

Introduction to the mass transportation theory; Applications (Pre-ultimate draft)

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

May 28, 2004

The Monge-Kantorovich mass transportation theory finds its roots in three influential papers. The first one is written by a great geometer, G. Monge [35]. The second and third one are due to Kantorovich [31] [30], who received a Nobel prize for related work in economics [36]. 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 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 kinetic theory and meteorology. These applications are fully developed in the following preprints, [11] [13] [16] [26], which you can download from my webpage at www.math.gatech.edu/gangbo/publications/.

The contribution on the mass of transportation theory by the author and its collaborators, to geometric inequalities as well as computer vision and the reflector problem, can be found at the same webpage.

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

*The author of these notes acknowledges 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.

1 Introduction

Assume that we are given a pile of sand occupying a region $X \subset \mathbf{R}^d$ and assume that a set $Y \subset \mathbf{R}^d$ consists of holes. 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 to move mass from x to Tx . It must satisfy a mass conservation condition which is that " T pushes ρ_o forward to ρ_1 " (see definition 3.1).

Monge problem Find a minimizer for

$$\inf_T \left\{ \int_X c(x, y) \rho_o(x) dx \mid T_{\#} \mu_o = \mu_1 \right\}. \quad (1)$$

In 1781, Monge conjectured that when $c(x, y) = |x - y|$ there exists an optimal map that transports the pile of sand to the holes. Two hundred years elapsed before Monge conjecture was proven by Sudakov in 1976, in [44]. It was recently discovered by Ambrosio [3], that Sudakov's proof contains a gap in the case $d > 2$. Before that gap was noticed, the proof of Monge conjecture was extended to higher dimensional spaces by Evans and the author in 1999, in [23]. The results in [44] and [23] were recently independently refined by Ambrosio [3], Trudinger–Wang [45] and Caffarelli–Feldman–McCann [12]. In a meanwhile, McCann and the author [26] [27] independently with Caffarelli [10], 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 solved in [26] [27].

One can generalize Monge problem to arbitrary measures μ and ν when there is no map T such that $T_{\#} \mu = \nu$. 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. As usually done, we denote by 2^Y the set of subsets on Y . We consider maps $T : X \rightarrow 2^Y$ and associate to each $x \in X$, a measure γ_x supported on Tx , and which tells how to split 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

$$\int_X \left[\int_{Tx} c(x, y) d\gamma_x(y) \right] d\mu_o(x).$$

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

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

The measure γ is to satisfy the mass conservation condition:

$$\mu[A] = \gamma[A \times Y], \quad \gamma[X \times B] = \nu[B]$$

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

In section 3, we introduce Kantorovich problem as a relaxation of Monge problem. To do that, we first extend the set $\mathcal{T}(\mu, \nu)$ of maps $T : X \rightarrow Y$ such that $T_{\#}\mu = \nu$, to a bigger set $\Gamma(\mu, \nu)$. Then, we extend the function $\rho_o \rightarrow I[\rho_o] := \int_X c(x, y)\rho_o(x)$ to a function \tilde{I} defined on $\Gamma(\mu, \nu)$ so that if $\mathcal{T}(\mu, \nu) \neq \emptyset$ then

$$\inf_{\mathcal{T}(\mu, \nu)} I = \inf_{\Gamma(\mu, \nu)} \tilde{I}.$$

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 formulate the mass transportation problem and under suitable assumptions, prove existence of solutions for both 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 study 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 kinetic theory, meteorology and optimal designed. Since we anticipate that the applications to geometric inequalities will be covered in parallel lectures given by N. Ghoussoub, we omit them here.

2 Prerequisite for the mass transportation theory

2.1 Convex Analysis: part I

The material of this section can be found in the books [20], [40]. 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 2.1. Let $X \subset Y$ be a convex subset of Y and let $\phi : X \rightarrow \mathbf{R} \cup \{+\infty\}$ be a real valued function.

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

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

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

$$\liminf_{n \rightarrow +\infty} \phi(x_n) \geq \phi(x_\infty),$$

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

Remark 2.2. Suppose that $\phi : X \rightarrow \mathbf{R} \cup \{+\infty\}$ and we defined $\bar{\phi} : Y \rightarrow \mathbf{R} \cup \{+\infty\}$ by

$$\bar{\phi}(x) = \begin{cases} \phi(x) & \text{if } x \in X \\ +\infty & \text{if } x \notin X. \end{cases} \quad (2)$$

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

Exercise 2.3. (i) Show that ϕ is lower semicontinuous if and only if its epigraph $\text{epi}(\phi) = \{(x, t) \mid \phi(x) \leq t\}$ 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 2.4. Assume that $X \subset Y$ is a convex set and that $\phi : X \rightarrow \mathbf{R} \cup \{+\infty\}$ is convex.

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

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

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

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

Remark 2.6. Note that ϕ and its natural extension have the same Legendre transform.

