^d]$.

Kantorovich problem. Find a minimizer for

$$\inf_{\gamma \in \Gamma(\mu_o, \mu_1)} \int_{\mathbf{R}^d \times \mathbf{R}^d} c(x,y) d\gamma(x,y). \tag{5}$$

Unlike Monge problem, existence of a minimizer for Kantorovich problem is an easy task. Let us denote by $\mathcal{P}(\mathbf{R}^d)$ the set of Borel probability measures on $\mathbf{R}^d$.

Theorem 2.4. *Assume that $\mu_o, \mu_1 \in \mathcal{P}(\mathbf{R}^d)$ and that $c : \mathbf{R}^d \times \mathbf{R}^d \to [0, +\infty)$ is continuous. Then, (5) admits a minimizer.*

Proof: The set $\Gamma(\mu_o, \mu_1)$ is a compact subset of $\mathcal{P}(\mathbf{R}^d \times \mathbf{R}^d)$ for the weak $*$ convergence. Thus, if $\{\gamma_n\}_{n=1}^\infty$ be a minimizing sequence in (5), extracting a subsequence if necessary, we may assume without loss of generality that $\{\gamma_n\}_{n=1}^\infty$ converges weak $*$ to some γ_∞ in $\Gamma(\mu_o, \mu_1)$. Let $R_r \in C_o(\mathbf{R}^d \times \mathbf{R}^d)$ be a function such that $0 \le R_r \le 1$ and $R_r \equiv 1$ on the ball of center 0 and radius $r > 0$. We have that

$$\begin{aligned}
\int_{\mathbf{R}^d \times \mathbf{R}^d} R_r(x,y) c(x,y) d\gamma_\infty(x,y) &= \lim_{n \to +\infty} \int_{\mathbf{R}^d \times \mathbf{R}^d} R_r(x,y) c(x,y) d\gamma_n(x,y) \\
&\le \lim_{n \to +\infty} \int_{\mathbf{R}^d \times \mathbf{R}^d} c(x,y) d\gamma_n(x,y) \\
&= \inf_{\gamma \in \Gamma(\mu_o, \mu_1)} \int_{\mathbf{R}^d \times \mathbf{R}^d} c(x,y) d\gamma(x,y) \qquad (6)
\end{aligned}$$

Letting r go to $+\infty$ in the first expression of (6) , we obtain that

$$\int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma_\infty(x,y) \leq \inf_{\gamma\in\Gamma(\mu_o,\mu_1)} \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y)$$

which proves that γ_∞ is a minimizer in (5) . This, concludes the proof of the theorem. QED.

Why is Kantorovich's problem a relaxation of Monge's problem? To each $T:\mathbf{R}^d\to\mathbf{R}^d$ such that $T_\#\mu_o=\mu_1$ we associate the measure γ_T defined on $\mathbf{R}^d\times\mathbf{R}^d$ by

$$\gamma_T[C]=\mu_o[\{x\in\mathbf{R}^d \mid (x,Tx)\in C\}].$$

Exercise 2.5. *Assume that $\mu_o,\mu_1\in\mathcal{P}(\mathbf{R}^d)$ and that c is a nonnegative continuous on $\mathbf{R}^d\times\mathbf{R}^d$. Define I on $\mathcal{T}(\mu_o,\mu_1)$ and $\bar I$ on $\Gamma(\mu_o,\mu_1)$ by*

$$I[T]=\int_{\mathbf{R}^d} c(x,Tx)d\mu_o(x), \qquad \bar I[\gamma]=\int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y)$$

(i) Prove that if $T_\#\mu_o=\mu_1$ then $\gamma_T\in\Gamma(\mu_o,\mu_1)$ and $I[T]=\bar I[\gamma_T]$.

(ii) Prove that if μ_o and μ_1 don't have atoms then $\{\gamma_T \mid T\in\mathcal{T}(\mu_o,\mu_1)\}$ is weak $*$ dense in $\Gamma(\mu_o,\mu_1)$.

(iii) Conclude that $\inf_{\mathcal{T}(\mu_o,\mu_1)} I=\inf_{\Gamma(\mu_o,\mu_1)}\bar I$.

A detailed proof of these statements can be found in [27].

2.2 Guessing a good dual to the Monge-Kantorovich problem

If $\bar I$ is a functional on a set Γ and J is a functional on a set $\mathcal{C}$ we say that $\inf_\Gamma \bar I$ and $\sup_{\mathcal{C}} J$ are dual problems to each other if

$$\inf_\Gamma \bar I=\sup_{\mathcal{C}} J.$$

In pratice, one is given for instance the variational problem $\inf_\Gamma \bar I$ to study and one needs to identify a dual problem $\sup_{\mathcal{C}} J$ that would be useful in understanding the primal problem. We use Kantorovich problem to illustrate these facts. We don't insist on the rigor of the arguments used in this subsection. We make them just to help readers who are unfamiliar with duality arguments, to understand how to guess "meaningful" dual problems.

Let $\mathcal{B}$ be the set of Borel measures on $\mathbf{R}^d\times\mathbf{R}^d$ and define $D:\mathcal{B}\to\mathbf{R}\cup\{+\infty\}$ by

$$D(\gamma)=\begin{cases}0 & \text{if} \quad \gamma\in\Gamma(\mu_o,\mu_1)\\ +\infty & \text{if} \quad \gamma\notin\Gamma(\mu_o,\mu_1).\end{cases}$$

In other words, if $\chi\in\{0,1\}$ is the characteristic function of $\Gamma(\mu_o,\mu_1)$ then, $D\in\{+\infty,0\}$ is the "infinite" characteristic function of $\Gamma(\mu_o,\mu_1)$. Define

$$L_\gamma(u,v)=\int_{\mathbf{R}^d} u(x)d\mu_o(x)-\int_{\mathbf{R}^d\times\mathbf{R}^d} u(x)d\gamma(x,y)+\int_{\mathbf{R}^d} v(y)d\mu_1(y)-\int_{\mathbf{R}^d\times\mathbf{R}^d} v(y)d\gamma(x,y)$$

on the set $C_o(\mathbf{R}^d) \times C_o(\mathbf{R}^d)$. Note that

$$D(\gamma) = \sup_{(u,v)\in C_o(\mathbf{R}^d)\times C_o(\mathbf{R}^d)} L_\gamma(u,v). \tag{7}$$

Indeed, if $\gamma \in \Gamma(\mu_o,\mu_1)$ then, $L_\gamma \equiv 0$. But if $\gamma \notin \Gamma(\mu_o,\mu_1)$ then, either

$$\int_{\mathbf{R}^d} u_o(x)d\mu_o(x) \neq \int_{\mathbf{R}^d\times\mathbf{R}^d} u_o(x)d\gamma(x,y)$$

for some u_o or

$$\int_{\mathbf{R}^d} v_o(y)d\mu_1(y) \neq \int_{\mathbf{R}^d\times\mathbf{R}^d} v_o(y)d\gamma(x,y)$$

for some v_o. Assume for instance that the first equality fails. Since we can substitute u_o by $-u_o$, we may assume without loss of generality that

$$\int_{\mathbf{R}^d} u_o(x)d\mu_o(x) > \int_{\mathbf{R}^d\times\mathbf{R}^d} u_o(x)d\gamma(x,y).$$

This yields that $L_\gamma(u_o,0) > 0$. Because L_γ is linear in its argument, we use that $L_\gamma(\lambda u_o,0) = \lambda L_\gamma(u_o,0)$ tends to $+\infty$ as λ tends to $+\infty$, to conclude that the supremum $D(\gamma) = +\infty$.

Clearly,

$$\inf_{\gamma\in\Gamma(\mu_o,\mu_1)} \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y) = \inf_{\gamma\in\mathcal{B}} D(\gamma) + \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y)$$

and so,

$$\begin{aligned}\inf_{\gamma\in\Gamma(\mu_o,\mu_1)} \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y) &= \inf_{\gamma\in\mathcal{B}} \sup_{(u,v)} \Big\{\int_{\mathbf{R}^d} u(x)d\mu_o(x) + \int_{\mathbf{R}^d} v(y)d\mu_1(y) \\ &+ \qquad \int_{\mathbf{R}^d\times\mathbf{R}^d} (c(x,y)-u(x)-v(y))d\gamma(x,y)\Big\} \qquad (8) \\ &= \sup_{(u,v)} \inf_{\gamma\in\mathcal{B}} \Big\{\int_{\mathbf{R}^d} u(x)d\mu_o(x) + \int_{\mathbf{R}^d} v(y)d\mu_1(y) \\ &+ \qquad \int_{\mathbf{R}^d\times\mathbf{R}^d} (c(x,y)-u(x)-v(y))d\gamma(x,y)\Big\} \qquad (9)\end{aligned}$$

To obtain that the expression in (8) and (9) are equal, we have used the minimax theorem since the functional

$$\int_{\mathbf{R}^d} u(x)d\mu_o(x) + \int_{\mathbf{R}^d} v(y)d\mu_1(y) + \int_{\mathbf{R}^d\times\mathbf{R}^d} (c(x,y)-u(x)-v(y))d\gamma(x,y)$$

is convex in the variables γ for (u,v) fixed and concave in the variables (u,v) for γ fixed.

Note that for (u,v) fixed

$$\inf_{\gamma\in\mathcal{B}} \int_{\mathbf{R}^d\times\mathbf{R}^d} (c(x,y)-u(x)-v(y))d\gamma(x,y) = \begin{cases} 0 & \text{if} \quad (u,v)\in\mathcal{C} \\ -\infty & \text{if} \quad (u,v)\notin\mathcal{C}, \end{cases} \tag{10}$$

where, $\mathcal{C}$ is the set of pairs (u, v) such that $u, v : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$ is upper semicontinuous and

$$u(x) + v(y) \leq c(x, y)$$

for all $x, y \in \mathbf{R}^d$. Indeed, if $(u, v) \in \mathcal{C}$ then

$$\gamma \to \int_{\mathbf{R}^d \times \mathbf{R}^d} (c(x, y) - u(x) - v(y)) d\gamma(x, y) \geq 0$$

is minimized by $\gamma_o \equiv 0$. But, if $(u, v) \notin \mathcal{C}$ then $c(x_o, y_o) - u(x_o) - v(y_o) < 0$ for a pair $(x_o, y_o) \in \mathbf{R}^d \times \mathbf{R}^d$. If we set $\gamma_\lambda = \lambda \delta_{(x_o, y_o)}$ then

$$\int_{\mathbf{R}^d \times \mathbf{R}^d} (c(x, y) - u(x) - v(y)) d\gamma_\lambda(x, y) \to -\infty$$

as λ tends to $+\infty$.

We combine (9) and (10) to conclude that

$$\inf_{\gamma \in \Gamma(\mu_o, \mu_1)} \int_{\mathbf{R}^d \times \mathbf{R}^d} c(x, y) d\gamma(x, y) = \sup_{(u,v) \in \mathcal{C}} \int_{\mathbf{R}^d} u(x) d\mu_o(x) + \int_{\mathbf{R}^d} v(y) d\mu_1(y).$$

The set $\mathcal{C}$ is convex and its extreme points are the (u, v) satisfying

$$u(x) = \inf_{y \in \mathbf{R}^d} c(x, y) - v(y), \qquad (x \in \mathbf{R}^d) \tag{11}$$

and

$$v(y) = \inf_{x \in \mathbf{R}^d} c(x, y) - u(x), \qquad (y \in \mathbf{R}^d). \tag{12}$$

The aim of next subsection is to study properties of the set of (u, v) satisfying (11) and (12).

2.3 Properties of "Extreme points of $\mathcal{C}$"

Throughout this subsection, we assume that there exists $h : \mathbf{R}^d \to [0, +\infty)$ such that

(H1) $c(x, y) = h(x - y)$

(H2) $\lim_{|z| \to +\infty} \frac{h(z)}{|z|} = +\infty$.

(H3) $h \in C^1(\mathbf{R}^d)$ is strictly convex

Definition 2.6. *If $u, v : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$ are upper semicontinuous, we define*

$$v^c(x) = \inf_{y \in \mathbf{R}^d} c(x, y) - v(y), \qquad (x \in \mathbf{R}^d) \tag{13}$$

and

$$u_c(y) = \inf_{x \in \mathbf{R}^d} c(x, y) - u(x), \qquad (y \in \mathbf{R}^d). \tag{14}$$

We call v^c the (upper) c-concave transform of v and u_c the (lower) c-concave transform of u.

Throughout this section, if $A \subset \mathbf{R}^d$ *and* $u : A \to \mathbf{R} \cup \{-\infty\}$ *we identify* u *with the function* $\tilde{u}$ *defined by*

$$\tilde{u}(x) = \begin{cases} u(x) & \text{if} \quad x \in A \\ -\infty & \text{if} \quad x \notin A. \end{cases}$$

Observe that in that case

$$(\tilde{u})_c(y) = \inf_{x \in A} c(x, y) - u(x), \qquad (y \in \mathbf{R}^d).$$

We define

$$\mathcal{C}_R := \{(u, v) \in C(\bar{B}_R) \times C(\bar{B}_R) \mid u(x) + v(y) \le c(x, y) \; \forall x, y \in \bar{B}_R\}.$$

Remark 2.7. *If* $(u, v) \in \mathcal{C}_R$ *and* $x \in \bar{B}_R$ *then*

$$u(x) \le c(x, y) - v(y) \tag{15}$$

for all $y \in \bar{B}_R$. *Minimizing the right handside of (15) over* $y \in \bar{B}_R$, *we obtain that* $u(x) \le v^c(x)$. *Hence,* $u \le v^c$. *Similarly,* $v \le u_c$.

Lemma 2.8. *Assume that* $c \in C(\mathbf{R}^d \times \mathbf{R}^d)$ *and that* $u, v \in C(\bar{B}_R)$. *Then (i)* $v \le (v^c)_c$ *and* $u \le (u_c)^c$.

(ii) If $v = u_c$ *then* $(v^c)_c = v$. *Similarly, if* $u = v^c$ *then* $u = (u_c)^c$.

Proof: 1. *For every* $x, y \in \bar{B}_R$ *we have that* $v^c(x) \le c(x, y) - v(y)$ *and so,*

$$v(y) \le c(x, y) - v^c(x).$$

Minimizing the right handside of the previous inequality over $x \in \bar{B}_R$ *we conclude that* $v \le (v^c)_c$. *The proof for* u *is done in a similar manner.*

2. *Note that if* $a \le b$ *then* $a_c \ge b_c$ *and so, the operation* $\cdot_c$ *reverses inequality. Assume that* $v = u_c$. *By (i) we have that* $v^c = (u_c)^c \ge u$ *and so,* $(v^c)_c \le u_c = v$. *This, combined with (i), gives that* $(v^c)_c = v$. *The proof for* u *is done in a similar manner.* *QED.*

Proposition 2.9. *Assume that* (H1), (H2) *and* (H3) *hold. Let* $B_R \subset \mathbf{R}^d$ *be the open ball of center* 0 *and radius* $R > 0$. *Assume that* $v \in C(\bar{B}_R)$, *that* $K \subset \mathbf{R}^d$ *is a convex compact and that* $u = v^c$. *Then*

(i) $Lip(u|_K) \le ||\nabla_x c||_{L^\infty(K \times B_R)}$.

(ii) If u *is differentiable at* x *then,* $Tx := x - \nabla h^*(\nabla u(x))$ *is the unique point* $y \in \bar{B}_R$ *at which the infimum in (13) is attained. We recall that here,* h^* *denotes the Legendre transform of* h.

(iii) If $H \in C(\bar{B}_R)$ *and* $u_r = (v + rH)^c$ *then,* $||u_r - u||_{L^\infty(K)} \le |r| \, ||H||_{L^\infty(B_R)}$. *If in addition* u *is differentiable at* x *then*

$$\lim_{r \to 0} \frac{(v + rH)^c(x) - v^c(x)}{r} = -H(Tx).$$

Proof: 1. *Let* $x_1, x_2 \in K$*. Assume without loss of generality that* $u(x_2) \geq u(x_1)$*. Let* $y_1 \in \bar{B}_R$ *be such that*

$$u(x_1) = c(x_1, y_1) - v(y_1), \tag{16}$$

in other words, y_1 *is a point at which the minimum is attained in (13) for* $x = x_1$*. Since* $y_1 \in \bar{B}_R$ *we have that*

$$u(x_2) \leq c(x_2, y_1) - v(y_1). \tag{17}$$

We combine (16) and (17) to obtain that

$$\begin{aligned} |u(x_2) - u(x_1)| = u(x_2) - u(x_1) &\leq c(x_2, y_1) - c(x_1, y_1) \\ &=< \nabla_x c(\bar{x}, y_1); x_2 - x_1 > \\ &\leq |x_2 - x_1| ||\nabla_x c||_{L^\infty(K \times B_R)}. \end{aligned}$$

We have used the mean–value–theorem in the latest expressions; the point $\bar{x}$ *is obtained as a convex combination of* x_1 *and* x_2*. This concludes the proof of (i).*

2. *Assume that* u *is differentiable at* x *and let* $y \in \bar{B}_R$ *be such that* $u(x) = c(x, y) - v(y)$*. Since* $(u, v) \in \mathcal{C}$ *we have that* $z \to l(z) := u(z) - c(z, y) + v(y) \leq 0$ *and so, the previous equality shows that* l *attains its maximum at* x*. Since* l *is differentiable at* x *we have that*

$$0 = \nabla l(x) = \nabla u(x) - \nabla_x c(x, y).$$

Hence,

$$\nabla u(x) = \nabla_x c(x, y) = \nabla h(x - y). \tag{18}$$

Because of (H2)*and* (H3)*, we have that* ∇h *is invertible and its inverse is* ∇h^**. This, together with (18) gives that* $y = x - \nabla h^*(\nabla u(x))$*. This proves (ii).*

3. *Assume that* $H \in C(\bar{B}_R)$ *and set* $u_r = (v + rH)^c$*. For* $x \in K$ *and each* r*, there exists* $y_r \in \bar{B}_R$ *such that*

$$u_r(x) = c(x, y_r) - v(y_r) - rH(y_r) \geq u(x) - rH(y_r). \tag{19}$$

In case $r = 0$ *we rewrite (19) as*

$$u_o(x) = c(x, y_o) - v(y_o) = c(x, y_o) - v(y_o) - rH(y_o) + rH(y_o) \geq u_r(x) + rH(y_o). \tag{20}$$

We use both (19) and (20) to obtain that

$$-rH(y_r) \leq u_r(x) - u(x) \leq -rH(y_o). \tag{21}$$

We obtain immediatly from (21) that

$$||u_r - u||_{L^\infty(K)} \leq |r| \; ||H||_{L^\infty(B_R)}. \tag{22}$$

By (22), $\{u_r\}_r$ *converges uniformly to* u *as* r *tends to* 0*.*

4. *Assume in addition that u is differentiable at x. We claim that*

$$\lim_{r\to 0} y_r = Tx.$$

To prove the latest statement, we have to prove that every subsequence of $\{y_r\}_r$ has a convergent subsequence which tends to Tx. Assume that $\{y_{r_n}\}_{n=1}^{\infty}$ is a subsequence of $\{y_r\}_r$. Let us extract from $\{y_{r_n}\}_{n=1}^{\infty}$ a converging subsequence that we still label $\{y_{r_n}\}_{n=1}^{\infty}$. Call $y \in \bar{B}_R$ the limit of $\{y_{r_n}\}_{n=1}^{\infty}$. Since v and c are continuous and $\{y_{r_n}\}_{n=1}^{\infty}$ converges to y, we deduce that the right handside of (19) tends to $c(x,y) - v(y)$ as n tends to $+\infty$. We use (21) to deduce that the left handside of (19) tends to $u_o(x)$ as n tends to $+\infty$. Consequently,

$$u_o(x) = c(x,y) - v(y).$$

By (ii), the previous equality yields that $y = Tx$. Since $\{y_{r_n}\}_{n=1}^{\infty}$ is an arbitrary subsequence of $\{y_r\}_r$ we conclude that $\lim_{r\to 0} y_r = Tx$.

We divide both sides of (21) by r and use that $\lim_{r\to 0} y_r = Tx$ to conclude the proof of (iii). QED.

Let

$$\mathcal{K}_R := \{(u,v) \in C(\bar{B}_R) \times C(\bar{B}_R) \mid u = v^c,\ v = u_c,\ u(0) = 0\}.$$

Note that $\mathcal{K}_R \subset \mathcal{C}_R$.

Lemma 2.10. *If* (H1), (H2)*and* (H3)*hold, then $\mathcal{K}_R$ is a compact subset of $C(\bar{B}_R) \times C(\bar{B}_R)$ for the uniform norm.*

Proof: *Define*

$$M_1 := ||\nabla_x c||_{L^\infty(\bar{B}_R \times \bar{B}_R)}, \qquad M_2 = ||c||_{L^\infty(\bar{B}_R \times \bar{B}_R)}$$

and let $(u,v) \in \mathcal{K}_R$. Since $u = v^c$, proposition 2.9 (i) gives that $Lip(u) \le M_1$. This, together with the fact that $u(0) = 0$ gives that

$$||u||_{W^{1,\infty}(B_R)} \le (1+R)M_1. \tag{23}$$

Since $v = u_c$, we use (23) to obtain that

$$||v||_{L^\infty(B_R)} \le ||c||_{L^\infty(\bar{B}_R \times \bar{B}_R)} + ||u||_{L^1(B_R)} \le M_2 + (1+R)M_1.$$

This, together with the fact that by proposition 2.9 (i) , $Lip(v) \le M_1$, yields that

$$||v||_{W^{1,\infty}(B_R)} \le (2+R)M_1 + M_2.. \tag{24}$$

In the light of Ascoli-Arzela theorem, (23) and (24) yield that $\mathcal{K}_R$ is a compact subset of $C(\bar{B}_R) \times C(\bar{B}_R)$ for the uniform norm.

2.4 Existence of a minimizer

Let $\mu_o, \mu_1 \in \mathcal{P}(\mathbf{R}^d)$ be such that

$$\mathrm{spt}\mu_o, \mathrm{spt}\mu_1 \subset B_R \tag{25}$$

and define

$$J[u,v] = \int_{\bar{B}_R} u d\mu_o + \int_{\bar{B}_R} v d\mu_1.$$

Lemma 2.11. *Assume that* (H1), (H2)*and* (H3)*hold. Then J admits a maximizer (u_o, v_o) over $\mathcal{C}_R$. If in addition $\mu_o << dx$ then $T_o x = x - \nabla h^*(\nabla u_o(x))$ is defined μ_o–almost everywhere and*

$$T_{o\#}\mu_o = \mu_1, \qquad u_o(x) + v_o(T_o x) = c(x, T_o x) \;\; \mu_o - \text{almost everywhere}.$$

Proof: 1. *Observe that if $(u,v) \in \mathcal{C}_R$ then remark 2.7 gives that $u \leq v^c$ and so, $J[u,v] \leq J[v^c, v]$. We repeat the same argument to obtain that $J[v^c, v] \leq J[v^c, (v^c)_c]$. Setting $(\bar{u}, \bar{v}) = (v^c, (v^c)_c)$, we have that $(\bar{u}, \bar{v}) \in \mathcal{C}_R$ and $J[u,v] \leq J[\bar{u}, \bar{v}]$. By lemma 2.8 we have that $(\bar{u})_c = \bar{v}$ and $(\bar{v})^c = \bar{u}$. Note in addition that*

$$J[\bar{u} - \bar{u}(0), \bar{v} + \bar{u}(0)] = J[\bar{u}, \bar{v}] + \bar{u}(0)(\mu_1(B_R) - \mu_o(B_R)) = J[\bar{u}, \bar{v}].$$

We have just proven that

$$\sup_{\mathcal{C}_R} J = \sup_{\mathcal{K}_R} J.$$

By lemma 2.10, $\mathcal{K}_R$ is a compact set for the uniform convergence. Thus, J admits a maximizer (u_o, v_o) over $\mathcal{K}_R$.

2. *Assume now that $\mu_o << dx$. Proposition 2.9 (i) gives that u_o is a Lipschitz function on $\bar{B}_R$ and so, it is differentiable μ_o–almost everywhere. Hence, the map $T_o x = x - \nabla h^*(\nabla u_o(x))$ is defined μ_o–almost everywhere. Proposition 2.9 (ii) gives that*

$$u_o(x) + v_o(T_o x) = c(x, T_o x) \tag{26}$$

μ_o–almost everywhere. Let $H \in C(\bar{B}_R)$ and define

$$u_r = (v_o + rH)^c, \qquad v_r = v_o + rH.$$

Note that $(u_r, v_r) \in \mathcal{C}_R$ and so, $J[u_r, v_r] \leq J[u_o, v_o]$. Hence, $\lim_{r\to 0}(J[u_r, v_r] - J[u_o, v_o])/r = 0$ provided that the limit exists. But, Proposition 2.9 (iii) gives that

$$\begin{aligned}\lim_{r\to 0} \frac{J[u_r, v_r] - J[u_o, v_o]}{r} &= \lim_{r\to 0} \int_{\bar{B}_R} \frac{u_r(x) - u_o(x)}{r} d\mu_o(x) + \int_{\bar{B}_R} H(y) d\mu_1(y) \\ &= \int_{\bar{B}_R} -H(T_o x) d\mu_o(x) + \int_{\bar{B}_R} H(y) d\mu_1(y).\end{aligned}$$

Thus,

$$\int_{\bar{B}_R} H(T_o x) d\mu_o(x) = \int_{\bar{B}_R} H(y) d\mu_1(y).$$

Since H is arbitrary, this proves that $T_{o\#}\mu_o = \mu_1$, which concludes the proof of the lemma. QED.

Remark 2.12 (Half of a way to duality). *We have the*

$$\sup_{\mathcal{C}} J \leq \inf_{\Gamma(\mu_o,\mu_1)} \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y) \leq \inf_{\mathcal{T}(\mu_o,\mu_1)} \int_{\mathbf{R}^d} c(x,Tx)d\mu_o(x). \tag{27}$$

Proof: 1. *If $(u,v)\in\mathcal{C}$ then $u(x)+v(y)\leq c(x,y)$ and so,*

$$\int_{\mathbf{R}^d\times\mathbf{R}^d} (u(x)+v(y))d\gamma(x,y) \leq \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y). \tag{28}$$

Since $\gamma\in\Gamma(\mu_o,\mu_1)$ we use exercise 2.2 (ii) to obtain that

$$\int_{\mathbf{R}^d} u(x)d\mu_o(x) = \int_{\mathbf{R}^d\times\mathbf{R}^d} u(x)d\gamma(x,y) \quad \text{and} \quad \int_{\mathbf{R}^d\times\mathbf{R}^d} v(y)d\gamma(x,y) = \int_{\mathbf{R}^d} v(y)d\mu_1(y).$$

This, together with (28) yields the first inequality in (27).

2. *To each $T\in\mathcal{T}(\mu_o,\mu_1)$ we associate the measure γ_T defined on $\mathbf{R}^d\times\mathbf{R}^d$ by*

$$\gamma_T[C] = \mu_o[\{x\in\mathbf{R}^d \mid (x,Tx)\in C\}].$$

Clearly, $\gamma_T\in\Gamma(\mu_o,\mu_1)$. Indeed, if we denote by C the characteristic function of C, the definition of γ_T gives that

$$\int_{\mathbf{R}^d\times\mathbf{R}^d} 1_C(x,y)d\gamma(x,y) = \int_{\mathbf{R}^d} 1_C(x,Tx)d\mu_o(x)$$

for arbitrary set $C\subset\bar{B}_R\times\bar{B}_R$. Hence, every function F which is a linear combination of characteristic functions, satisfies

$$\int_{\mathbf{R}^d\times\mathbf{R}^d} F(x,y)d\gamma(x,y) = \int_{\mathbf{R}^d} F(x,Tx)d\mu_o(x).$$

Standard approximation arguments yield that the previous identity holds for every $F\in C(\bar{B}_R\times\bar{B}_R)$. In particular,

$$\int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y) = \int_{\mathbf{R}^d} c(x,Tx)d\mu_o(x).$$

We conclude that

$$\begin{aligned}\inf_{\mathcal{T}(\mu_o,\mu_1)} \int_{\mathbf{R}^d} c(x,Tx)d\mu_o(x) &= \inf_{\mathcal{T}(\mu_o,\mu_1)} \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma_T(x,y)\\ &\geq \inf_{\Gamma(\mu_o,\mu_1)} \int_{\mathbf{R}^d\times\mathbf{R}^d} c(x,y)d\gamma(x,y).\end{aligned}$$

This proves the second inequality in (27). *QED.*

Lemma 2.13. *Assume that* (H1), (H2)*and* (H3)*hold. Assume that* $\mu_o << dx$, *that* $T_{o,\#}\mu_o = \mu_1$ *and that there exists an upper semicontinuous function* $v_o : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$ $T_o x = x - \nabla h^*(\nabla u_o(x))$, *where* $u_o = v_o^c$. *Assume that* $v_o \equiv -\infty$ *outside the support of* μ_1. *Then,* T_o *minimizes*

$$\inf_{T \in \mathcal{T}(\mu_o,\mu_1)} \int_{\mathbf{R}^d} c(x, Tx) d\mu_o(x), \tag{29}$$

and so, all the expressions in (27) coincide.

Proof: *In order to simplify the proof of this lemma, we assume in addition that the support of* μ_1 *is bounded and so, there exists a ball of radius* $R < +\infty$ *such that*

$$\mathrm{spt}\mu_1 \subset B_R.$$

As a consequence, $v_o \equiv -\infty$ *on the complement of* B_R. *By proposition 2.9 (i),* u_o *is locally Lipschitz and so, it is differentiable* μ_o*–almost everywhere (since* $\mu_o << dx$ *). We use proposition 2.9 (ii) to obtain that*

$$u_o(x) + v_o(T_o x) = c(x, T_o x) \tag{30}$$

μ_o*–almost everywhere. We integrate both sides of (30) and use that* $T_{o\#}\mu_o = \mu_1$ *to obtain that*

$$J[u_o, v_o] = \int_{\mathbf{R}^d} c(x, T_o x) d\mu_o(x) \tag{31}$$

Because $u_o = v_o^c$, *we have that* $(u_o, v_o) \in \mathcal{C}$. *By remark 2.12, since equality holding in (31), we must have that* T_o *minimizes (29) and that all the expressions in (27) coincide.* *QED.*

Theorem 2.14. *Assume that* (H1), (H2)*and* (H3)*hold and that* $h(z) = l(|z|)$ *for some function* l *(As one can see in [29], the last and new assumption we have just made, is not important but is meant here to make simple statements). Assume that* $\mu_o << dx$. *Then, (29) admits a minimizer. If the expression in (29) is finite then, (i) the minimizer* T_o *in (29) is unique and*

$$\inf_{\mathcal{T}(\mu_o,\mu_1)} \int_{\bar{B}_R} c(x, Tx) d\mu_o(x) = \inf_{\Gamma(\mu_o,\mu_1)} \int_{\bar{B}_R \times \bar{B}_R} c(x, y) d\gamma(x, y). \tag{32}$$

(iii) If in addition $\mu_1 << dx$ *then* T_o *is invertible on* $\mathbf{R}^d$ *up to a set of zero* μ_o*–measure.*

Proof: *Theorem 2.14 is fundamental in many applications of the Monge-Kantorovich theory. The case* $p = 2$ *was first proven by Brenier in [6]. The general case was independently proven by Caffarelli [11], Gangbo and McCann [29] under assumptions more general than the above one. To make the proof simpler, we further assume that the supports of* μ_o *and* μ_1 *are bounded so that, there exists a ball of radius* $R < +\infty$ *such that*

$$\mathrm{spt}\mu_o, \mathrm{spt}\mu_1 \subset B_R.$$

1. *According to lemma 2.11, J admits a maximizer (u_o, v_o) over $\mathcal{C}_R$, $T_o x = x - \nabla h^*(\nabla u_o(x))$ is defined μ_o–almost everywhere and*

$$T_{o\#}\mu_o = \mu_1, \qquad u_o(x) + v_o(T_o x) = c(x, T_o x) \;\; \mu_o \;\; \textit{almost everywhere.}$$

These, together with lemma 2.13 prove that T_o is a minimizer for the Monge problem (29) and (32) holds.

2. *Assume that T_1 is another minimizer of (29). Then, we must have that*

$$J[u_o, v_o] = \int_{\bar{B}_R} c(x, T_o x) d\mu_o(x) \geq \int_{\bar{B}_R} (u_o(x) + v_o(T_1 x)) d\mu_o(x) = J[u_o, v_o].$$

Thus, $u_o(x) + v_o(T_1 x) = c(x, T_1 x)$ μ_o–almost everywhere. But, proposition 2.9 (ii) gives that the equation $u_o(x) + v_o(y) = c(x, y)$ admits the unique solution $y = T_o x$ for μ_o–almost every x. This proves that $T_1 x = T_o x$ μ_o–almost everywhere.

3. *If in addition $\mu_1 << dx$ then by symmetry, the Monge problem*

$$\inf_{S \in \mathcal{T}(\mu_1, \mu_o)} c(Sy, y) d\mu_1(y)$$

admits a unique solution S_o and we have that

$$S_{o\#}\mu_1 = \mu_o, \qquad u_o(S_o y) + v_o(y) = c(S_o y, y) \;\; \mu_1 \;\; \textit{almost everywhere.}$$

This proves that $S_o(T_o(x)) = x$ μ_o–almost everywhere and $T_o(S_o(y)) = y$ μ_1–almost everywhere. This concludes the proof of the theorem. *QED.*

3 The Wasserstein distance

Assume that μ_o and μ_1 are two probability measures on $\mathbf{R}^d$ and that $0 < p < +\infty$. We define

$$W_p^p(\mu_o, \mu_1) := \frac{1}{p} \inf_{\gamma} \{ \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^p d\gamma(x, y) : \; \gamma \in \Gamma(\mu_o, \mu_1) \}. \tag{33}$$

We have in theorem 2.14 that when $\mu_o << dx$ then

$$W_p^p(\mu_o, \mu_1) := \frac{1}{p} \inf_{T} \{ \int_{\mathbf{R}^d} |x - T(x)|^p d\mu_o(x) : \; T_{\#}\mu_o = \mu_1 \}. \tag{34}$$

In fact, one can improve that result to obtain that when μ_o and μ_1 don't have atoms then (34) holds. The proof of the statement can be found in [4] and [27].

Remark 3.1. *Assume that $c(x, y) = \frac{|x-y|^2}{2}$ and that $u, v : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$. Set $\psi(x) = \frac{|x|^2}{2} - u(x)$ and $\phi(y) = \frac{|y|^2}{2} - v(y)$. We have that $u = v^c$ if and only if*

$$u(x) = \inf_{y \in \mathbf{R}^d} \frac{|x - y|^2}{2} - v(y) = \frac{|x|^2}{2} + \inf_{y \in \mathbf{R}^d} - < x; y > + \frac{|y|^2}{2} - v(y).$$

This is equivalent to $\psi(x) = \sup_{y \in \mathbf{R}^d} < x; y > - \phi(y) = \phi^*(y)$. In that case

$$u(x) + v(\nabla \psi(x)) = 1/2 |x - \nabla \psi(x)|^2$$

Combining lemma 2.13, theorem 2.14 and remarks 3.1, 3.5, we conclude the following corollary.

Corollary 3.2. *Assume that $\mu_o, \mu_1 \in \mathcal{P}(\mathbf{R}^d)$ are of bounded second moments, that $\mu_o << dx$ and that $c(x,y) = \frac{|x-y|^2}{2}$. Then (i) there exists a convex, lower semicontinuous function on $\mathbf{R}^d$, whose gradient minimizes*

$$\inf_T \{ \int_{\mathbf{R}^d} |x - T(x)|^2 d\mu_o(x) : \; T_\# \mu_o = \mu_1 \}.$$

(ii) Conversely, if ψ is a convex, lower semicontinuous function on $\mathbf{R}^d$, such that $\nabla\psi_\# \mu_o = \mu_1$ then, $\nabla\psi$ minimizes

$$\inf_T \{ \int_{\mathbf{R}^d} |x - T(x)|^2 d\mu_o(x).$$

Exercise 3.3. *. (See also [41])) Assume that $1 \le p < +\infty$, that $\{\mu_n\}_{n=1}^\infty \subset \mathcal{P}(\mathbf{R}^d)$ and that $\mu_\infty \in \mathcal{P}(\mathbf{R}^d)$. Prove that the following are equivalent*

(i) $\lim_{n \to +\infty} W_p(\mu_n, \mu_\infty) = 0$.