Exercise 2.7. Assume that $\phi : Y \rightarrow \mathbf{R} \cup \{+\infty\}$ 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

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

(iii) Say whether or not the following hold:

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

Definition 2.8. 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 2.9. Show that $Z \subset \mathbf{R}^d \times \mathbf{R}^d$ is cyclically monotone if and only if there exists a convex function $\phi : \mathbf{R}^d \rightarrow \mathbf{R} \cup \{+\infty\}$ such that $Z \subset \partial\phi$.

Exercise 2.10. Assume that $\Omega \subset \mathbf{R}^d$ is an open, convex set and that $\phi : \Omega \rightarrow \mathbf{R}$ 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 $x_n \rightarrow x_\infty$ 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 [2]: for almost every x_o , $\nabla\phi(x_o)$ exists and there exists a symmetric matrix A such that

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

(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 [1]. The proof of (iv) is the most difficult one and we refer the reader to [2].

Exercise 2.11. Assume that $\phi : \mathbf{R}^d \rightarrow \mathbf{R}$ is convex. Show that ϕ is strictly convex if and only if ϕ^* is differentiable everywhere on $\{x \mid \phi(x) < +\infty\}$.

Exercise 2.12. 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 $u, v : \mathbf{R}^d \rightarrow \mathbf{R} \cup \{-\infty\}$, not identically $-\infty$ we define

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

- (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 2.12 except that now, we have lost the property that K, L are compact, by replacing them by \mathbf{R}^d .

Exercise 2.13. 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 $u, v : \mathbf{R}^d \rightarrow \mathbf{R} \cup \{-\infty\}$, not identically $-\infty$ we define

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

- (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 .

2.2 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

$$\delta_{z_o}[B] = \begin{cases} 1 & \text{if } z_o \in B \\ 0 & \text{if } z_o \notin B \end{cases} \quad (3)$$

for $B \subset Z$.

(b) If Z is a subset of \mathbf{R}^d and $\rho : Z \rightarrow [0, +\infty]$ 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. That measure is said to have ρ as a density and to be absolutely continuous with respect to Lebesgue measure.

Exercise 2.14. *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 2.15. *A Borel measure μ on X is said to have x_o as an atom if $\mu\{x_o\} > 0$.*

Exercise 2.16. *Suppose that μ is a Borel measure on X . Show that the set of atoms of μ is countable.*

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 2.17. *(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 [42], Theorem 16 for its proof.

Proposition 2.18. *Let μ be a finite Borel measure on a complete separable metric space X . Assume that μ 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 2.19. *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 2.20. *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 2.21. *Assume that $Z = X \times Y$ and that $\gamma \in \mathcal{P}(Z)$. The first and second marginals of γ are the measures $\text{proj}_X \gamma$ defined on X and $\text{proj}_Y \gamma$ defined on Y by*

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

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

Definition 2.22. *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 2.23. *Assume that $1 < p < +\infty$, that $\{\gamma_n\}_{n=1}^\infty \subset \mathcal{P}(\mathbf{R}^d)$ and that $\{M_p[\gamma_n]\}_{n=1}^\infty$ 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 2.24. *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 $F \in C_o(\mathbf{R}^d)$.

3 Formulation of the mass transport problems

3.1 The original Monge-Kantorovich problem

Monge problem is in terms of transport maps which we next define. To be able to formulate Kantorovich problem, we need to replace the transport maps by special multivalued maps, coupled with a family of measure. We call those *transport scheme*.

Definition 3.1 (Transport maps and schemes). *Assume that μ is a measure on X and that ν is a measure on Y . (i) We say that $T : X \rightarrow Y$ transports μ onto ν and we write $T_{\#}\mu = \nu$ if*

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

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) A measure γ on $X \times Y$ has μ and ν as its marginals if $\mu = \text{proj}_X \gamma$ and $\nu = \text{proj}_Y \gamma$. We write that $\gamma \in \Gamma(\mu, \nu)$ and call γ a transport scheme for μ and ν .

Remark 3.2. (i) Note that (4) expresses a mass conservation condition between the two measures.

(ii) Given two measures μ and ν , proposition 2.18 gives a sufficient condition for the existence of a map T that transports μ onto ν . It is easy to see that such a map may not exist in general. For instance, assume that x, y, z are three distinct elements of X . Set $\mu = 1/2(\delta_x + \delta_y)$ and $\nu = 1/3(\delta_x + \delta_y + \delta_z)$. Then there is no map T that transports μ onto ν .

(iii) If $\gamma \in \Gamma(\mu, \nu)$, (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 $K \subset X \times Y$ such that $\gamma[K] = \gamma[X \times Y]$.

Kantorovich problem. Find a minimizer for

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y). \quad (5)$$

Why is Kantorovich's problem a relaxation of Monge's problem? To each $T : X \rightarrow Y$ such that $T_{\#}\mu = \nu$ we associate the measure γ_T defined on $X \times Y$ by

$$\gamma_T[C] = \mu[\{x \in X \mid (x, Tx) \in C\}].$$

Exercise 3.3. Assume that μ and ν are two probability measures on \mathbf{R}^d and that c is a nonnegative continuous on $\mathbf{R}^d \times \mathbf{R}^d$. Define I on $\mathcal{T}(\mu, \nu)$ and \tilde{I} on $\Gamma(\mu, \nu)$ by

$$I[T] = \int_X c(x, Tx) d\mu(x), \quad \tilde{I}[\gamma] = \int_{X \times Y} c(x, y) d\gamma(x, y)$$

(i) Prove that if $T_{\#}\mu = \nu$ then $\gamma_T \in \Gamma(\mu, \nu)$ and $I[T] = \tilde{I}[\gamma_T]$.