(ii) $\{\mu_n\}_{n=1}^\infty$ converges weak $*$ to μ_∞ and $\int_{\mathbf{R}^d} |x|^p d\mu_n(x)$ converges to $\int_{\mathbf{R}^d} |x|^p d\mu_\infty(x)$.

We denote the p–moment of a measure $\mu \in \mathcal{P}(\mathbf{R}^d)$ by

$$M_p(\mu) = \int_{\mathbf{R}^d} |x|^p d\mu(x).$$

Remark 3.4. *Let δ_{z_o} be the dirac mass concentrated at a point $z_o \in \mathbf{R}^d$. Note that $\Gamma(\mu, \delta_{z_o})$ contains only one element which is the product measure $\mu \times \delta_{z_o}$. The interpretation of this is that there is only one strategy to transport the mass carried by μ to the point z_o. We conclude that*

$$W_p^p(\mu, \delta_{z_o}) = 1/p \int_{\mathbf{R}^d} |x - z_o|^p d\mu(x),$$

which is the p–moment (up to a multiplicative constant) of μ with respect to z_o.

Remark 3.5. *Assume that μ_o and μ_1 have bounded p–moments and $\gamma \in \Gamma(\mu_o, \mu_1)$. Then (i) for $1 \le p < +\infty$, we use that $z \to |z|^p$ is a convex function on $\mathbf{R}^d$ to deduce that*

$$\int_{\mathbf{R}^d} |x-y|^p d\gamma(x,y) \le 2^{p-1} \int_{\mathbf{R}^d} (|x|^p + |y|^p) d\gamma(x,y) = 2^{p-1} (\int_{\mathbf{R}^d} |x|^p d\mu_o(x) + \int_{\mathbf{R}^d} |y|^p d\mu_1(y)) < +\infty.$$

For $0 < p < 1$ we use that $z \to |z|^p$ is a metric on $\mathbf{R}^d$ to have that

$$\int_{\mathbf{R}^d} |x-y|^p d\gamma(x,y) \le \int_{\mathbf{R}^d} (|x|^p + |y|^p) d\gamma(x,y) = (\int_{\mathbf{R}^d} |x|^p d\mu_o(x) + \int_{\mathbf{R}^d} |y|^p d\mu_1(y)) < +\infty.$$

Lemma 3.6. *For $1 \le p < +\infty$, W_p is a metric on $\mathcal{P}^p := \mathcal{P}(\mathbf{R}^d) \cap \{\mu \mid M_p(\mu) < +\infty\}$. For $0 < p \le 1$, W_p^p is a metric on $\mathcal{P}^p$.*

Proof: **1.** *The facts that W_p is symmetric and that $W_p(\mu_o, \mu_1) = 0$ if and only if $\mu_o = \mu_1$ are readily checked.*

2. *Assume that $1 \leq p < +\infty$ and let us then prove the triangle inequality. Let $\mu_o, \mu_1, \mu_2 \in \mathcal{P}^d$. Assume first that these three measures are absolutely continuous with respect to Lebesgue measure. By theorem 2.14, there exist $T_o, T_1 : \mathbf{R}^d \to \mathbf{R}^d$ such that $T_{o\#}\mu_o = \mu_1$, $T_{1\#}\mu_1 = \mu_2$,*

$$W_p^p(\mu_o, \mu_1) = 1/p \int_{\mathbf{R}^d} |x - T_o x|^p d\mu_o(x) = 1/p ||\mathbf{id} - T_o||^p_{L^p(\mu_o)}$$

and

$$W_p^p(\mu_1, \mu_2) = 1/p \int_{\mathbf{R}^d} |y - T_1 y|^p d\mu_1(y) = 1/p ||\mathbf{id} - T_1||^p_{L^p(\mu_1)},$$

where **id** *is the identity map on $\mathbf{R}^d$. Note that if $T := T_1 \circ T_o$ then $T_{\#}\mu_o = \mu_2$. Thus , using the triangle inequality for $|| \cdot ||_{L^p(\mu_o)}$ we have that*

$$W_p(\mu_o, \mu_2) \leq (1/p)^{\frac{1}{p}} ||\mathbf{id} - T_o||_{L^p(\mu_o)} \leq (1/p)^{\frac{1}{p}} \Big(||\mathbf{id} - T_o||_{L^p(\mu_o)} + ||T_o - T||_{L^p(\mu_o)} \Big). \tag{35}$$

Observe that since $T_{o\#}\mu_o = \mu_1$, we have that

$$||T_o - T||^p_{L^p(\mu_o)} = \int_{\mathbf{R}^d} |T_o x - T_1(T_o x)|^p d\mu_o(x) = \int_{\mathbf{R}^d} |y - T_1 y|^p d\mu_1(y) = p W_p^p(\mu_1, \mu_2).$$

This, together with (35) yields that

$$W_p(\mu_o, \mu_2) \leq W_p(\mu_o, \mu_1) + W_p(\mu_1, \mu_2),$$

which proves the triangle inequality.

Since the set of measures in $\mathcal{P}^d$, that are absolutely continuous is a dense set of $\mathcal{P}^d$ for the weak $*$ topology, we use exercise 3.3 to conclude that the triangle inequality holds for general measures in $\mathcal{P}^d$. The proof of the triangle inequality for $0 < p \leq 1$ is obtained in the same manner. *QED.*

Definition 3.7 (The Wasserstein distance). *When $1 \leq p < +\infty$, we call W_p the p–Monge-Kantorovich distance. When $0 < p \leq 1$, we call W_p^p the p–Monge-Kantorovich distance. When $p = 2$, W_2 is called the Wasserstein distance.*

3.1 Geodesics of W_2

Let x_o and x_1 be two distinct points in $\mathbf{R}^d$ and let $\mu_o = \delta_{x_o}$, respectively $\mu_1 = \delta_{x_1}$ be the dirac masses concentrated at x_o , respectively x_1. One may think of several way of interpolating between μ_o and μ_1 is an continuous way. One interpolation could be

$$\bar{\mu}_t = (1 - t)\mu_o + t\mu_1.$$

From the mass transportation (and maybe fluids mechanic) point of view, that interpolation is not "natural" in the sense that we originally started two measures whose supports contain exactly one point, whereas the support of each interpolating measure has two points. From the mass transportation point of view, the next interpolation is interesting since we will see that it is a geodesic with respect to W_2*. That second interpolation between* μ_o *and* μ_1 *is given by*

$$\mu_t = \delta_{x_t} \tag{36}$$

where $x_t = (1-t)x_o + tx_1$*.*

In the light of remark 3.4, each one of the distances $W_2(\mu_o, \mu_t)$*,* $W_2(\mu_t, \mu_1)$ *and* $W_2(\mu_o, \mu_1)$ *is straightforward to compute. Clearly,* $W_2(\mu_o, \mu_t) + W_2(\mu_t, \mu_1) = W_2(\mu_o, \mu_1)$ *and so,* $t \to \mu_t$ *is a geodesic for* W_2*. For general measures* $\mu_o, \mu_1 \in \mathcal{P}(\mathbf{R}^d)$ *there is an analoge of (36), obtained by interpolating between each point* x *in the support of* μ_o *and each point in the support of* μ_1*. This leads us to define the maps*

$$\Pi^t(x,y) = (x, (1-t)x + ty), \qquad \Pi_t(x,y) = ((1-t)x + ty, y).$$

Let $\gamma_o \in \Gamma(\mu, \nu)$ *be a minimizer in* $W_2^2(\mu_o, \mu_1)$ *in the sense that*

$$W_2^2(\mu_o, \mu_1) := 1/2 \int_{\mathbf{R}^d \times \mathbf{R}^d} |x-y|^2 d\gamma_o(x,y).$$

Recall that, to define a measure on $\mathcal{P}(\mathbf{R}^d \times \mathbf{R}^d)$*, it suffices to know how to integrate functions* $F \in C_o(\mathbf{R}^d \times \mathbf{R}^d)$ *with respect to that measure. We define* γ^t *and* γ_t *on* $\mathbf{R}^d \times \mathbf{R}^d$ *by*

$$\int_{\mathbf{R}^d \times \mathbf{R}^d} F(x,y) d\gamma^t(x,y) = \int_{\mathbf{R}^d \times \mathbf{R}^d} F(\Pi^t(x,y)) d\gamma_o(x,y)$$

and

$$\int_{\mathbf{R}^d \times \mathbf{R}^d} F(x,y) d\gamma_t(x,y) = \int_{\mathbf{R}^d \times \mathbf{R}^d} F(\Pi_t(x,y)) d\gamma_o(x,y).$$

In other words, $\gamma^t = \Pi^t_{\#}\gamma_o$ *and* $\gamma_t = \Pi_{t\#}\gamma_o$*. Note that the first marginal of* γ^t *is* μ_o*. Indeed,*

$$\int_{\mathbf{R}^d \times \mathbf{R}^d} F(x) d\gamma^t(x,y) = \int_{\mathbf{R}^d \times \mathbf{R}^d} F(x) d\gamma_o(x,y) = \int_{\mathbf{R}^d} F(x) d\mu_o(x).$$

We define μ_t *to be the second marginal of* γ^t *:*

$$\int_{\mathbf{R}^d} G(y) d\mu_t(y) = \int_{\mathbf{R}^d \times \mathbf{R}^d} G(y) d\gamma^t(x,y) = \int_{\mathbf{R}^d \times \mathbf{R}^d} \int_{\mathbf{R}^d} G((1-t)x + ty) d\gamma_o(x,y).$$

Observe that the first marginal of γ_t *is* μ_t *since ,*

$$\int_{\mathbf{R}^d \times \mathbf{R}^d} F(x) d\gamma_t(x,y) = \int_{\mathbf{R}^d \times \mathbf{R}^d} F((1-t)x + ty) d\gamma_o(x,y) = \int_{\mathbf{R}^d} F(y) d\mu_t(y).$$

We conclude that

$$W_2^2(\mu_o,\mu_t) \le 1/2 \int_{\mathbf{R}^d\times\mathbf{R}^d} |x-y|^2 d\gamma^t(x,y) = 1/2 \int_{\mathbf{R}^d\times\mathbf{R}^d} |x-(1-t)x+ty|^2 d\gamma_o(x,y)$$
$$= t^2/2 \int_{\mathbf{R}^d\times\mathbf{R}^d} |x-y|^2 d\gamma_o(x,y)$$
$$= t^2 W_2^2((\mu_o,\mu_1). \tag{37}$$

Similarly, we use that $\gamma_t \in \Gamma(\mu_t,\mu_1)$ to obtain that

$$W_2^2(\mu_t,\mu_1) \le (1-t)^2 W_2^2((\mu_o,\mu_1). \tag{38}$$

We combine (37) and (38) to conclude that

$$W_2(\mu_o,\mu_t) + W_2(\mu_t,\mu_1) \le W_2(\mu_o,\mu_1).$$

This, together with the fact that W_2 satisfies the triangle inequality yields that

$$W_2(\mu_o,\mu_t) + W_2(\mu_t,\mu_1) = W_2(\mu_o,\mu_1)$$

and

$$W_2(\mu_o,\mu_t) = tW_2(\mu_o,\mu_1),\ W_2(\mu_t,\mu_1) = (1-t)W_2(\mu_o,\mu_1).$$

Hence, $t \to \mu_t$ is a geodesic for W_2, which is parametrized by the arc length.

3.2 Connecting paths of minimal energy

Fact 1. *(Calculus of variations) Assume that ρ_o is a probability density of bounded support K, that is bounded away from 0 on K. Assume K is a set of smooth boundary whose interior is Ω. Define*

$$\bar{E}[v] = \int_{\mathbf{R}^d} \frac{|v|^2}{2}\rho_o(x)dx = \int_\Omega \frac{|v|^2}{2}\rho_o(x)dx \qquad (\ v \in L^2(\Omega,\mathbf{R}^d,\rho_o)\ ).$$

Assume that $\phi \in W^{1,2}(\Omega,\rho_o)$, that $v \in L^2(\Omega,\mathbf{R}^d,\rho_o)$, that

$$\mathrm{div}(\rho_o v) = \mathrm{div}(\rho_o \nabla\phi) \qquad in \quad \Omega$$

and

$$< v; \mathbf{n} > = < \nabla\phi; \mathbf{n} > \qquad on \quad \partial\Omega,$$

where $\mathbf{n}$ is the unit outward normal to $\partial\Omega$. Using the fact that $z \to |z|^2/2$ is strictly convex, we have that

$$\frac{|v|^2}{2} > \frac{|\nabla\phi|^2}{2} + < \nabla\phi, v - \nabla\phi >, \tag{39}$$

except at the points where $v = \nabla\phi$, *in which case, equality holds. We multiply both sides of (39) by* ρ_o *and integrate the subsequent equation over* Ω *to obtain that*

$$\begin{aligned}\bar{E}[v] &\geq \bar{E}[\nabla\phi] + \int_\Omega < \nabla\phi, v - \nabla\phi > \rho_o dx \\ &= \bar{E}[\nabla\phi] - \int_\Omega \phi \text{div}[(v - \nabla\phi)\rho_o dx = \bar{E}[\nabla\phi]. \end{aligned} \tag{40}$$

To obtain the last equality, we have used that $\text{div}(\rho_o v) = \text{div}(\rho_o \nabla\phi)$ *in* Ω *and* $< v; \mathbf{n} >=< \nabla\phi; \mathbf{n} >$ *on* $\partial\Omega$. *Inequality in (40) is strict unless,* $v = \nabla\phi$ ρ_o*–almost everywhere. Hence, given a function* $f : \Omega \to \mathbf{R}$, *the energy* $\bar{E}$ *is minimized over the set*

$$\{v \in L^2(\Omega, \mathbf{R}^d, \rho_o) \mid f = -\text{div}(\rho_o v)\}$$

by a unique v_o, *which is characterized by the fact that it satisfies* $v_o = \nabla\phi$ *for some function* ϕ, *provided that we know that such a* ϕ *exists (an argument for the existence of* ϕ *will be given soon, under the necessary and sufficient condition that* f *is of null average).*

Fact 2. *(Calculus of variations) Assume that* ρ_o *and* ρ_1 *are probability densities on* $\mathbf{R}^d$. *For each map* $T : \mathbf{R}^d \to \mathbf{R}^d$ *such that* $T_\# \rho_o = \rho_1$, *we define*

$$\mathcal{G}_T = \{\mathbf{g} : [0,1] \times \mathbf{R}^d \to \mathbf{R}^d \mid \mathbf{g}(0,x) = x, \;\; \mathbf{g}(1,x) = Tx, \quad \mathbf{g}(\cdot, x) \in W^{1,2}((0,1), \mathbf{R}^d) \;\}$$

and introduce the energy functional

$$E[\mathbf{g}] = \int_{\mathbf{R}^d} \Big(\int_0^1 \frac{|\dot{\mathbf{g}}|^2}{2} dt \Big) \rho_o dx$$

By Jensen inequality , if $\mathbf{g} \in \mathcal{G}_T$ *and we set* $\mathbf{g}_T(t,x) = (1-t)x + tTx$ *then*

$$\int_0^1 \frac{|\dot{\mathbf{g}}|^2}{2} dt \geq |Tx - x|^2 = \int_0^1 \frac{|\dot{\mathbf{g}}_T|^2}{2} dt.$$

This, together with (34) yields that

$$\inf_{\mathbf{g}} \{E[\mathbf{g}] \mid \mathbf{g}(1, \cdot)_\# \rho_o = \rho_1\} = \inf_T \{E[\mathbf{g}_T] \mid T_\# \rho_o = \rho_1\} = W_2^2(\rho_o, \rho_1). \tag{41}$$

Fact 3. *(Fluids mechanic) Assume that* $v : [0,1] \times \mathbf{R}^d \to \mathbf{R}^d$ *is a smooth vector field and associate to it the flow* $\mathbf{g}$ *defined by*

$$\begin{cases} \dot{\mathbf{g}}(t,x) = & v(t, \mathbf{g}(t,x)) \quad t \in [0,1], \;\; x \in \mathbf{R}^d \\ \mathbf{g}(0,x) = & x, \quad x \in \mathbf{R}^d. \end{cases} \tag{42}$$

Then $\mathbf{g}$ *is also smooth and invertible. Conversely, if* $\mathbf{g} : [0,1] \times \mathbf{R}^d \to \mathbf{R}^d$ *is smooth, invertible and satisfies* $\mathbf{g}(0,x) = x$ *then* v *defined by*

$$v(t,y) = \dot{\mathbf{g}}(t, \mathbf{g}^{-1}(t,y))$$

is such that (42) holds. Assume next that (42) holds and that $\rho(t,\cdot)$ is a probability density. One can check that the following are equivalent

(i) $\mathbf{g}(t,\cdot)_{\#}\rho_o = \rho(t,\cdot)$

(ii) $\frac{\partial \rho(t,\cdot)}{\partial t} + \mathrm{div}(\rho(t,\cdot)v) = 0$ *and* $\rho(0,\cdot) = \rho_o$.

It is straightforward to verify that

$$\int_0^1 dt \int_{\mathbf{R}^d} \frac{|\dot{\mathbf{g}}(t,x)|^2}{2}\rho_o(x)dx = \int_0^1 dt \int_{\mathbf{R}^d} \frac{|v(t,y)|^2}{2}\rho(t,y)dy$$

Assuming that the support of ρ_o was compact for establishing fact 1, was imposed to make our arguments easy and was not necessary. For a complete proof of facts 1–3, in a large generality, we refer the reader to [14].

Corollary 3.8 (Conclusions reached from Facts 1– 3 and and (41).). *Assume that ρ_o and ρ_1 are two probability densities. Then,*

$$W_2^2(\rho_o, \rho_1) = \inf_{v,\rho(t,\cdot)} \{ \int_0^1 dt \int_{\mathbf{R}^d} \frac{|v(t,y)|^2}{2}\rho(t,y)dy \mid \frac{\partial \rho(t,\cdot)}{\partial t} + \mathrm{div}(\rho(t,\cdot)v) = 0\}.$$

Furthermore, the infimum is attained in by a vector field $v_o(t,y) = \nabla\phi(t,y)$.