(ii) Prove that if μ and ν don't have atoms then $\{\gamma_T \mid T \in \mathcal{T}(\mu, \nu)\}$ is weak * dense in $\Gamma(\mu, \nu)$.

(iii) Conclude that $\inf_T I = \int_{\gamma} \tilde{I}$.

A detailed proof of these statements can be found in [25]

3.2 A dual to the Monge-Kantorovich problem

(still need to be written up. Will be covered during the lectures)

3.3 Existence of a minimizer

(Not completed. Will be covered during the lectures)

Theorem 3.4 (Existence of a unique minimizer in (7)). If $p > 1$ and μ, ν are absolutely continuous with respect to Lebesgue measure then, (7) admits a unique minimizer. In particular if $p = 2$, the minimizer is characterized by the fact that it is the gradient of the convex function ϕ that satisfies $(\nabla\phi)_{\#}\mu = \nu$.

Proof: Theorem 3.4 is fundamental in many applications of the Monge-Kantorovich theory. The case $p = 2$ was first proved by Brenier in [5]. The general case was independently proved by Caffarelli [10], and Gangbo & McCann [27].

4 The Wasserstein distance

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

$$W_p^p(\mu, \nu) := \inf_{\gamma} \left\{ \int_{\mathbf{R}^d \times \mathbf{R}^d} |x - y|^p d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\}. \quad (6)$$

When μ and ν don't have atoms then

$$W_p^p(\mu, \nu) := \inf_{\mathbf{s}} \left\{ \int_{\mathbf{R}^d} |x - \mathbf{s}(x)|^p d\mu(x) : \mathbf{s}_{\#}\mu = \nu \right\}. \quad (7)$$

Definition 4.1 (The Wasserstein distance). . It is well-known that W_p is a metric for $p \geq 1$ and W_p^p is a metric for $0 < p \leq 1$. When $p = 2$, W_2 is called the Wasserstein distance [39].

5 Example of cost functions; fluids mechanic and dynamical systems

Many mechanical systems can be described via a lagrangian $L : \mathbf{R}^d \times \mathbf{R}^d \rightarrow \mathbf{R}$, defined on the phase space $\mathbf{R}^d \times \mathbf{R}^d$. Customarily, $L(x, \cdot)$ satisfies some convexity assumptions and $L(\cdot, v)$ satisfies suitable growth or periodicity conditions. To make the reader get to the main point faster, we omit specify here to specify the conditions on $L \in C^r(\mathbf{R}^d \times \mathbf{R}^d)$, which can be found 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

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

The Hamiltonian H is continuous and by (H3), $H(x, \cdot)$ is continuously differentiable. Also, the map

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

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. Its inverse

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

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

$$c(T, x_o, x_T) = \inf_{\sigma} \left\{ \int_0^T L(\sigma, \dot{\sigma}) dt \mid \sigma(0) = x_o, \sigma(T) = x_T \right\}, \quad (8)$$

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

$$\inf_{\mathbf{r} \# \rho_o = \rho_1} \int_{\mathbf{R}^d} c(1, x, \mathbf{r}(x)) \rho_o(x) dx = \inf_{\mathbf{g}(\cdot, \cdot)} \left\{ \int_0^1 dt \int_{\mathbf{R}^d} L(x, \mathbf{g}(t, x)) \rho_o(x) dx \right\} \quad (9)$$

where the infimum is performed over the set of $\mathbf{g} : [0, 1] \times \mathbf{R}^d \rightarrow \mathbf{R}^d$ such that $\mathbf{g}(0, x) = x$ and $\mathbf{g}(1, \cdot)_{\#}\rho_o = \rho_1$. The expression at the left handside of (9) is $W_{\bar{c}}$ where $\bar{c} = c(1, \cdot, \cdot)$.

When (9) admits a unique minimizer $\bar{\mathbf{g}}$ (see proposition 7.1 for a condition on L that ensures such properties), we define the path

$$\bar{\rho}(t, \cdot) = \bar{\mathbf{g}}(t, \cdot)_{\#}\rho_o \quad (10)$$

When $L(x, v) = |v|^p/p$ for some $p \in (0, +\infty)$ then $t \rightarrow \bar{\rho}(t, \cdot)$ is a geodesic for the Wasserstein distance (see [4] and [37]). The passage from Lagrangian to Eulerian coordinates is done through the ODE

$$\dot{\mathbf{g}}(t, x) = \mathbf{V}(t, \mathbf{g}(t, x)), \quad \mathbf{g}(0, x) = x. \quad (11)$$

We combine (9) and (11) to obtain that

$$W_{\bar{c}}(\rho_o, \rho_1) = \inf_{\rho(\cdot, \cdot), \mathbf{V}} \left\{ \int_0^1 dt \int_{\mathbf{R}^d} L(y, \mathbf{V}(t, y)) \rho(t, y) dy \right\}, \quad (12)$$

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

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

When $L(x, v) = |v|^2/2$, one can recognize the expression in (12) to be the minimal kinetic energy spent by the system during its evolution, from time $t = 0$ to time $t = T$, given in Lagrangian form.

6 Applications

6.1 The Kinetic Fokker-Planck equations

Before formulating the kinetic Fokker-Planck equations (KFPE) we introduce few definitions.

Definition 6.1 (Maxwellians and local Maxwellians). *Assume that f is a probability density on the phase space $\mathbf{T}^d \times \mathbf{R}^d$. Define*

$$\rho(t, x) = \int_{\mathbf{R}^d} f(t, x, v) dv, \quad F(t, x, v) = \frac{f(t, x, v) dv}{\rho(t, x)}.$$

The bulk velocity and the temperature associated to f are:

$$u_F(x) := \int_{\mathbf{R}^d} \mathbf{v} F(x, \mathbf{v}) d\mathbf{v}, \quad \theta_F(x) = 1/d \int_{\mathbf{R}^d} |\mathbf{v} - \mathbf{u}_F(x)|^2 F(x, \mathbf{v}) d\mathbf{v}. \quad (13)$$

- The Maxwellian with bulk velocity $\mathbf{u} \in \mathbf{R}^d$ and temperature $\theta > 0$ is

$$M_{\mathbf{u},\theta}(\mathbf{v}) := (2\pi\theta)^{-d/2} e^{-\frac{|\mathbf{v}-\mathbf{u}|^2}{2\theta}}.$$

- The space dependent Maxwellian with bulk velocity $\mathbf{u}_F(x)$ and temperature $\theta_F(x)$ is denoted $M_{F(x,\cdot)}$.
- Finally, the local Maxwellian corresponding to f on the phase space is M_f defined by

$$M_f(x, \mathbf{v}) = \rho(x) M_{F(x,\cdot)}(\mathbf{v}).$$

- We call \mathcal{M} the set of local Maxwellians.

The kinetic Fokker-Planck equations are

$$\partial_t f(t, x, \cdot) + \mathbf{v} \cdot \nabla_x f(t, x, \cdot) = \mathcal{L}_p(f(t, x, \cdot)). \quad (14)$$

Here,

$$\mathcal{L}_p(F(x, \cdot))(\mathbf{v}) = \theta_F^p(x) \operatorname{div}_{\mathbf{v}} \left[F(x, \cdot) \nabla_{\mathbf{v}} \ln \left(\frac{F(x, \cdot)}{M_F(x, \cdot)} \right) \right]$$

$\nabla_{\mathbf{v}}$, and $\operatorname{div}_{\mathbf{v}}$ are the gradient and the divergence with respect to the \mathbf{v} variables and p is a parameter.

Physicists expect (14) to admit solutions that conserve their total energy:

$$\int_{\mathbf{T}^d \times \mathbf{R}^d} \frac{|v|^2}{2} f(t, x, v) dx dv.$$

Because, the solutions obtained so far are not known to be "smooth enough" checking the energy conservation property remains an important challenge in kinetic theory.

Chief advantage of working with the Wasserstein distance. The Wasserstein distance has the advantage of requiring very little in the way of estimates on spacial regularity. To be more precised, the expression $\partial_t f + \mathbf{v} \nabla_x f$ is the limit as h tends to 0 of

$$\frac{f(t+h, x+h\mathbf{v}, \mathbf{v}) - f(t, x, \mathbf{v})}{h}.$$

But, in terms of the Wasserstein distance W_2 on $\mathcal{P}(\mathbf{T}^d \times \mathbf{R}^d)$, we have that

$$W_2^2(f(t+h, x+h\mathbf{v}, \mathbf{v}), f(t, x, \mathbf{v})) \leq h^2 E(f_o).$$

That first order approximation of $f(t+h, x+h\mathbf{v}, \mathbf{v})$ by $f(t, x, \mathbf{v})$ in terms of the Wasserstein distance, means that we don't need to worry about existence of $\partial_t f$ and $\mathbf{v} \cdot \nabla_x f$ separately.

Remark 6.2 (Homogeneous vs. inhomogeneous Fokker-Planck). (a) When $f_o(x, \mathbf{v}) \equiv f_o(\mathbf{v})$, one seeks for a solution f of (14) that satisfies $f(t, x, \mathbf{v}) \equiv f(t, \mathbf{v})$. In that case, (14) becomes linear, and its study becomes much more simpler; (14) is then called the homogeneous Fokker-Planck equation. Jordan-Kinderlehrer-Otto were the first to show in [29], that the scheme (3) of lemma 6.3, produces an approximate solution to the homogeneous Fokker-Planck equation. (1), (2) and (4) were not needed there.