In fact, explicit computations in [14] show that if ψ is a convex function such that $(\nabla\psi)_{\#}\rho_o = \rho_1$ and ψ^* is its Legendre transform, then

$$v_o(t,y) = \frac{y - \nabla\psi_t^*(y)}{t}, \qquad \text{where} \quad \psi_t(x) = (1-t)\frac{|x|^2}{2} + t\psi(x).$$

Definition of Tangent spaces associated to elements of $\mathcal{M} = \mathcal{P}^2$. *We recall that $\mathcal{P}^2$ denotes the set of Borel probability measures on $\mathbf{R}^d$, of bounded second moments. For $\mu_o \in \mathcal{M}$, we would like to describe the tangent space $\mathbf{T}_{\mu_o}\mathcal{M}$ at μ_o, to make $\mathcal{M}$ a Riemannian manifold. A description for general measures μ_o needs a lot of care, but is possible (see [5]). To make our definiton easy, we are going to assume for instance that $\mu_o = \rho_o dx$ and that ρ_o is bounded away from 0 on its support K.*

If $t \to \rho(t,\cdot)$ is a path in the set of probability densities on $\mathbf{R}^d$ such that $\rho(0,\cdot) = \rho_o$ then

$$\frac{d}{dt}\int_{\mathbf{R}^d} \rho(t,x)dx = \frac{d}{dt}1 = 0.$$

This shows that the tangent vector $\frac{\partial \rho(0,\cdot)}{\partial t}$ at $\rho_o = \rho(0,\cdot)$ must be of null average. Conversely, assume that f is a function of null average. Assume in addition that the support of f is contained in the support of ρ_o. If we set $\rho(t,x) = \rho_o(x) + tf$, we obtain that $t \to \rho(t,\cdot)$ is a path in the set of probability densities on $\mathbf{R}^d$, for small t and $\frac{\partial \rho(0,\cdot)}{\partial t} = f$. This completes the argument that the tangent space at ρ_o is "exactly" the set of functions on $\mathbf{R}^d$ of null average. Given a function f on $\mathbf{R}^d$ of null average, there are infinitely many vectors field v such that

$$f = -\mathrm{div}(\rho_o v). \tag{43}$$

We select the special solution of (43) that is of the form

$$v_o = \nabla \phi_o$$

for some function ϕ. *In other words,*

$$f = -\text{div}(\rho \nabla \phi_o). \tag{44}$$

To show that such a ϕ_o *exists, one may use the following direct argument (of the calculus of variations) by assuming again that* f *is supported by* K. *Let* Ω *denotes the interior of* K. *We minimize* $\int_\Omega |\nabla \phi|^2 \rho_o - \phi f$ *over* $W_o^{1,2}(\Omega)$ *and obtain a unique minimizer* ϕ_o. *The Euler-Lagrange equations of that minimization problem is (44). Thanks to (44), we may identify the tangent space at* ρ_o, *which is the set of functions* f *of null average, with the set*

$$\mathbf{T}_{\rho_o}\mathcal{M} = \{\nabla \phi \mid \nabla \phi \in L^2(\mathbf{R}^d, \mathbf{R}^d, \rho_o)\ \}.$$

Assume that $f_1, f_2 \in \mathbf{T}_{\rho_o}\mathcal{M}$ *are two functions on* $\mathbf{R}^d$ *of null average. We define the inner product between* f_1 *and* f_2 *to be* $\mathbf{T}_{\rho_o}$

$$< f_1; f_2 >_{\rho_o} = 1/2 \int_{\mathbf{R}^d} < \nabla \phi_1, \nabla \phi_2 > \rho_o dx$$

where,

$$f_i = -\text{div}(\rho_o \nabla \phi_i) \qquad (i = 1, 2).$$

By corollary 3.8 we have that if ρ_1 *is another probability density of* $\mathbf{R}^d$ *then*

$$W_2^2(\rho_o, \rho_1) = \inf_{\rho(t,\cdot)} \{ < \frac{\partial \rho(t,\cdot)}{\partial t}; \frac{\partial \rho(t,\cdot)}{\partial t} >_{\rho(t,\cdot)} \ \mid \rho(o,\cdot) = \rho_o,\ \rho(1,\cdot) = \rho_1 \}.$$

This proves that the Wasserstein distance is consistent with the inner products $< \cdot ; \cdot >_\rho$.

The exponential map on $\mathcal{M} = \mathcal{P}(\mathbf{R}^d)$. *Let* $e \in C_o^2(\mathbf{R}^d)$. *Then for* t *small enough,* $x \to \frac{|x|^2}{2} + te = \psi_t(x)$ *is a strictly convex function since* $\nabla^2 \psi_t = I + t\nabla^2 e \geq I/2$ *for* t *small enough, and so,* $\nabla \psi_t^*$, *the inverse of* $\nabla \psi_t$, *exists and is Lipschitz for these values of* t. *As a consequence, the push forward by* $\nabla \psi_t$ *of a measure which is absolutely continuous with respect to Lebesgue measure, is itself absolutely conitnuous with respect to Lebesgue. Let* ρ_o *be a probablity density on* $\mathbf{R}^d$ *and define*

$$\rho(t,\cdot) = \nabla \psi_{t\#} \rho_o.$$

The path $t \to \rho(t,\cdot)$ *is a geodesic in* $\mathcal{P}(\mathbf{R}^d)$ *and by corollary 3.2,*

$$W_2^2(\mu_o, \mu_t) = 1/2 \int_{\mathbf{R}^d} |x - \nabla \psi_t|^2 \rho_o dx.$$

We set

$$\rho(t,\cdot) = exp_{\rho_o}(t \nabla e)$$

where we have identified ∇e with the tangent vector $f = -\mathrm{div}(\rho_o \nabla e)$.

Dilatations and translations on $\mathcal{M}$. *Assume that $\lambda > 0$ and that $\mathbf{u} \in \mathbf{R}^d$. We define the dilatation operator from $\mathcal{P}(\mathbf{R}^d)$ into $\mathcal{P}(\mathbf{R}^d)$ by*

$$\mu \to \mu^\lambda = D^\lambda_\# \mu.$$

where $D^\lambda x = \lambda x$. In other words,

$$\mu^\lambda[A] = \mu[\frac{A}{\lambda}]$$

Let δ_0 be the dirac mass at the origin. We use remark 3.4 twice to obtain that

$$W_2^2(\mu^\lambda, \delta_0) = 1/2 \int_{\mathbf{R}^d} |y|^2 d\mu^\lambda(y) = 1/2 \int_{\mathbf{R}^d} |D^\lambda x|^2 d\mu = \lambda^2/2 \int_{\mathbf{R}^d} |x|^2 d\mu = \lambda^2 W_2^2(\mu, \delta_0)$$

Note that $\mu \to \mu^\lambda$ can be viewed as an extension of D^λ from $\mathbf{R}^d$ to $\mathcal{P}(\mathbf{R}^d)$ ($\mathbf{R}^d$ can be considered as a subset of the $\mathcal{P}(\mathbf{R}^d)$ through the imbedding $x \to \delta_x$).

We next define the translation operator from $\mathcal{P}(\mathbf{R}^d)$ into $\mathcal{P}(\mathbf{R}^d)$ by

$$\mu \to \mu^{\mathbf{u}}[A] = T^{\mathbf{u}}_\# \mu,$$

where $T^{\mathbf{u}} x = x - \mathbf{u}$. In other words, $\mu^{\mathbf{u}}[A] = \mu[A - \mathbf{u}]$. Let $\mu, \nu \in \mathcal{P}(\mathbf{R}^d)$ and assume that

$$W_2^2(\mu, \nu) = 1/2 \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^2 d\gamma(x, y)$$

for some $\gamma \in \Gamma(\mu, \nu)$. Then $\gamma_1 = (T^{\mathbf{u}} \times T^{\mathbf{u}})_\# \gamma \in \Gamma(\mu^{\mathbf{u}}, \nu^{\mathbf{u}})$. Hence,

$$\begin{aligned} W_2^2(\mu^{\mathbf{u}}, \nu^{\mathbf{u}}) \leq 1/2 \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^2 d\gamma_1(x, y) &= 1/2 \int_{\mathbf{R}^d \times \mathbf{R}^d} |T^{\mathbf{u}} x - T^{\mathbf{u}} y|^2 d\gamma(x, y) \\ &= 1/2 \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^2 d\gamma(x, y) \\ &= W_2^2(\mu, \nu). \end{aligned} \tag{45}$$

Since $\gamma = (T^{-\mathbf{u}} \times T^{-\mathbf{u}})_\# \gamma_1$, an analoge reasoning yields that

$$W_2^2(\mu, \nu) \leq W_2^2(\mu^{\mathbf{u}}, \nu^{\mathbf{u}}). \tag{46}$$

We combine (45) and (46) to conclude that $W_2^2(\mu, \nu) = W_2^2(\mu^{\mathbf{u}}, \nu^{\mathbf{u}})$. Note that $\mu \to \mu^{\mathbf{u}}$ can be viewed as an extension of $T^{\mathbf{u}}$ from $\mathbf{R}^d$ to $\mathcal{P}(\mathbf{R}^d)$. This justify why we call $\mu \to \mu^{\mathbf{u}}[A]$ a translation.

3.3 differentiating the entropy functional in $\mathcal{M}$

We learn how to differentiate functions with respect to W_2. Assume that $H : \mathcal{M} \to \mathbf{R}$ and we want to compute the gateaux derivatives of H at a point ρdx. If we choose a "direction" f and to treat $\rho + rf$ as a density function so that the derivative of $H(\rho + rf)$ with respect to r at the 0 gives directional derivates then, we have to worry that there maybe points where $\rho + rf$ becomes negative. A better alternative for describing paths in $\mathcal{P}(\mathbf{R}^d)$, that originate at $\mu_o \in \mathcal{P}(\mathbf{R}^d)$ is to choose vector fields $\xi \in C^1_o(\mathbf{R}^d, \mathbf{R}^d)$ and set

$$\mu_r = (\mathbf{id} + r\xi)_{\#}\mu_o \in \mathcal{P}(\mathbf{R}^d).$$

These arguments are well developed in many works, including [5] [14] [16] [35]. To avoid listing the assumptions these functionals must satisfy to be differentiable, we choose to work with the following examples. First, we define

$$W(x) = \frac{|x - \mathbf{u}|^2}{2}, \qquad (x \in \mathbf{R}^d)$$

and

$$A(t) = t \ln t, \qquad t \geq 0.$$

Then we consider the functionals

$$\rho \to P(\rho) = \int_{\mathbf{R}^d} W\rho\, dx, \qquad S(\rho) = \int_{\mathbf{R}^d} A(\rho) dx$$

on the set of probability densities. The convexity of S along geodesics of $\mathcal{P}(\mathbf{R}^d)$ is ensured if one assumes that $t \to t^d A(t^{-d})$ is convex and nonincreasing while P is convex along geodesics of $\mathcal{P}(\mathbf{R}^d)$ if and only if W is convex on $\mathbf{R}^d$. Furthermore, the greatest lower bound of the Hessian $\nabla^2 W$ in $\mathbf{R}^d$ coincides with the greatest lower bound of the Hessian $\nabla^2_{\mathcal{P}(\mathbf{R}^d)} P$ in $\mathcal{P}(\mathbf{R}^d)$ (see [16]). To verify these claims, we are to compute the first and second derivatives of $r \to S(\rho^r), P(\rho^r)$ when, $r \to \rho^r$ is a geodesic. The next computations show that $\frac{d}{dr}S(\rho^r)$, $\frac{d}{dr}P(\rho^r)$ and $\frac{d^2}{dr^2}P(\rho^r)$ are straightforward to compute. The second derivative $\frac{d^2}{dr^2}S(\rho^r)$ is more involved, and we refer the reader to [35] for that.

Let ρ_1 be a probability density on $\mathbf{R}^d$, let $\xi \in C^1_o(\mathbf{R}^d, \mathbf{R}^d)$ and define

$$\rho_1^r = (\mathbf{id} + r\xi)_{\#}\rho_1,$$

where $\mathbf{id}$ is the identity map. Since $T_r := \mathbf{id} + r\xi$ is smooth and one-to-one for r small, we have that

$$\rho_1^r(T_r x) det[\nabla T_r x] = \rho_1(x) \tag{47}$$

for these r.

Note that

$$P(\rho_1^r) = \int_{\mathbf{R}^d} W(y)\rho_1^r(y) dy = \int_{\mathbf{R}^d} W(T_r x)\rho_1(x) dx.$$

Differentiating, we obtain that

$$\frac{d}{dr}P(\rho_1^r) = \int_{\mathbf{R}^d} < \nabla W(T_r x); \xi(x) > \rho_1(x)dx \tag{48}$$

and

$$\frac{d^2}{dr^2}P(\rho_1^r) = \int_{\mathbf{R}^d} < \nabla^2 W(T_r x)\xi(x); \xi(x) > \rho_1(x)dx \geq \int_{\mathbf{R}^d} |\xi(x)|^2 \rho_1(x)dx. \tag{49}$$

In particular, if $\xi = \nabla e$, *is the gradient of a function* e, *as already observed in the previous subsection while defining the exponential map on* $\mathcal{M}$, $r \to \rho_1^r$ *is a geodesic for* W_2 *and the inequality in (49) proves that the the Hessian of* $\nabla^2_{\mathcal{P}(\mathbf{R}^d)} P$ *in* $\mathcal{P}(\mathbf{R}^d)$ *is bounded below by* 1. *For instance, if we assume that* $||\nabla^2 e||_{L^\infty(\mathbf{R}^d)}$ *is small enough then,* $x \to |x|^2/2 + e(x) = T_1 x$ *is convex and*

$$\int_{\mathbf{R}^d} |\xi(x)|^2 \rho_1(x)dx = 2W_2^2(\rho_1, T_{1\#}\rho_1).$$

This, together with (49) gives that

$$\frac{d^2}{dr^2}P(\rho_1^r)|_{r=0} = \int_{\mathbf{R}^d} |\xi(x)|^2 \rho_1(x)d = 2W_2^2(\rho_1, T_{1\#}\rho_1).$$

Remark 3.9. *We have proven that the Hessian of* P *with respect to* W_2 *is bounded below by* 1 *whereas,* P *is merely affine with respect to the* L^2 *metric and its Hessian is null with respect to that metric.*

We again use that $T_{r\#}\rho_1 = \rho_1^r$ *to obtain that*

$$S(\rho_1^r) = \int_{\mathbf{R}^d} \rho_r(y)\ln(\rho_1^r(y))dy = \int_{\mathbf{R}^d} \rho_1(x)\ln(\rho_1^r(T_r x))dx.$$

This, together with (47) gives that

$$\begin{aligned} S(\rho_1^r) = \int_{\mathbf{R}^d} \rho_1(x)\ln\frac{\rho_1(x)}{det[\nabla T_r x]}dx &= S(\rho_1) - \int_{\mathbf{R}^d} \rho_1(x)\ln det[I + r\nabla\xi]dx \\ &= S(\rho_1) - \int_{\mathbf{R}^d} \rho_1(x)\ln(1 + r\mathrm{div}\xi + o(r))dx \\ &= S(\rho_1) - r\int_{\mathbf{R}^d} \rho_o(x)\mathrm{div}\xi dx + o(r). \end{aligned} \tag{50}$$

This proves that

$$\frac{d}{dr}S(\rho_1^r)|_{r=0} = -\int_{\mathbf{R}^d} \rho_1(x)\mathrm{div}\xi dx = \int_{\mathbf{R}^d} < \nabla\rho_1(x); \xi > dx. \tag{51}$$

Assume that ρ_o, ρ_1 *are probability densities on* $\mathbf{R}^d$, *of bounded second moments. We use corollary 3.2 to conclude that there exists a convex function* ϕ *on* $\mathbf{R}^d$ *such that* $\nabla\phi_{\#}\rho_1 = \rho_o$ *and*

$$W_2^2(\rho_o, \rho_1) = 1/2\int_{\mathbf{R}^d} |y - \nabla\phi|^2 \rho_1 dy = \ 1/2\int_{\mathbf{R}^d} |x - \nabla\psi|^2 \rho_o dx, \tag{52}$$

where ψ is the Legendre transform of ϕ so that, by symmetry, $\nabla\psi_{\#}\rho_o = \rho_1$. Since $T_{r\#}\rho_1 = \rho_1^r$, we conclude that $(T_r \circ \nabla\psi)_{\#}\rho_o = \rho_1^r$. Hence,

$$\begin{aligned} W_2^2(\rho_o, \rho_1^r) &\leq 1/2 \int_{\mathbf{R}^d} |x - T_r \circ \nabla\psi(x))|^2 \rho_o dx \\ &= 1/2 \int_{\mathbf{R}^d} |x - \nabla\psi(x) - r\xi \circ \nabla\psi(x)|^2 \rho_o dx \\ &= 1/2 \int_{\mathbf{R}^d} |x - \nabla\psi(x)|^2 \rho_o dx - r \int_{\mathbf{R}^d} < x - \nabla\psi(x), \xi(\nabla\psi(x)) > \rho_o dx \\ &+ r^2/2 \int_{\mathbf{R}^d} |\xi(\nabla\psi(x))|^2 \rho_o dx. \end{aligned} \tag{53}$$

We use (52) and (53) to obtain that

$$W_2^2(\rho_o, \rho_1^r) \leq W_2^2(\rho_o, \rho_1) - r \int_{\mathbf{R}^d} < \nabla\phi(y) - y; \xi(y) > \rho_1 dy + +r^2/2 \int_{\mathbf{R}^d} |\xi(y)|^2 \rho_1 dy.$$

Hence,

$$\limsup_{r \to 0^+} \frac{W_2^2(\rho_o, \rho_1^r) - W_2^2(\rho_o, \rho_1)}{r} \leq - \int_{\mathbf{R}^d} < \nabla\phi(y) - y; \xi(y) > \rho_1 dy. \tag{54}$$

In fact, condition under which equality holds in (54), have been studied in [5].