(b) When $f_o(x, \mathbf{v}) \not\equiv f_o(\mathbf{v})$ (14) is referred to as the inhomogeneous Fokker-Planck. In that case, the evolutive system is decidedly, a nonlinear second order differential equation with nonlocal coefficients. These coefficients may be very small. I am not aware of any classical tools that could then be used to obtain existence of solutions that conserve energy.

Entropy functionals. Assume that $f \in \mathcal{P}^a(\mathbf{T}^d \times \mathbf{R}^d)$ and that, as in (13), F is the conditional velocity distribution associated to f . The entropy of $F(x, \cdot)$ is

$$S(F(x, \cdot)) = \int_{\mathbf{R}^d} F(x, \mathbf{v}) \ln(F(x, \mathbf{v})) d\mathbf{v} \quad (15)$$

and the entropy of f is

$$H(f) = \int_{\mathbf{T}^d \times \mathbf{R}^d} f(x, \mathbf{v}) \ln(f(x, \mathbf{v})) dx d\mathbf{v}.$$

The relative entropies of $G \in \mathcal{P}^a(\mathbf{R}^d)$ with respect to $M_{F(x, \cdot)}$, and the relative entropies of f with respect to M_f are

$$S(G|M_{F(x, \cdot)}) = \int_{\mathbf{R}^d} G \ln\left(\frac{G}{M_{F(x, \cdot)}}\right) d\mathbf{v}, \quad H(f|M_f) = \int_{\mathbf{T}^d \times \mathbf{R}^d} f \ln\left(\frac{f}{M_f}\right) dx d\mathbf{v}. \quad (16)$$

Lemma 6.3 (A discrete scheme). Fix an initial phase space density f_o defined on the phase space $\mathbf{T}^d \times \mathbf{R}^d$, and fix a time step $h > 0$. Inductively define f_k in terms of f_{k-1} through the following algorithm:

(1) First, run the streaming: define $\tilde{f}_k(x, \mathbf{v}) = f_{k-1}(x - h\mathbf{v}, \mathbf{v})$.

(2) Define $\rho_k(x) = \int_{\mathbf{R}^d} \tilde{f}_k(x, \mathbf{v}) d\mathbf{v}$ and the precollision conditional velocity distribution \tilde{F}_k by $\rho_k(x) \tilde{F}_k(x, \mathbf{v}) = \tilde{f}_k(x, \mathbf{v})$.

(3) Now run the collisions through steepest descent of the relative entropy: for each x , let $F_k(x, \cdot)$ minimize the functional

$$G \rightarrow \frac{W_2^2(G, \tilde{F}_k(x, \cdot))}{\theta_{\tilde{F}_k(x, \cdot)}^p} + hS(G|M_{\tilde{F}_k(x, \cdot)})$$

over the set $\mathcal{P}_a(\mathbf{R}^d)$ of probability density functions on \mathbf{R}^d .

(4) Finally, reconstruct f_k through $f_k(x, v) = \rho_k(x)F_k(x, \mathbf{v})$.

Definition 6.4 (Approximate solutions to the kinetic Fokker-Planck).

Define a time dependent phase space probability density $f^{(h)}(t, x, \mathbf{v})$ through

$$f^{(h)}(t, x, \mathbf{v}) = f_k(x - (t - kh)\mathbf{v}, \mathbf{v}) \quad \text{for } t \in [t_k, t_{k+1}),$$

and by convention, $t_k = kh$.

Theorem 6.5. Assume that $p = 1$ and $T > 0$. Assume that the sixth moment $\int_{\mathbf{T}^d \times \mathbf{R}^d} f_o(x, \mathbf{v}) |\mathbf{v}|^6 dx d\mathbf{v}$ of the initial probability density f_o is finite. Then the set $\{f^{(h)}\}$ is weakly compact in $L^1((0, T) \times \mathbf{T}^d \times \mathbf{R}^d)$ and a subsequence of $\{f^{(h)}\}$ converges weakly to an $f(t, \cdot, \cdot) \in \mathcal{P}^a(\mathbf{T}^d \times \mathbf{R}^d)$. Furthermore,

(i) $\partial_t f^{(h)} + x \cdot \nabla_{\mathbf{v}} f^{(h)} - \mathcal{L}_1(f^{(h)})$ tends to 0 in the distributional sense.

(ii) f is a solution to the kinetic Fokker-Planck equations (14) and the conservation of energy holds:

$$\int_{\mathbf{T}^d \times \mathbf{R}^d} \frac{|v|^2}{2} f(t, x, v) dx dv = \int_{\mathbf{T}^d \times \mathbf{R}^d} \frac{|v|^2}{2} f(0, x, v) dx dv.$$

That theorem is proved in [13].

Open problem 1. For parameters $\epsilon > 0$, consider equation

$$\partial_t f_\epsilon + \mathbf{v} \cdot \nabla_x f_\epsilon = \frac{1}{\epsilon} \mathcal{L}_p(f_\epsilon) \quad (17)$$

that admits a solution since we could solve (14). As ϵ tends to 0 does $\{f_\epsilon\}_\epsilon$ tends to a limit f where ρ, \mathbf{u}_F satisfy the Euler equations of compressible fluids

$$\partial_t(\rho \mathbf{u}_F) + \operatorname{div}_x(\rho \mathbf{u}_F \otimes \mathbf{u}_F) = -\nabla_x p, \quad \partial_t \rho + \operatorname{div}_x(\rho \mathbf{u}_F) = 0.$$

Here ρ, F and \mathbf{u}_F are given through $f \equiv f(t, x, \mathbf{v})$ as in (13) and Definition 6.1.