4 Applications I: The linear Fokker-Planck equations

The application of the Monge Kantorovich theory to the linear Fokker-Planck equations was discovered by Jordan-Kinderlehrer-Otto in [31]. We also refer the reader to [15] where a more general Fokker-Planck equations were studied. Throughout this section, we define

$$\mathcal{P}_a^2(\mathbf{R}^d) := \{\rho \in L^1(\mathbf{R}^d), \rho \geq 0, \int_{\mathbf{R}^d} \rho(x) dx = 1, \int_{\mathbf{R}^d} |x|^2 \rho(x) dx < +\infty \}$$

and

$$E[\rho] = \int_{\mathbf{R}^d} (\rho \ln \rho + W\rho) dx,$$

where $W(x) = |x - \bar{\mathbf{u}}|^2/2$ and $\bar{\mathbf{u}} \in \mathbf{R}^d$.

We assume that $\rho_o \in \mathcal{P}_a^2(\mathbf{R}^d)$ and consider the linear Fokker-Planck equation

$$\begin{cases} \frac{\partial\rho}{\partial t} + \operatorname{div}\Big(\nabla\rho + \rho(x - \bar{\mathbf{u}})\Big) = 0 \\ \rho(0, \cdot) = \rho_o. \end{cases} \tag{55}$$

The following algorithm was used in [31] to solve (55). Fix $h > 0$ a time step size. Define inductively ρ_{k+1} to be the unique solution of the variational problem

$$\inf_{\rho \in \mathcal{P}_a^2(\mathbf{R}^d)} W_2^2(\rho_k, \rho) + hE[\rho] \tag{56}$$

and set

$$\rho^h(t,x) = \begin{cases} \rho_o(x) & \textit{if} \quad t=0, x \in \mathbf{R}^d \\ \rho_{k+1} & \textit{if} \quad t \in (t_k, t_{k+1}], x \in \mathbf{R}^d. \end{cases} \tag{57}$$

Here, we have set $t_k = kh$. When h tends to 0 then $\{\rho^h\}_{h>0}$ converges weakly to the solution of (55). The novel here is not the existence of solutions of (55) but, the scheme used to construct these solutions.

4.1 Properties of the entropy functional

Definition 4.1 (Local Maxwellians). *When $\rho \in \mathcal{P}_a^2(\mathbf{R}^d)$, we define its average $\mathbf{u}_\rho$ and its variance θ_ρ by*

$$\mathbf{u}_\rho = \int_{\mathbf{R}^d} x\rho(x)dx, \qquad \theta_\rho = \frac{1}{d}\int_{\mathbf{R}^d} |x-\mathbf{u}|^2\rho.$$

(ii) If $\theta > 0$ and $\mathbf{u} \in \mathbf{R}^d$ we define the local Maxwellians

$$M_{\mathbf{u},\theta}(x) = \frac{1}{\sqrt{2\pi\theta}^d} e^{-\frac{|x-\mathbf{u}|^2}{2\theta}}.$$

Remark 4.2. *(i) One can check that $M_{\mathbf{u},\theta}$ is a probability density whose average is $\mathbf{u}$ and whose variance is θ.*

(ii) Note that if $\rho \in \mathcal{P}_a^2(\mathbf{R}^d)$ then its average $\mathbf{u}$ and its variance θ satisfy

$$\int_{\mathbf{R}^d} |x|^2\rho(x)dx = d\theta + |\mathbf{u}|^2. \tag{58}$$

(iii) Under the assumption in (ii), if $\bar{\mathbf{u}} \in \mathbf{R}^d$ then

$$\begin{aligned}\int_{\mathbf{R}^d} \frac{|x-\bar{\mathbf{u}}|^2}{2}\rho(x)dx = \int_{\mathbf{R}^d} \frac{|x-\mathbf{u}+\mathbf{u}-\bar{\mathbf{u}}|^2}{2}\rho(x)dx &= \int_{\mathbf{R}^d} \frac{|x-\mathbf{u}|^2}{2}\rho(x)dx + \frac{|\bar{\mathbf{u}}-\mathbf{u}|^2}{2} \\ &= \frac{d}{2}\theta + \frac{|\bar{\mathbf{u}}-\mathbf{u}|^2}{2}. \end{aligned} \tag{59}$$

Lemma 4.3. *Assume that $\rho \in \mathcal{P}_a^2(\mathbf{R}^d)$ is a probability density on $\mathbf{R}^d$ whose average is $\mathbf{u}$ and whose variance is θ. Then (i)*

$$\rho\ln\rho \geq \rho\ln M_{\mathbf{u},\theta} + \rho - M_{\mathbf{u},\theta} = -\frac{d}{2}\rho\Big(\ln(2\pi\theta) + \frac{|x-\mathbf{u}|^2}{d\theta}\Big) + \rho - M_{\mathbf{u},\theta}, \tag{60}$$

(ii)

$$|\rho\ln\rho| \leq \rho\ln\rho + M_{\mathbf{u},\theta} + \frac{d}{2}\rho\Big(|\ln(2\pi\theta)| + \frac{|x-\mathbf{u}|^2}{d\theta}\Big) \tag{61}$$

(iii)

$$S(\rho) \geq S(M_{\mathbf{u},\theta}) = -\frac{d}{2}(\ln(2\pi\theta)+1). \tag{62}$$

Proof: *Since* $t \to A(t) = t \ln t$ *is convex, we have that* $A(t) \geq A(t_o) + A'(t_o)(t - t_o)$. *Applying that inequality to*

$$t = \rho, \quad t_o = M_{\mathbf{u},\theta}$$

we obtain (i). To obtain (iii), we simply integrate both sides of (i) over $\mathbf{R}^d$ *and check that all the inequalities that led to (i) are in fact equalities when* $\rho = M_{\mathbf{u},\theta}$.

Note that (ii) holds if $\rho \ln \rho \geq 0$. *Next, if* $\rho \ln \rho \leq 0$, *then the positive part of* $\rho \ln \rho$ *is null and so, by (i), its negative part satisfies*

$$(\rho \ln \rho)_- \leq \frac{d}{2}\rho\Big(|\ln(2\pi\theta)| + \frac{|x - \mathbf{u}|^2}{d\theta}\Big) + M_{\mathbf{u},\theta}.$$

This concludes the proof of (ii). QED.

Corollary 4.4. *(i) If* $\rho \in \mathcal{P}_a^2(\mathbf{R}^d)$ *then its entropy* $S(\rho)$ *is well defined and is finite.*

(ii) For each constant $C < +\infty$ *there exists a constant* k_C *depending only on* C *and on* $\bar{\mathbf{u}}$ *(the average that occurs in the definition of* W *and hence the definition of* E*), such that the inclusion* $\mathcal{K}_C \subset \mathcal{K}'_C$ *holds for*

$$\mathcal{K}_C := \{\rho \in \mathcal{P}_a^2(\mathbf{R}^d) \mid E[\rho] \leq C\}, \qquad \mathcal{K}'_C := \{\rho \in \mathcal{P}_a^2(\mathbf{R}^d) \mid \frac{1}{k_C} \leq |\mathbf{u}_\rho|^2, \theta_\rho \leq k_C\}.$$

(iii) If $C < +\infty$ *then, there exists a constant* l_C *depending only on* C *and on* $\bar{u}$ *such that*

$$\mathcal{K}_C \subset \{\rho \in \mathcal{P}_a^2(\mathbf{R}^d) \mid \int_{\mathbf{R}^d} |\rho \ln \rho| dx \leq l_C\}.$$

Hence, $\mathcal{K}_C$ *is weakly compact in* $L^1(\mathbf{R}^d)$.

Proof: *Lemma 4.3 gives (i). We use remark 4.2 and lemma 4.3 (iii), to obtain that if* $\rho \in \mathcal{P}_a^2(\mathbf{R}^d)$ *and* $E[\rho] \leq C < +\infty$ *then*

$$-\frac{d}{2}(\ln(2\pi\theta_\rho) + 1) + \frac{d}{2}\theta_\rho + \frac{|\mathbf{u}_\rho - \bar{\mathbf{u}}|^2}{2} \leq C. \tag{63}$$

Thus,

$$-\frac{d}{2}(\ln(2\pi\theta_\rho) + 1) + \frac{d}{2}\theta_\rho \leq C. \tag{64}$$

Since

$$\lim_{a \to +\infty} -\frac{d}{2}(\ln(2\pi a) + 1) + \frac{d}{2}a = +\infty \quad \text{and} \quad \lim_{a \to 0^+} -\frac{d}{2}(\ln(2\pi a) + 1) + \frac{d}{2}a = +\infty,$$

(64) implies that there exists a constant k *depending only on* C *such that*

$$\frac{1}{k} \leq \theta_\rho \leq k. \tag{65}$$

We combine (63) and (65) to conclude the proof of (ii).

If $\rho \in \mathcal{P}_a^2(\mathbf{R}^d)$ and $E[\rho] \leq C < +\infty$ then, (ii) of this corollary and lemma 4.3 (ii) give that

$$\int_{\mathbf{R}^d} |\rho \ln \rho| dx \leq C + 1 + \frac{d}{2}\Big(1 + \max_{a \in [\frac{1}{k_C}, k_C]}\Big) =: l_C,$$

which proves the inclusion claimed in (iii). Because $t \to t \ln t$ grows faster than linearly as t tends to $+\infty$, (iii) proves that $\mathcal{K}_C$ is weakly closed in $L^1(\mathbf{R}^d)$. QED.

4.2 The algorithm in (56) is well-posed

In this subsection, we set

$$I_k[\rho] = W_2^2(\rho_k, \rho) + hE[\rho].$$

Theorem 4.5. *Assume that $\rho_o \in \mathcal{P}_a^2(\mathbf{R}^d)$. Then inductively, there is a unique ρ_{k+1} minimizer in (56). Furthermore,*

(i) $E[\rho_{k+1}] \leq E[\rho_o]$.

(ii) Let ϕ_{k+1} be a convex function on $\mathbf{R}^d$ such that $\nabla \phi_{k+1\#} \rho_{k+1} = \rho_k$, as ensured by corollary 3.2 . We have that $\nabla \phi_{k+1}(y) = y + h(\nabla \ln \rho_{k+1} + \nabla W)$ on the support of ρ_{k+1}.

Proof: 1. *Let $\{\rho^n\}_{n=1}^\infty$ be a minimizing sequence in (56). We may assume without loss of generality that*

$$W_2^2(\rho^n, \rho_k) + hE[\rho^n] = I_k[\rho^n] \leq I_k[\rho_k] = W_2^2(\rho_k, \rho_k) + hE[\rho_k] = E[\rho_k]. \tag{66}$$

Corollary 4.4 and (66) yield that $\{\rho^n\}_{n=1}^\infty$ admits a subsequence we still label $\{\rho^n\}_{n=1}^\infty$, that is weakly convergent to some $\rho_{k+1} \in \mathcal{P}_a^2(\mathbf{R}^d)$. By exercise 3.3, $W_2^2(\cdot, \rho_k)$ is weakly lower semi-continuous on $L^1(\mathbf{R}^d)$. Standard arguments give that E is also weakly lower semicontinuous on $L^1(\mathbf{R}^d)$. Hence, I_k is weakly lower semicontinuous on $L^1(\mathbf{R}^d)$ and so,

$$\inf_{\rho \in \mathcal{P}_a^2(\mathbf{R}^d)} I_k[\rho] = \liminf_{n \to +\infty} I_k[\rho^n] \geq I_k[\rho_{k+1}].$$

This, proves that ρ_{k+1} is a minimizer in (56).

2. *It is easy to check that $W_2^2(\cdot, \rho_k)$ and P are convex. Since S is strictly convex, we conclude that $I_k = W_2^2(\cdot, \rho_k) + hE$ is strictly convex. This proves that (56) cannot admit another minimizer $\bar\rho \neq \rho_{k+1}$ since otherwise we would have that*

$$I_k[\frac{\rho_{k+1} + \bar\rho}{2}] < \frac{1}{2}(I_k[\rho_{k+1}] + I_k[\bar\rho]) = I_k[\rho_{k+1}],$$

which would contradict the minimality of I_k at ρ_{k+1}.

3. *Because $\rho_k, \rho_{k+1} \in \mathcal{P}_a^2(\mathbf{R}^d)$, corollary 3.2 gives the existence of a convex function ϕ_{k+1} on $\mathbf{R}^d$ such that $\nabla \phi_{k+1\#} \rho_{k+1} = \rho_k$ and*

$$W_2^2(\rho_k, \rho_{k+1}) = 1/2 \int_{\mathbf{R}^d} |y - \nabla \phi_{k+1}(y)|^2 \rho_{k+1}(y) dy. \tag{67}$$

Let $\xi \in C_o^1(\mathbf{R}^d, \mathbf{R}^d)$ and define

$$\rho_{k+1}^r = (\mathbf{id} + r\xi)_{\#}\rho_{k+1}.$$

We have that $\rho_{k+1}^r \in \mathcal{P}_a^2(\mathbf{R}^d)$. Using the partial derivatives computed in (48), (51), (54) and make the substitution

$$\rho_o \to \rho_k, \qquad \rho_1 \to \rho_{k+1}$$

and using the fact that ρ_{k+1} is a minimizer in (56), we obtain that

$$\begin{aligned} 0 \leq \limsup_{r \to 0^+} \frac{I_k[\rho_{k+1}^r] - I_k[\rho_{k+1}]}{r} \leq & - \int_{\mathbf{R}^d} < \nabla\phi_{k+1}(y) - y; \xi(y) > \rho_{k+1}(y)dy \\ & + h \int_{\mathbf{R}^d} < \nabla\rho_{k+1}(y); \xi(y) > dy \\ & + h \int_{\mathbf{R}^d} < \nabla W(y); \xi(y) > \rho_{k+1}(y)dy. \end{aligned} \tag{68}$$

We substitute ξ by $-\xi$ and use the fact that the expressions at the right handside of (68) depends linearly on ξ to conclude that

$$\begin{aligned} \int_{\mathbf{R}^d} < \nabla\phi_{k+1}(y) - y; \xi(y) > \rho_{k+1}(y)dy = & h \int_{\mathbf{R}^d} < \nabla\rho_{k+1}(y); \xi(y) > dy \\ & + h \int_{\mathbf{R}^d} < \nabla W(y); \xi(y) > \rho_{k+1}(y)dy \end{aligned} \tag{69}$$

for every $\xi \in C_o^1(\mathbf{R}^d, \mathbf{R}^d)$. This proves (ii). QED.

Lemma 4.6. *Assume that $\rho_o \in \mathcal{P}_a^2(\mathbf{R}^d)$ and let ρ_k be defined inductively as in theorem 4.5. Then, (i)*

$$E[\rho_k] \leq E[\rho_{k-1}] \leq \cdots \leq E[\rho_o].$$

(ii) There exists a constant C depending only on $\bar{\mathbf{u}}$ and ρ_o such that

$$\frac{1}{C} \leq \theta_{\rho_k} \leq C.$$

(iii) There exists a constant j_C that depends only on $\bar{\mathbf{u}}$ and ρ_o such that

$$\sum_{k=0}^{N-1} W_2^2(\rho_k, \rho_{k+1}) \leq h j_C$$

Proof: 1. *Since ρ_{k+1} is a minimizer in (56) we have that*

$$W_2^2(\rho_{k-1}, \rho_k) + hE[\rho_k] = I_{k-1}[\rho_k] \leq I_{k-1}[\rho_{k-1}] = hE[\rho_{k-1}]. \tag{70}$$

Hence,

$$E[\rho_k] \leq E[\rho_{k-1}] \leq \cdots \leq E[\rho_o],$$

which proves (i). We use (i) and corollary 4.4 (ii) to conclude that (ii) holds.

2. *We sum up both sides of (70) over $k = 0, \cdots, N-1$, to obtain that*

$$\sum_{k=0}^{N-1} W_2^2(\rho_k, \rho_{k+1}) \leq h(E[\rho_o] - E[\rho_N]) \leq h(E[\rho_o] - S(\rho_N)). \tag{71}$$

By (ii) and lemma 4.3 ,

$$S(\rho_N) \geq -\frac{d}{2}(\ln(2\pi\theta_\rho) + 1) \geq -\frac{d}{2} \max_{a \in [\frac{1}{C}, C]} \{\ln(2\pi a) + 1\}. \tag{72}$$

We combine (71) and (72) to conclude the proof of (iii). *QED.*

Theorem 4.7. *Assume that $\rho_o \in \mathcal{P}_a^2(\mathbf{R}^d)$ and that $T > 0$. We introduce an integer parameter N and set $h = T/N$. Let $\{\rho^h\}_{h>0}$ be defined as in (57). Then, $\{\rho^h\}_{h>0}$ is weakly compact in $L^1((0,T) \times \mathbf{R}^d)$ and converges weakly to the solution of (55).*

Proof: *It suffices to show that any arbitrary subsequence of $\{\rho^h\}_{h>0}$ has itself a subsequence that converges weakly to the solution of (55) and show that (55) admits a unique solution. Here, we will skip the uniqueness proof which can be found in standard partial differential equations books. We also refer the reader to [31] and [39] where the uniquenes of solution in (55) is obtained by using that E is uniformly convex along geodesics of W_2. The curious reader could consult [1] where a large class of parabolic equations was studied, using cost functions which don't lead to a metric. Recall that (FPE) stands for Fokker-Planck Equations.*

1. In what sense is $\rho_k \to \rho_{k+1}$ discretizing the (FPE). *Let $\eta \in C_o^2(\mathbf{R}^d)$. As in theorem 4.5, let be a ϕ_{k+1} be a convex function on $\mathbf{R}^d$ such that $(\nabla\phi_{k+1})_\# \rho_{k+1} = \rho_k$. We have that*

$$\begin{aligned}\int_{\mathbf{R}^d} (\rho_k - \rho_{k+1})\eta dx &= \int_{\mathbf{R}^d} \eta(x)\rho_k(x)d - \int_{\mathbf{R}^d} \eta(y)\rho_{k+1}(y)dy \\ &= \int_{\mathbf{R}^d} \eta(\nabla\phi_{k+1}(y))\rho_{k+1}(y)dy - \int_{\mathbf{R}^d} \eta(y)\rho_{k+1}(y)dy.\end{aligned} \tag{73}$$

Set

$$l(t) = \eta\Big((1-t)y + t\nabla\phi_{k+1}\Big)$$

so that

$$l'(t) =< \nabla\eta((1-t)y + t\nabla\phi_{k+1}); \nabla\phi_{k+1} - y > \tag{74}$$

and

$$l"(t) =< \nabla^2\eta((1-t)y + t\nabla\phi_{k+1})(\nabla\phi_{k+1} - y); \nabla\phi_{k+1} - y > \tag{75}$$

Taylor's expansion gives that

$$\eta(\nabla\phi_{k+1}) - \eta(y) = l(1) - l(0) = l'(0) + \int_0^1 dt \int_0^t l"(s)ds$$

We combine (73), (74) and (75) to conclude that

$$\int_{\mathbf{R}^d} (\rho_k - \rho_{k+1})\eta dx = \int_{\mathbf{R}^d} \nabla\eta; \nabla\phi_{k+1} - y > \rho_{k+1} dy$$
$$+ \int_{\mathbf{R}^d} \int_0^1 dt \int_0^t ds < \nabla^2\eta((1-t)y + t\nabla\phi_{k+1})(\nabla\phi_{k+1} - y); \nabla\phi_{k+1} - y > \rho_{k+1} dy. \tag{76}$$

1. a. Let us better understand the link between (76) and the (FPE). *We are next going to see how (76), in some sense, means that*

$$\frac{\rho_{k+1} - \rho_k}{h} = \triangle\rho_{k+1} + \operatorname{div}\Big(W\rho_{k+1}\Big) + 0(h). \tag{77}$$

Before making a rigorous argument, the following formal reasoning will help develop a better intuition. Note that using (67) we have the following estimate of the last term in (76)

$$|\int_{\mathbf{R}^d} \int_0^1 dt \int_0^t ds < \nabla^2\eta((1-s)y + s\nabla\phi_{k+1})(\nabla\phi_{k+1} - y); \nabla\phi_{k+1} - y > \rho_{k+1} dy|$$
$$\leq \frac{1}{2} ||\nabla^2\eta||_{L^\infty(\mathbf{R}^d)} \int_{\mathbf{R}^d} |\nabla\phi_{k+1} - y)|^2 \rho_{k+1} dy = ||\nabla^2\eta||_{L^\infty(\mathbf{R}^d)} W_2^2(\rho_k, \rho_{k+1}). \tag{78}$$

Theorem 4.5 (ii) tells us that we can make the substitution $\nabla\phi_{k+1}(y) = y + h(\nabla \ln\rho_{k+1} + \nabla W)$ ρ_{k+1}–almost everywhere in (76) and use (78) to obtain that

$$|\int_{\mathbf{R}^d} (\rho_k - \rho_{k+1})\eta dx - h\int_{\mathbf{R}^d} < \nabla\eta; (\nabla \ln\rho_{k+1} + \nabla W) > \rho_{k+1} dy|$$
$$\leq ||\nabla^2\eta||_{L^\infty(\mathbf{R}^d)} W_2^2(\rho_k, \rho_{k+1}) \leq h||\nabla^2\eta||_{L^\infty(\mathbf{R}^d)} (E[\rho_k] - E[\rho_{k+1}]). \tag{79}$$

If $E[\rho_k] - E[\rho_{k+1}]) = 0(h)$ then (79) gives (77). In fact, one does not attempt to prove that the "time pointwise" estimate (77) holds, to conclude the proof of this theorem. It is enough to have that (77) holds with "high probability". These comments are made rigorous in the next paragraph.

2. Establishing the rigorous estimates. *Let $\eta \in C_o^2([0,T) \times \mathbf{R}^d)$, set*

$$t_k = kh, \quad \eta_k = (t_k, \cdot).$$

We have that

$$\int_0^T dt \int_{\mathbf{R}^d} \frac{\partial\eta}{\partial t} \rho^h dx = \sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^d} \frac{\partial\eta}{\partial t} \rho_{k+1}\, dx = \sum_{k=0}^{N-1} \int_{\mathbf{R}^d} (\eta_{k+1} - \eta_k)\rho_{k+1} dx$$
$$= \sum_{k=0}^{N-1} \int_{\mathbf{R}^d} (\rho_k - \rho_{k+1})\eta_k dx + \int_{\mathbf{R}^d} (\rho_N\eta_N - \rho_o\eta_o)\, dx \tag{80}$$

The last equality in (80) was obtained by rearranging the terms of the previous expression. We next use that $\eta_N \equiv 0$, *combine (76) , (80) and substitute* $\nabla\phi_{k+1} - y$ *by* $h\nabla(\ln\rho_{k+1} + W)$ *to conclude that*

$$\int_0^T dt \int_{\mathbf{R}^d} \frac{\partial\eta}{\partial t}\rho^h dx = h\sum_{k=0}^{N-1} < \nabla\eta; \nabla(\ln\rho_{k+1} + W) > \rho_{k+1}dy - \int_{\mathbf{R}^d} \rho_o\eta_o dx$$

$$+ \sum_{k=0}^{N-1} \int_{\mathbf{R}^d}\int_0^1 dt \int_0^t ds < \nabla^2\eta((1-s)y + s\nabla\phi_{k+1})(\nabla\phi_{k+1} - y); \nabla\phi_{k+1} - y > \rho_{k+1}dy. \quad (81)$$

Next, observe that

$$\int_0^T dt \int_{\mathbf{R}^d} < \nabla\eta; \nabla(\ln\rho^h + W) > \rho^h dx = \sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^d} < \nabla\eta; \nabla(\ln\rho_{k+1} + W) > \rho_{k+1}dx$$

$$= \sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^d} < \nabla\eta - \nabla\eta_k; \nabla(\ln\rho_{k+1} + W) > \rho_{k+1}dx$$

$$+ h\sum_{k=0}^{N-1}\int_{\mathbf{R}^d} < \nabla\eta_k; \nabla(\ln\rho_{k+1} + W) > \rho_{k+1}dx$$

$$= \sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^d} < \nabla\eta - \nabla\eta_k; \frac{\nabla\phi_{k+1} - y}{h}) > \rho_{k+1}dx$$

$$+ h\sum_{k=0}^{N-1}\int_{\mathbf{R}^d} < \nabla\eta_k; \nabla(\ln\rho_{k+1} + W) > \rho_{k+1}dx \quad (82)$$

Set

$$B_T := (0,T) \times \mathbf{R}^d.$$

We combine (81) and (82) to obtain that

$$
\begin{aligned}
&|\int_0^T dt \int_{\mathbf{R}^d} (\frac{\partial \eta}{\partial t} + \triangle \eta - < \nabla W; \nabla \eta >) \rho^h dx + \int_{\mathbf{R}^d} \rho_o \eta_o dx| \\
&= |\int_0^T dt \int_{\mathbf{R}^d} (\frac{\partial \eta}{\partial t} - < \nabla \ln \rho^h + \nabla W; \nabla \eta >) \rho^h dx + \int_{\mathbf{R}^d} \rho_o \eta_o dx| \\
&= |\sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} dt \int_{\mathbf{R}^d} < \nabla \eta - \nabla \eta_k; \frac{\nabla \phi_{k+1} - y}{h} > \rho_{k+1} dx \\
&+ \int_{\mathbf{R}^d} \int_0^1 dt \int_0^t ds < \nabla^2 \eta((1-s)y + s\nabla \phi_{k+1})(\nabla \phi_{k+1} - y); \nabla \phi_{k+1} - y > \rho_{k+1} dy| \\
&\leq \frac{h^2}{2} ||\nabla^2_{t,x} \eta||_{L^\infty(B_T)} \sum_{k=0}^{N-1} \int_{\mathbf{R}^d} |\frac{\nabla \phi_{k+1} - y}{h}| \rho_{k+1} dy + ||\nabla^2_{xx} \eta||_{L^\infty(B_T)} \sum_{k=0}^{N-1} W_2^2(\rho_k, \rho_{k+1}) \\
&\leq h ||\nabla^2 \eta||_{L^\infty(B_T)} \Big(j_C + \sqrt{\frac{N h j_C}{2}} \Big) = h ||\nabla^2 \eta||_{L^\infty(B_T)} \Big(j_C + \sqrt{\frac{T j_C}{2}} \Big) \qquad (83)
\end{aligned}
$$

To obtain (83), we have exploited (78) and used the fact that

$$
\begin{aligned}
\sum_{k=0}^{N-1} \int_{\mathbf{R}^d} |\frac{\nabla \phi_{k+1} - y}{2h}| \rho_{k+1} dy &= \frac{1}{h\sqrt{2}} \sum_{k=0}^{N-1} \int_{\mathbf{R}^d} |\frac{\nabla \phi_{k+1} - y}{\sqrt{2}}| \rho_{k+1} dy \\
&\leq \frac{\sqrt{N}}{h\sqrt{2}} \sqrt{\sum_{k=0}^{N-1} \int_{\mathbf{R}^d} \frac{|\nabla \phi_{k+1} - y|^2}{2} \rho_{k+1} dy} \\
&= \frac{\sqrt{N}}{h\sqrt{2}} \sqrt{\sum_{k=0}^{N-1} W_2^2(\rho_k, \rho_{k+1})} \\
&\leq \frac{\sqrt{N}}{h\sqrt{2}} \sqrt{h j_C} = \frac{\sqrt{T j_C}}{h\sqrt{2}}. \qquad (84)
\end{aligned}
$$

Here, j_C is the constant given by lemma 4.6.

3. Weak compactness of $\{\rho^h\}_{h>0}$ in $L^1((0,T) \times \mathbf{R}^d)$. *Lemma 4.6 gives existence of a constant $C_1 \in (0, +\infty)$ independent of h such that*

$$
E[\rho^h(t, \cdot)] \leq C_1, \qquad \frac{1}{C_1} \leq \theta_{\rho^h(t,\cdot)} \leq C_1
$$

for all $t \in [0,T]$. This, together with corollary 4.4 gives existence of a constant $C_2 \in (0, +\infty)$ independent of h such that

$$
\sup_{t \in [0,T]} \int_{\mathbf{R}^d} |\rho^h \ln \rho^h| dx \leq C_2.
$$

This proves that $\{\rho^h\}_{h>0}$ *is weakly compact in* $L^1((0,T)\times\mathbf{R}^d)$ *and so, up to a subsequence, it converges weakly to some* $\rho\in\mathcal{P}_a^2(\mathbf{R}^d)$. *Letting* h *go to* 0 *in (83), we conclude that* ρ *is a solution of (55).* *QED.*

5 Application II: Semigeostrophic equations

The semi-geostrophic *systems are systems of partial differential equations which have two main attractions. First of all, they model how fronts arise in large scale weather patterns. Second, they describe a* 3*-D free boundary problem which is an approximation of the* 3*-D Euler equations of incompressible fluids, in a rotating coordinate frame around the* Ox_3*-axis where, the effects of rotation dominate. Establishing existence of solutions to the* 3*-D Euler equations of incompressible fluids, for large times, remains a challenge in partial differential equations. The* semi-geostrophic *systems keep many interesting feactures of the* 3*-D Euler equations incompressible and its time-space discretization via the Monge-Kantorovich theory raises a lot of hope for making substantial progress which could help better understanding the original* 3*-D Euler equations (see the review paper by L.C. Evans [24] or [7]). For keeping this manuscript short, in this section, we omit details and proofs, which the reader will find in [18] .*

The semi-geostrophic *systems were first introduced by Eliassen [23] in 1948, and rediscovered by Hoskins [30] in 1975. Since then, they have been intensively studied by geophysicists (e.g. [19], [20], [43], [45]). One of the main contributions of Hoskins was to show that the* semi-geostrophic system, *could be solved in particular cases by a coordinate transformation which then allows analytic solutions to be obtained. The subsequent new system is, at least formally, equivalent to the original* semi-geostrophic *system, and has the advantage to be more tractable. Hoskins claimed that, in the new system, the mechanisms for the formation of fronts in the atmosphere could be modelled analytically.*

Here, we consider a particular case of the semi-geostrophic *systems, the so-called the* semi-geostrophic shallow water system *(SGSW). We skip its derivation that can be found in [18]. In the system below,* h *represents the height of the water above a fixed region* Ω *and is related to what is called the generalized geopotential function*

$$P(t,x)=|x|^2/2+h(t,x),\quad (t\in[0,+\infty),\ \ h(t,x)\geq 0).$$

Let $P^*(t,\cdot)$ *be the* Legendre transform *of* $P(t,\cdot)$. *It is related to the geostrophic velocity* $\mathbf{w}$ *by*

$$\alpha:=\nabla P(t,\cdot)_{\#}\rho(t,\cdot),\quad \mathbf{w}(t,y)=J(y-\nabla P^*(t,y)). \tag{85}$$

The semigeostrophic shallow water in dual variables *are*

$$\begin{cases} (i) & \frac{\partial\alpha}{\partial t}+\operatorname{div}(\alpha\mathbf{w})=0 \quad \textit{in the weak sense in} \quad [0,T]\times\mathbf{R}^2\\ (ii) & \mathbf{w}(t,y):=J(y-\nabla P^*(t,y)), \quad \textit{in} \quad [0,T]\times\mathbf{R}^2\\ (iii) & P(t,x):=|x|^2/2+h(t,x), \quad \textit{in} \quad [0,T]\times\Omega\\ (iv) & \alpha(t,\cdot):=\nabla P(t,\cdot)_{\#}h(t,\cdot), \quad t\in[0,T]\\ (v) & \alpha(0,\cdot)=\alpha_o \quad \textit{in} \quad \mathbf{R}^2. \end{cases} \tag{86}$$

A time discretized scheme for solving the SWGS. *We fix a time step size $\delta > 0$. We consider the Hamiltonian*

$$H(\alpha) := 1/2 \min_{\eta \in \mathcal{P}_a(\Omega)} \{W_2^2(\alpha, \eta) + \int_\Omega \eta^2 dx\}. \tag{87}$$

Step 1. *Given $\alpha_k \in \mathcal{P}_a(\mathbf{R}^d)$, we substitute α in (87) and define h_k to be the unique minimizer of $H(\alpha_k)$. Let P_k be a convex function such that $(\nabla P_k)_\# h_k = \alpha_k$. The Euler-Lagrange equations of (87) give that*

$$P_k(x) = |x|^2/2 + h_k(x).$$

Step 2. *We set $\mathbf{w}_k(y) := J(y - \nabla P_k^*(y))$ where P_k^* denotes the Legendre transform of P_k. We solve the system of equations*

$$\begin{cases} \frac{\partial \alpha}{\partial t} + \mathrm{div}(\alpha \mathbf{w}_k) = 0 \qquad in \quad [k\delta, (k+1)\delta] \times \mathbf{R}^2 \\ \\ \alpha(k\delta, \cdot) := \alpha_k \qquad in \quad \mathbf{R}^2 \end{cases}$$

and set $\alpha_{k+1} = \alpha((k+1)\delta, \cdot)$.

An approximate solution of the SGS. *We define $\alpha^h(k\delta, \cdot) = \alpha_k$ and extend $\alpha^\delta(t, \cdot)$ on $(k\delta, (k+1)\delta)$ by linearly interpolating between α_k and α_{k+1}. In [18] we show the following theorem.*

Theorem 5.1 (Main existence result). *Assume that $1 < r < +\infty$, and that $\alpha_o \in L^r(B_r)$ is a probability density whose support is strictly contained in B_r, and let B_R be the ball of center 0, and radius $R := r(1+T)$. There exists $h \in L^\infty((0,T); W^{1,\infty}(\Omega))$ which is a limit point of $\{h^\delta\}_{\delta > 0}$ such that $h(t, \cdot) \in \mathcal{P}_a(\Omega)$. Furthermore, there exist a function $\alpha \in L^\infty((0,T); L^r(\mathbf{R}^d))$, such that (α, h) is a stable solution of (86) and*

$$W_1(\alpha(s_2, \cdot), \alpha(s_1, \cdot)) \leq C|s_1 - s_2|.$$

for all $s_1, s_2 \in [0,T]$. Here C is a constant that depends only on the initial data.

Open problem . Degenerate "hamiltonian" structure and uniqueness. *No standard method apply for studying uniqueness of solution for the SGS. The success of the current effort made by [5] to develop a rigorous tool that associate a riemannian structure to the Wasserstein distance is a step toward finding a systematic way of studying uniquess of solutions of some systems. Using that riemannian structure, we made more precised the degenerate "hamiltonian" structure of the SGS which we next explain: let $\mathcal{M}$ be the set of probability measures on $\mathbf{R}^d$. If $\omega_o \in \mathcal{M}$, the tangent space at ω_o is $T_{\omega_o}\mathcal{M} = \{f \mid \int_{\mathbf{R}^d} f dx = 0\}$. To each $f \in T_{\omega_o}\mathcal{M}$ we associate ψ_f defined by the PDE $-\mathrm{div}(\omega_o \nabla \psi_f) = f$. The inner product of $f, g \in T_{\omega_o}\mathcal{M}$ suggested by [39] is*

$$< f, g >_{\omega_o} = \int_{\mathbf{R}^d} \omega_o \nabla \psi_f \cdot \nabla \psi_g dx.$$

We propose to introduce the skew-symmetric form

$$\beta_{\omega_o}(f,g) = \int_{\mathbf{R}^d} \omega_o J \nabla \psi_f \cdot \nabla \psi_f dx,$$

where J *is the standard simplectic matrix such that* $-J^2$ *is the identity matrix. For instance if the physical domain is* Ω *is time independent, the SGS consists in finding* $t \to \omega(t,\cdot)$ *satisfying for all* $f \in T_{\omega(t,\cdot)}\mathcal{M}$,

$$< \dot{\omega}, f >= \beta_{\omega_o}\Big(\frac{\delta H}{\delta f}, f\Big). \tag{88}$$

Uniqueness of solution in (88) will be straightfoward to establish if H *was a smooth function. The question is to know how much we could exploit the fact that* H *is only semiconcave with respect to* W_2*. For which initial condition* $\omega(0,\cdot)$ *the variations of* H *matters only in some directions? This leads to the problem of understanding the kernel of* $\beta_{\omega_o}(f,\cdot)$*. When* $d = 2$*, the kernel of* $\beta_{\omega_o}(f,\cdot)$ *is the set of* g *such that* ω_o *and* ψ_g *have the same level set. This means that there exists a function a monotone function on* β *such that* $\psi_g(x) = -\beta(\omega(x))$*. Hence, for a convex function* A*, we have that* $A' = \beta$*. A flow along degenerate directions is given by*

$$\partial_t \omega = \mathrm{div}\Big[\omega \nabla \Big(A'(\omega)\Big)\Big]. \tag{89}$$

The question is to know how much (89) contributes to the understanding of (88).