Open problem 2. Under what assumptions on the initial density f_o is the solution of (14) unique? A more modest but, still difficult question is: assume that we are given two initial velocity distributions f_o and \tilde{f}_o on the phase space $\mathbf{T}^d \times \mathbf{R}^d$, and let $f^{(h)}$ and $\tilde{f}^{(h)}$ be the corresponding solution obtained in lemma 6.5. Do we control the Wasserstein distance $W_2(f^{(h)}, \tilde{f}^{(h)})$ in terms of $W_2(f_o, \tilde{f}_o)$.

Comments. The point is to understand the stability of our algorithm. The answer to problem 2 can be readily shown in the case of the homogeneous Fokker-Planck system [37]. aaa

6.2 Semigeostrophic equations

The *semi-geostrophic* systems are new to the mathematical community, but are attracting more and more attention because of their similarities with the Euler equations of incompressible fluids (see the review paper by L.C. Evans [22] or [6]).

The *semi-geostrophic* systems are used by meteorologists to model how fronts arise in large scale weather patterns. It is a 3-D free boundary problem which is an approximation of the 3-D Euler equations of incompressible fluid, in a rotating coordinate frame around the Ox_3 -axis where, the effects of rotation dominate. It was first introduced by Eliassen [21] in 1948, and rediscovered by Hoskins [28] in 1975. Since then, it has been intensively studied by geophysicists (e.g. [17], [18], [41], [43]). 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 *semigeostrophic shallow water system* (SGSW). We skip its derivation that can be found in [16]. 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

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

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

$$\alpha := DP(t, \cdot) \# \rho(t, \cdot), \quad \mathbf{w}(t, y) = J(y - DP^*(t, y)). \quad (18)$$

The *semigeostrophic shallow water in dual variables* are

$$\left\{ \begin{array}{l} (i) \quad \frac{\partial \alpha}{\partial t} + \operatorname{div}(\alpha \mathbf{w}) = 0 \quad \text{in the weak sense in } [0, T] \times \mathbf{R}^2 \\ (ii) \quad \mathbf{w}(t, y) := J(y - DP^*(t, y)), \quad \text{in } [0, T] \times \mathbf{R}^2 \\ (iii) \quad P(t, x) := |x|^2/2 + h(t, x), \quad \text{in } [0, T] \times \Omega \\ (iv) \quad \alpha(t, \cdot) := DP(t, \cdot) \# h(t, \cdot), \quad t \in [0, T] \\ (v) \quad \alpha(0, \cdot) = \alpha_o \quad \text{in } \mathbf{R}^2. \end{array} \right. \quad (19)$$

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

$$H(\alpha) := 1/2 \min_{\eta \in \mathcal{P}_\alpha(\Omega)} \{W_2^2(\alpha, \eta) + \int_{\Omega} \eta^2 dx\}. \quad (20)$$

Step 1. Given $\alpha_k \in \mathcal{P}_a(\mathbf{R}^d)$, we substitute α in (20) 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 (20) give that

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

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

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

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 [16] we show the following theorem.

Theorem 6.6 (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 (19) 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 3. 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 [4] to develop a rigorous tool that associate a riemannian structure to the Wasserstein distance is a step toward finding a systematic way of studying uniqueness 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 $-\operatorname{div}(\omega_o \nabla \psi_f) = f$. The inner product of $f, g \in T_{\omega_o}\mathcal{M}$ suggested by [37] is

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

We propose to introduce the skew-symmetric form

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

where J is the standard symplectic 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 $t \rightarrow \omega(t, \cdot)$ satisfying for all $f \in T_{\omega(t, \cdot)} \mathcal{M}$,

$$\langle \dot{\omega}, f \rangle = \beta_{\omega_o} \left(\frac{\delta H}{\delta f}, f \right). \quad (21)$$

Uniqueness of solution in (21) will be straightforward 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

$$\partial_t \omega = \operatorname{div} \left[\omega \nabla \left(A'(\omega) \right) \right]. \quad (22)$$

The question is to know how much (22) contributes to the understanding of (21).