6 Example of cost functions; fluids mechanic and dynamical systems

Many mechanical systems can be described via a lagrangian $L : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$*, defined on the phase space* $\mathbf{R}^d \times \mathbf{R}^d$*. Customarily,* $L \in C^r(\mathbf{R}^d \times \mathbf{R}^d)$*,* $L(x,\cdot)$ *satisfies some convexity assumptions and* $L(\cdot, v)$ *satisfies suitable growth or periodicity conditions that we call (A1)–(A4), in the appendix. Now, we introduce a Hamiltonian associated to* L*, the so-called Legendre transform of* $L(x,\cdot)$*. For* $(x,p) \in \mathbf{R}^d \times \mathbf{R}^d)$ *we set*

$$H(x,p) = \sup_{v \in \mathbf{R}^d} \{v \cdot p - L(x,v)\}, \qquad ((x,p) \in \mathbf{R}^d \times \mathbf{R}^d).$$

The Hamiltonian H *is continuous and by* (A3)*,* $H(x,\cdot)$ *is continuously differentiable. Also, the map*

$$(x,v) \to (x, \nabla_v L(x,v)) = \mathbf{T}(x,v)$$

is of class $C^{r-1}(\mathbf{R}^d \times \mathbf{R}^d)$ *and its restriction to* $\mathbf{T}^d \times \mathbf{R}^d$ *is one-to-one. It inverse*

$$(x,p) \to (x, \nabla_p H(x,p)) = \mathbf{S}(x,p)$$

is then of class $C^{r-1}(\mathbf{R}^d \times \mathbf{R}^d)$*. This proves that* H *is in fact of class* $C^r(\mathbf{R}^d \times \mathbf{R}^d)$*.*

One studies the kinematics and dynamics of these systems through the action

$$c(T, x_o, x_T) = \inf_\sigma \{ \int_0^T L(\sigma, \dot\sigma) dt \mid \sigma(0) = x_o, \sigma(T) = x_1 \}, \tag{90}$$

where the infimum is performed over the set $AC(T; x_o, x_T)$ of $\sigma : [0, T] \to \mathbf{R}^d$ that are absolutely continuous and that satisfy the endpoint constraint $\sigma(0) = x_o$, $\sigma(T) = x_T$. By rescalling, we may assume that $T = 1$.

Given two probability densities ρ_o and ρ_1 on $\mathbf{R}^d$, the Monge-Kantorovich problem is then

$$\inf_{\mathbf{r}_\# \rho_o = \rho_1} \int_{\mathbf{R}^d} c(1, x, \mathbf{r}(x)) \rho_o(x) dx = \inf_{\mathbf{g}(\cdot,\cdot)} \Big\{ \int_0^1 dt \int_{\mathbf{R}^d} L(x, \mathbf{g}(t, x)) \rho_o(x) dx \Big\} \tag{91}$$

where the infimum is performed over the set of $\mathbf{g} : [0, 1] \times \mathbf{R}^d \to \mathbf{R}^d$ such that $\mathbf{g}(0, x) = x$ and $\mathbf{g}(1, \cdot)_\# \rho_o = \rho_1$. The expression at the left handside of (91) is $W_{\bar c}$ where $\bar c = c(1, \cdot, \cdot)$.

When (91) admits a unique minimizer $\bar{\mathbf{g}}$ (see proposition 8.1 for a condition on L that ensures such properties), we define the path

$$\bar\rho(t, \cdot) = \bar{\mathbf{g}}(t, \cdot)_\# \rho_o \tag{92}$$

When $L(x, v) = |v|^p/p$ for some $p \in (0, +\infty)$ then $t \to \bar\rho(t, \cdot)$ is a geodesic for the Wasserstein distance (see [5] and [39]). The passage from Lagrangian to Eulerian coordinates is done through the ODE

$$\dot{\mathbf{g}}(t, x) = \mathbf{V}(t, \mathbf{g}(t, x)), \qquad \mathbf{g}(0, x) = x. \tag{93}$$

We combine (91) and (93) to obtain that

$$W_{\bar c}(\rho_o, \rho_1) = \inf_{\rho(\cdot,\cdot),\, \mathbf{V}} \Big\{ \int_0^1 dt \int_{\mathbf{R}^d} L(y, \mathbf{V}(t, y)) \rho(t, y) dy \Big\}, \tag{94}$$

where the infimum is performed over the set of pairs $(\rho, \mathbf{V})$ such that

$$\rho(0, \cdot) = \rho_0, \rho(1, \cdot) = \rho_1 \quad \text{and} \quad \frac{\partial \rho}{\partial t} + \mathrm{div}(\rho \mathbf{V}) = 0.$$

When $L(x, v) = |v|^2/2$, one can recognize the expression in (94) to coincide with the one in corollary 3.8.

7 Prerequisite for the mass transportation theory

7.1 Convex Analysis in $\mathbf{R}^d$

The material of this section can be found in the books [22], [42]. The solutions to the exercises in this section appear as theorems, lemma, and propositions in these books. Throughout this section Y, is a Banach space.

Definition 7.1. *Let $X \subset Y$ be a convex subset of Y and let $\phi : X \to \mathbf{R} \cup \{+\infty\}$ be a real valued function.*

(i) ϕ is said to be convex if ϕ is not identically $+\infty$ and

$$\phi((1-t)x + ty) \le (1-t)\phi(x) + t\phi(y)$$

for all $t \in [0,1]$ and all $x, y \in X$.

(ii) ϕ is said to be strictly convex if ϕ is not identically $+\infty$ and

$$\phi((1-t)x + ty) < (1-t)\phi(x) + t\phi(y)$$

for all $t \in (0,1)$ and all $x, y \in X$ such that $x \neq y$.

(iii) ϕ is said to be lower semicontinuous on X if

$$\liminf_{n \to +\infty} \phi(x_n) \ge \phi(x_\infty),$$

for every sequence $\{x_n\}_{n=1}^{+\infty} \subset X$ converging to $x_\infty \in X$.

Remark 7.2. *Suppose that $\phi : X \to \mathbf{R} \cup \{+\infty\}$ and we defined $\bar{\phi} : Y \to \mathbf{R} \cup \{+\infty\}$ by*

$$\bar{\phi}(x) = \begin{cases} \phi(x) & \text{if} \quad x \in X \\ +\infty & \text{if} \quad x \notin X. \end{cases} \tag{95}$$

Note that $\bar{\phi}$ is convex. We refer to it as the natural convex extension of ϕ.

Exercise 7.3. *(i) Show that ϕ is lower semicontinuous if and only if its epigraph* $\mathrm{epi}(\phi) = \{(x,t) \mid \phi(x) \le t\}$ *is closed.*

(ii) Show that ϕ is convex if and only if its epigraph is a convex set.

(iii) Is there any extra assumption one needs to impose on X for (i) and (ii) to hold?

Definition 7.4. *Assume that $X \subset Y$ is a convex set and that $\phi : X \to \mathbf{R} \cup \{+\infty\}$ is convex.*

(i) The subdifferential of ϕ is the set $\partial\phi \subset X \times Y$ that consists of the (x,y) such that

$$\phi(\mathbf{z}) \ge \phi(x) + y \cdot (\mathbf{z} - x)$$

for all $\mathbf{z} \in X$.

(ii) If $(x,y) \in \partial^c\phi$ we say that $y \in \partial\phi(x)$. If $E \subset X$ we denote by $\partial\phi(E)$ the union of the $\partial\phi(x)$ such that $x \in E$.

Definition 7.5. *Assume that $X \subset Y$ and that $\phi : X \to \mathbf{R} \cup \{+\infty\}$ is not identically $+\infty$. The Legendre transform of ϕ is the function $\phi^* : Y \to \mathbf{R} \cup \{+\infty\}$ defined by*

$$\phi^*(y) = \sup_{x \in X}\{x \cdot y - \phi(x)\}.$$

Remark 7.6. *Note that ϕ and its natural extension have the same Legendre transform.*

Exercise 7.7. *Assume that $\phi : Y \to \mathbf{R} \cup \{+\infty\}$ is convex and lower semicontinuous.*

(i) Show that ϕ^* is convex and lower semicontinuous (in particular ϕ^* is not identically $+\infty$).

(ii) Show that $\phi = \phi^{**} = C\phi$ where

$$C\phi = \sup\{g \mid g \leq \phi,\ g \quad convex\}.$$

(iii) Say whether or not the following hold:

$$(x, y) \in \partial\phi \Longleftrightarrow (y, x) \in \partial\phi^*$$

(iv) Conclude the $\nabla\phi(\nabla\phi^*(x)) = x$ whenever $y := \nabla\phi^*(x)$ exists and ϕ is differentiable at y.

Definition 7.8. *A subset $Z \subset Y \times Y$ is said to be cyclically monotone if for every natural number n, for every $\{(x_i, y_i)\}_{i=1}^n \subset Z$ and every permutation σ of n letter, we have that*

$$\sum_{i=1}^n |x_i - y_i|^2 \leq \sum_{i=1}^n |x_i - y_{\sigma(i)}|^2.$$

Exercise 7.9. *Show that $Z \subset \mathbf{R}^d \times \mathbf{R}^d$ is cyclically monotone if and only if there exists a convex function $\phi : \mathbf{R}^d \to \mathbf{R} \cup \{+\infty\}$ such that $Z \subset \partial\phi$.*

Exercise 7.10. *Assume that $\Omega \subset \mathbf{R}^d$ is an open, convex set and that $\phi : \Omega \to \mathbf{R}$ is convex. Then*

(i) ϕ is continuous on Ω. The gradient map $\nabla\phi$ is defined almost everywhere and is a Borel map.

(ii) If $(x_n, y_n) \in \partial\phi$ and $x_n \to x_\infty$ in Ω, then every subsequence of $\{y_n\}_{n=1}^\infty$ admits a subsequence that converges to some $y_\infty \in \partial\phi(x)$. Conclude that $\partial\phi$ is closed.

(iii) The function ϕ is twice differentiable almost everywhere in the sense of Alexandrov [3]: for almost every x_o, $\nabla\phi(x_o)$ exists and there exists a symmetric matrix A such that

$$\phi(x_o + h) = \phi(x_o) + < \nabla\phi(x_o), h > + \frac{1}{2} < Ah; h > + o(|h|^2).$$

(iv) Differentiability of ϕ fails only on a rectifiable set of dimension less than or equal to $d - 1$.

The proofs of (i) and (ii) is easy while the proof of (iii) needs a little bit more thinking and can be found in [2]. The proof of (iv) is the most difficult one and we refer the reader to [3].

Exercise 7.11. *Assume that $\phi : \mathbf{R}^d \to \mathbf{R}$ is convex. Show that ϕ is strictly convex if and only if ϕ^* is differentiable everywhere on $\{x \mid \phi^*(x) < +\infty\}$.*

Exercise 7.12. *Assume that $c \in C^1(\mathbf{R}^d \times \mathbf{R}^d)$ and that $K, L \subset \mathbf{R}^d$ are compact sets. For $u, v : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$, not identically ∞ we define*

$$u^c(y) = \inf_{k \in K}\{c(k, y) - u(k)\}, \qquad v_c(x) = \inf_{l \in L}\{c(x, l) - v(l)\}.$$

(i) Show that u^c and v_c are Lipschitz.

(ii) Show that if $v = u^c$ then $(v_c)^c = v$.

(iii) Determine the class of u for which $(u^c)_c = u$.

The next exercise is very similar to exercise 7.12 except that now, we have lost the property that K, L are compact, by replacing them by $\mathbf{R}^d$.

Exercise 7.13. *Assume that $c \in C^1(\mathbf{R}^d \times \mathbf{R}^d)$ and that $c(z) = l(|z|)$ for a function l that grows faster than linearly as $|z|$ tends to $+\infty$. For $u, v : \mathbf{R}^d \to \mathbf{R} \cup \{-\infty\}$, not identically ∞ we define*

$$u^c(y) = \inf_{k \in \mathbf{R}^d} \{c(k, y) - u(k)\}, \qquad v_c(x) = \inf_{l \in \mathbf{R}^d} \{c(x, l) - v(l)\}.$$

(i) Show that u^c and v_c are locally Lipschitz.
(ii) Show that if $v = u^c$ then $(v_c)^c = v$.
(iii) Say whether or not $(u^c)_c = u$ for arbitrary u.

7.2 Measure Theory

Throughout this section X, Y and Z are Banach spaces. We denote by $\mathcal{P}(Z)$ the set of probability measures on Z. Most of the statements below stated for $X \subset \mathbf{R}^d$ are still valid if we substitute $\mathbf{R}^d$ by a polish space.

Material we assume that you know and which we don't recall

1. The definition of a measure (nonnegative), a Borel measure and a Radon measure on Z. Definition of a probability measure on Z.

2. The total variation of $\gamma \in \mathcal{P}(Z)$ is $\gamma[Z]$.

3. The definition of the weak $*$ convergence on the set of measure.

4. The definition of $L^p(Z, \gamma)$ for $1 \leq p \leq +\infty$ and γ a measure on Z.

Examples of measures *(a) Assume that $z_o \in Z$. The dirac mass at z_o is the measure δ_{z_o} defined by*

$$\delta_{z_o}[B] = \begin{cases} 1 & \text{if} \quad z_o \in B \\ 0 & \text{if} \quad z_o \notin B \end{cases} \tag{96}$$

for $B \subset Z$.

(b) If Z is a subset of $\mathbf{R}^d$ and $\rho : Z \to [0, +\infty]$ is a Borel function whose total mass is 1, we define $\mu := \rho dx$ by

$$\mu[B] = \int_B \rho dx,$$

for all $B \subset Z$ Borel set. The measure μ is said to have ρ as a density and to be absolutely continuous with respect to Lebesgue measure.

Exercise 7.14. *Suppose that $X \subset \mathbf{R}^d$. (i) Show that every probability measure $\mu \in \mathcal{P}(X)$ is the weak $*$ limit of a convex combination of dirac masses.*

(ii) Show that every probability measure $\mu \in \mathcal{P}(X)$ is the weak $*$ limit of a sequence of measures that are absolutely continuous with respect to Lebesgue measure.

Definition 7.15. *A Borel measure μ on X is said to have x_o as an atom if $\mu\{x_o\} > 0$.*

Exercise 7.16. *Suppose that μ is a Borel measure on X and that $\mu[X] = 1$. Show that the set of atoms of μ is countable.*

Hint. *Show that $\{x \in X | \frac{1}{n} \leq \mu\{x\} < \frac{1}{n-1}\}$ has at most n elements.*

For these lectures, we don't expect you to master the next definition and the proposition that follows but, since they are considered basic facts in measure theory, we include them here.

Definition 7.17. *(i) We denote by $\mathcal{B}(X)$ the Borel sigma algebra on the metric space X.*

(ii) Assume that μ is a Borel measure on X and ν is a Borel measure on Y. We say that $(X, \mathcal{B}(X), \mu)$ is isomorphic to $(Y, \mathcal{B}(Y), \nu)$ if there exists a one-to-one map T of X onto Y such that for all $A \in \mathcal{B}(X)$ we have $T(A) \in \mathcal{B}(Y)$ and $\mu[A] = \nu[T(A)]$, and for all $B \in \mathcal{B}(Y)$ we have $T^{-1}(B) \in \mathcal{B}(X)$ and $\mu[T^{-1}(B)] = \nu[B]$. For brevity we say that μ is isomorphic to ν.

The next proposition is an amazing result that is considered a basic fact in measure theory. We refer the reader to the book by Royden [44], Theorem 16 for its proof.

Proposition 7.18. *Let μ be a finite Borel measure on a complete separable metric space X. Assume that ν has no atoms and $\mu[X] = 1$. Then $(X, \mathcal{B}(X), \mu)$ is isomorphic to $([0,1], \mathcal{B}([0,1]), \lambda_1)$, where λ_1 stands for the one-dimensional Lebesgue measure on $[0,1]$.*

Definition 7.19. *Assume that γ is a measure on Z and that $Z' \subset Z$. The restriction of γ to Z is the measure $\gamma|_{Z'}$ defined on Z' by*

$$\gamma|_{Z'}[C] = \gamma[C \cap Z']$$

for all $C \subset Z$.

Exercise 7.20. *Assume that $Z' \subset Z$ and that γ' is a measure on Z'. Define*

$$\gamma[C] = \gamma'[C \cap Z']$$

for all $C \subset Z$. Is there any condition we must impose on Z' for γ to be a measure on Z?

Definition 7.21. *Assume that $Z = X \times Y$ and that $\gamma \in \mathcal{P}(Z)$. The first and second marginals of γ are the measures $\mathrm{proj}_X\gamma$ defined on X and $\mathrm{proj}_Y\gamma$ defined on Y by*

$$\mathrm{proj}_X\gamma[A] = \gamma[A \times Y], \qquad \mathrm{proj}_Y\gamma[B] = \gamma[X \times B],$$

for all $A \subset X$ and all $B \subset Y$.

Definition 7.22. *If $\gamma \in \mathcal{P}(Z)$ and $1 \leq p < +\infty$, the p–moment of γ is*

$$M_p[\gamma] = 1/p \int_Z ||\mathbf{z}||^p d\gamma(\mathbf{z}).$$

Exercise 7.23. *Assume that $1 < p < +\infty$, that $\{\gamma_n\}_{n=1}^\infty \subset \mathcal{P}(\mathbf{R}^d)$ and that $\{M_p[\gamma]\}_{n=1}^\infty$ is a bounded set. Show that there exists a subsequence of $\{\gamma_n\}_{n=1}^\infty$ that converges weak $*$ in $\mathcal{P}(\mathbf{R}^d)$.*

Warning. *The limit of the subsequence must be not only a measure but a probability measure.*

Exercise 7.24. *Assume that $\gamma, \bar{\gamma}$ are two Borel probability measures on $\mathbf{R}^d$. Show that $\gamma[C] = \bar{\gamma}[C]$ for every Borel set if and only if*

$$\int_Z F(\mathbf{z}) d\gamma(\mathbf{z}) = \int_Z F(\mathbf{z}) d\bar{\gamma}(\mathbf{z})$$

for all $F \in C_o(\mathbf{R}^d)$.

8 Appendix

Throughout this section $L : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ is a continuous functions such that

(A1) L(x+k,v)=L(x,v) for each $(x, v) \in \mathbf{R}^d \times \mathbf{R}^d$ and each $k \in \mathbf{Z}^d$.

We assume that L is smooth enough in the sense that there exists an integer $r > 1$ such that

(A2) $L \in C^r(\mathbf{R}^d \times \mathbf{R}^d)$.

We also impose that the Hessian matrix is positive:

(A3) $(\frac{\partial^2 L}{\partial v_i \partial v_j}(x, v)) > 0$

in the sense that its eigenvalues are all positive. We need the following uniform superlinear growth condition:

(A4) For every $A > 0$ there exists a constant $\delta > 0$ such that $\frac{L(x,v)}{||v||} \geq A$

for every $x \in \mathbf{R}^d$ and every v such that $||v|| \geq \delta$. In particular, there exists a real number $B(L)$ such that for every $(x, v) \in \mathbf{R}^d \times \mathbf{R}^d$, we have that

$$L(x, v) \geq ||v|| - B(L).$$

A continuous function $L : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ satisfying (A1–A4) *is called a lagrangian. In many mechanical systems, the Lagrangian $L(x, \cdot)$ does not go faster than exponentially as v tends to $+\infty$: there is a constant $b(L) \in \mathbf{R}$ such that*

(A5) $L(x, v) \leq e^{||v||} - b(L) - 1$ for each $(x, v) \in \mathbf{R}^d \times \mathbf{R}^d$.

Now, we introduce a Hamiltonian associated to L, the so-called Legendre transform of $L(x, \cdot)$. For $(x, p) \in \mathbf{R}^d \times \mathbf{R}^d)$ we set

$$H(x, p) = \sup_{v \in \mathbf{R}^d} \{v \cdot p - L(x, v)\}, \qquad ((x, p) \in \mathbf{R}^d \times \mathbf{R}^d).$$

The Hamiltonian H is continuous and by (A3)*, $H(x, \cdot)$ is continuously differentiable. Also, the map*

$$(x, v) \to (x, \nabla_v L(x, v)) = \mathbf{T}(x, v)$$

is of class $C^{r-1}(\mathbf{R}^d \times \mathbf{R}^d)$ and its restriction to $\mathbf{T}^d \times \mathbf{R}^d$ is one-to-one. It inverse

$$(x, p) \to (x, \nabla_p H(x, p)) = \mathbf{S}(x, p)$$

is then of class $C^{r-1}(\mathbf{R}^d \times \mathbf{R}^d)$. This proves that H is in fact of class $C^r(\mathbf{R}^d \times \mathbf{R}^d)$. These arguments are standard and can be found in [34] pp 1355.

If $(x,v) \in \mathbf{R}^d \times \mathbf{R}^d$ *and* $p = \nabla_v L(x,v)$, *because both* $L(x,\cdot)$ *and* $H(x,\cdot)$ *are convex and Legendre transform of each other then*

$$v = \nabla_p H(x,p), \qquad \nabla_x L(x,v) = -\nabla_x H(x,p). \tag{97}$$

One studies the kinematics and dynamics of these systems through the action

$$c(T,x_o,x_T) = \inf_\sigma \{ \int_0^T L(\sigma,\dot\sigma)dt \mid \sigma(0) = x_o, \sigma(T) = x_T\}, \tag{98}$$

where the infimum is performed over the set $AC(T;x_o,x_T)$ *of* $\sigma : [0,T] \to \mathbf{R}^d$ *that are absolutely continuous and that satisfy the endpoint constraint* $\sigma(0) = x_o$, $\sigma(T) = x_T$.

In the light of (A3) *and* (A4), *there exists* $\sigma_o \in AC(T;x_o,x_T)$ *that is a minimizer in (98) and* σ_o *satisfies the Euler-Lagrange equations*

$$\frac{d}{dt}[\nabla_v L(\sigma_o(t),\dot\sigma_o(t))] = \nabla_x L(\sigma_o(t),\dot\sigma_o(t)), \qquad 0 < t < T. \tag{99}$$

The infimum in (98) represents the cost for transporting a unit mass from x_o *to* x_T *during the time interval* $T > 0$. *There maybe several* σ_o *minimizer in (98), if the minimum is performed other* $AC(T;x_o,x_T)$. *Therefore, the differential equation (99) may have multiple solutions in* $AC(T;x_o,x_T)$. *It is natural to ask if given* $(x_o,v) \in \mathbf{R}^d \times \mathbf{R}^d$, *(99) has a unique solution* σ *for all* $t \in \mathbf{R}$, *when we prescribe* $\sigma(0) = x_o$, *and* $\dot\sigma(0) = v$. *We briefly recall what is known about the initial value problem and how one ensures existence of a* flow $\Phi : \mathbf{R} \times \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}^d \times \mathbf{R}^d$ *associated to the Lagrangian L, defined by* $\Phi(t,x,v) = (\phi(t,x,v), \dot\phi(t,x,v))$ *where* ϕ *satisfies*

$$\frac{d}{dt}[\nabla_v L(\sigma(t),\dot\sigma(t))] = \nabla_x L(\sigma(t),\dot\sigma(t)), \quad (\sigma(0),\dot\sigma(0)) = (x,v). \tag{100}$$

Here, we have set $\sigma(t) = \phi(t,x,v)$ *and have temporarily drop the argument* (x,v) *in* $\phi(t,x,v)$, *to make the text more readable Define*

$$p(t) = \nabla_v L(\sigma(t),\dot\sigma(t))$$

so that by (97), we have that (100) is equivalent to

$$\dot\sigma(t) = \nabla_p H(\sigma(t),p(t)) \qquad \dot p(t) = -\nabla_x H(\sigma(t),p(t)) \qquad \sigma(0) = x, \quad p(0) = \nabla_v L(v,v). \tag{101}$$

Now (101) is a standard initial value problem and so, it admits a unique maximal solution on an open interval (t_-,t_+). *That solution satisfies the conservation law*

$$H(\sigma(t),p(t)) = H(\sigma(0),p(0)), \qquad (t \in (t_-,t_+)). \tag{102}$$

As a byproduct, (100) admits a unique maximal solution on the same interval (t_-,t_+). *Set* $q = \nabla_v L(x,v)$. *We display the dependence in* (x,q) *and in* (x,v) *and introduce the flow:*

$$\Phi(t,x,v) = (\sigma(t),\dot\sigma(t)), \qquad \Phi(0,x,q) = (x,v).$$

together with the so-called dual-flow Φ^* *:*

$$\Phi^*(t,x,q) = (\sigma(t), p(t)), \qquad \Phi(0,x,q) = (x,q).$$

This terminology of dual flow is justified by the following fact:

$$\Phi(t,x,v) = \mathbf{S} \circ \Phi^* \circ \mathbf{T},$$

where $\mathbf{S}$ *and* $\mathbf{T}$ *are the diffeomorphisms defined through the functions* L *and* H *that are Legendre dual of each other.*

As in [34] we can ensure that the completness assumption $t_- = -\infty$ *and* $t_+ = +\infty$ *holds. For that it is enough to impose that* L *satisfies* (A5) *so that*

$$H(x,p) \geq ||p||Log||p|| + b(L) + 1 \geq ||p|| + b(L). \tag{103}$$

If (103) holds then by (102) we have that

$$||p(t)|| \leq \bar{c} := H(\phi(0), p(0)) - b(L). \tag{104}$$

We combine (101) and (104) to have that

$$||\dot{\phi}(t)|| \leq ||\nabla_p H||_{L^\infty(\mathbf{T}^d \times \bar{B}_{\bar{c}}(0))} \tag{105}$$

where $B_{\bar{c}}(0)$ *is the open ball in* $\mathbf{R}^d$ *of center* 0 *and radius* $\bar{c}$*. Consequently,* $||p(\cdot)|| + ||\phi(\cdot)||$ *are locally bounded in time. This shows that* $t_- = -\infty$ *and* $t_+ = +\infty$*. Consequently, under the completeness assumption which we make in the sequel, the flow* Φ *is well-defined for all* $t \in \mathbf{R}$*. Furthermore, it satisfies*

$$\Phi(t+s,x,v) = \Phi(t, \Phi(s,x,v)), \qquad ((t,s) \in \mathbf{R} \times \mathbf{R},\ (x,v) \in \mathbf{R}^d \times \mathbf{R}^d). \tag{106}$$

This is a byproduct of the uniqueness property of solutions of (100). Also if $T > 0$ *and* $\Phi(T,x,v) = \Phi(0,x,v)$ *then* $\Phi(\cdot,x,v)$ *must be periodic of period* T*.*

$$\Phi(t+T,x,v) = \Phi(t,x,v), \qquad ((t,x,v) \in \mathbf{R} \times \mathbf{R}^d \times \mathbf{R}^d). \tag{107}$$

In the next proposition, we assume that

$$L(x,v) = l(v) + W(x),$$

that W *is* $\mathbf{T}^d$*-periodic of class* C^2*, and that there exists a number* $e_l > 0$ *such that*

$$< \nabla^2 l(v)a; a > \geq e_l ||a||^2$$

for all $a \in \mathbf{R}^d$*. We show that for small times* $T > 0$ *there exists a unique opimal path* $\sigma_{x,y}$ *that minimizes* $\sigma \to \int_0^T L(\sigma, \dot{\sigma})ds$ *over* $AC(T,x,y)$*. Let us denote by* e_W *the smallest eigenvalue of* $\nabla^2 W$*, and let* c_1 *be the Poincare constant on* $(0,1)$*, defined to be the largest number* c_1 *such that*

$$c_1 \int_0^1 b^2 ds \leq \int_0^1 \dot{b}^2 ds$$

for all $b \in C_0^1(0,1)$*.*

Proposition 8.1 (Uniqueness of paths connecting two points). *Assume $T>0$ and that $e_W+\frac{e_l c_1}{T^2}>0$. For every $x,y\in\mathbf{R}^d$ there exists a unique $\sigma_o\in AC(T,x,y)$ that satisfies the Euler-Lagrange equation (99). Therefore, σ_o is the unique minimizer of $\sigma\to K[\sigma]=\int_0^T L(\sigma,\dot\sigma)ds$ over $AC(T,x,y)$.*

Proof: *Assume that $\sigma_o\in AC(T,x,y)$ satisfies (99). We write Taylor approximation of $L(\sigma,\dot\sigma)$ around $(\sigma_o,\dot\sigma_o)$, use that satisfies the Euler-Lagrange equation (99) and that $0=\sigma(0)-\sigma_o(0)=\sigma(t)-\sigma_o(T)$ to conclude that*

$$K[\sigma]-K[\sigma_o]\geq\int_0^T(e_W|\sigma-\sigma_o|^2+e_l|\dot\sigma-\dot\sigma_o|^2)ds\geq\int_0^T(e_W+\frac{e_lc_1}{T^2})|\sigma-\sigma_o|^2ds.$$

This concludes the proof of the proposition. *QED.*

References

[1] *M. Agueh. PhD Dissertation 2002. School of Mathematics Georgia Institute of Technology.*

[2] *G. Alberti. On the structure of singular sets of convex functions.* Calc. Var. Partial Differential Equations **2**, *17–27 (1994).*

[3] *A.D. Aleksandrov. Existence and uniqueness of a convex surface with a given integral curvature.* C.R. (Doklady) Acad. Sci. URSS (N.S.) **35**, *131–134 (1942).*

[4] *L. Ambrosio. Lecture notes on optimal transport problems. To appear with Proceedings of a Centro Internazionale Matematico Estivo, Summer School in the Springer-Verlag Lecture Notes in Mathematics Series. a*

[5] *L. Ambrosio, N. Gigli and G. Savaré. Gradient flows in metric spaces and the Wasserstein spaces of probability measures. (Preprint).*

[6] *Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs.* C.R. Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.*

[7] *J.D. Benamou and Y. Brenier. Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampeère equations/transport problem.* SIAM J. Appl. Ana. Math. **58**, *no 5, 1450–1461 (1998).*

[8] *L.A. Caffarelli. The regularity of mappings with a convex potential.* J. Amer. Math. Soc. **5** *(1992) 99–104.*

[9] *L.A. Caffarelli. Boundary regularity of maps with convex potentials.* Comm. Pure Appl. Math. **44** *(1992) 1141–1151.*

[10] *L.A. Caffarelli. Boundary regularity of maps with convex potentials — II.* Annals of Mathematics (2) **144** *(1996) 453–496.*

[11] *L. Caffarelli. Allocation maps with general cost functions. In P. Marcellini et al, editor,* Partial Differential Equations and Applications*, number 177 in Lecture Notes in Pure and Appl. Math., pages 29–35. Dekker, New York, 1996.*

[12] *E.A. Carlen and W. Gangbo. Constrained Steepest Descent in the 2–Wasserstein Metric.* Annals of Mathematics. **157***, 807–846, 2003.*

[13] *L.A. Caffarelli, M. Feldman, R.J. McCann. Constructing optimal maps in Monge's transport problem as a limit of strictly convex costs.* J. Amer. Math. Soc. **15** *(2002) 1–26.*

[14] *Constrained steepest descent in the 2–Wasserstein metric.* Annals of Mathematics. **157** *(2003) 807–846. Preprint www.math.gatech.edu/ gangbo/publications/.*

[15] *On the solution of a model Boltzmann Equation via steepest descent in the 2–Wasserstein metric. To appear in* Archive for Rational Mech. and Analysis. *Preprint www.math.gatech.edu/ gangbo/publications/.*

[16] *D. Cordero-Erausquin, W. Gangbo and C. Houdre. Inequalities for generalized entropy and optimal transportation. To appear in* AMS Contemp. Math. *Preprint www.math.gatech.edu/ gangbo/publications/.*

[17] *P.Cloke and M.J.P. Cullen. A semi-geostrophic ocean model with outcropping .* Dyn. Atmos. Oceans*, 21:23–48, 1994.*

[18] *M.J.P. Cullen and W. Gangbo. A variational approach for the 2-dimensional semi-geostrophic shallow water equations.* Arch. Ration. Mech. Anal. **156** *(2001) 241–273.*

[19] *M.J.P. Cullen and R.J. Purser. An extended lagrangian theory of semigeostrophic frontogenesis.* J. Atmosph. Sciences **41***, 1477–1497 (1984).*

[20] *M.J.P. Cullen and R.J. Purser. Properties of the lagrangian semigeostrophic equations.* J. Atmosph. Sciences *vol 40,* **17***, 2684–2697 (1989).*

[21] *M.J.P. Cullen and I. Roulstone. A geometric model of the nonlinear equilibration of two-dimensional Eady waves.* J. Atmos. Sci.*,* **50***, 328-332, 1993.*

[22] *B. Dacorogna.* Direct Methods in the Calculus of Variations. *Springer-Verlag, 1989.*

[23] *A Eliassen. The quasi-static equations of motion. Geofys. Publ.,* **17***, No 3, 1948.*

[24] *L.C. Evans. Partial differential equations and Monge-Kantorovich mass transfer.* Int. Press, Current Dev. Math*,* **26***, 26–78, 1997.*

[25] *L. C. Evans and W. Gangbo. Differential equations methods for the Monge-Kantorovich mass transfer problem Mem. Amer. Math. Soc. #654* **137** *(1999) 1–66.*

[26] *W. Gangbo. An elementary proof of the polar factorization of vector-valued functions.* Arch. Rational Mech. Anal.*, 128:381–399, 1994.*

[27] *W. Gangbo. The Monge mass transfer problem and its applications. NSF-CBMS Conference on the Monge-Ampere equation: applications to geometry and optimization, July 09–13 1997. Contemporary Mathematics, Vol 226, 79–103, (1999).*

[28] *W. Gangbo and R.J. McCann. Optimal maps in Monge's mass transport problem.* C.R. Acad. Sci. Paris Sér. I Math. **321** *(1995) 1653–1658.*

[29] *W. Gangbo and R.J. McCann. The geometry of optimal transportation.* Acta Math. **177** *(1996) 113–161.*

[30] *B.J. Hoskins. The geostrophic momentum approximation and the semi-geostrophic equations.* J. Atmosph. Sciences **32***, 233–242, 1975.*

[31] *Jordan, Kinderleher and Otto. The variational formulation of the Fokker–Planck equation.* SIAM Jour. Math Anal.*,* **29** *1–17, 1998.*

[32] *L. Kantorovich. On the translocation of masses.* C.R. (Doklady) Acad. Sci. URSS (N.S.), *37:199–201, 1942.*

[33] *L. Kantorovich. On a problem of Monge (In Russian).* Uspekhi Math. Nauk.*, 3:225–226, 1948.*

[34] *J.N. Mather. Variational construction of connecting orbits.* Ann. Inst. Fourier, Grenoble, **43** *no 5, (1993), 1349–1386.*

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

[36] *R.J. McCann. Existence and uniqueness of monotone measure-preserving maps.* Duke Math. J.*,* **80** *309–323, 1995.*

[37] *G. Monge. Mémoire sur la théorie des déblais et de remblais.* Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704 (1781).*

[38] *B.S. Katz, editor.* Nobel Laureates in Economic Sciences: a Biographical Dictionary*. Garland Publishing Inc., New York, 1989.*

[39] *F. Otto. The geometry of dissipative evolution equations: the porous medium equation.* Comm. P.D.E.*,* **26** *156–186 2001.*

[40] *J. Pedlosky.* Geophysical Fluid Dynamics*. Springer-Verlag (1982).*

[41] *S.T. Rachev and L. Rüschendorf.* Mass Transportation Problems*; Vol. I & II. Probability and its Applications, Springer Verlag, Heidleberg (1998).*

[42] *R.T. Rockafellar.* Convex Analysis*. Princeton University Press, Princeton, 1970.*

[43] *C.G. Rossby, et al. Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacement of the semi-permanent centers of actions.* J. Marine Res. **2**, *38–55 (1939).*

[44] *Royden.* Real Analysis, *3rd Ed., Macmillan, 1988.*

[45] *H. Stommel. The westward intensification of wind-driven ocean currents.* Trans. Amer. Geoph. Union **29**, *202–206 (1948).*

[46] *V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. Proc. Steklov Inst. Math.* **141** *(1979) 1–178.*

[47] *N. Trudinger and X.-J. Wang. On the Monge mass transfer problem.* Calc. Var. Partial Differential Equations **13**, *19–31. (2001)*