7 Appendix

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

$$(H1) \quad L(x+k, v) = L(x, v) \text{ for each } (x, v) \in \mathbf{R}^d \times \mathbf{R}^d \text{ and each } k \in \mathbf{Z}^d.$$

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

$$(H2) \quad L \in C^r(\mathbf{R}^d \times \mathbf{R}^d).$$

We also impose that the Hessian matrix is positive:

$$(H3) \quad \left(\frac{\partial^2 L}{\partial v_i \partial v_j} (x, v) \right) > 0$$

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

$$(H4) \quad \text{For every } A > 0 \text{ there exists a constant } \delta > 0 \text{ such that } \frac{L(x, v)}{\|v\|} \geq A$$

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 $L : \mathbf{R}^d \times \mathbf{R}^d \rightarrow \mathbf{R}$ satisfying (H1–H4) 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

$$(H5) \quad L(x, v) \leq e^{\|v\|} - b(L) - 1 \text{ for each } (x, v) \in \mathbf{R}^d \times \mathbf{R}^d.$$

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

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

The Hamiltonian H is continuous and by (H3), $H(x, \cdot)$ is continuously differentiable. Also, the map

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

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. Its inverse

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

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 [32] 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

$$v = \nabla_p H(x, p), \quad \nabla_x L(x, v) = -\nabla_x H(x, p). \quad (23)$$

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

$$c(T, x_o, x_T) = \inf_{\sigma} \left\{ \int_0^T L(\sigma, \dot{\sigma}) dt \mid \sigma(0) = x_o, \sigma(T) = x_T \right\}, \quad (24)$$

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

In the light of (H3) and (H4), there exists $\sigma_o \in AC(T; x, x_T)$ that is a minimizer in (24) and σ_o satisfies the Euler-Lagrange equations

$$\frac{d}{dt} [\nabla_v L(\sigma_o(t), \dot{\sigma}_o(t))] = \nabla_x L(\sigma_o(t), \dot{\sigma}_o(t)), \quad 0 < t < T. \quad (25)$$

The infimum in (24) represents the cost for transporting a unit mass from x_o to x_T during the time interval $T > 0$. There may be several σ_o minimizer in (24), if

the minimum is performed other $AC(T; x_o, x_T)$. Therefore, the differential equation (25) may have multiple solutions in $AC(T; x_o, x_T)$. It is natural to ask if given $(x_o, v) \in \mathbf{R}^d \times \mathbf{R}^d$, (25) 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* $\Phi : \mathbf{R} \times \mathbf{R}^d \times \mathbf{R}^d \rightarrow \mathbf{R}^d \times \mathbf{R}^d$ associated to the Lagrangian L , defined by $\Phi(t, x, v) = (\phi(t, x, v), \dot{\phi}(t, x, v))$ where ϕ satisfies

$$\frac{d}{dt}[\nabla_v L(\sigma(t), \dot{\sigma}(t))] = \nabla_x L(\sigma(t), \dot{\sigma}(t)), \quad \sigma(0) = (x, v). \quad (26)$$

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 (23), we have that (26) is equivalent to

$$\dot{\sigma}(t) = \nabla_p H(\sigma(t), p(t)) \quad \dot{p}(t) = -\nabla_x H(\sigma(t), p(t)) \quad \sigma(0) = x, \quad p(0) = \nabla_v L(x, v). \quad (27)$$

Now (27) 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

$$H(\sigma(t), p(t)) = H(\sigma(0), p(0)), \quad (t \in (t_-, t_+)). \quad (28)$$

As a byproduct, (26) 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:

$$\Phi(t, x, v) = (\sigma(t), \dot{\sigma}(t)), \quad \Phi(0, x, q) = (x, v).$$

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

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

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 [32] we can ensure that the completeness assumption $t_- = -\infty$ and $t_+ = +\infty$ holds. For that it is enough to impose that L satisfies (H5) so that

$$H(x, p) \geq \|p\| \text{Log} \|p\| + b(L) + 1 \geq \|p\| + b(L). \quad (29)$$

If (29) holds then by (28) we have that

$$\|p(t)\| \leq \bar{c} := H(\phi(0), p(0)) - b(L). \quad (30)$$

We combine (27) and (30) to have that

$$\|\dot{\phi}(t)\| \leq \|\nabla_p H\|_{L^\infty(\mathbf{T}^d \times \bar{B}_{\bar{c}}(0))} \quad (31)$$

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 in that case we must have 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

$$\Phi(t+s, x, v) = \Phi(t, \Phi(s, x, v)), \quad ((t, s) \in \mathbf{R} \times \mathbf{R}, (x, v) \in \mathbf{R}^d \times \mathbf{R}^d). \quad (32)$$

This is a byproduct of the uniqueness property of solutions of (26). Also if $T > 0$ and $\Phi(T, x, v) = \Phi(0, x, v)$ then $\Phi(\cdot, x, v)$ must be periodic of period T .

$$\Phi(t+T, x, v) = \Phi(t, x, v), \quad ((t, x, v) \in \mathbf{R} \times \mathbf{R}^d \times \mathbf{R}^d). \quad (33)$$

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

$$\langle \nabla^2 l(v)a; a \rangle \geq e_l \|a\|^2$$

for all $a \in \mathbf{R}^d$. We show that for small times $T > 0$ there exists a unique optimal path $\sigma_{x,y}$ that minimizes $\sigma \rightarrow \int_0^T L(\sigma, \dot{\sigma}) ds$ 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 7.1 (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 (25). Therefore, σ_o is the unique minimizer of $\sigma \rightarrow K[\sigma] = \int_0^T L(\sigma, \dot{\sigma}) ds$ over $AC(T, x, y)$.*

Proof: Assume that $\sigma_o \in AC(T, x, y)$ satisfies (25). We write Taylor approximation of $L(\sigma, \dot{\sigma})$ around $(\sigma_o, \dot{\sigma}_o)$, use that satisfies the Euler-Lagrange equation (25) and that $0 = \sigma(0) - \sigma_o(0) = \sigma(T) - \sigma_o(T)$ to conclude that

$$K[\sigma] - K[\sigma_o] \geq \int_0^T (e_W |\sigma - \sigma_o|^2 + e_l |\dot{\sigma} - \dot{\sigma}_o|^2) ds \geq \int_0^T (e_W + \frac{e_l c_1}{T^2}) |\sigma - \sigma_o|^2 ds.$$

This concludes the proof of the proposition.

QED.

References

- [1] G. Alberti. On the structure of singular sets of convex functions. *Calc. Var. Partial Differential Equations* **2**, 17–27 (1994).
- [2] 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).
- [3] 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
- [4] L. Ambrosio, N. Gigli and G. Savaré. Gradient flows in metric spaces and the Wasserstein spaces of probability measures. (Preprint).
- [5] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.
- [6] J.D. Benamou and Y. Brenier. Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère equations/transport problem. *SIAM J. Appl. Ana. Math.* **58**, no 5, 1450–1461 (1998).
- [7] L.A. Caffarelli. The regularity of mappings with a convex potential. *J. Amer. Math. Soc.* **5** (1992) 99–104.
- [8] L.A. Caffarelli. Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* **44** (1992) 1141–1151.
- [9] L.A. Caffarelli. Boundary regularity of maps with convex potentials — II. *Ann. of Math. (2)* **144** (1996) 453–496.

- [10] 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.
- [11] E.A. Carlen and W. Gangbo. Constrained Steepest Descent in the 2–Wasserstein Metric. *Annals of Mathematics*. **157**, 807–846, 2003.
- [12] 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.
- [13] 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/.
- [14] 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/.
- [15] P.Cloke and M.J.P. Cullen. A semi-geostrophic ocean model with outcropping . *Dyn. Atmos. Oceans*, 21:23–48, 1994.
- [16] 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.
- [17] M.J.P. Cullen and R.J. Purser. An extended lagrangian theory of semi-geostrophic frontogenesis. *J. Atmosph. Sciences* **41**, 1477–1497 (1984).
- [18] M.J.P. Cullen and R.J. Purser. Properties of the lagrangian semigeostrophic equations. *J. Atmosph. Sciences* vol 40, **17**, 2684–2697 (1989).
- [19] 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.
- [20] B. Dacorogna. *Direct Methods in the Calculus of Variations*. Springer-Verlag, 1989.
- [21] A Eliassen. The quasi-static equations of motion. *Geofys. Publ.*, **17**, No 3, 1948.
- [22] L.C. Evans. Partial differential equations and Monge-Kantorovich mass transfer. *Int. Press, Current Dev. Math*, **26**, 26–78, 1997.

- [23] 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.
- [24] W. Gangbo. An elementary proof of the polar factorization of vector-valued functions. *Arch. Rational Mech. Anal.*, 128:381–399, 1994.
- [25] 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).
- [26] 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.
- [27] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.* **177** (1996) 113–161.
- [28] B.J. Hoskins. The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmosph. Sciences* **32**, 233–242, 1975.
- [29] Jordan, Kinderlehrer and Otto. The variational formulation of the Fokker-Planck equation. *SIAM Jour. Math Anal.*, **29** 1–17, 1998.
- [30] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
- [31] L. Kantorovich. On a problem of Monge (In Russian). *Uspekhi Math. Nauk.*, 3:225–226, 1948.
- [32] J.N. Mather. Variational construction of connecting orbits. *Ann. Inst. Fourier, Grenoble*, **43** no 5, (1993), 1349–1386.
- [33] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.* **128** 153–179, 1997.
- [34] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, **80** 309–323, 1995.
- [35] 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).

- [36] B.S. Katz, editor. *Nobel Laureates in Economic Sciences: a Biographical Dictionary*. Garland Publishing Inc., New York, 1989.
- [37] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. P.D.E.*, **26** 156–186 2001.
- [38] J. Pedlosky. *Geophysical Fluid Dynamics*. Springer-Verlag (1982).
- [39] S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems; Vol. I & II. Probability and its Applications*, Springer Verlag, Heidleberg (1998).
- [40] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [41] 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).
- [42] Royden. *Real Analysis*, 3rd Ed., Macmillan, 1988.
- [43] H. Stommel. The westward intensification of wind-driven ocean currents. *Trans. Amer. Geoph. Union* **29**, 202–206 (1948).
- [44] V. N. Sudakov. Geometric problems in the theory of infinite-dimensional probability distributions. *Proc. Steklov Inst. Math.* **141** (1979) 1–178.
- [45] N. Trudinger and X.-J. Wang. On the Monge mass transfer problem. *Calc. Var. Partial Differential Equations* **13**, 19–31. (2001)